



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Directed evolution and structural
analysis of an OB-fold domain towards a
specific binding reagent.

A thesis submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy

in

Biological Sciences

at

The University of Waikato

by

John Durand Steemson



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2011

Acknowledgements

First and foremost I would like to thank my supervisor, Assoc. Prof. Vickery Arcus, who was endlessly patient and encouraging in equal measure, thank you Vic, I couldn't have wished for a better supervisor. My second supervisor also deserves thanks, Dr. J. Shaun Lott. Thanks to Dr. Jansna Rankonjac for the gift of the phage display vectors and expertise, Dr. Matthias Baake who started this project, and Dr. Randall Watson who joined in for a while as well.

I couldn't have done it without the support of my lab mates in the C2 labs; with special word for Dr. Judith burrows, without whom the lab would fall over. Matt, Marisa and Svendh, you were my official-unofficial support group and I love you all!

Huge thanks to Dr. Mark Liddament for last minute advice, my dear mother Katherine Ransom for the proof-reading services, my Dad for the care package and well wishes, my Grandmother Betty for the financial support over my entire university education as well as when I needed it most and to Dr. Jo Peace for all-round moral support and general awesomeness.

I was funded with a Waikato University Doctoral scholarship, without which I could not have done it. I also received ongoing support from the AgResearch Forage Biotechnology Group.

The Maurice Wilkins Centre at Auckland University, headed by Prof. Ted Baker, has been welcoming in the extreme, and most of the work in this thesis would not have been done without access to their x-ray crystallography equipment. An extra word for Ivan Ivanovic, Richard Bunker and Graham Bailey from that group, who all helped me a lot with the various instruments that I used.

Abstract

Interactions between proteins are a central concept in biology, and understanding and manipulation of these interactions is key to advancing biological science. Research into antibodies as customised binding molecules provided the foundation for development of the field of protein “scaffolds” for molecular recognition, where functional residues are mounted on to a stable protein platform. Consequently, the immunoglobulin domain has been describes as “nature’s paradigm” for a scaffold, and has been widely researched to make engineered antibodies better tools for specific applications. However, limitations in their use have lead to a number of non-immunoglobulin domains to be investigated as customisable scaffolds, to replace or complement antibodies. To be considered a scaffold, a protein domain must show an evolutionarily conserved hydrophobic core in diverse functional contexts. The study presented here investigated the oligosaccharide/oligonucleotide-binding (OB) fold as scaffold, which is a 5-standed β -barrel seen in diverse organisms with no sequence conservation. The term “Obody” was coined to describe engineered OB-folds. This thesis examined a previously engineered Obody with affinity for lysozyme ($K_D = 40 \mu\text{M}$) in complex with its ligand by x-ray crystallography (resolution 2.75 Å) which revealed the atomic details of binding. Affinity maturation for lysozyme was undertaken by phage display directed evolution. Gene libraries were constructed by combinatorial PCR incorporating site-specific randomised codons identified by examination of the structure in complex with lysozyme, or by random generation of point mutations by error-prone PCR. Overall a 100-fold improvement in affinity was achieved ($K_D = 600 \text{ nM}$). To investigate the structural basis of the affinity maturation, two further Obody-lysozyme complexes were solved by x-ray crystallography, one at a K_D of 5 μM (resolution 1.96 Å), one at 600 nM (resolution 1.86 Å). Analysis of the structures revealed changes in individual residue arrangements, as well as rigid-body changes in the relative orientation of the Obody and lysozyme molecules in complex. Directed evolution of Obodies as protein binding reagents remains a challenge, but this study demonstrates their potential. The structures presented here will contribute invaluable insights for the future design of improved Obodies.

Table of Contents

Acknowledgements.....	2
Abstract.....	3
Table of Contents.....	4
Table of Figures.....	10
Table of Tables.....	12
Abbreviations.....	13
1 General Introduction.....	15
1.1 Proteins as scaffolds for functional residues.....	15
1.2 Scaffolds for General Molecular Recognition.....	16
1.2.1 Antibodies.....	16
1.2.1 Constrained Peptide Carriers.....	21
1.2.2 Ig-like domains as scaffolds.....	22
1.3 Non-Immunoglobulin-like Scaffolds.....	24
1.3.1 Identification of a scaffold candidate.....	24
1.3.2 Anticalins.....	26
1.3.3 Affibodies.....	27
1.3.4 Repeat Proteins.....	28
1.4 The OB-fold as a Scaffold.....	29
1.5 Directed Evolution Technology.....	32
1.5.1 Diversity Generation.....	33
1.5.2 In vivo selection methods.....	34
1.5.3 In vitro selection methods.....	38
1.6 Protein-Protein Interactions.....	40
1.6.1 Geometric complementarity and polar bonds.....	41
1.6.2 Hydrophobic interactions.....	42

1.6.3	Hot spots	43
1.6.4	Thesis Outline	44
2	Materials and Methods	45
2.1	General methods	45
2.1.1	Clone Storage	45
2.1.2	Primer Design	45
2.1.3	Sequence Analysis and Alignment	45
2.1.4	Common Buffers.....	45
2.2	DNA Manipulation and Analysis	46
2.2.1	PCR	46
2.2.2	Agarose Gel Electrophoresis.....	47
2.2.3	DNA Purification	48
2.2.4	DNA Cloning	49
2.2.5	Alkaline Lysis Plasmid Purification	51
2.3	Microbiology	52
2.3.1	Antibiotics	52
2.3.2	Media	52
2.3.3	Electrocompetent Cells – General Cloning.....	54
2.3.4	Electroporation transformation of E. coli	54
2.4	Gene Libraries and Selection.....	55
2.4.1	Phagemid vector.....	55
2.4.2	Gene Library Construction.....	55
2.4.3	Phage Libraries	59
2.4.4	Phage Display	62
2.4.5	Western Blot Protein Detection	65
2.4.6	Surface Plasmon Resonance	67
2.4.7	Isothermal Titration Calorimetry	68
2.4.8	Differential Scanning Fluorimetry	69

2.5	Protein Expression and Purification	70
2.5.1	SDS - Polyacrylamide Gel Electrophoresis (PAGE)	70
2.5.2	General Protein Expression.....	72
2.5.3	Immobilised Metal Affinity Chromatography	73
2.5.4	GSH-Affinity Chromatography	74
2.5.5	rTEV Cleavage.....	75
2.5.6	Size-Exclusion Chromatography	75
2.6	Protein Crystallography	77
2.6.1	Sitting Drop Method	77
2.6.2	Crystal Diffraction Screening	77
2.6.3	Data Collection	78
2.6.4	Data Processing.....	78
3	Structure of an engineered Obody complex.....	80
3.1	Introduction.....	80
3.1.1	Preliminary Obodies Work	80
3.2	Results.....	82
3.2.1	Expression and Purification	82
3.2.2	Crystallisation and Data Collection	83
3.2.3	Model building and Refinement	84
3.2.1	The Obody-HEWL Complex	87
3.2.2	The Obody-HEWL interface.....	93
3.2.3	Comparison with the initial library model	95
3.2.4	Comparison with Other HEWL complexes	95
3.2.5	Structure-Based Library Design.....	98
3.3	Discussion.....	100
3.3.1	Structural Validation of Library Design	100
3.3.2	Affinity maturation	101
4	Gene Library Design and Selection	103

4.1	Introduction.....	103
4.1.1	Outline.....	103
4.2	Results.....	104
4.2.1	Library nomenclature	104
4.2.2	Assessment of phagemid pR _{psp2}	104
4.2.3	paOB3 wt template modification	106
4.2.4	Structure-based affinity maturation – L8 10m.....	108
4.2.5	Structure-based affinity maturation – 6m library.....	118
4.2.6	Random Mutation Library	120
4.2.7	Limited Codon Library	126
4.2.8	Thermal stability of engineered Obodies	130
4.3	Discussion.....	132
4.3.1	Structure-based design	132
4.3.2	Error-Prone Affinity Maturation	135
4.3.3	Limited Codon Library	138
4.3.4	Limitations of the phage-display format.....	139
4.3.5	Surface Plasmon Resonance	140
4.3.6	Conclusions	140
5	Structural Analysis of Engineered OB-folds	142
5.1	Introduction.....	142
5.2	Results - L8 10m L10	142
5.2.1	Expression and Purification	142
5.2.2	Crystallisation and Data Collection	143
5.2.3	Model building and Refinement	144
5.2.4	Library Mutations	147
5.2.5	Interface Residues	148
5.2.6	Summary	152
5.3	Results - L200EP 06	154

5.3.1	Expression and Purification	154
5.3.2	Crystallisation and Data Collection	154
5.3.3	Model building and Refinement	155
5.3.4	Library Mutations	158
5.3.5	Description of the Obody Interface.....	164
5.3.6	Three-structure comparison	166
5.4	Discussion.....	167
5.4.1	Affinity maturation - L8 10m L10	167
5.4.2	Affinity maturation – L200EP 06	170
5.4.3	Conclusion	172
6	General Discussion	173
6.1.1	Structural Validation and Library Improvement.....	174
6.1.2	Structure-based affinity maturation for HEWL	175
6.1.3	Third generation structure-based affinity maturation	175
6.1.4	Third generation random affinity maturation.....	177
6.1.5	Conclusions	179
6.1.6	Summary	180
A1	Appendix	182
A1.1	Electrophoresis standards	182
A1.1.1	Precision Plus Protein Standards.....	182
A1.1.2	1 kb Plus DNA Standards.....	182
A1.2	Vectors.....	183
A1.2.1	pProEx Htb	183
A1.2.2	pRpsp2.....	183
A1.3	Oligonucleotides	184
A1.4	Sequencing Data	186
A1.4.1	OB-fold domain from aspartyl tRNA-synthetase.....	186
A1.4.2	Unselected Variant 13mRL U81	187

A1.4.3	L8 10m Library	188
A1.4.4	L10 6m library	190
A1.4.5	Error-prone library	191
A1.5	rTEV Protease.....	195
A1.5.1	Reagents	195
A1.5.2	Expression Method.....	196
A1.6	Crystal Screen	197
A1.7	IUPAC Nucleotide codes.....	199
A2	Appendix – digital media	200
	References.....	201

Table of Figures

Figure 1.1 Antibody construction.....	17
Figure 1.2 Engineered FN3 domain	23
Figure 1.3 Engineered Non-Ig Domains.....	26
Figure 1.4 OB-fold diversity	31
Figure 2.1 Combinatorial gene assembly schematic	57
Figure 3.1 Sample purification of 13mRL L8.	83
Figure 3.2 13mRL L8-HEWL complex crystal	84
Figure 3.3 Interface Contacts vs. Library Mutations.....	89
Figure 3.4 Intermolecular Polar Bonds	91
Figure 3.5 Obody predicted hotspots	93
Figure 3.6 Overlay of the Obody monomers	94
Figure 3.7 Superposition of 13mRL L8 with the library modelled domain	95
Figure 3.8 Comparison with other known HEWL binders.....	97
Figure 3.9 Selected face residues for the L8 10m library	99
Figure 4.1 Library nomenclature conventions.....	104
Figure 4.2 paOB3 template modification.....	107
Figure 4.3 L8 10m library design	109
Figure 4.4 L8 10m library gel fragments	111
Figure 4.5 Panning results with L8 10m phage library	112
Figure 4.6 Unique sequences from the L8 10m library	113
Figure 4.7 Raw SPR data	115
Figure 4.8 Affinity analysis of variants from the L8 10m library	116
Figure 4.9 L10 6m library design	118
Figure 4.10 L10 6m library assembly.....	119
Figure 4.11 L10 6m library panning titre.....	120
Figure 4.12 L200EP library panning titre	121
Figure 4.13 Mutation clustering in L200 EP variants	123
Figure 4.14 Biacore analysis of L200EP round five variants.....	125
Figure 4.15 13mLC library plan.....	127
Figure 4.16 13mLC assembly process	128
Figure 4.17 13mLC panning results.....	129
Figure 4.18 Differential scanning fluorimetry thermal stability curves	130
Figure 5.1 L8 10m L10 purification	143
Figure 5.2 L8 10m L10-HEWL complex crystals and diffraction.....	144
Figure 5.3 L8 10mL10 mutations	148

Figure 5.4 Varied modes of L8 10mL10 Obody-HEWL complexation.....	149
Figure 5.5 Water-filled cavity at the L8 10m L10-HEWL interface	152
Figure 5.6 Purification of L200EP 06.....	154
Figure 5.7 Crystals of L200EP 06 in complex with HEWL	155
Figure 5.8 S29H and A56R.....	159
Figure 5.9 Mutations in L200EP 06.....	160
Figure 5.10 M37K Mutation	161
Figure 5.11 Loop mutations and their effect on the interface.....	163
Figure 5.12 FADE surface complementarity at the interface	165
Figure 5.13 Overlay of all three Obodies and their cognate HEWLs	166

Table of Tables

Table 2.1 Restriction enzymes used.....	49
Table 2.2 Bacterial strains used.....	54
Table 3.1 13mRL L8 complex structure statistics.....	86
Table 3.2 Intermolecular polar bonds between 13mRL L8 and HEWL	90
Table 3.3 Hot spot predictions for the 13mRL L8 HEWL complex.....	92
Table 3.4 Comparison of Obody-HEWL interface.....	98
Table 4.1 Effect of dephosphorylated vector on transformation.....	105
Table 4.2 Affinity for HEWL of L8 10m library variants.....	115
Table 4.3 Affinities from L200EP library.....	126
Table 5.1 L8 10m L10-HEWL complex structure statistics.....	146
Table 5.2 Polar bonds across the asymmetric unit	151
Table 5.3 Hot Spot predictions for the L8 10mL10 interface	151
Table 5.4 L200EP 06-HEWL complex structure statistics	157
Table 5.5 Polar bonds at the L200EP-HEWL interface	164
Table 5.6 Hot spot predictions of the L200EP 06-HEWL interface.....	165
Table 5.7 Comparative interface statistics.....	167

Abbreviations

Amp	Ampicillin
APS	Ammonium persulfate
aspRS	Aspartyl tRNA synthetase
CDR	Complementarity determining region
cfu	Colony forming units
c-terminal	Carboxy terminal
CTLA-4	Cytotoxic T-lymphocyte antigen 4
DARPin	Designed ankyrin repeat protein
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribose nucleic acid
dNTP	Deoxynucleotide triphosphate
dsDNA/RNA	Double stranded DNA or RNA
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediamine tetra-acetic acid
Fab	Antibody “antibody” fragment
FN3	Fibronectin type-III domain
Fv	Antibody “variable” fragment
gIII	Gene three from M13 bacteriophage
GSH	Glutathione reduced
GSSH	Glutathione oxidised
GST	Glutathione-S-transferase
H-bond	Hydrogen bond
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEWL	Hen egg white lysozyme
Ig	Immunoglobulin
IMAC	Immobilised metal affinity chromatography
IPA	Isopropan-2-ol
IPTG	Isopropyl β -D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
kb	Kilo-base
kDa	Kilo-dalton
MHC	Major histocompatibility complex
moi	Multiplicity of infection
MPa	Mega-pascals
MPEG	Methoxypolyethylene glycol
MQ	Molecular biology grade water
mRNA	Messenger RNA
NCS	Non-crystallographic symmetry
n-terminal	Amino terminal
OB-fold	Oligosaccharide/oligonucleotide binding fold
OD₆₀₀	Optical density at 600 nm
PAD	Protein assay dye
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PDB	Protein databank

PEG	Polyethylene glycol
pfu	Plaque forming units
pIII	Protein translated from gIII of M13 bacteriophage
r.m.s.d.	Root mean squared deviation
RNA	Ribonucleic acid
RNAse A/RA	Ribonuclease A
rTEV	Recombinant tobacco etch virus
RU	Response unit
scFab	Single chain Fab fragment
scFv	Single chain Fab fragment
SDS	Sodium dodecyl sulfate
SPA	Staphylococcal protein A
SPR	Surface plasmon resonance
ssDNA/RNA	Single stranded DNA or RNA
TAE	Tris-acetate-EDTA
TBS	Tris Buffered Saline
TCR	T-cell receptor
TDP	Transducing particles
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
tRNA	Transfer ribose nucleic acid
UV₂₈₀	Ultra-violet (280 nm)
wt	Wild-type
βme	β-mercaptoethanol

1 General Introduction

1.1 Proteins as scaffolds for functional residues

Molecular recognition is a central concept in biology. Cellular metabolism is critically dependant on specific interactions between proteins and their substrates or ligands, interactions which range from transient regulatory adjustments to practically irreversible binding. Although a growing body of evidence is showing the importance of natively unfolded proteins (which typically fold on binding), this process depends generally on the versatility of proteins to adopt well-defined folds in order to present the chemistry required.

Compared to the number of possible DNA coding sequences there are relatively few (~2500) unique folds known. Many folds are disproportionately populated, which suggests that they occupy especially useful minima in the energy landscape, or that they were discovered first by evolution and have had longer to diversify. Sequence conservation, which is often used as a proxy for fold conservation, implies homology, but in the absence of detectable sequence conservation, structural comparison is used, as by the Structural Classification of Proteins (SCOP) database, to identify analogous or homologous proteins which possess similar folds (Murzin *et al.* 1995; Lo Conte *et al.* 2000; Lo Conte *et al.* 2002; Andreeva *et al.* 2004; Andreeva *et al.* 2008). For example, the known structures showing an immunoglobulin (Ig)-like fold are divided into 28 superfamilies, each defined by a different set of distinct structural and sequence features.

A common fold that is used for many different functions indicates evolutionary utility, in the sense that it can accommodate many mutations while retaining stability, allowing it to either diverge widely or be independently discovered by convergent evolution. Regardless of proposed evolutionary origin, the unifying fold of a cluster of superfamilies can therefore be considered a “scaffold” on which different functions and binding sites can be mounted through the introduction of point mutations, extended loops and even entire domain insertions. Based on this concept, proteins amenable to engineering can, in principle, be identified and manipulated to perform useful

functions (Ku and Schultz 1995). Engineering efforts utilising the scaffold concept essentially fall into two main categories; tailored enzyme catalysis and custom molecular recognition.

In the first part of this chapter I introduce the concept of the protein scaffold, its application to antibody engineering and the immunoglobulin (Ig) fold, expansion to alternative non-Ig domains and propose a new alternative scaffold domain, the oligosaccharide/oligonucleotide (OB)-fold. The second part of this chapter introduces the most powerful technique currently employed in protein engineering, namely directed evolution. Lastly, specifics of protein-protein interactions are discussed, with reference to discovering the important residues of a newly engineered interface.

1.2 Scaffolds for General Molecular Recognition

Applications in diagnostics, therapeutics and experimental biology require tools for specific and sensitive binding to targets of interest. Traditionally, antibodies are the gold standard for customisable molecular recognition. But while the antibody format has proved especially useful in many contexts, limitations of antibodies along with the development of molecular techniques for manipulation of mutant gene libraries have prompted a range of protein scaffolds to be investigated as alternatives.

1.2.1 *Antibodies*

Antibodies are multi-chain complexes of immunoglobulin (Ig) fold domains, stabilised by inter- and intra-molecular disulfide bonds. The Ig-fold consists of a densely packed β -sandwich of 7-9 strands (Bork *et al.* 1994) and has been described as nature's paradigm for a scaffold. At each end of the domain the β -strands are linked in a Greek-key motif with loops of varying lengths. In antibodies, these loops form the complementarity-determining regions (CDR) for specific recognition of antigens. Although antibodies come in several isotypes, the most common form in biotechnology is IgG, which consists of two four-domain heavy chains and two two-domain light chains, linked covalently with disulfide bridges (Figure 1.1 A).

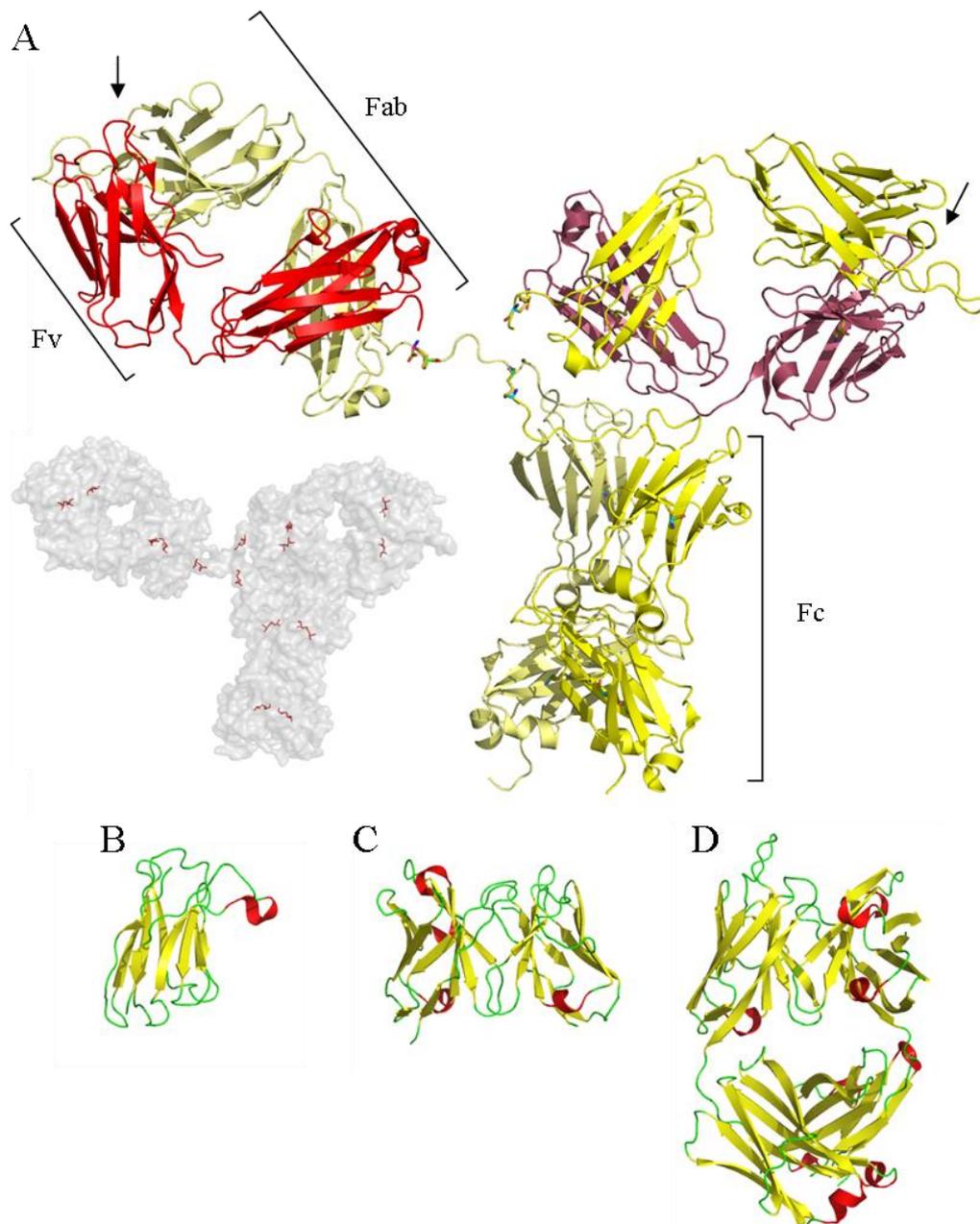


Figure 1.1 Antibody construction

(A) An IgG antibody (PDB accession 1HZH) displaying the classic “T” shape, with the different fragments labelled. The heavy chains are coloured in yellow, the light chains in red and purple. Each intact IgG antibody consists of two Fab and one Fc fragment. The Fab (antibody) fragment is the scaffold on which the CDR loops are mounted (indicated by arrows) and is further divided into the Fv (variable) region, where almost all of the variation between antibodies is located. The Fc region doesn’t contain variable residues and is the binding site for other immune-associated proteins. The inset grey surface model of the same molecule shows the 14 disulfide bonds present in this antibody (red). The other three molecules show the major structural formats of engineered antibodies, coloured according to secondary structure; (B) a VH camelid domain (PDB accession 3K81), (C) a single-chain Fv fragment (PDB accession 1DZB) and (D) an Fab fragment (PDB accession 1MRF).

The immune system generates diversity in the CDR loops, mounted on the variable domains, by shuffling gene segments (denoted VDJ) and somatic hypermutation (French *et al.* 1989; Rajewsky 1996). The initial germ line-encoded generation gives rise to a population of 10^7 - 10^8 lymphocytes containing unique sequences which are routinely exposed to potential antigens. On repeated exposure to an antigenic ligand, a subset of positively-selected lymphocytes is re-mutated to generate tighter, more specific antibodies. The Ig-fold can accommodate these drastic sequence changes and remain stable by a virtual decoupling of folding from loop sequence and length, although surface turns have been shown to influence stability of a folded domain (Predki *et al.* 1996; Nagi and Regan 1997). The scaffold residues of the Ig-domain, rather than being directly involved in binding, serve to support and impose conformational restrictions on the loop residues (AlLazikani *et al.* 1997). Loop sequence plasticity is highlighted by CDR grafting experiments, where affinity for a ligand can be stably transferred by transplant of the CDR loops from one antibody to another (Kettleborough *et al.* 1991; Foote and Winter 1992). The first true use of antibodies as a scaffold produced fully “humanised” antibodies by transplanting both heavy and light chain CDR regions of a mouse antibody into a human IgG scaffold (Riechmann *et al.* 1988a).

1.2.1.1 Limitations

Before molecular techniques to produce synthetic genetic diversity were developed, the lymphocyte repertoire and *in vivo* selection occurring naturally in the immune system was the most readily accessible source of selectable diversity for production of molecular recognition reagents. Consequently, antibodies became the default format for binding reagents. While successful in many ways, *in vivo* immunisation and harvesting of anti-serum posed several problems, particularly in therapeutic contexts.

The polyclonal nature of antiserum can result in a high level of non-specific, or otherwise unwanted, reactivity. This prompted development of technology to produce monoclonal isolates, both to increase specificity and to isolate a particular binding profile to a particular sequence (Kohler and Milstein 1975). Further, while the immune complement is extensive, it is the result of a negative selection for anti-self activity, which necessarily excludes

many otherwise potentially useful sequences. PCR amplified B-cell derived phage libraries sought to increase the diversity for selection beyond that available *in vivo* (Marks *et al.* 1991; Gram *et al.* 1992; Hoogenboom and Winter 1992; Akamatsu *et al.* 1993; Vaughan *et al.* 1996). Potential target ligands were also limited by their toxicity, serum lability and their ability to elicit a sufficiently vigorous immune response. *In vitro* immunisation of lymphocytes sought to bypass the problems of antigen toxicity and *in vivo* degradation (Borrebaeck *et al.* 1988).

Finally, natural antibodies are large, multi-chain constructs, requiring post-translational assembly and multiple cloning events to introduce required specificity, as well as specialised cell lines for expression. Because of the high cost of expression in tissue culture, alternative hosts for antibody expression have been investigated (Skerra 1993; Verma *et al.* 1998) and today antibody libraries are routinely produced in microbial expression/selection hosts, as well as completely *in vitro* systems. Tissue penetration and antigenicity of a protein therapeutic is directly related to molecular weight, so different formats were investigated to improve antibodies as drugs. Murine monoclonal antibodies tend to draw an immune response when administered as a drug, reducing effectiveness (Kuus-Reichel *et al.* 1994). Murine/human chimeric antibodies incorporating recombinant human heavy chains and native mouse light chains were developed to reduce antigenicity (Jones *et al.* 1986; Better *et al.* 1988). Smaller but still functional antibody fragments were explored as an alternative to intact antibodies, where an intact light chain and the variable domain and first constant domain of a heavy chain are expressed (Fab) or just the variable domains (Fv) (Riechmann *et al.* 1988b)(Figure 1.1 A,D).

The non-covalent association of co-expressed domains in Fv fragments reduces stability compared to larger fragments incorporating disulfide bonds (Reiter *et al.* 1994a; Reiter *et al.* 1994b). These stability and ease-of-manipulation limitations were partially circumvented by fusion of the light and heavy chain variable regions for production as a single polypeptide (scFv, Figure 1.1 C) (Huston *et al.* 1988; Bird and Walker 1991), although in practise these constructs exhibited a tendency to aggregate, probably because of transient dissociation of the two domains when not covalently stabilised (Reiter *et al.* 1994a; Reiter *et al.* 1994b). Recently, aggregation-resistant scFv

antibodies were developed by shuffling fragments of antibodies known to be thermally resistant (Christ *et al.* 2007). Stand-alone variable domains from heavy chains (V_H) were also trialled, but were plagued by problems of aggregation and non-specific adsorption, presumably due to exposure of the relatively hydrophobic dimerisation face (Ward *et al.* 1989) (Glockshuber *et al.* 1990). The later discovery of naturally occurring heavy chains without light chain partners in camelids (camels and llamas, denoted VHH) revitalised this effort (Hamerscaterman *et al.* 1993; Ghahroudi *et al.* 1997; Muyldermans 2001). Efforts in overcoming these limitations and applications of products of recent antibody engineering developments have been the subject of recent reviews (Weiner and Carter 2005; Almagro *et al.* 2006; Sidhu and Fellouse 2006; Zafir-Lavie *et al.* 2007).

1.2.1.2 Immune system independence

All of these developments mentioned above have resulted in a gradual dissociation of antibody selection from the immune system. Today, artificially-generated diversity can be introduced into synthetic, consensus-designed Ig domains by biasing libraries towards sequences that allow for the maximum proportion of correctly folded proteins while maintaining diversity (Knappik *et al.* 2000).

Perhaps the single most important development in non-immune antibody generation was phage display technology (Smith 1985). Whereas the immune system (and other cell-based systems) performs selection by expression of an antibody on the surface of a cell, in phage display the selectable unit is a virus displaying a fusion protein, allowing selection for binding outside of the immune system (Clackson *et al.* 1991; Hawkins *et al.* 1992; Figini *et al.* 1994; Nissim *et al.* 1994; Winter *et al.* 1994), initially with antibody libraries based on “primed” B-cell diversity complements and later with completely naïve diversity. Compared to cell-based systems, phage libraries are more robust, cheaper and capable of handling much higher levels of diversity. Phage display is described in more detail in section 1.5, along with other directed evolution methods. Drawing on evolutionary principles and modelled on the immune system, phage display technology, coupled with techniques for artificial introduction of diversity, has been instrumental in the development of non-

immune directed evolution as an engineering technique for producing specific binding proteins. In addition to selection for binding, aspects of the scaffold itself can be evolved for other properties (Worn and Pluckthun 2001), such as thermostability (Ueda *et al.* 2004). Critically, once the immune system was removed as the limiting factor, so too was the reliance on antibodies and the Ig fold as the sole scaffold, and over the last twenty years the rapid increase in available structural information has provided a fertile source of potential scaffolds for investigation.

1.2.1 *Constrained Peptide Carriers*

Perhaps the simplest molecular recognition scaffold is a rigid independently folding domain displaying a single peptide loop. This expression context compared to a free peptide allows for some protection against proteolytic attack and imposes a degree of conformational restriction, reducing entropic penalties by loss of degrees of freedom on binding (Oneil *et al.* 1992). Phage display of oligopeptide libraries was used to identify antibody epitopes and ligand mimics for specific targets (Smith 1985; Cortese *et al.* 1994; Burritt *et al.* 1996) and this technology was adapted to generate new enzyme-specific inhibitors, using a Kunitz domain displayed as a stable protein scaffold to support a selected peptide (Markland *et al.* 1991). A Kunitz domain is a small, disulfide-stabilised fold, commonly found in protease inhibitors where an extended loop is inserted in the active site of the target protease (Marquart *et al.* 1983). To alter enzyme specificity, target-biased mutations were introduced into the active loop of bovine pancreatic trypsin inhibitor and selected for inhibition of human neutrophil elastase (Roberts *et al.* 1992a; Roberts *et al.* 1992b). Similar experiments have used Kunitz domain scaffolds based on Alzheimer's amyloid β -precursor inhibitor (Dennis and Lazarus 1994a; Dennis and Lazarus 1994b), human lipoprotein-associated coagulation inhibitor (Markland *et al.* 1996a; Markland *et al.* 1996b) and human pancreatic secretory trypsin inhibitor (Rottgen and Collins 1995) by introducing much broader artificial diversity in one or more surface loops.

Other scaffold proteins have been adapted to display constrained, functional peptides, including staphylococcal nuclease as an application-

specific inhibitor for dissection of biological pathways (Norman *et al.* 1999), tendamistat (McConnell and Hoess 1995) and the broad-spectrum protease inhibitor ecotin (Wang *et al.* 1995).

1.2.2 *Ig-like domains as scaffolds*

The same molecular library techniques which expanded antibody research so considerably also allowed other domains to be examined for potential as useful binding reagents. The goal was not necessarily to produce a direct replacement for antibodies, but to supplement the complement of antibodies with constructs that could be produced in greater quantities at lower cost and did not rely on disulfide bonds for stability, but were able to replicate binding at least as well as an antibody. In what could be seen as a natural progression, proteins with Ig-like domains were investigated as scaffolds in an effort to overcome some of the limitations of the antibody context.

Human Cytotoxic T-Lymphocyte associated antigen 4 (CTLA-4) was investigated as a peptide carrier due to its monomeric nature and close homology to the human antibody variable domain (Nuttall *et al.* 1999). A library utilising a single CDR-like loop was generated, focusing on integrin specificity with moderate success (Hufton *et al.* 2000). In a similarly targeted project, CDR loops of the T-cell receptor (TCR) were randomised to gain insight into TCR-peptide-MHC complexes (Holler *et al.* 2001).

1.2.2.1 **Fibronectin type III domain**

Fibronectin is a large, multi-domain serum protein involved in extracellular matrix adhesion and wound healing. It consists of repeating modules of independently folding domains of three types, of which type III (FN3) is a monomeric Ig-like β -sandwich (Baron *et al.* 1991). While the topology of FN3 domains is very similar to Ig domains found in antibodies, with exposed loops analogous to CDR loops, it has two fewer β -strands and does not contain an intra-molecular disulfide bond. Based on the diversity of observed function and the tangible advantages over classical antibody libraries, the tenth FN3 domain from fibronectin (Figure 1.2) was investigated by Koide and co-workers as a scaffold for the presentation of structurally constrained peptides in CDR-analogous loops (Koide *et al.* 1998).

The authors constructed a synthetic, codon-optimised gene for expression and display testing, on which was based the subsequent generation of the library for selection. Diversity was introduced into two of the three CDR-like loops, identified by sequence alignment, using oligonucleotides with degenerate codons. From a transformed library with $\sim 10^8$ members, binders to the model target human ubiquitin were enriched by phage display, with solubility and stability profiles worse or no better than the template domain. Affinity was not quantitatively determined, but competition ELISA suggested a K_D of 5 μM . Several strategies were employed to improve affinity. For example, classical antibodies increase binding by an increase in avidity. This principle was replicated with a non-covalent pentameric construct with affinity for integrins, oligomerised via a so-called assembly domain which improved apparent affinity from 2 μM to 10 nM (Duan *et al.* 2007).

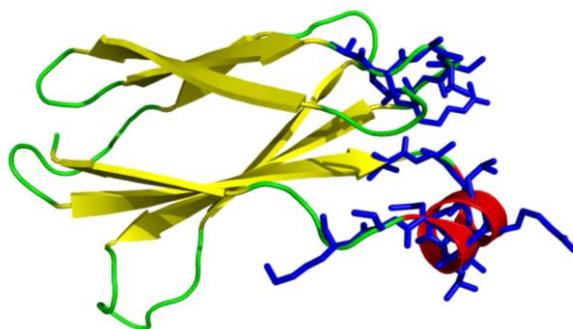


Figure 1.2 Engineered FN3 domain

The Fibronectin type 3 domain, shown here as a cartoon diagram, is an Ig-like fold, with two fewer β -strands. At one end of the domain are two loops analogous to CDR loops found in antibody Fv regions. The CDR-like loop residues are highlighted with blue stick diagrams. Structure coordinates are from PDB accession 3K2M.

The FN3 domain has been further developed as a scaffold and labelled monobodies, with a highest reported affinity of 1 pM (Hackel *et al.* 2008). Studies have shown the potential of the loops available for mutation (Koide *et al.* 2001) and basic improvements of the template protein have been made for increased stability (Batori *et al.* 2002). As stability of the FN3 domains is not dependent on the redox state of the solvent, they can be produced as active intracellular species. For example, monobodies were used to probe conformation state variation in the estrogen receptor bound to different ligands

by yeast 2-hybrid, where receptors in an unbound state and in complex with two different ligands were distinguished (Koide *et al.* 2002).

1.3 Non-Immunoglobulin-like Scaffolds

Even as technology improves to produce better antibodies, it has become apparent that antibody-based (and Ig-like) tools may not be the best option in every scenario. For example, reliance of Ig domains on disulfide bonds for stability limits their utility as intracellular reagents, due to the reducing environment in the cytoplasm, although examples of functional intracellular antibodies (“intrabodies”) have been produced (Worn *et al.* 2000) and an ongoing research effort to improve intrabody performance has met with considerable success (Stocks 2004). Also, the CDR-loop format of the binding region probably represents an incomplete sample of possible binding interfaces, and may be a sub-optimal choice for a custom binder, depending on the target ligand and the projected circumstances for use (Hudson and Souriau 2003). As a result, alternative scaffolds drawn from other fold families have been investigated for their utility as a tool for general molecular recognition duties.

1.3.1 *Identification of a scaffold candidate*

It is not always possible to predict with any accuracy the effect of mutations on a given protein, and therefore its utility as a scaffold, as the sequence-fold link has yet to be fully elucidated. Ig-like scaffolds share an advantage in that CDR-like loops may be readily identified as potential sites for mutation by comparison with antibody domains. However, when widening the scope to non-immunoglobulin domains, a new paradigm for introducing diversity must be considered, particularly when dealing with secondary structure-based interaction surfaces.

Skerra (2000) identified proxy characteristics that a protein fold family should possess as a minimum to be considered a viable scaffold for engineering as a generic molecular recognition template: A structurally super-imposable hydrophobic core across the fold family, where a sufficiently densely-packed core will contribute enough to the free energy of folding that the exterior can be mutated almost at will and still fold correctly; a solvent-exposed pocket or

face, ideally spatially separated from the hydrophobic core to minimise disruption from mutations. In addition the fold should have roles in biochemically diverse contexts, ideally using the same section of the fold to effect the function.

These three core characteristics can be considered basic requirements, with any additional properties application specific, such as low antigenicity and very small size in a therapeutic protein, or thermostability and pH tolerance in a research reagent. A later review outlined further ideal characteristics, dependant on the intended function (Binz and Pluckthun 2005). A relatively small, robust scaffold would possess increased diffusivity, decreased potential for antigenicity and reduced chances of non-specific interactions. Single chain scaffolds have obvious benefits over multi-domain constructs in relative ease of engineering and handling. Single domain scaffolds are even better, especially with specifically selected fusion partners for effector or detection functions. A universally useful fold should also lack disulfide bonds to allow for expression and activity under reducing conditions, facilitating efficient bacterial production (Nord *et al.* 1997) and potential intracellular expression.

In recent years a number of candidate domains have been investigated as template for use as molecular recognition scaffolds, with varying degrees of success. For a comprehensive review, see (Binz *et al.* 2005). Examples include ankyrin repeat domains (Binz *et al.* 2003; Binz *et al.* 2004), the Z domain from staphylococcal protein A (Nord *et al.* 1995; Nord *et al.* 1997), the lipocalin fold (Beste *et al.* 1999), zinc finger repeats (Rebar and Pabo 1994), a carbohydrate binding domain (Gunnarsson *et al.* 2004; Gunnarsson *et al.* 2006a; Gunnarsson *et al.* 2006b), armadillo repeat proteins (Parmeggiani *et al.* 2008), neocarzinostatin (Heyd *et al.* 2003) and an SH3 domain (Mouratou *et al.* 2007; Krehenbrink *et al.* 2008). It should be noted that the last example mentioned, the SH3 domain, was initially incorrectly identified as an OB-fold. The domain in question is classified by SCOP as an SH3 domain, and it was referred to as such in a recent review of scaffold engineering (Gebauer and Skerra 2009).

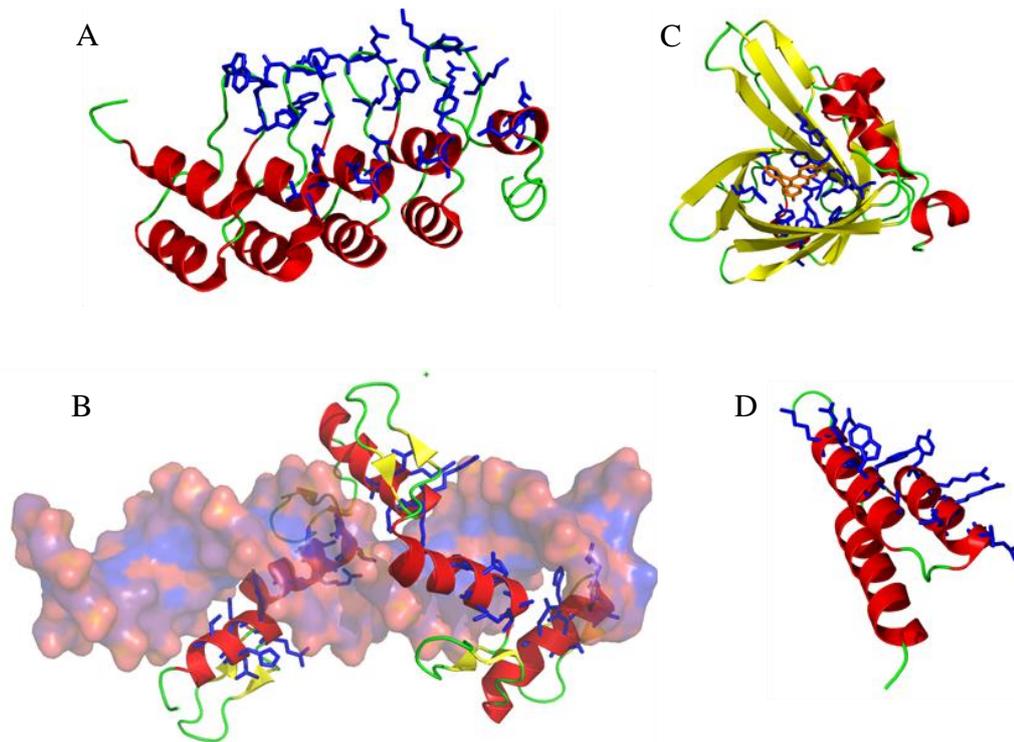


Figure 1.3 Engineered Non-Ig Domains

Four of the most successful non-immunoglobulin scaffold domains engineered for specific binding, shown as cartoons. Residues mutated in the library producing the particular example shown are coloured dark blue. DARPin domains (A) utilise a combination of varied residues (primarily in one loop) and a shuffled ankyrin repeat format to generate a protein binding surface. This example has 3 repeat modules and two capping modules (Schweizer *et al.* 2007). Zinc finger repeats (b) use a similar process to produce a DNA sequence-specific binder (bound DNA shown as surface representation), but the varied residues are located on the helix of each zinc finger. The structure shown actually has six domains, but one is disordered (Segal *et al.* 2006). Anticalins (c) mount functional residues to form a hapten binding pocket, in this case with fluorescein (orange) bound (Korndorfer *et al.* 2003), while Affibodies (d) utilise the combined surface of two helices to bind protein targets (Eigenbrot *et al.*).

1.3.2 Anticalins

Whereas most other examples of engineered proteins are searching for binding to various biological macromolecules, anticalins are specialised for high-affinity hapten binding (Skerra 2001). The domain is based on the lipocalin fold, which is part of a well-conserved structural superfamily with little sequence conservation (Flower 1996). The fold itself is an eight-stranded β -barrel with a deep hapten binding pocket (Figure 1.3C). Family members are primarily involved in small molecule transport and sequestration, such as fatty acids (Young *et al.* 1994a) and retinol (Cowan *et al.* 1990), with one member forming part of the human complement system (Ortlund *et al.* 2002).

In the first libraries, the authors randomised a series of loops at the binding site of the template gene (a bacterial bilin receptor), which form a pocket at one end of the β -barrel. Phage display selection yielded variants with mid nM-range affinity for model targets fluorescein (Beste *et al.* 1999) and digoxigenin (Schlehuber *et al.* 2000). In addition to the binding pocket, surface loops were also varied and selected, generating a molecule with specificity for two separate ligands (Schlehuber and Skerra 2001) and the potential as a drug-delivery tool was immediately recognised (Schlehuber and Skerra 2005). Accordingly, a human scaffold was developed along the same lines, based on apolipoprotein D, and successfully panned against a model protein target, with a resulting K_D of $\sim 2 \mu\text{M}$ (Vogt and Skerra 2004).

Because of their demonstrated ability to bind simultaneously both small molecule ligands and protein ligands, engineered anticalins have been proposed as drug-focusing vehicles and toxin scavengers, as well as the more traditional binding-mediated receptor modulation (Skerra 2007b).

1.3.3 *Affibodies*

Staphylococcal protein A (SPA) is a surface receptor which binds to the Fc region of IgG. A synthetic version of the Z domain from SPA was used as a template scaffold for production of specific binding proteins, termed an affibody (Nord *et al.* 1995; Nord *et al.* 1997). The domain is very small, with only 58 residues, forming a closely packed three helix bundle with the Fc binding region on the surface of two of the helices (Figure 1.3D).

In contrast to the CDR loops of Ig and Ig-like domains the alpha helical nature of the binding face provides a much larger degree of conformational rigidity. In principle this reduces degrees of freedom entropy loss on binding compared to a restrained loop, but this benefit may be partially offset by limitations of residue choice due to negative effects on domain stability. First generation affibodies from a naïve phage library showed affinity for a protein target of $\sim 2 \mu\text{M}$ (Nord *et al.* 1997).

In an analogous process to antibody maturation, a second generation affibody domain was affinity-matured by selective re-randomisation of approximately half of the binding face, resulting in a 100-fold improvement in affinity for its ligand (Gunneriusson *et al.* 1999). A lesser though still

significant improvement was gained by dimerisation of selected domains (Gunneriusson *et al.* 1999; Steffen *et al.* 2005). The highest reported affinity achieved using affibodies is 22 pM, to the HER2 cancer marker (Orlova *et al.* 2006). Affibodies have been successfully deployed in affinity-dependent applications, such as tumour imaging and targeting (Orlova *et al.* 2006; Steffen *et al.* 2006; Engfeldt *et al.* 2007; Magnusson *et al.* 2007; Tolmachev *et al.* 2007b; Vernet *et al.* 2008) and protein purification or analysis (Nord *et al.* 2000) (Eklund *et al.* 2002; Eklund *et al.* 2004; Renberg *et al.* 2007).

1.3.4 **Repeat Proteins**

A recent development in protein engineering for molecular recognition has been the use of protein repeats as a modular architecture for entirely synthetic scaffolds (Forrer *et al.* 2003), where the interaction surface composition as well as size can be varied. Scaffolds based on ankyrin repeats domains (Binz *et al.* 2003) and leucine-rich repeats (Stumpp *et al.* 2003) have proved remarkably successful in terms of achieved protein binding affinity and stability.

1.3.4.1 **DARPinS**

Designed ankyrin repeat proteins (DARPinS) were constructed from modules of helix-turn-helix consensus designed repeats, with variable residues located on the β -turn loops between the helices (Binz *et al.* 2003). Each domain contains between two and six such modules, with N and C-terminal capping modules (Figure 1.3A). Combinatorial libraries of DARPin domains were constructed by trinucleotide randomisation of six residues on the surface then random ligation into an expression vector, giving variation in repeat residue complement and repeat order. DARPinS have proved a highly stable platform, with extensive hydrogen bonding evident between surface residues, including those in the variable β -turns (Kohl *et al.* 2003). Specific binders were selected by ribosome display to nM-range affinities to protein targets (Binz *et al.* 2004). Selected DARPinS have since been demonstrated in practical uses, such as co-crystallisation (Warke and Momany 2007; Bandejas *et al.* 2008; Grubisha *et al.* 2010), enzyme inhibitors (Amstutz *et al.* 2005; Kohl *et al.* 2005) and as potential therapeutics (Zahnd *et al.* 2007), with affinities in the low nM to pM range.

1.3.4.2 Zinc Finger Domains

Zinc finger domains are metal-stabilised small modular units of approximately 30 residues which represent the most common DNA binding motif identified (Jacobs 1992). They were investigated as scaffolds for specific nucleic acid binding by consensus-driven designation of critical structural residues, based on the concept that all zinc fingers share a common binding mode in the major groove of the DNA double helix ((Jacobs 1992), Figure 1.3B). Using modestly randomised constructs of two or three repeats, binders were generated by phage display for both DNA (Rebar and Pabo 1994) and RNA (Friesen and Darby 1998) sequences with affinities in the nM range. Subsequently it was shown that high-affinity binding (19 nM) could be achieved with only a single zinc finger (Friesen and Darby 2001), making it amongst the smallest engineered domains currently known.

1.4 The OB-fold as a Scaffold

The OB-fold is a 5-stranded β -barrel domain arranged in a Greek-key motif (Murzin 1993). It commonly presents an external concave binding face mounted directly on the β -sheet of the barrel and in most cases where the domain is present the same face is used for binding. A survey of the SCOP database displays many OB-folds which are heavily modified with additional loops or entire new domains inserted (Murzin *et al.* 1995). Examples of OB-folds can be found in diverse organisms, including archea, yeast and mammals, with no detectable sequence conservation across the superfamily. They boast a diverse range of natural ligands, including proteins, oligonucleotides and oligosaccharides (Figure 1.4). Affinity data on OB-folds for their natural ligands is sparse, but shows nM-range or better (Theobald *et al.* 2003). These combined factors led to the suggestion that the OB-fold is ancient and tolerant to mutation, with an easily adaptable binding face (Murzin 1993). More recently, hydrogen exchange studies have shown that an OB-fold can be correctly folded by multiple pathways, suggesting that the fold is not reliant on certain seed folding nuclei (Watson *et al.* 2007). Rather, correct folding comes from a “diffuse network of interactions” making up the hydrophobic core, which are distributed throughout the primary structure, stabilising the entire fold.

Ultimately the unifying feature of the OB-fold is topology, with architecture capable of supporting a very wide range of sequences and modifications while still folding correctly. Examples of the diversity of OB-fold proteins include ssDNA-binding in the oncogene BRCA2 (Yang *et al.* 2002), anticodon recognition in aspartyl- and lysyl-tRNA synthetases (Commans *et al.* 1995; Schmitt *et al.* 1998; Rees *et al.* 2000; Moulinier *et al.* 2001), telomere end binding domain from the yeast protein Cdc13 (Mitton-Fry *et al.* 2002; Mitton-Fry *et al.* 2004; Wuttke *et al.* 2004) and the cell-surface oligosaccharide binding domain of the shiga toxin from *Shigella dysenteriae*, as well as related AB5 toxins (Fraser *et al.* 1994; Stein *et al.* 1994; Zhang *et al.* 1995; Fraser *et al.* 2004). Nucleic acid/OB-fold interactions have been the subject of a recent review (Theobald *et al.* 2003). OB-fold domains also mediate protein-protein interactions in superantigens during bacterial attack on the human immune system (Arcus *et al.* 2000; Arcus *et al.* 2002). These observations suggest that the versatility of the OB-fold might be emulated *in vitro* through protein engineering for tailor-made molecular recognition, potentially involving protein, nucleic acid and oligosaccharide targets (Arcus 2002). In addition, the β -sheet mounted character of the binding face represents relatively unexplored territory in engineered protein-protein interactions, with most examples using constrained loops and helical residues.

The central aim of the research for this thesis is to develop the OB-fold as a specific, high-affinity binding reagent by selection from a combinatorial library by phage display, for which we have chosen the term “Obody”. Based on previous proof-of-principle work, the research reported in this thesis shows affinity maturation for a model target by both rational and random methods, biophysical characterisation of binding and the engineered domains themselves, and structural analysis of an Obody during the affinity maturation process to reveal the nature of the binding faces developed.

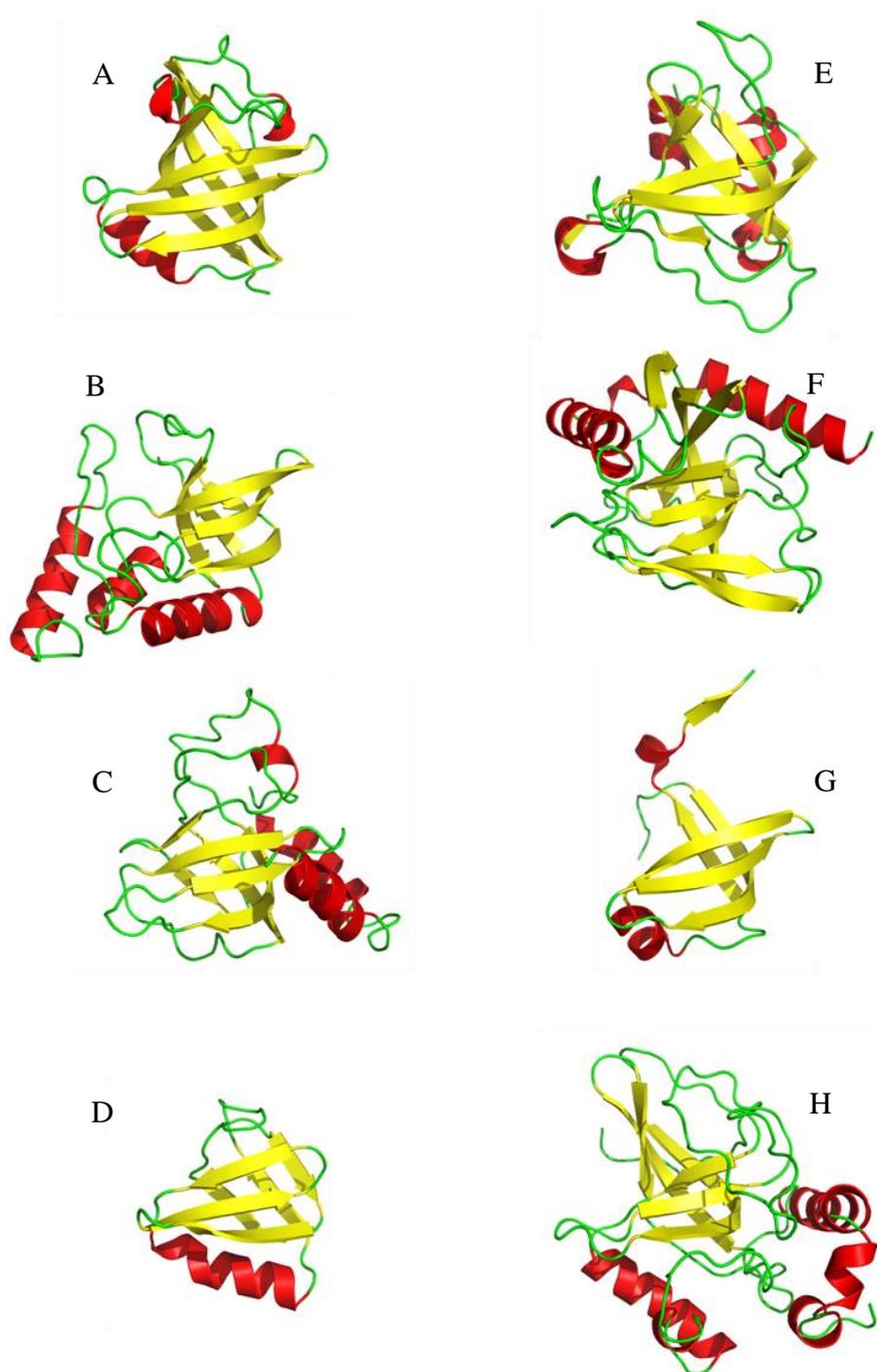


Figure 1.4 OB-fold diversity

All structures are oriented so that the first three β -strands of the barrel are visible to the front. Although they display wide variation in ancillary additions, including helix insertions and additional β -strands, the core five-stranded fold is clearly recognisable despite no detectable sequence conservation. These examples exhibit diversity in biochemical function, as well as wide occurrence of OB-folds in evolutionarily diverse organisms; (a) anticodon-binding domain from *Escherichia coli* aspartyl tRNA synthetase (Moulinier *et al.* 2001), (b) OB-fold from staphylococcal nuclease (Hynes and Fox 1991), (c) oncogene BRCA2 (Yang *et al.* 2002), (d) OB-fold “B” subunit from shiga toxin (Fraser *et al.* 1994), (e) human TIMP-1 inhibitory domain (Iyer *et al.* 2007), (f) domain from *Thermus thermophilus* inorganic pyrophosphatase (Teplyakov *et al.* 1994), (g) a homohexameric molybdate binding protein from *Sporomusa ovate* (Wagner *et al.* 2000). (h) DNA-binding domain from yeast *cdc13* (Mitton-Fry *et al.* 2004).

1.5 Directed Evolution Technology

Rational approaches to protein engineering have proved successful, especially where detailed structural information is available. For example, critical residues for folding and stability have been determined by structural consensus in a limited number of well known examples, most pertinently the Ig fold (Knappik *et al.* 2000) and the ankyrin repeat domain (Mosavi *et al.* 2002). Semi-rational approaches have also been used, for example loop exchange between enzymes of the same fold to introduce targeted variation (Park *et al.* 2006; Ochoa-Leyva *et al.* 2009).

However, a rational or semi-rational approach cannot always be applied. In a situation where information about the target ligand or substrate is limited, rational choices for design of a binding region are not always available. To deal with the lack of target data, an evolutionary screening process has been adopted. Developments in combinatorial genetics and selection techniques have allowed researchers to efficiently manipulate genes of interest towards a particular goal by utilising the evolutionary principles of selection from a diverse population, to increase the frequency of an “allele” with desirable properties. Examples include new and modified enzymes (Arnold and Volkov 1999; Hibbert *et al.* 2005; Park *et al.* 2006; Seelig and Szostak 2007), antibodies and antibody fragments (Huse *et al.* 1989; Gram *et al.* 1992; Griffiths *et al.* 1994), synthetic domains (Braisted and Wells 1996; Knappik *et al.* 2000; Binz *et al.* 2003) and *de novo* proteins (Keefe and Szostak 2001; Chaput and Szostak 2004). Typically these processes use a combination of rational design of a template scaffold and combinatorial screening of specifically or randomly mutated variants for a particular function, hence the term “directed” evolution.

Discovery of binding to an arbitrary target where structural or functional information isn't available requires a great deal of diversity from which to select in order to be successful. The vertebrate immune system was the immediately available source of selectable diversity, so in conjunction with hybridoma technology (Kohler and Milstein 1975), monoclonal custom antibodies became the gold standard for specific detection. While a great deal of success was garnered this way, limitations in the toxicity and

immunogenicity of targets drove development of alternative methods for selection from diversity. Genotype-phenotype linking techniques, such as yeast display (Boder and Wittrup 1997), bacterial display (Charbit *et al.* 1986; Charbit *et al.* 1988) and especially phage display (Smith 1985; Scott and Smith 1990) allowed selection of combinatorial libraries of peptide antigens, and later, antibodies (Figini *et al.* 1994; Nissim *et al.* 1994; Winter *et al.* 1994) outside of the immune system.

Three factors must be present in gene library design and selection for affinity. First, diversity must be introduced in a manner that does not destroy the template fold, but creates a randomised, solvent-exposed face, pocket or loop. Second, phenotypic selection requires a physical linkage between the phenotype (folded protein) and genotype. Lastly, because of the large numbers of unique sequences in gene libraries (up to 10^{13} in the largest libraries), successfully selected genes must be able to be amplified while compartmentalised from the rest of the library. This preserves the phenotype-genotype link across multiple rounds of selection which promotes the chances of isolation of high-affinity binder from the remainder of the library.

1.5.1 *Diversity Generation*

The introduction of diversity into a domain for future selection is affected by several competing factors, such as the nature of the binding region, potential targets and library size limits. The optimum strategy is not obvious when considering a new domain. A variety of methods have been developed for efficiently mutating loop regions, but relatively few deal with secondary structure-based regions. *In vitro* recombination, termed DNA shuffling or sexual PCR (Stemmer 1994; Coco *et al.* 2001; Stemmer 2001) has been used extensively to generate chimera libraries from two or more related genes. This approach, inspired by the recombination that takes place in lymphocyte diversification, has been subject to various refinements such as the staggered extension process (Zhao *et al.* 1998), and degenerate oligonucleotide shuffling (Gibbs *et al.* 2001).

While DNA shuffling works well where a family of related genes are available, as found in enzyme families, introduction of site-specific naïve diversity to form a defined binding face or pocket is better done by other

means. Two main methods are employed to produce diverse libraries with site-specific mutant residues: PCR incorporation of oligonucleotides with randomised codons into a synthetic gene (Derbyshire *et al.* 1986), or whole gene synthesis of as many different variants as practically possible. Consideration must also be given to the possible residues at each selected position, where, for example, an inappropriately placed proline may be detrimental to fold stability or affinity. Similarly, cysteines and stop codons are usually excluded, if possible, by different combinations of possible nucleotides at each position in the codon. The recent development of codon-based oligonucleotide synthesis (Yanez *et al.* 2004) has allowed researchers to specify a unique codon subset at each mutational position, but this method remains expensive.

Perhaps the single most common method for introducing diversity is the very simple error prone PCR method (Cadwell and Joyce 1992), where a non-proof reading polymerase enzyme is induced into a higher error rate than normal during amplification. Errors are promoted by the addition of Mn^{2+} to reduce enzyme specificity, and spiking the reaction with disproportionate amounts of two of the four nucleotides, promoting miss-incorporation. This results in a distribution of mutation rates that can be difficult to calculate accurately, but can be modelled as a stochastic distribution of single nucleotide mutations (Moore and Maranas 2000; Pritchard *et al.* 2005). A gene mutated in this manner covers a broad, sparse sampling of sequence space, making it ideal for maturation of an existing function or biophysical characteristic.

1.5.2 *In vivo selection methods*

To ensure that the genotype-phenotype linkage is stable across several generations, the gene must be compartmentalised in some manner. The immune system achieves this by storing immune diversity in dedicated cell lines, expressing antibodies and related proteins on the cell surface for selection. This process has been replicated with alternative hosts such as bacteria (Freudl 1989; Skerra 1993) and yeast (Boder and Wittrup 1997). *In vivo* display methods are separated into two broad categories: cell surface display, where a peptide or protein library is displayed as the fusion of the extracellular domain of a membrane protein and the cell is the selectable unit;

and phage display, where a secreted phage particle bearing a fusion protein is the selectable unit.

1.5.2.1 Phage Display

Phage display has been used in phenotypic selection strategies for 30 years, starting with peptide display for identifying epitopes (Smith 1985; Scott and Smith 1990), then expanding into whole proteins. Most pertinent to this study are antibodies (Winter *et al.* 1994), as well as new scaffolds, such as monodies (Koide *et al.* 1998) and affibodies (Nord *et al.* 1997). Phage libraries have been built, for various purposes, using lambda (Huse *et al.* 1989), M13 (Sidhu 2000) and T7 (Dai *et al.* 2008) phage.

Generally, a gene library is inserted into an expression plasmid modified to contain a phage packaging signal and origin, or “phagemid”. Individual library members are expressed as the fusion product of a phage coat protein, engineered so that the introduced protein is solvent exposed. After transformation, the bacterial cells are induced for fusion expression and infected with a “helper” phage to begin production of transducing particles (TDP); that is, phage containing a copy of the phagemid instead of the phage genome. The result is, ideally, a diverse population of TDPs displaying at least one fusion product, with the protein-encoding gene in the encapsulated plasmid.

M13 filamentous phage are stable and can easily be produced in large quantities with the appropriate bacterial host. Samples can remain almost 100% infectious after months in solution at 4°C, are highly resistant to proteolytic attack and can withstand a wide pH range without losing viability (Barbas III *et al.* 2001). Compared to the *in vitro* technologies, phage display imposes relatively little set up cost in terms of specialised equipment or expensive reagents. However, there are two major limitations. Firstly, the upper limit of the naïve gene library is determined by successful transformation rates. Consequently, library sizes larger than 10^9 - 10^{10} are possible, but rarely achieved. Secondly, because of the highly robust nature of M13 phage, cross-contamination issues are a common hazard when working with multiple libraries.

A number of different formats for selection using filamentous phage display have been investigated, the most important of which are pIII and pVIII fusions. Major coat protein pVIII is the most abundant in M13, forming the bulk of the protein coat (Armstrong *et al.* 1981; Grant *et al.* 1981). Thus a library fused to it can expect many copies of the fusion protein displayed on each phage, greatly increasing avidity-driven selection pressure and allowing variants with low affinities to be discovered (Sidhu *et al.* 2000). However, this approach is greatly limited by phage assembly processes, and has recently fallen out of favour.

Minor coat protein pIII, responsible for infection, is present in only low numbers, from three to five per particle (Grant *et al.* 1981), so selection pressure is geared more towards affinity of individual variants compared to pVIII display systems. Display of phagemid-encoded pIII fusion protein can be promoted by deletion of the native pIII gene in the rescuing helper phage, ensuring that all copies present come only from the phagemid (Rakonjac *et al.* 1997) (Rondot *et al.* 2001). However even at that level, multivalent display is thought to allow retention of moderate-affinity variants, even where higher-affinity, but lower avidity, variants are present (Cwirla *et al.* 1990).

“Monovalent” pIII fusion libraries, so called because the library fusion protein must compete with native pIII derived from the helper phage for inclusion into phage particles, result in a mixed population from each cell, containing on average one or less fusion pIII proteins per TDP (Lowman *et al.* 1991). The expectation is that a lower display count will increase the chances of high affinity variant being discovered due to increased emphasis on selection for affinity of individual domains.

As pIII is responsible for infection of a new host cell, infectivity has been used as a selection filter with pIII fusion libraries, independent of, or in addition to, the primary selection for binding or other property. As infectivity is dependent on the close proximity of the three pIII domains to each other, efficient selection can be produced by linking protein stability with pIII domain association. This has been shown by “rescue” with a separate construct containing the required pIII N-terminal domains and a binding site, to which those library variants remaining folded after a thermal challenge can bind

(Spada *et al.* 1997). The same effect has been observed by cloning the gene library between pIII domains and selecting with proteolysis, where proteolytically sensitive variants will lose infectivity (Kristensen and Winter 1998; Sieber *et al.* 1998). Using the same principle, selective infection approaches have been shown to enhance selection for binding in monovalent phage display. Introduction of a helper phage with trypsin-sensitive pIII (Kristensen and Winter 1998) and treatment with trypsin following affinity-based selection of fusion reduces background by biasing reinfection to those particles displaying a phagemid-derived (and therefore trypsin resistant) pIII (Goletz *et al.* 2002).

1.5.2.2 Cell-surface display

Similar to phage display, cell-surface display of heterologous proteins and peptides relies on expression as a chimeric “receptor” with the extracellular domain of a membrane protein, primarily in bacteria but also in yeast. The methodology was first conceived as a tool for probing membrane protein arrangements (Charbit *et al.* 1986), and to look for antigenic epitope peptides (Agterberg *et al.* 1987; Charbit *et al.* 1988), but has been expanded into whole proteins, including scFv libraries (Fuchs *et al.* 1991; Francisco *et al.* 1993a; Gunneriusson *et al.* 1996) and enzymes (Francisco *et al.* 1993b). Perhaps the most popular use for bacterial display is in development of live vaccines, where the bacteria itself acts as an adjunct to promote activation of an immune response to a displayed antigen (Nguyen *et al.* 1993; Stover *et al.* 1993; Nguyen *et al.* 1995; Georgiou *et al.* 1997). Maximum library sizes are small compared to phage display (10^5 - 10^6 , depending on the system used), and affinity isolation of cells is generally more problematic. The attraction of cell-surface display seems to lie in the great range of options available, in terms of surface-exposed proteins for insertion of a gene library. Cell-surface display has been the subject of a recent review (Jostock and Dubel 2005).

1.5.3 *In vitro selection methods*

1.5.3.1 **Ribosome Display**

Ribosome display was developed as an entirely cell-free method for evolutionary selection, to avoid transformation and the bottleneck it imposes on maximum diversity. First developed by Mattheakis et al (1994) to screen for peptide epitopes, the authors claimed library sizes of 10^{12} members. Hanes and co-workers showed its use for the display of antibody fragments (Hanes and Pluckthun 1997; Schaffitzel *et al.* 1999). The technique involves cell free expression, where the ribosomes are stalled on messenger RNA (mRNA) library transcripts by the absence of a stop codon to prevent peptide release and a combination of a sudden drop in temperature with the addition of chloramphenicol to stabilise the complex. Thus, the mRNA transcript library, translating ribosomes and unreleased polypeptides form a complex (or “polysome”) which can be selected for affinity for a new ligand. After selection the mRNA is recovered and amplified by reverse transcription PCR. The manual style of amplification between rounds lends itself to efficient, controlled introduction of small numbers of mutations without the necessity of building a new library from scratch. While ribosome display is a very powerful technique and has been used to select various binding proteins (in some cases to pM-range affinities), including antibodies (Rothe *et al.* 2006) and DARPins (Binz *et al.* 2004), it remains technically very challenging, involving multiple steps of reverse-transcription, and handling of large quantities of RNA.

1.5.3.2 **mRNA/DNA display**

These display technologies, like ribosome display, removed cells as the method of compartmentalisation. Here the nascent polypeptide is usually covalently coupled to the nucleic acid it was translated from, forming an *in vitro* virion suitable for phenotypic selection, as opposed to the much larger non-covalent polysome of ribosome display.

mRNA display was developed first (Nemoto *et al.* 1997; Roberts and Szostak 1997). The model library was built and modified for translation and selection, starting with a 3` puromycin and a DNA oligo linker. Synthetic mRNA containing the gene of interest and a 5` untranslated region (UTR) was then ligated to complete the “virion” genome. To produce the protein encoded

by each mRNA, *in vitro* synthesis was used. Coupling of the peptide to its template was achieved by the ribosome stalling as it reached the DNA portion of the template, allowing the 3` puromycin, a tRNA-mimic antibiotic, to enter the ribosomal A site and form an amide bond with the C-terminal residue of the peptide.

DNA display was developed to overcome the inherent susceptibility to enzymatic degradation that comes with the use of RNA as a template. Tabuchi et al (2001) used the same puromycin-dependant covalent link to the polypeptide. However, the puromycin was instead attached 5` to a ssDNA oligonucleotide that annealed to the 3` end of the mRNA library. After translation and formation of the puromycin-peptide amide bond, the oligo was extended to form a double stranded cDNA-peptide virion. Although the virion contained a dsDNA gene, the library used separately transcribed & labelled mRNA as the input into a cell-free expression system for production of the protein for display.

In both of the cases above, the covalent attachment mechanism also served as a compartmentalisation mechanism, preventing the template gene from separating from the translated polypeptide. Alternatively, emulsion droplet encapsulation in tandem with other linking methods allowed the translation complex to dissociate naturally before assembling into virions for selection. In this vein, Doi and co-workers used biotinylated DNA to express a streptavidin fusion product, making a high-affinity, non-covalent linkage (Doi and Yanagawa 1999; Yonezawa *et al.* 2003). Another method expressed the encoded gene as the fusion of Hae III methylase and coupled the DNA to a covalent bond-forming inhibitor (Bertschinger and Neri 2004). While conceivably either of these techniques could be used in conjunction with mRNA display, their major advantage lies in the more stable dsDNA format of the template.

Display and selection using *in vitro* virions has been used in selection of peptide aptamers (Wilson *et al.* 2001), FN3 binding domains (Xu *et al.* 2002; Olson *et al.* 2008; Liao *et al.* 2009) and *de novo* proteins (Keefe and Szostak 2001).

1.5.3.3 Other *in vitro* methods

Recently a number of other techniques have been developed for selection of proteins of interest from large, diverse gene libraries. Atomic force microscopy, developed as a method for visualising and quantifying molecular-scale interactions, has been adapted for screening libraries looking for functional DNA aptamers (Miyachi *et al.* 2010). In this case, no compartmentalisation is required, as the selection is based on binding by the DNA itself. Display and evolution using proteins attached to microbeads has also been used in engineering, where single beads are isolated in oil emulsions for *in vitro* synthesis for evolution of enzyme function (Tawfik and Griffiths 1998; Ghadessy *et al.* 2001). This technique was adapted for selection of binding proteins by linking successful binding to association with fluorescein, allowing discrimination by virtue of increased fluorescence, and sorted using a flow cytometer (Feldhaus *et al.* 2003).

1.6 Protein-Protein Interactions

Interfaces between proteins have been an area of intense research over decades. Virtually all proteins contact other proteins at some point in their life, ranging from transient signal transduction, to practically irreversible binding. Understanding regulatory networks is therefore reliant to some extent on comprehension of the nature of binding sites. Although the biophysics of affinity of one protein for its binding partner are not yet fully understood, it can be broken down to a combination of factors: electrostatic, hydrophobic, polar (including hydrogen bonds, van Der Waals interactions) and geometric.

Evolutionary conservation has been used as a method for finding important residues for binding (Lichtarge *et al.* 1996; Zhou and Shan 2001; Lichtarge and Sowa 2002; Ma *et al.* 2003) and structural conservation is a common component of prediction algorithms. In stark contrast, newly engineered domains have no equivalent evolutionary relationships in the sequence databases from which to draw conclusions. Consequently, computational analysis of an engineered domain in complex with its non-native ligand becomes an exercise in docking in reverse, where a binding site is known, but the determinants of binding are not. Based on known protein-

protein interaction properties, new interfaces can be computationally analysed to determine the structural basis of binding.

1.6.1 *Geometric complementarity and polar bonds*

Each potential interaction between binding partners is heavily influenced by the ability of the individual components to be presented in an orientation that is favourable, i.e. that the interaction does not require a residue to deviate from its allowable conformations. Residues which form a surface that is complemented closely in a binding partner are more likely to be binding hot spots (Li *et al.* 2004), and interface geometric complementarity is currently used as an evaluation of docking quality (Mitchell *et al.* 2001; Ban *et al.* 2006).

Electrostatic interactions and hydrogen bonds are particularly important in determining interaction specificity, as shown by evolutionary preservation of intermolecular bonds in the conservation of function (Xu *et al.* 1997b). Electrostatics are commonly accepted as the first long-range orientation filter for attraction between two potential partners (Schreiber and Fersht 1995) and are especially important in protein-DNA interfaces (Shanahan *et al.* 2004). While both these types of interaction can contribute to the overall affinity of a complex, any contributed decrease in the free energy of binding is offset to some degree by the desolvation effect, defined as the free energy increase due to unsatisfied polar bonds following occlusion of solvent during binding. Indeed, theoretical studies indicate that intermolecular electrostatic bonds are actually most often destabilising due to the high desolvation penalty (Novotny and Sharp 1992; Hendsch and Tidor 1994; Sheinerman and Honig 2002; Dong and Zhou 2006).

To some extent the pattern recognition function of hydrogen bond networks is similar in both intra- and inter-molecular interactions (Xu *et al.* 1997a). Where they differ is in the degree to which they can be optimised. Norel and co-workers (Norel *et al.* 1995; Norel *et al.* 1999) showed that geometric complementarity can be used successfully as the sole docking criteria for finding binding sites, even when using structures of monomers not solved in the presence of their ligand. This strongly implies that, at least in the cases examined, the binding surface shape is largely determined before binding, essentially restricting unsatisfied hydrogen bonds on the interface of a

folded structure to a rigid-body search. Consequently, bonds formed between proteins present a much broader range of angles and distances than found in other contexts (Lin *et al.* 1995; Xu *et al.* 1997b), impacting negatively on bond stability as it diverges from the theoretical ideal. There are cases where large conformational shifts can be seen on binding, most obviously in domain-swapping dimers or natively unfolded proteins (e.g. β -crystallin on calcium binding (Srivastava *et al.* 2010)), but it has been noted that antibody-antigen complexes exhibit comparatively little conformational changes caused by binding (Lo Conte *et al.* 1999), and that amongst the available degrees of freedom for a particular residue, a strongly binding but infrequently sampled conformational aspect contributes poorly to the free energy of binding (Gallicchio *et al.* 2010). These data support the idea that pre-arrangement of binding residues into favourable conformations is a general property of high affinity protein interactions.

1.6.2 *Hydrophobic interactions*

Hydrophobic interactions influence binding via exclusion of solvent from hydrophobic residues resulting in a free energy loss on binding. Patch theory of protein interactions posits that interfaces are influenced by hydrophobic complementarity, organised into discrete “patches” on the surface of each protein (Korn and Burnett 1991), surrounded by more polar residues that may remain solvent exposed on binding. Detection of these patches was proposed as a method for identification of possible interaction sites (Young *et al.* 1994b), with significant success (Jones and Thornton 1997). A general gain in atomic hydrophobicity was found to be correlated with an increased chance of finding particular residues in the core of binding patches, which tend to be enriched for hydrophobic and aromatic residues (Lo Conte *et al.* 1999). A survey of protein interfaces showed that interfaces of 2000 Å² or less tend to have only one hydrophobic patch, with each patch defined as a collection of atoms with an average hydrophobicity less than the protein surface average, surrounded by a rim of residues more representative of the surface average (Chakrabarti and Janin 2002). This conclusion was supported by a larger, more recent study (Yan *et al.* 2008).

1.6.3 *Hot spots*

It has become evident that the contributions to binding by individual residues involved at interfaces are not uniform, even within the same patch. Using alanine scanning mutagenesis, calculation of contributions to binding at a single residue resolution show disproportionate changes in the free energy of binding (Castro and Anderson 1996; Clackson *et al.* 1998). These positions, labelled hot spots, are critical for binding and are often highly conserved, aromatic and closely packed in the interface (Li *et al.* 2004).

Hot spot prediction by various methods based on structural data can provide useful information about an interface. Virtual alanine scanning gives a change in free energy difference ($\Delta\Delta G$) of binding estimate and can be used qualitatively to predict critical residues in a protein-ligand interface by the *in silico* mutation of residues to alanine (Massova and Kollman 1999). This method has been implemented in the Robetta web server, which combines computational alanine scanning with Rosetta structure prediction (Kim *et al.* 2004). Unsurprisingly, prediction of hot spot residues tends to become more accurate when different metrics are used in combination. For example, the KFC web server (Darnell *et al.* 2007) combines shape complementarity calculated by fast atomic density evaluation (FADE) (Mitchell *et al.* 2001) and satisfaction of biochemical bonds in conjunction with virtual alanine scanning (calculated by Robetta as a separate input) to make hot spot predictions. Using these three methods, the authors claim an accuracy rate of 72%. In an alternative model, change in accessible solvent area has also been included (Tuncbag *et al.* 2009; Tuncbag *et al.* 2010). A major drawback of most predictive computational models is that explicit solvent molecules are not included in the analysis. Although protein-protein interfaces are typically hydrophobic and mostly “dry”, a recent structure survey showed that water does play an important role in many interfaces (Rodier *et al.* 2005); waters buried in interfaces may not always contribute to binding directly, but may play a cooperative role in stability of individual monomers or residues (Reichmann *et al.* 2008). A study incorporating ordered solvent molecules in the prediction showed an incremental improvement in accuracy over less comprehensive benchmark models (Li and Li 2010).

1.6.4 *Thesis Outline*

This thesis, divided into six chapters, describes the development of an Obody, derived from a nucleic acid binding domain, engineered for affinity for hen egg white lysozyme (HEWL). Chapter two details the methods used in the subsequent chapters. Chapter three presents the first structure of an Obody, in complex with its ligand, and discusses the binding determinants. Chapter four describes the directed evolution by phage display of three focused and one naïve gene library based on the same domain, with a view to improve Obody binding to a model protein target. Chapter five presents two further crystal structures as products of two of gene libraries in chapter four, followed by a general discussion in chapter six.

2 Materials and Methods

2.1 General methods

2.1.1 *Clone Storage*

Isolated clones and naïve library stocks were stored either as transformed glycerol stocks indefinitely at -80°C, or as purified plasmid indefinitely at -20°C. Streaked clones on ampicillin-containing agar plates were stored at 4°C for a maximum of one month, then discarded.

2.1.2 *Primer Design*

Primers for dissection of the template gene and incorporation of mutagenic oligonucleotides were designed using Vector NTI Advance (Invitrogen, Carlsbad CA, U.S.A.). The primers were selected for position, with little room for optimisation of primer pairs. In some cases, the original template was modified to accommodate more efficient priming. A list of primers used in the course of this study is available in Appendix A1.3.

2.1.3 *Sequence Analysis and Alignment*

Sequence data were stored and manipulated using Geneious (Biomatters, Auckland, New Zealand) and Vector NTI Advance (Invitrogen).

2.1.4 *Common Buffers*

2.1.4.1 **Phosphate-buffered saline (PBS) pH 7.4**

10 mM Na₂HPO₄

1.76 mM KH₂PO₄

2.7 mM KCl

137 mM NaCl

The PBS recipe was taken from Sambrook and Russel (2001), prepared as a 10x stock and diluted as needed.

2.1.4.2 Wash buffer (PBS-T)

1x PBS pH 7.4

0.1% (v/v) Tween-20

2.2 DNA Manipulation and Analysis

2.2.1 PCR

All PCRs involved with library production and cloning were performed using Platinum Pfx DNA Polymerase (Invitrogen), using the standard recipe formulation as provided with the product (Enzyme at 1 U per 50 μ L reaction, 1.5 mM MgSO₄, dNTPs at 0.2 mM each, primers at 100 nM each), except that the provided 10x buffer was used routinely at 1.5x concentration in the reactions. Screening PCRs were performed using Taq DNA Polymerase (Invitrogen) according to the information provided with the product (Enzyme at 0.2 U per 50 μ L reaction, 1 mM Mg Cl₂, dNTPs at 0.2 mM each, primers at 100 nM each).

2.2.1.1 Nested PCR

Nested PCR was used to attach 3' and 5' attB recombination sites for Gateway cloning (section 2.2.4.4). The procedure was as follows: the template gene was amplified in a 25 μ L standard PCR reaction, with 100nM of each gene-specific primer (Oligos 155/156 were used for all Obody cloning, appendix A1.3) for 25 cycles with an annealing temperature of 55 °C. This first primer pair added a short (12 bp) linker at either end of the insert for the second step. A second reaction was then made, with 100 nM each of the generic gateway adapter primers (sequence adapted from (Moreland *et al.* 2005)), which anneal to the linkers added in the first step, with 1 μ L of unpurified reaction mix as the template. The reaction was cycled 25 times, with an annealing temperature of 45 °C in the first step, and 55 °C in the second. PCR product was purified by gel extraction.

2.2.2 *Agarose Gel Electrophoresis*

All DNA electrophoresis was performed using horizontal-format submerged gels, using Mini-sub Cell GT Cells. (Bio-Rad Laboratories Pty, California, U.S.A.)

2.2.2.1 **TAE 50x Stock, 1 L**

242 g	tris(hydroxymethyl)aminomethane (tris)
57.1 mL	glacial acetic acid
18.6 g	ethylenediamine tetra acetic acid (EDTA)

All agarose gels were run in 1x Tris-Acetate-EDTA (TAE)

2.2.2.2 **Gel Preparation**

Agarose gels varied between 0.8 and 2% (w/v) and were always prepared fresh in 1x TAE buffer. The agarose was melted with intermittent stirring in a microwave, then cooled to <50°C and poured into the caster. The gel was allowed at least 20 minutes to set then transferred, complete with casting tray, into the running apparatus. The gel was submerged in 1xTAE before loading samples and run at 100 V for 45 min

2.2.2.3 **DNA loading Dye 10x**

0.05% (w/v)	Bromothymol Blue
0.25% (w/v)	Xylene cyanol
30% (v/v)	Glycerol

2.2.2.4 **DNA Standards**

To estimate the size of double stranded DNA (dsDNA) fragments on an agarose gel, the 1KB Plus DNA ladder (Invitrogen) was used (Appendix A1.1.2). The commercial stock was diluted 1:10 in 1x TAE with 1x DNA loading dye and stored at -20°C until needed.

2.2.2.5 **DNA Detection**

Detection of DNA on an agarose gel used ethidium bromide or SYBr Safe (Invitrogen). Small-scale gels for confirmation of a correctly amplified PCR fragment were stained for ~10 min in a 0.1% (w/v) ethidium bromide solution and visualised with a UV transilluminator and a digital camera.

Preparative gels for purification of PCR products were run with 1x SYBr Safe in the gel and visualised on a Safe Imager blue light (470 nm) transilluminator (Invitrogen) for excision of DNA bands for purification.

2.2.3 DNA Purification

2.2.3.1 High Pure PCR Purification Kit

PCR products were purified either straight from the PCR mix or from excised gel bands using a High Pure PCR Product Purification Kit (Roche, Basel, Switzerland). The High Pure kit was also used to purify DNA following enzyme treatment. Binding buffer was 3 M guanidine-thiocyanate, 10 mM Tris pH 6.6, 5 % ethanol. Wash buffer was 80% (v/v) ethanol, 20 mM NaCl, 2 mM Tris pH 7.5. Elution buffer was 10 mM Tris pH 8.5.

For agarose gel extraction, the required band was excised with a clean scalpel and weighed in a 1.5 mL centrifuge tube. Binding buffer was added to the gel piece at 300 μ L per 100 mg agarose. The agarose was melted by shaking at 50°C. After melting, isopropan-2-ol (IPA) was added at 150 μ L per 100 mg agarose and mixed thoroughly. The solution was loaded on to a silica spin-column provided in the kit and centrifuged at maximum speed for 1 min. The column was washed with 500, then 200 μ L of wash buffer, taking care to keep the column dry after spinning. The DNA was eluted with 50 μ L of elution buffer into a clean micro centrifuge tube.

For purification without gel extraction, the method was the same, except for the initial binding conditions: Binding buffer was added at 500 μ L per 100 μ L of DNA solution and IPA at 250 μ L per 100 μ L.

2.2.3.2 “Freeze ‘n Squeeze” Gel Extraction

DNA fragments less than 70 bp were purified by Freeze and Squeeze, due to poor retention on the High Pure PCR Product columns. The required band was excised from an agarose gel with a clean scalpel, wrapped in a small square of parafilm and incubated at -80°C for 15 min. The frozen gel pieces were partially unwrapped and squeezed between thumb and finger, without touching the gel piece. The drops containing the DNA were collected by pipette as they emerged and transferred to a new 1.5 mL centrifuge tube.

2.2.3.3 Isopropyl Alcohol Precipitation

The sample containing DNA to be precipitated was acidified by the addition of 1:10 volume of 3 M sodium acetate pH 4.8. The DNA was precipitated by the addition of an equal volume of 100 % analytical grade IPA and incubation on ice for at least 1 hr. Precipitate was harvested at 13,000 g for 10 min at 4°C, washed three times with 70% (v/v) ethanol, dried at 37°C for 30 min and re-dissolved in 10 mM Tris pH 8.0.

2.2.4 DNA Cloning

2.2.4.1 Restriction Enzymes

Restriction enzyme digestion was carried out at 37°C for at least 1 hr, at a concentration of at least 1U enzyme per 1 µg DNA. After digestion, samples were purified using the High Pure PCR kit.

Table 2.1 Restriction enzymes used

The symbol ^ in the sequence denotes the position of the enzymatic cleavage. Note that the React buffers distributed with Invitrogen restriction enzymes have been discontinued.

Name	Cut site	Buffer
NcoI	5` C^CATGG	REact 3†
NotI	5` GC^GGCCGC	REact 3†
BamHI	5` G^GATCC	REact 3†
EcoRI	5` G^AATTC	REact 3†
BsrGI	5` T^GTACA	NEB 2‡

† Invitrogen	‡ New England Biolabs
50mM Tris-HCl, pH 8.0	10 mM Tris-HCl pH 7.4
10mM MgCl ₂	50 mM NaCl
100mM NaCl	10 mM MgCl ₂
	1 mM Dithiothreitol

2.2.4.2 Restriction/Ligation

Basic cloning was carried out using ligation of “sticky end” restricted inserts and vectors with T4 DNA ligase (Invitrogen). After restriction enzyme digestion, vectors were treated with shrimp alkaline phosphatase (SAP)(Roche). SAP removes 5` phosphates, reducing vector self-ligation.

Ligation reactions were performed according to the manufacturer's specifications; 30 fmol of vector, 90 fmol on insert, 1 U of ligase per 20 μ L reaction. Reactions were incubated for at least 6 hr at room temperature, or overnight at 14°C before transforming 1 μ L of undiluted reaction mix.

2.2.4.3 **pProEx Htb**

For general cloning, expression and sequence confirmation, pProEx Htb (Invitrogen) was the standard vector used. All library constructs were cloned into this vector using BamHI and EcoRI restriction sites, amplified by primer pair 005/006. pProEx Htb is a non-T7 vector, expressing the inserted gene with an N-terminal recombinant tobacco etch virus protease (rTEV) cleavable His-tag for immobilised metal affinity chromatography purification. See Appendix A1.2.1 for cloning site map.

2.2.4.4 **Gateway Cloning**

Some gene cloning for expression was done using the Gateway system (Invitrogen). This system uses site-directed, direction specific *in vitro* recombination (Hartley *et al.* 2000). Inserts for cloning were amplified by nested PCR (section 2.2.1.1) to introduce attB recombination sites, an rTEV cleavage site and a 3' TGA stop codon. The amplified constructs were inserted into host "donor" vector pDONR221 (Invitrogen) by BP reaction, transformed into a DH5 α *E. coli* host, plasmid mini-prepped and the presence of insert confirmed by digestion with BsrGI. Successful clones were used in LR reactions for insertion into expression ("destination") vector pDEST15 (Invitrogen), which was again transformed into a DH5 α host. The successfully transformed clones were plasmid prepped, sequenced and re-transformed into BL21 (DE3) for expression.

2.2.5 Alkaline Lysis Plasmid Purification

Large scale plasmid preparation was performed using an up-scaled alkaline lysis method, combined with PEG-precipitation, modified from Sambrook & Russel (Sambrook and Russel 2001).

2.2.5.1 GTE Resuspension Buffer

50 mM Glucose

25 mM Tris pH 8

10 mM EDTA pH 8

Sterile filtered, stored at 4°C

2.2.5.2 Lysis Buffer

1% (w/v) SDS

0.2 M NaOH

Prepared fresh from separate 10x stocks.

2.2.5.3 Method

1. Inoculate a 100 mL 2YT culture with a single colony from a freshly streaked plate and grow overnight.
2. Harvest the cells and discard the supernatant, removing as much of it as possible.
3. Resuspend the pellet in 8 mL of cold GTE, ensuring there are no clumps of cells remaining.
4. Add 12 mL of lysis buffer and mix thoroughly by gentle inversion. Do not allow lysis to proceed for longer than 5 min
5. Add 12 mL of 3M sodium acetate pH 4.8 and immediately mix by inversion to prevent localised precipitation. Incubate on ice for 10 min.
6. Centrifuge at 15,000 g for 10 min at 4°C and decant the supernatant. Add an equal volume of isopropan-2-ol (IPA) and incubate on ice for 30 min.
7. Pellet the precipitate at 15,000 g for 30 min at 4°C. Decant the supernatant and rinse the pellet once with 70% (w/v) ethanol.

8. Dissolve the pellet in 1 mL 10m M tris pH 8.0, centrifuge 1 min at 13,000 g to remove insoluble material, then extract with 500 μ L phenol/chloroform four times, or until no further precipitate is visible. Extract twice with 500 μ L chloroform or until no further precipitate is visible.
9. Precipitate the DNA by adding one tenth volume of 3 M potassium acetate pH 4.8, then an equal volume of IPA. Incubate on ice for 30 min and pellet at 13,000 g for 10 min at 4°C.
10. Decant the supernatant, rinse the pellet twice with 70% ethanol and air-dry at 37°C for 1 hr. Re-dissolve the pellet in 192 μ L of 10mM Tris pH 8.0 and transfer to a clean microfuge tube.
11. Add 48 μ L of 4 M NaCl and 240 μ L of 13% (w/v) Polyethylene glycol 8000 (PEG 8K) and mix thoroughly. Incubate on ice for 30 min, then pellet the precipitate at 13,000 g for 30 min at 4°C.
12. The pellet formed is diffuse and difficult to dissolve. Dissolve in 250 μ L 10mM Tris pH 8.0 by scraping with a clean pipette tip and incubation at 50°C

2.3 Microbiology

2.3.1 *Antibiotics*

Antibiotics were stored as 1000x stocks at -20°C

100 mg/mL Ampicillin

50 mg/mL Kanamycin

2.3.2 *Media*

2.3.2.1 **5x M9 Salts**

233 mM Na₂HPO₄

110 mM KH₂PO₄

42.7 mM NaCl

9.3 mM NH₄Cl

M9 salts were autoclaved and stored at room temperature

2.3.2.2 **M9 Minimal Media**

1x M9 Salts
7 mM MgSO₄
0.3 mM Thiamine
44 mM Glucose
0.1 mM CaCl₂

2.3.2.3 **2YT**

16 g/L Tryptone
10 g/L Yeast Extract
5 g/L NaCl

2.3.2.4 **SOC Media**

20 g/L Tryptone
5 g/L Yeast extract
10 mM NaCl
2.5 mM KCl
5 mM MgSO₄
10 mM MgCl₂
20 mM Glucose

SOC media was prepared and autoclaved without glucose, which was added from a 40% (w/v) stock before use

2.3.2.5 **Agar plates**

Plates were prepared by supplementing the required nutrient media mix with 15 g/L granulated agar and autoclaving. The agar/media was kept at 50°C until used.

2.3.2.6 Bacterial Strains

XL1 Blue and DH5 α were used essentially interchangeably for the purposes of basic cloning and non-T7 expression, except in the case of Gateway cloning where DH5 α was used exclusively. BL21 was used for expression with T7 vectors, and TG1 was used in all phage production.

Table 2.2 Bacterial strains used

Name	Phenotype
TG1	supE thi-1 Δ (lac-proAB) Δ (mcrB-hsdSM)5 (r _K ⁻ m _K ⁻) [F' traD36 proAB lacI ^q Z Δ M15]
XL1 Blue	endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F[::Tn10 proAB ⁺ lacI ^q Δ (lacZ)M15] hsdR17(r _K ⁻ m _K ⁺)
DH5α	F' thi-1 endA1 recA1 relA1 gyrA96 Φ 80lacZ Δ M15 Δ (lacZYA-argF)U169, hsdR17(r _K ⁻ m _K ⁺), λ -
BL21 (DE3)	F' ompT gal dcm lon hsdS _B (r _B ⁻ m _B ⁻) λ (DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

2.3.3 Electrocompetent Cells – General Cloning

E. coli cells were prepared for single-gene cloning using the following method:

1. Pick a single colony from a freshly streaked plate and inoculate a 5 mL LB culture.
2. The following morning inoculate 500 mL of LB in a 2L baffled flask with the entire 5 mL overnight culture and grow at 37°C to OD⁶⁰⁰ of 0.6
3. Harvest the cells by centrifugation at 3000 g, taking care to remove as much of the supernatant as possible, then gently resuspend in 500 mL of ice-cold sterile-filtered 10% glycerol.
4. Repeat step 3 twice, each time reducing the resuspension volume to 250 mL, then 20 mL.
5. Harvest the cells in a 50 mL falcon tube at 3000 g and gently resuspend the pellet in 1.5 mL of ice-cold sterile-filtered 10% glycerol.
6. Aliquot the prepared suspension into 1.5 mL cryotubes, 50 μ L per tube, and flash-freeze in liquid nitrogen. Store the frozen aliquots at -80°C.

2.3.4 Electroporation transformation of *E. coli*

E. coli cells were transformed using the following method:

1. Thaw an aliquot of electrocompetent cells on ice, add 1 μ L of DNA to be transformed and mix gently by pipetting.

2. Transfer the aliquot of cells into a pre-cooled 0.2 cm GenePulser cuvette (Bio-Rad) and pulse at 2.5 V. Immediately dilute the transformed cells with 1 mL of SOC media and mix.
3. Incubate the diluted cells at 37°C for 30 min, then spread 50 µL on a selective agar plate. Adjust plating volume for amount of DNA and cell competency so as to get individual colonies.

2.4 Gene Libraries and Selection

2.4.1 *Phagemid vector*

All library screening was done using phagemid pRpsp2 (Beekwilder *et al.* 1999). Genes were inserted using restriction enzymes NcoI and NotI. pRpsp2 expresses the inserted gene as a fusion protein of the C-terminus of M13 minor coat protein pIII, with the C-terminal peptide from human Myc (cMyc) as an epitope tag located between pIII and the fused Obody. Expressed protein was targeted for secretion with an N-terminal pelB leader sequence, a *sec* pathway signal. See Appendix A1.2.2 for a map of the vector cloning site.

2.4.2 *Gene Library Construction*

2.4.2.1 **Fragment Generation and Nomenclature**

Gene libraries were constructed by PCR dissection and combinatorial re-assembly using internal and flanking primer pairs, which are listed in appendix A1.3. Each primer pair amplifies either an unmodified section of the template gene, or was used to double-strand and amplify a mutagenic single-stranded oligonucleotide. Gene fragments were labelled according to the flanking primers used during amplification and suffixed with the number of random codons it incorporated. For example, the product of a reaction using primers 005 and 006 which contains four mutant codons was labelled 005/006 4m.

Mutational oligonucleotides were purchased as single-stranded templates, with specific codons randomised by replacing explicit bases with a random base (N) or a defined subset of the full four possibilities (Appendix A1.7). Before assembly into an intact gene, each oligo was first “filled in” to produce a double-stranded template, then amplified with flanking primers to promote

dominance of correctly-sized fragments with the correct overlaps for assembly with adjacent fragments.

2.4.2.2 **Fragment Assembly**

Assembly reactions were designed using a “super primer” principle (Figure 2.1), where the sense strand of the 5` template fragment and the antisense strand of the 3` template fragment act as primers for each other, resulting in a double stranded product fragment, and two unincorporated single strands from the template fragments. Inclusion of a flanking primer pair that anneals to the 3` end of each residual single-strand regenerates double stranded template fragments and allows the cycle to begin again. Like any PCR based procedure, this method is not truly random and bias towards fragments which assemble first was expected.

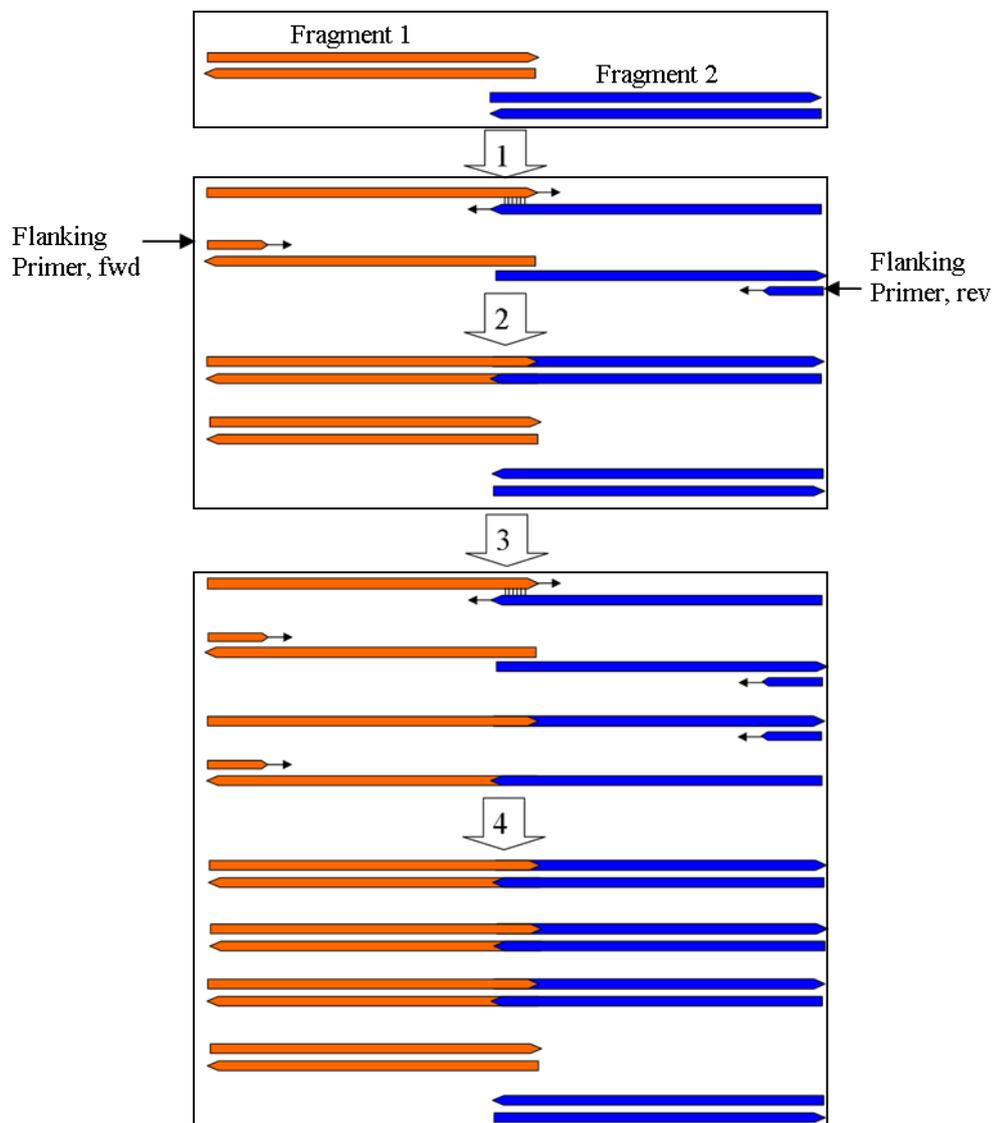


Figure 2.1 *Combinatorial gene assembly schematic*

In this hypothetical example, the PCR reaction contains two gene fragments that overlap. In the first cycle of denaturation and annealing (1), two species are produced; the first is the sense strand from fragment 1 annealed to the complementary antisense strand from fragment 2. The second is the two remaining strands which can anneal, but do so at their respective 5` ends and therefore cannot use each other as templates during the extension step. In addition, flanking primers complementary to the 3` ends of this second species are annealed. The first extension step (2) produces three species; a full length product and regenerated copies of the original two fragments. Subsequent denaturation, annealing (3) and extension (4) cycles result in further production of newly combined fragments, regeneration product of the original fragments, as well as amplification of those full length products produced in previous cycles.

Adjacent fragments were combined in equimolar amounts for assembly along with flanking primers in a PCR mix and cycled. The annealing temperature was varied according to which flanking primer pair was being used. For assembly reactions involving fragments with randomised codons, the amount of template fragments added into the reaction, as well as the size of the

reaction, was informed by the maximum theoretical diversity of the product fragment. Where possible, the amount of input fragments was set higher than the maximum theoretical diversity. Upper limits on the achievable diversity were imposed by PCR reaction sizes, the amount of mutational oligos available and practical concerns in handling large quantities of PCR fragments. When combining two fragments where both contain mutational oligos, the required quantity (in mols) was calculated as the product of the maximum theoretical diversity of both fragments, although practically the maximum output in mols from a reaction was limited by the amount of input flanking primers. Standard Platinum Pfx PCR assembly reactions always contained 50 pmol of primer pair per 50 μL reaction volume, which corresponds to a theoretical maximum of 3.011×10^{12} molecules. As the largest PCR reactions used were 500 μL , the theoretical maximum diversity of any library assembled this way is 3.011×10^{13} .

The number of cycles for each assembly reaction was varied according to the concentration of the fragments being assembled, so as not to continue cycling once all of the primers in the mix had been incorporated into assembled gene fragments. This calculation was performed assuming that the concentration of assembled fragments after the first cycle is equal to the molarity of each fragment before cycling, and that each cycle after that doubles the number of assembled fragments.

Where large reaction volumes were needed, 25 μL pilot reactions were run first to test the competency of the individual fragments for assembly. Product fragments were assessed by the presence of a dominant (80% or greater) band of the correct size on an agarose gel. Large-scale PCR reactions were separated into 25 or 50 μL aliquots for cycling. Correctly assembled fragments were excised from a preparative-scale agarose gel and purified on a HiPure PCR Product Purification column (Roche).

2.4.2.3 Error-prone PCR library generation

Error-prone (EP)-PCR (Cadwell and Joyce 1992) was used to generate a randomly mutated library for affinity maturation to a HEWL target. EP-PCR was performed using *Taq* DNA polymerase and primer pair 005 and 006. Errors were induced by the introduction of 0.5 mM MnCl_2 to reduce polymerase specificity. Also, deoxynucleotide-triphosphates thymidine and

cytidine were disproportionately increased, from 0.2 mM to 1 mM, to increase the likelihood of miss-incorporation. MgCl₂ was increased to 7 mM reduce primer specificity. To reduce the chances of non-specific products, the template gene was introduced as a freshly prepared PCR product, at 10 fmol per 100 µL reaction. Taq polymerase concentration was also increased, to 5 U per reaction. The mixture was split into 10 µL aliquots for cycling to prevent dominance of an early mutation in the library as a whole.

The PCR product was gel purified and the whole amount amplified with phagemid cloning primers (192 and 040) to produce quantities amenable to library cloning. The final PCR product library was gel purified and digested with NcoI and NotI for cloning into the pRsp2 phagemid.

2.4.3 *Phage Libraries*

2.4.3.1 **Library Cloning**

Gene library cloning was done using an up-scaled classic restriction/ligation technique, using T4 DNA ligase (Invitrogen). The supplier-recommended recipe for a single 20 µL ligation reaction was used as a small-scale test to calculate the absolute number of transformable clones produced by the ligation. Before ligation, vector was freshly digested with NcoI and NotI, purified with a HiPure PCR Product Purification Kit (Roche) and desphosphorylated. Insert and plasmid were added at a ratio of 3:1 to a final concentration of 9 pmol insert and 3 pmol plasmid per 2 mL reaction. The ligation reactions were incubated overnight at 14°C.

Ligated DNA was recovered by centrifuge ultrafiltration, using a Microcon 50,000 Da cut-off (Millipore, Massachusetts, U.S.A.). The sample was exchanged into MQ H₂O by repeated dilution and concentration, then spun off of the ultrafiltration membrane. DNA was extracted by IPA precipitation (see section 2.2.3.3), then dissolved in 200 µL 10 mM Tris pH 8.0 and purified a final time using a High Pure PCR Product Purification kit (Roche). DNA concentration was measured by nanodrop.

2.4.3.2 Library Transformation

Gene libraries were always transformed into freshly prepared TG1 *E. coli* electrocompetent cells, prepared using the following method:

1. A single colony from a freshly streaked M9 minimal media agar plate was picked and grown overnight in 5 mL M9 minimal media broth.
2. Two 2 L baffled conical flasks were rinsed in hot water and autoclaved full of MQ water to removed residual detergents. Each sterile flask was shaken with 400 mL of 2YT inoculated with 2 mL of the overnight culture at 30°C, to an OD₆₀₀ of 0.4-0.45.
3. After attaining the correct OD₆₀₀ the cultures were split into 50 mL conical centrifuge tubes and incubated on ice for 2 hrs, then centrifuged at 3,000 g for 10 min to gently pellet the cells. The supernatant was removed by aspiration.
4. The cell pellets were very gently resuspended in 10 mL of ice-cold 1 mM HEPES pH 7.4, 10% glycerol by pipetting, making sure to keep them on ice as much as possible. Each pellet was diluted to 50 mL in the same buffer, incubated for 10 min on ice and centrifuged 3,000 g for 15 min. The supernatant was removed by aspiration.
5. Step 4 was repeated, except resuspension was only by flicking and swirling the buffer, no pipetting. Again, keeping the cells on ice as much as possible.
6. Each pellet was resuspended in 5 mL ice-cold 10% glycerol and pooled in groups of 4, to get 20 mL aliquots in each tube, always with care to keep the buffer cold. These aliquots were diluted to 50 mL with cold 10% glycerol and incubated on ice for 10 min.
7. The cells were harvested for a final time by centrifugation at 3,000 g for 15 min. The supernatant was removed by aspiration and the pellet very gently resuspended in 450 µL of 10% glycerol per tube. The final cell suspension was used in 50 µL aliquots for Electroporation.

Following preparation of the electrocompetent cells, the volume was estimated and purified ligation DNA added to them to a ratio of 1 μL DNA per 50 μL cells. One 50 μL aliquot was kept separate for competency estimation.

The library was transformed according to the general method given in section 2.3.4, except that after dilution in 1 mL SOC, the whole transformed culture was transferred into a sterile 50 mL centrifuge tube and cuvette then rinsed twice with 1 mL SOC, pooling the rinses with the first lot of cells from the cuvette.

This process was repeated until all of the competent cells containing the ligation DNA were transformed; typically 30-40 individual transformations, at which point the whole culture was diluted to approximately 100 mL with SOC and incubated at 37°C for 20 min. To estimate the transformation efficiency, 10 μL of the culture was taken and immediately diluted into 90 μL of ice-cold 10% glycerol. The total cell count was estimated by titration using a 10-fold dilution series spotted on to 2YT agar plates supplemented with 100 $\mu\text{g}/\text{mL}$ ampicillin (2YT/amp plates). Total cell count was used as a proxy for library size, assuming that every transformant is unique and that the sample was taken before there was significant growth in the culture.

Electro-competency was measured by transforming a separate 50 μL aliquot of cells with 5 ng of empty pRpsp2 and titration of transformed cell count to calculate the number of expected colony forming units (cfu) per μg of DNA transformed.

After sampling, the transformed library was pelleted by centrifugation at 3500 g for 10 min, then gently resuspended in ~ 5 mL 2YT and spread on to four 230 mm 2YT/amp plates. The inoculated plates were incubated overnight at 37°C and the resulting bacterial lawn scraped off gently with a sterile plastic spatula. The cells were resuspended in ~ 20 mL 2YT, then diluted 1:2 with sterile 50% glycerol. The library transformants were flash frozen in liquid nitrogen and stored at -80°C as 1.5 mL aliquots.

2.4.4 *Phage Display*

2.4.4.1 **Blocking Buffer**

1x PBS pH 7.4

10 mg/mL Bovine serum albumin (BSA)(Invitrogen)

2.4.4.2 **5x Phage Precipitation Solution**

20% (w/v) Polyethylene Glycol (PEG) 8000 (Sigma)

15% (w/v) NaCl

2.4.4.3 **Phage Titration**

Phage counts were measured using TG1 as reporter bacteria. A single colony was picked from a fresh TG1 agar plate and inoculated into 3 mL 2YT, which was allowed to grow to saturation (4-6 hr) at 37°C. A titre plate consists of 5 mL of “soft” agar (2YT with ~0.7% agar) inoculated with 150 µL of the saturated TG1 culture, layered on to the surface of a standard 2YT 1.5% agar plate, supplemented with ampicillin for transducing particle (TDP) counts, or no antibiotic for plaque forming units (PFU) counts. The prepared plates were allowed to dry standing open in a laminar flow cabinet for 30 min before use.

The input, experimental and control phage samples from panning experiments were serially diluted 100-fold in sterile PBS or MQ water, up to 10⁻¹⁰-fold for input phage, 10⁻⁸-fold for eluted experimental and control samples. 10 µL of each dilution was spotted on to a single titre plate, spaced out so that the spots did not overlap. The plates were allowed to stand uncovered in a laminar flow hood until the spots were absorbed, then covered and incubated at 37°C overnight. Phage titres per mL were calculated per by counting colonies (in the case of TDP titration) or plaques (in the case of helper phage titration) in the highest dilution where individual colonies or plaques were distinguishable.

2.4.4.4 **Vcsm13 Helper Phage**

Vcsm13 (Agilent Biotechnology) is a derivative of m13 phage, with a kanamycin selectable marker. Helper phage stocks were produced at least once every 6 months and stored at 4°C as 1 mL aliquots.

To produce a working stock, the source stock was titred (see section 2.4.4.3 for phage titre method) on non-selective media to produce single plaques. A single plaque was picked, inoculated into a 5 mL culture of TG1 in 2YT media at OD₆₀₀ 0.5 and allowed to stand for 1 hr at 37°C. The infected culture was transferred into a 2 L baffled conical flask with 500 mL 2YT with 50 µg/L Kanamycin and grown overnight at 37°C.

The following day the cells were pelleted by centrifugation for 10 min at 10,000 *g*, and the supernatant decanted. Dissolved phage were precipitated from the supernatant with 0.4% PEG 8000/0.3% NaCl for 2 hr on ice. The precipitated phage were harvested at 15,000 *g* for 20 min, re-dissolved in 20 mL sterile PBS and centrifuged at 15,000 *g* for 15 min to remove residual cells. The supernatant was filtered to 0.45 µm and precipitated with 1x phage precipitation solution overnight at 4°C. The precipitated phage were harvested at 15,000 *g* for 20 min, re-dissolved in 20 mL PBS and titred. The phage stock was aliquoted into 1 mL lots and stored at 4°C.

2.4.4.5 **KM13 helper phage**

KM13 is a trypsin-sensitive helper phage (Kristensen and Winter 1998). Briefly, a linker peptide containing a trypsin recognition site was added to the loop joining the second and third domains of minor coat protein pIII. Although it was first conceived as a tool for proteolytic selection for stability, KM13 seems to benefit ligand-driven selection as well (Goletz *et al.* 2002).

Preparation of KM13 helper phage was identical to the method described for Vcsm13 in section 2.4.4.4. After preparation, the titred stocks were tested for trypsin sensitivity by incubation with 1 mg/mL trypsin in PBS supplemented with 1 mM CaCl₂ for 2 hr at 37°C then re-titred. The freshly prepared phage stock was judged sufficiently trypsin sensitive if a 10⁶-fold reduction in infectivity was observed after treatment. If the required drop was not seen, the preparation was discarded and new stock prepared.

2.4.4.6 Phage Library Production

To generate the input phage for the first round of panning, a single aliquot of library transformants was thawed on ice. The entire aliquot was used to inoculate 1 L of 2YT/ampicillin, which was grown in a baffled 2 L conical flask at 37°C to an OD₆₀₀ of 0.5 then infected with Vcsm13 helper phage, with multiplicity of infection (m.o.i.) of 20. The infected culture was incubated standing at 37 °C for 30 min, centrifuged at 4,000 g for 15 min, resuspended in 1 L of fresh 2YT/amp/kan and grown overnight at 37°C with constant shaking.

Phage secreted into the media supernatant were purified by centrifugation of the overnight culture at 10,000 g for 20 min and decanting the supernatant into a sterile 1 L Schott bottle containing 40 g PEG 8000 and 30 g NaCl for precipitation. The cell pellet was discarded. The precipitating phage were incubated on ice for 1-2 hr, then centrifuged at 15,000 g for 20 min in sterile 500 mL centrifuge bottles. The supernatant was carefully removed and the centrifuge bottles allowed to stand inverted for 10 min to drain excess precipitant solution. Phage pellets were dissolved in 2 mL PBS and transferred to a 2 mL microfuge tube. The centrifuge bottles were rinsed with a further 2 mL PBS, the rinse pooled with the first aliquot. The re-dissolved phage sample was centrifuged at 13,000 g for 15 min and filtered to 0.45 µm with a syringe filter. After filtration, the phage were precipitated a second time by the addition of phage precipitation solution to 1x and incubated overnight at 4°C. Immediately before use, the phage were pelleted at 15,000 g for 20 min and dissolved in 2 mL PBS.

2.4.4.7 Immunotube Immobilisation

Nunc immunotubes (ThermoFisher Scientific, Waltham MA, U.S.A.) were used as the only method of ligand immobilisation for panning experiments. Ligand in 2 mL PBS was incubated with constant shaking overnight at room temperature in parafilm-sealed immunotubes. Each tube containing ligand was matched to a control tube with an equal concentration of BSA. After the overnight incubation, each tube was blocked for 3 hrs in 4 mL blocking buffer (section 2.4.4.1), then rinsed five times with 5 mL PBS.

2.4.4.8 **Panning**

Before panning, the 2 mL input phage was diluted with sterile PBS (750 μ L into 2 mL) into two identical aliquots. The diluted phage solution was pre-adsorbed with 1 mg/mL BSA for 3 hrs with constant inversion, then added to the rinsed immunotubes, one into the ligand tube, the other into the BSA control tube, and incubated with constant gentle shaking for 1 hr at room temperature. After discarding the unbound phage supernatant, each tube was washed 10 times with PBS-T, then 10 times with PBS. Bound phage were eluted using either 1 mg/mL trypsin in 2 mL PBS with 1 mM CaCl_2 or 1 mg/mL ligand for 1 hr with constant gentle shaking. Eluted phage were decanted from the immunotubes, titred and stored at 4°C.

2.4.4.9 **Single Clone Isolation**

Single phage isolation was done on phage titre plates, as described in section 2.4.4.3. The phage sample was diluted in 10-fold serial dilutions, to get between 1 and 100 phage per 250 μ L. Three titre plates per sample were infected with 250 μ L of phage solutions at different dilutions, to ensure that some plates have single colonies in sufficient number. The plates were incubated overnight at 37°C, colonies picked and grown overnight for plasmid-preparation.

2.4.5 ***Western Blot Protein Detection***

2.4.5.1 **Transfer buffer**

25 mM Tris.Cl
192 mM Glycine
20% (v/v) Methanol
0.01% Sodium dodecyl sulfate (SDS)

2.4.5.2 **Ponceau Red Stain**

0.2% (w/v) Ponceau Red
1% (v/v) Glacial Acetic Acid

2.4.5.3 Membrane Blocking Buffer

1x PBS pH 7.4

10% (w/v) Skim milk powder

0.1% (v/v) Tween-20

2.4.5.4 Blotting and Detection Method

Western blots use labelled antibody probes to detect specific proteins bound to a membrane. The primary antibody binds directly to the blotted protein; the secondary antibody binds to Fc region of the primary and is conjugated to horse radish peroxidase (HRP) which provides the detection signal.

An SDS-PAGE gel of the samples for detection, pre-stained ladder and control samples were run and then added to the western blot transfer “sandwich” along with a square of nitrocellulose membrane large enough to cover the whole gel. The blot assembly was run at 100 V for 1 hr with constant stirring, with an ice-pack to keep the buffer cool. All western transfers were performed using a Mini Transblot Cell (Bio-Rad). After transferring, the membrane was removed and stained for 5 min in enough Ponceau Red Stain to cover it completely, then rinsed with tap water until bands were visible. The positions of the lanes and the bands of the ladder were marked with a pencil, and the membrane blocked for 1 hr with constant shaking in Blocking Buffer.

After blocking, primary antibody was administered diluted in 2 mL of Blocking Buffer spotted on to a glass plate. The membrane was carefully layered on to the spotted antibody, covered and incubated at 21°C for 1 hr. Unbound antibody was removed with five washes of 5 min each with PBS-T. Secondary antibody was administered in the same way as the primary, with the same wash steps afterwards.

Detection was achieved by chemiluminescence, using Pierce SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific, Rockford, USA), a binary peroxide substrate. The substrate components were mixed 1:1 then diluted 1:2, layered on to the membrane and allowed to incubate for 1 min. Excess substrate was carefully removed with blotting paper and the

membrane covered with a transparency. Visualisation was done with a 5 min exposure in a Fuji Intelligent Dark Box II, with a Fuji Las 1000 digital camera.

2.4.6 *Surface Plasmon Resonance*

All surface plasmon resonance (SPR) work was carried out using a Biacore 3000 SPR Instrument (GE Healthcare). SPR measures changes in the angle of diffraction off of the surface of a gold chip, brought about by mass variations on the opposite side. Output was in response units (RU).

2.4.6.1 **Running Buffer**

10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
(HEPES) pH 7.4

150 mM NaCl

3 mM EDTA

0.005 % (v/v) Tween-20

2.4.6.2 **Chip Surface Preparation**

For analysis of Obody-ligand interaction, CM5 chips were used, which covalently attach the ligand via amine groups on the protein. The exact concentration of ligand varied, but the general method was controlled by the Biacore software.

Two surfaces were prepared in tandem, one with ligand the other as a reference blank. First, the surface was activated by injection of 1-ethyl 3-(3-dimethylaminopropyl)-carbodiimide hydrochloride, followed by N-hydroxysuccinimide. The ligand was injected into only one flow cell, typically at 10 µg/mL for 30 s at 5 µL/min. The remaining active surface was deactivated with 1 M ethanolamine, and the final response for the chip calculated to indicate the level of bound ligand. The target bound-surface response was 100-150 RU.

2.4.6.3 **SPR Analysis**

Affinity determination on the Biacore 3000 was performed using a 2-fold dilution series of the OBody, starting at a minimum concentration of 10-fold higher than the affinity, if it had previously been calculated. Each dilution was

injected at 15 $\mu\text{L}/\text{min}$, until equilibrium at the chip. The surface was regenerated with running buffer for 5 min after each injection. Each experiment utilised two flow-cells, one with bound ligand, the other as a reference. The final response curves for each injection were the reference subtracted from the flow cells with bound ligand. The K_D was calculated by plotting the maximum response at each concentration vs. OBody concentration in mol/L and fitting a Langmuir saturation binding curve to the data (Equation 2.1), using GraphPad Prism modelling software. Fits with different stoichiometries of Obody and ligand were compared to test for applicability of a 1:1 binding model.

$$y = \frac{(Rmax \cdot x)}{(K_D + x) + mx + c}$$

Equation 2.1 Langmuir binding curve, fitted using Graphpad Prism

This equation assumes the curve is the sum of a non-linear specific response and a linear non-specific response that increases with concentration, where Rmax is the maximum response, K_D is the dissociation constant. For the linear component, m is the gradient and c is the y-intercept of the linear portion.

2.4.7 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) was carried out using a MicroCal VP-ITC instrument (GE Healthcare). ITC measures binding enthalpy by measurement of the required input of energy to bring a known volume of ligand solution back to the base temperature after each injection during titration of a concentrated analyte into a dilute ligand of known volume.

As this technique is very sensitive to dilution effects, each protein was exhaustively dialysed into a single batch of PBS to minimise buffer differences. Two titrations were performed for each experiment; a control titration of concentrated analyte (HEWL at 350 μM) into buffer giving the enthalpy of dilution, and an experimental titration of the same analyte into Obody solution at $\sim 30 \mu\text{M}$. The control titration curve was subtracted from the experimental titration curve and the results analysed using Origin ITC Analysis software (Origin Lab, Northampton MA, USA).

The results were plotted as the molar ratio of the analyte and ligand on as the x variable, with change in energy in kcal/mol analyte on the y axis and fitted to Equation 2.2. Analysis was done using Origin ITC analysis software to give the enthalpy of binding and dissociation constant.

$$y = \frac{Rmax}{1 + e^{h(m-x)}}$$

Equation 2.2 ITC binding curve, fitted using Origin ITC

This equation was fitted and the subsequent calculations carried out to determine the binding constant by Origin ITC analysis software, where Rmax is the maximum change in enthalpy (ΔH), h is the Hill Slope and m is the x-intercept at the point of inflection. The x intercept gives the stoichiometry of binding, and the K_D is derived from the Hill Slope, taking into account the concentration of the two protein constituents.

2.4.8 *Differential Scanning Fluorimetry*

A method for determining the T_m of proteins using fluorescence has recently been published (Ericsson *et al.* 2006; Niesen *et al.* 2007). Briefly, the protein is mixed with a dye which fluoresces when bound to hydrophobic residues then subjected to a temperature gradient using a qPCR thermocycler. Here, Obodies were tested using SYPro Orange (Invitrogen) and a Corbett Rotor-Gene RT-PCR machine (Corbett Life Science, Concorde NSW, Australia). As the protein denatures, fluorescence increases, ideally in a sigmoidal cooperative unfolding curve, but usually until it precipitates and fluorescence drops, resulting in a discrete peak instead of a sigmoidal curve. Before fitting, the data is trimmed to exclude points past the peak maximum, forcing the fitting program to extrapolate. A curve is fitted using Equation 2.3, giving the T_m . This method is straightforward and seems to agree with more rigorous methods of determining T_m , such as differential scanning calorimetry.

Equation 2.3 Fluorescence melt curve equation

Where yMin and yMax are the minimum and maximum fluorescence values, h is Hill's slope and T_m is the point of inflection.

$$y = yMin + \frac{yMax - yMin}{1 + 10^{h(Tm-x)}}$$

2.5 Protein Expression and Purification

2.5.1 *SDS - Polyacrylamide Gel Electrophoresis (PAGE)*

2.5.1.1 **Coomassie Blue Stain**

0.05% (w/v) Coomassie Brilliant Blue R-250

25% (v/v) IPA

10% (v/v) Acetic acid

2.5.1.2 **Destain**

10% Acetic acid

2.5.1.3 **Resolving Gel:**

12.1 mL ddH₂O

10 mL 30% (w/v) Acrylamide (37.5:1 acrylamide: N,N'-methylene-bis-acrylamide)

7.5 mL 1.5 M Tris pH 8.8

0.3 mL 10% (w/v) SDS

15 µL N,N,N',N'-tetramethylethylenediamine (TEMED)

150 µL 10% (w/v) Ammonium persulfate (APS)

2.5.1.4 **Stacking Gel:**

17 mL ddH₂O

4.25 mL 30% Acrylamide

3.125 mL 1.0 M Tris pH 6.8

0.25 mL 10% (w/v) SDS

15 µL TEMED

150 µL APS

2.5.1.5 Gel Preparation

SDS-PAGE gels were made in batches of 5 and stored at 4°C for up to 1 month. The resolving gel was made first and poured into the 5-gel caster, then covered with a layer of IPA to prevent bubbles setting at the top of the gel. Once the resolving gel was set and the IPA drained, stacking gel was made and poured on top along with well-forming combs.

2.5.1.6 Denaturing Protein Loading Dye

10 mL 1M Tris pH 6.8
8 mL glycerol
16 mL 10% (w/v) SDS
4 mL β -mercaptoethanol
1 mL 1% bromophenol blue

2.5.1.7 Protein Standards

To estimate the size of protein bands on gels stained with coomassie as well as a gel quality marker, Precision Plus Protein Unstained Standard (Bio-rad Laboratories Pty, California, U.S.A.) was used. For gels to be used for western blotting, BenchMark Prestained Ladder (Invitrogen Corporation, California, U.S.A.) was also used to validate a successful blot. See appendix A1.1 for marker weights.

2.5.1.8 Protein Detection

All protein gel visualisation was done using coomassie blue stain. After removing the gel from the running apparatus and placing it in a suitable container, enough stain was added to completely cover the gel, then heated for ~20 s in a microwave at its highest setting. The gel and stain was allowed to cool down with constant shaking, then the stain decanted. The gel was rinsed in tap water and again covered in destain. After heating on high for ~20 s, a rolled paper towel was wadded at one end of the container to absorb the coomassie dye. The gel and destain was allowed to cool with constant shaking, with destain and paper towel replaced if necessary.

2.5.2 *General Protein Expression*

All engineered Obodies worked with over the course of this study behaved identically during purification, and purification was treated as a medium-throughput operation. Unless otherwise stated, protein production followed a general method. A single colony was picked from a freshly streaked plate and used to inoculate a small (~5 mL) 2YT overnight starter culture, with the appropriate antibiotic. The following day, 1 L of 2YT culture in a 2 L conical baffled flask was inoculated with 1 mL of the overnight culture and grown at 37°C with constant shaking to an optical density at 600 nm (OD₆₀₀) ~0.6. IPTG was added to 1mM and the induced cultured grown overnight in the same conditions.

The following day, the cells were centrifuged at 10,000 *g* for 20 min at 4°C, the supernatant discarded and the pellet re-suspended in ~40 mL PBS, with one Complete EDTA-free Protease Inhibitor Tablet (Roche). Lysis was achieved by sonication; 6 rounds of 30 s, with 30 s pause between them.

The lysate was fractionated at 15,000 *g* for 20 min at 4°C and the supernatant decanted into a fresh 50 mL tube. The insoluble pellet was re-suspended in 40 mL PBS and sampled for a later gel if necessary. The soluble fraction was filtered to 0.2 µm with sterile Minisart luer-lock syringe filters (Sartorius Stedium Biotech GmbH, Goettingen, Germany), then loaded on to the appropriate column, according to the affinity tag in use.

Soluble protein concentration was calculated by UV absorbance at 280 nm, giving concentration based on the extinction coefficient for individual proteins (Equation 2.4).

Equation 2.4 Beer-Lambert Law

The Beer-Lambert law, where *A* is absorbance at 280 nm, ϵ is extinction coefficient in M⁻¹cm⁻¹, *b* is pathlength in cm and *c* is concentration in mol/L

$$A = \epsilon bc$$

2.5.3 *Immobilised Metal Affinity Chromatography*

Immobilised metal affinity chromatography (IMAC) was used to purify His-tagged fusion proteins. All purifications were done using a General Electric (GE) Healthcare 5 mL HisTrap FF chelating column charged with Ni²⁺.

2.5.3.1 **Column Preparation**

Before protein binding, the column was washed with 5 column volumes of water. 5 mL of 100 mM EDTA pH 8.0 was pushed into the column and allowed to incubate for 5 min. The column was washed with PBS until no blue colour was visible, charged with 5 mL of 100 mM NiCl₂, then washed again with 5 column volumes of PBS.

Following lysis, soluble fractionation and filtration, the lysate was slowly pushed through the column with a syringe, collecting the flow-through.

2.5.3.2 **FPLC Purification**

IMAC purifications of new proteins were run using Fast Protein Liquid Chromatography (FPLC) instruments, either an ÄKTA Prime Plus or ÄKTA Basic (GE Healthcare, Waukesha WI, U.S.A.), which have 280 nm UV absorbance (UV₂₈₀) and conductivity inline monitors. Before attaching the protein-charged column the instrument was equilibrated into PBS (Buffer A) then PBS + 1 M imidazole pH 7.4 (Buffer B). After equilibration the system was rinsed with Buffer A only and the column attached and washed with 1% Buffer B, effectively a PBS + 10 mM imidazole wash. Once the UV₂₈₀ trace reached a plateau, a gradient of increasing levels of imidazole from 10 mM to 500 mM was run (1% -50% Buffer B) over 100 mL, collecting 2 mL fractions along the full length of the gradient. Peaks visible in the UV₂₈₀ trace were analysed by SDS-PAGE.

2.5.3.3 **Manual Purification**

IMAC purifications of known proteins were performed using syringes, without the use of an FPLC instrument. Based on the UV₂₈₀ trace from FPLC runs, the imidazole levels required to elute the protein from the column were estimated. Protein levels in flow-through were monitored using Protein Assay Dye Reagent (PAD) (Bio-Rad), diluted 1:5 in MQ water. Fractions were

assayed in a 96-well microplate by mixing 5-20 μL of sample with 150 μL of diluted PAD and looking for development of a blue colour, indicating the presence of protein.

Obody purification was done in 3 steps: first, the column was washed with 5 volumes of PBS +10 mM imidazole to remove non-specifically bound proteins; second, the column was washed with multiple column volumes of PBS + 100 mM imidazole in 2 mL fractions until no further protein was detected coming off the column using PAD. Bound protein was eluted with PBS + 250 mM imidazole in 2 mL fractions, assayed using the Bio-Rad dye to determine when elution had finished.

2.5.4 *GSH-Affinity Chromatography*

2.5.4.1 **Column Preparation**

Glutathione-S-transferase (GST) binds specifically to reduced glutathione (GSH). All GSH-affinity purifications were performed using GSTrap HP 5 mL columns (GE Healthcare). Before use, the column was washed four times with three column volumes of alternating acidic/basic buffers (25 mM Acetate pH 4.5/25 mM N-Cyclohexyl-2-aminoethanesulfonic acid (CHES) pH 9.5), 5 column volumes of a high salt wash (500 mM NaCl), then equilibrated into PBS + 1mM β -mercaptoethanol (βme).

2.5.4.2 **Purification Method**

GSH-affinity chromatography was done either on an FPLC instrument or manually by syringe, but the method was essentially identical in both cases. Following loading the column with protein, it was washed with at least 5 column volumes of PBS, or until no protein is detectable in the flow through (by UV_{280} trace on an FPLC or by PAD assay if done manually). Bound protein was eluted with at least 5 column volumes of PBS + 10 mM GSH + 1mM βme in 2 mL fractions, again monitoring as above for the presence of protein in each fraction. If no peak was detected, fractions were checked on an SDS-PAGE gel.

2.5.5 *rTEV Cleavage*

Recombinant Tobacco Etch Virus (rTEV) protease was used to specifically cleave between the protein of interest and fusion affinity tag (See Appendix A1.5 for cut site and production method). Following affinity purification, selected fractions were pooled and estimated for total protein concentration. rTEV was added to a ratio of 1 mg:20 mg of fusion protein and incubated at 21°C for 1 hr, then overnight at 4°C. Cleavage was tested by before/after comparison on SDS-PAGE.

In the case that there was GSH present in the buffer with the fusion protein, rTEV digestion was carried out in a 6-8,000 Da cut-off dialysis tube during dialysis into 1 L of PBS with constant stirring at 4°C overnight.

2.5.6 *Size-Exclusion Chromatography*

Purification by size-exclusion was typically the second step in the purification process, following affinity purification and removal of the affinity tag. It was always performed on an ÄKTA FPLC. Size-exclusion chromatography separates a complex mixture according to hydrodynamic radius, which is often used as a proxy for molecular weight. The column bed consists of beads with a known average pore size. Smaller proteins are able to enter more pores than larger, so are retained for longer on the column. Proteins too large to fit into any pores in the column elute first, in what is known as the “void” volume. Actual resolution of the different sized proteins occurs between the void volume and the buffer front from the injected sample, which contains the smallest constituents.

2.5.6.1 **Preparative Grade Columns**

A Superdex 75 16/60 size exclusion prep-grade column (GE Healthcare) was used during the second step to remove residual contaminating *E. coli* proteins, cleaved affinity tag and uncleaved fusion protein as well as the added rTEV. Before use the column was washed with 1 column volume of filtered ultra pure water and equilibrated with 1 column volume of sterile-filtered PBS. After rTEV treatment the digested sample was concentrated using Vivaspin 5,000 Da cut-off centrifugal ultrafiltration devices (GE Healthcare) to 5 mL or less, filtered to 0.2 µm and injected on to the column. Fractions were collected

(2 mL) after allowing the void volume to be discarded. Analysis was performed by inline UV₂₈₀ absorbance trace.

2.5.6.2 Analytical Grade Columns

A Superdex 75 10/300 size exclusion analytical grade column (GE Healthcare) was used as the final “polishing” step of purification, to remove aggregates and degraded proteins. The analytical grade columns are smaller, contain smaller superdex beads, run at higher pressure and provide better resolution than the prep-grade columns. Before use the column was washed with 1 column volume of water and equilibrated into sterile-filtered PBS. The sample to be purified was concentrated to 1 mL or less, filtered to 0.2 μm and injected in successive runs of ~200 μL lots. Each run was collected in 500 μL fractions and analysed by inline UV₂₈₀ absorbance trace.

2.6 Protein Crystallography

2.6.1 *Sitting Drop Method*

All crystallisation screens were performed using a sitting-drop vapour diffusion format in Intelliplate (Hampton) 96-well trays.

2.6.1.1 **Naïve Screens**

A standard five-plate (480 conditions) screen was employed as the primary naïve screen for initial crystallisation conditions (Moreland *et al.* 2005), using a MultiPROBE II HT/EX liquid handling robot (Perkin-Elmer) to mix and dispense the screens, and a Cartesian HoneyBee (Genome Solutions) robot to lay down 100 nL protein + 100 nL mother liquor drops. Completed plates were sealed and incubated at 18°C. Individual conditions were screened manually under a microscope for the presence of crystals or promising conditions.

2.6.1.2 **Crystal Optimisation**

Fine screens were designed around promising conditions from the initial naïve screens to improve crystal quality and increase the number of crystals available for experimentation. They were constructed in a two- or three-dimensional array depending on the individual condition, varying salt concentration, precipitant concentration or pH. Protein and mother liquor drops were combined in a 1 µL + 1 µL format and incubated at 18°C.

2.6.2 *Crystal Diffraction Screening*

The cryogenic temperatures at which protein crystal x-ray diffraction is typically done (100 K) requires measures to prevent the formation of ice crystals, which can severely interfere with processing the protein-derived diffraction pattern. This necessitates inclusion of a cryoprotectant with the mother liquor so that on freezing it forms a non-crystalline and therefore minimally diffracting vitreous “glass” (Garman and Schneider 1997). To test for an appropriate cryoprotectant, the mother liquor condition for the desired condition was supplemented with a range of concentrations of glycerol, from 5% to 30% in 5% steps. The cryoprotectant was selected on the basis of an

absence of ice ring diffraction, with minimum deviation for the original condition.

Before mounting and testing for diffraction, each crystal was briefly (1 min) soaked in a series of increasingly concentrated cryoprotectants, starting at 5% glycerol and increasing in 5% steps until the desired concentration is reached. The crystal was then scooped on to a nylon mounting loop, flash cooled in liquid nitrogen and tested for diffraction with 5 min, 1° phi exposure at 0° and 90° relative to the initial orientation. The images were indexed and an appropriate collection strategy determined using Mosflm (Leslie 1992).

2.6.3 *Data Collection*

2.6.3.1 **Home Source Data Collection**

Home source datasets were collected at 100 K using the Mar x-ray generator at the Maurice Wilkins Centre (University of Auckland, New Zealand). The instrument generates x-rays (λ 1.54179 Å) with a Rigaku rotating copper anode and records diffraction patterns using a Mar2300 image plate.

2.6.3.2 **Synchrotron Data Collection**

Synchrotron datasets were collected at 100 K at the Stanford Synchrotron Radiation Lightsource (SSRL) macromolecular crystallography beamline 9-1 and the Australian Synchrotron MX1 beamline.

2.6.4 *Data Processing*

Data images were indexed and integrated using MOSFLM (Leslie 1992) or XDS (Kabsch 2010). Data reduction was carried out using SCALA (Evans 2006) and TRUNCATE (French and Wilson 1978) from the CCP4 suite (Bailey 1994).

2.6.4.1 **Molecular Replacement Phasing**

Phasing of the crystallographic data was done by molecular replacement, using either MOLREP (Vagin and Teplyakov 1997) or Phaser (McCoy *et al.* 2007) as provided as part of the CCP4 suite (Bailey 1994).

2.6.4.2 Model Building and Validation

Model building was done exclusively with Coot molecular modelling software (Emsley and Cowtan 2004), including placement of ordered solvent molecules and validation of the final model. Refinement was done using alternately Refmac (Murshudov *et al.* 1997) and Phenix (Adams *et al.* 2010) restrained refinement. Phenix restrained refinement with simulated annealing was used early in the refinement to remove model bias. Refined models were analysed for biophysical characteristics using the Robetta virtual alanine scanning server (Kim *et al.* 2004), EBI PDBePISA interface prediction and analysis server (Krissinel and Henrick 2007), the KFC hot spot prediction server (Darnell *et al.* 2007) and the PROTORP interface comparison server (Reynolds *et al.* 2009).

3 Structure of an engineered Obody complex

3.1 Introduction

Structural studies provide unparalleled levels of information about a protein or complex and can be considered essential for an engineering project. Much of the information about major binding domains is based on high-resolution structural data, allowing the researchers to make informed hypotheses about the position and utility of residues, down to conformational bias and atomic composition of an interface. Evolution of new binding proteins raises questions about the nature of the newly evolved face that cannot be answered by sequence data alone. Biochemical and biophysical techniques can provide answers, such as relative stability of engineered domains, individual contribution of residues to binding, affinity and specificity. Crystallographic and NMR protein structures provide atomic-resolution information, allowing the estimation of many of the important aspects of binding, as well as a basis for rationally improving binding with future libraries.

Compared to the number of engineered binding proteins investigated, there are relatively few structures of engineered protein-protein complexes. Hogbom et al claim to have published the first such structure, with an affibody in complex with its wild-type (wt) template protein, the Z domain from staphylococcal protein A (Hogbom *et al.* 2003). Other examples include an Anticalin in complex with CTL-A (Schonfeld *et al.* 2009) and a DARPin in complex with maltose binding protein (Binz *et al.* 2004).

3.1.1 *Preliminary Obodies Work*

Previously, a phage library of randomised Obodies, based on the anticodon recognition domain (which adopts an OB-fold) of aspartyl-tRNA synthetase (aspRS) from *Pyrobaculum aerophilum* (Genbank accession AE009441), was selected for affinity for hen egg white lysozyme (HEWL). For a full description of the initial library design and selection, see the attached manuscript (Appendix A2). This particular domain from aspRS was selected from a panel of 8 OB-fold domains and was therefore denoted paOB3. Briefly, this library randomised 13 positions along the first three anti-parallel β -strands of the OB-fold β -barrel and four residues in a loop between β -strands 4 and

five (4/5 loop), giving a total of 17 randomised positions. Lacking structural data for paOB3, the was domain was modelled using the Swissprot homology modelling server (Peitsch 1995; Arnold *et al.* 2006; Kiefer *et al.* 2009)

The general location of the binding face for the initial library was selected by comparison of the modelled domain with protein and nucleic acid binding OB-folds, specifically staphylococcal enterotoxin B (Swaminathan *et al.* 1992), streptococcal enterotoxin C (Roussel *et al.* 1997), yeast single-stranded telomeric binding domain Cdc13 (Mitton-Fry *et al.* 2002; Mitton-Fry *et al.* 2004) and the N-terminal OB-fold of Asp-tRNA synthetases from *Pryococcus kodakarensis* (Schmitt *et al.* 1998) and *Escherichia coli* (Rees *et al.* 2000). All of these proteins use an OB-fold to mediate some of their interactions, though not always via structurally equivalent residues. In order to create a randomised binding face with minimum perturbation of the structure, thereby preserving the hydrophobic core, specific positions on the OB-fold domain were identified for randomisation as codons on the basis of side chain interactions with its tRNA substrate, as determined using the crystal structure of the *E. coli* aspRS-tRNA complex (Eiler *et al.* 1999).

By restricting the final nucleotide in each randomised codon to only two possibilities (NNK, where K= T/G) a full complement of 20 amino acids was allowed, but the UGA and UAA stop codons eliminated. The third stop codon, UAG or “amber”, was suppressed the by supE phenotype of the TG1 *E. coli* host used for phage display work, substituting a glutamine instead (Inokuchi *et al.* 1979). Single letter nucleotide codes are as per the IUPAC abbreviations (Appendix A1.7) This library was called 13mRL to reflect the number and distribution of randomised codons (13 on the β -sheet (13m), 4 in the randomised loop (RL)) and was panned against various model targets, the most successful of which was HEWL. Selected variants were analysed for affinity using surface plasmon resonance (SPR) and the variant with the highest measured affinity (13mRL L8, 40 μ M) was used in crystal trials in the presence of HEWL.

This chapter presents the crystallographic structure of that Obody in complex with HEWL and analysis of the interface for binding determinants. First, it is shown that the engineered Obody utilises its randomised face to bind

the target protein. Second, the interface is analysed for individual binding components; hydrogen bonds, hydrophobic patches, shape complementarity and binding hot spots. Lastly, the complex structure is used to construct a semi-rational library for the purposes of affinity maturation.

3.2 Results

3.2.1 *Expression and Purification*

The 13mRL HEWL-selected variant L8 was cloned into destination vector pDEST15 (Invitrogen) using the Gateway cloning system (section 2.2.4.4), giving an expression construct with an N-terminal glutathione-S-transferase (GST) fusion tag, cleavable with rTEV protease. The fusion protein was purified by hand on a 5 mL HiTrap GSH column (GE Healthcare) (see section 2.5.4 for method). The eluted protein (typically ca. 20 mL) was pooled and dialysed at 4°C in 6,000 Da cut-off dialysis tubing into 1 L PBS, with rTEV protease (section 2.5.5).

The following day the dialysed sample was concentrated using a Vivaspin 5,000 Da cut-off ultrafiltration device, down to approximately 2.5 mL. The entire sample was filtered through a 0.2 µm syringe filter and re-purified on a Superose 75 16/60 preparative grade size exclusion column, to remove the GST tag and rTEV.

As a final polishing step to remove residual small molecule contaminants, the Obody peak fraction was pooled and concentrated to approximately 1 mL with the same ultrafiltration device as above, then purified a final time using a Superose 75 10/300 analytical grade size exclusion column (Figure 3.1).

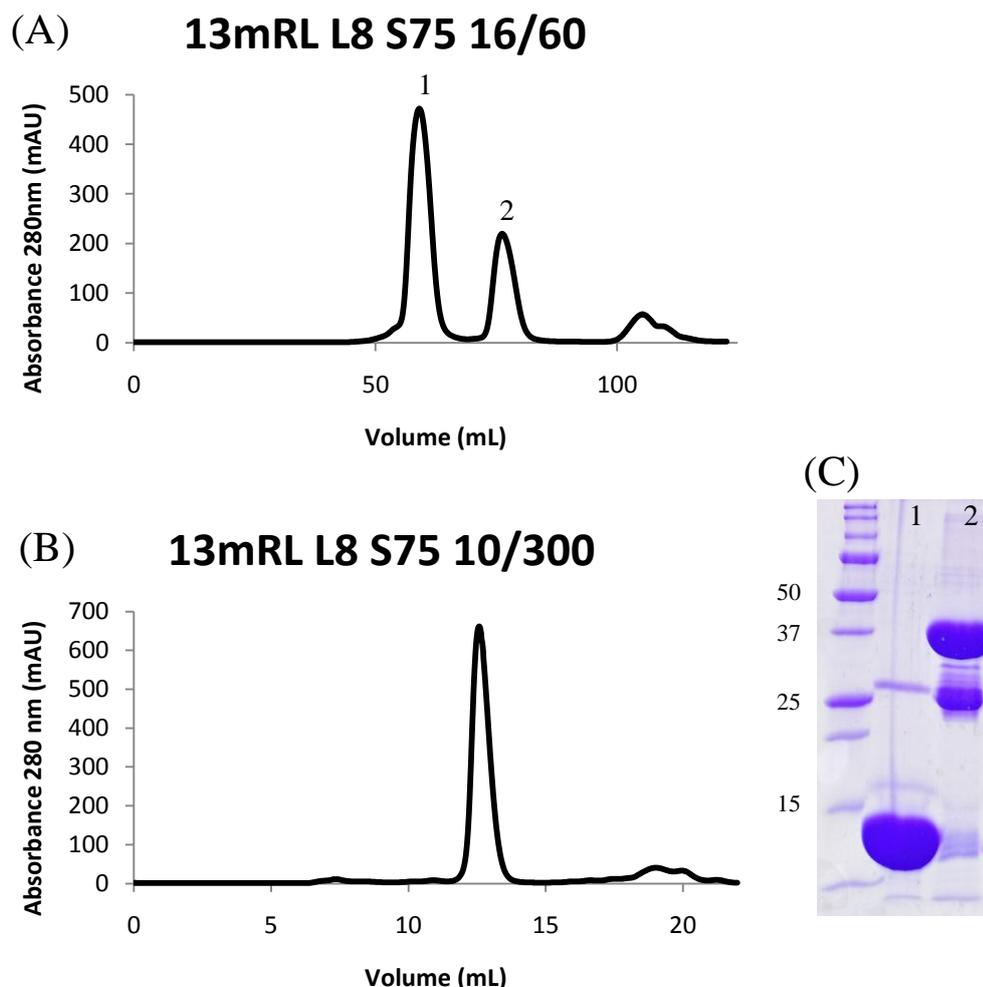


Figure 3.1 Sample purification of 13mRL L8.

(A) Preparative scale size exclusion chromatography trace, showing elution of Glutathione-S-Transferase (GST, peak 1) and the digested Obody (peak 2) after treatment with rTEV protease. (B) Analytical-scale size exclusion chromatography trace of the Obody, re-purified after the preparative scale size exclusion column. (C) SDS-PAGE gel showing purified cleaved Obody (lane 1, expected size 12,500 Da) and intact Obody-GST fusion protein (lane 2, expected size 40,800 Da). Protein markers are labelled in kDa.

3.2.2 Crystallisation and Data Collection

Obody 13mRL L8 was screened for crystallisation conditions at 40 mg/mL in PBS pH 7.4 with equimolar HEWL (Roche/Sigma) in 10 mM sodium acetate pH 4.8, using a 480 condition screen (Moreland *et al.* 2005). To form the complex, the two proteins were mixed at high concentration, with 13mRL L8 at 100 mg/mL and HEWL at 120 mg/mL, then diluted to bring the Obody to 40 mg/mL. A single diffracting crystal was observed in a condition containing the mother liquor 0.2 M HEPES pH 7.3, 7% methyl polyethylene glycol (MPEG) 5000, diffracting to a resolution of approximately 2.7 Å (Figure 3.2).

Using the crystal from the wide screen condition above, a 360° dataset was collected with a phi of 0.5° per image, for a total of 720 images, on the home x-ray source at the Maurice Wilkins Centre. The same crystal was used to collect a second dataset with the same parameters at the Stanford Synchrotron Radiation Lightsource (SSRL). Both datasets were comparable in terms of maximum resolution, but the SSRL dataset was of considerably higher quality across all parameters (Table 3.1). Both showed the same P4₁2₁2 spacegroup and very similar unit cell parameters.

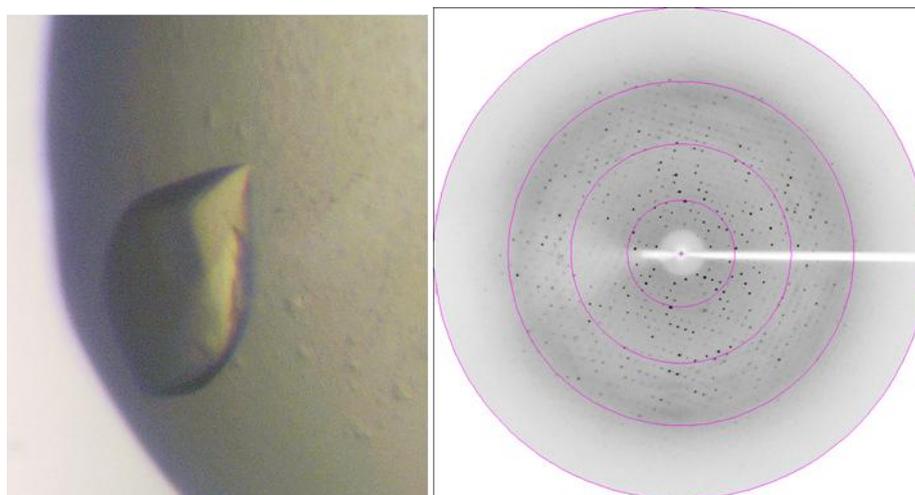


Figure 3.2 13mRL L8-HEWL complex crystal

(left) Grown in 0.2 M HEPES pH 7.3, 7% MPEG 5000 and measuring approximately 20 μm across at its smallest axis from this perspective, this crystal diffracted to $\sim 2.7 \text{ \AA}$ and was used to collect 720 images at the Maurice Wilkins Centre X-ray facility and another 720 at the SSRL beamline. A sample image from the homesource dataset is shown on the right. Starting from the centre of the image, resolution circles mark 10.4, 5.2, 3.5 and 2.6 \AA .

3.2.3 *Model building and Refinement*

The home source data was phased by molecular replacement using HEWL structure 193L and the OB-fold domain of aspartyl-tRNA synthetase from *Pyrococcus kodakarensis* (PDB accession 1B8A, pairwise sequence identity 39%). Using MolRep to look for two copies of each monomer, the phasing yielded an R-factor of 52.7%, although only one copy of lysozyme and two OB-folds were found. As the Matthews coefficient suggested the presence of two copies of an entity the size of an Obody-HEWL complex (2.36 Da/\AA^2 , solvent content 47.8%, $P = 0.99$) and as the single HEWL appeared to be in complex with one of the OB-fold monomers, a second HEWL was introduced

by treating the OB-fold-HEWL complex as a rigid body and superimposing the OB-fold portion of that complex on to the second OB-fold monomer. The two-complex model of the asymmetric unit was subjected to simulated annealing using Phenix, which dropped the R-factor below 50%.

The OB-fold model was replaced with the correct 13mRL L8 sequence by manual residue mutation and refined by alternating rounds of Refmac restrained refinement and Phenix minimisation with simulated annealing. Because of the moderate resolution, strict non-crystallographic symmetry was imposed on the main chain atoms between the monomers of both 13mRL L8 and HEWL. In addition, bond length and angle restriction weightings were increased as moderate resolution data reduced the ability of the model to accurately detect unusual conformations. The final R-free was 34%, though this does not reflect a completely refined structure. The complexes were labelled AD and BC, reflecting the chain ID, where chain A and B are HEWL and C and D are Obodies.

A second dataset from the same crystal, collected at the SSRL, was used to further refine the structure complex derived from the home source data. The complexes were refined as described above, resulting in an R-factor of 22.9% and R-free of 29%. The final model was analysed with PROCHECK and the Ramachandran plot showed no residues in disallowed or generously allowed regions, 11.6% in allowed regions and 88.4% in favoured regions. Refinement statistics are summarised in Table 3.1. Data was cut off at 2.75 Å for refinement due to a deteriorating *R*merge statistic at the highest resolution shell.

Table 3.1 13mRL L8 complex structure statistics

Figures in brackets represent the highest resolution shell

	Home Source	SSRL
Data Collection & Integration		
Space Group	p4 ₁ 2 ₁ 2	p4 ₁ 2 ₁ 2
Unit cell (Å)		
a	76.585	76.759
b	76.585	76.759
c	166.15	166.344
α, β, γ (°)	90	90
Wavelength (Å)	1.54179	0.95666
Resolution Limits (Å)	50 - 2.8 (2.872-2.8)	34.9 - 2.69 (2.76 -2.69)
Measured Reflections	267,952	144,772
Unique Reflections	12,301	16,010
Multiplicity	21	9
Completeness (%)	95.3 (54.75)	99.15 (92.8)
Rmerge (%) †	4.2 (59.0)	7.3 (54.3)
<I/σ(I)>	50.14 (1.6)	32.5 (4.2)
Wilson B (Å ²)	85	65
Mosaicity (°)	0.6	0.6
Molecular Replacement‡		
Correlation Coefficient	66.2	-
R-factor (%)	52.7	
Refinement		
Resolution (Å)	25 - 2.8 (2.87 - 2.8)	27.5 – 2.75 (2.82-2.75)
Reflections	11,629 (479)	12,789 (908)
R _{work} (%)*	26.5 (38.0)	22.9 (26.5)
R _{free} (%)*	34.0 (48.0)	29.64 (37.6)
Free reflections	597 (22)	667 (55)
Refined Atoms	3319	3379
r.m.s.d. bond lengths (Å)	0.012	0.013
r.m.s.d. bond angles (°)	1.541	1.452
<B factor> (Å ²)	66.94	51.15

† $R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - \langle I \rangle|}{\sum I_{\text{obs}}}$

‡ Ref. (Vagin and Teplyakov 1997)

* $R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$

3.2.1 *The Obody-HEWL Complex*

The three-dimensional structure for the complex between the 13mRL variant L8 and HEWL shows in atomic detail the specific interactions between the two proteins. The expression construct sequence was 115 residues in length, including the 109 domain residues, a two-residue extension in the 2/3 loop and an N-terminal four-residue remnant of an rTEV cleavage site. The final model had two Obody-HEWL complexes in the asymmetric unit, with HEWL chains A and B in complex with Obody chains D and C respectively. For each Obody monomer the first 22 residues were not visible; these are not part of the core OB-fold domain. In the full-length aspRS protein, the residues loop around the OB-fold domain and interact with the aspRS catalytic domain. The C-terminus of the core OB-fold is visible, but both monomers lack density for the final 10 residues of the construct. Four residues in the 4/5 loop were targeted for randomisation, but in both of the Obody monomers this loop is very poorly ordered, with only the backbone atoms visible. The selected sequence for these positions in this clone is GVGR, with no density visible for the arginine or valine side chains. In both cases the side chains were added with zero occupancy past C β and set to the most favourable rotamer that resulted in no clashes.

The Obody binds to HEWL by wrapping its binding face around the surface of the protein and inserting the C-terminal end of β -strand 1 and the short loop between it and β -strand 2 (residues 32-35) into the active site cleft of the enzyme (Figure 3.3). Of the 17 residues which were randomised to construct the library, ten are intimately involved at the protein-protein interface. These ten residues combine with a further 8 wild-type residues to form the interface.

3.2.1.1 **The Interface**

The interface between Obody 13mRL L8 and HEWL buries 838 and 805 \AA^2 of the Obody in complexes AD and BC respectively (as determined by PDBePISA (Krissinel and Henrick 2007)), and is evenly divided between atoms with hydrophobic or hydrophilic character. Although individually the C α superposition of the HEWL and Obody monomers in each complex shows effectively identical structures, the angle of interaction is noticeably different.

Overlaying the Obody monomers reveals changes in the cognate HEWL monomers of up to 1.0 Å in portions most distal to the interface, with lesser shifts of as little as 0.3 Å close to the HEWL active site. This corresponds to a binding angle difference of approximately 5 ° between the two complexes. This small degree of flexibility may be a consequence of the modest binding affinity of this complex, but the moderate data quality may also have introduced uncertainties sufficient to account for this difference. Between the Obody β-strand inserted into the HEWL active site is a large unfilled cavity (Figure 3.3E), which probably negatively impacts binding affinity. Although there is no evidence for ordered waters in the cavity at this resolution, it seems likely that they are present.

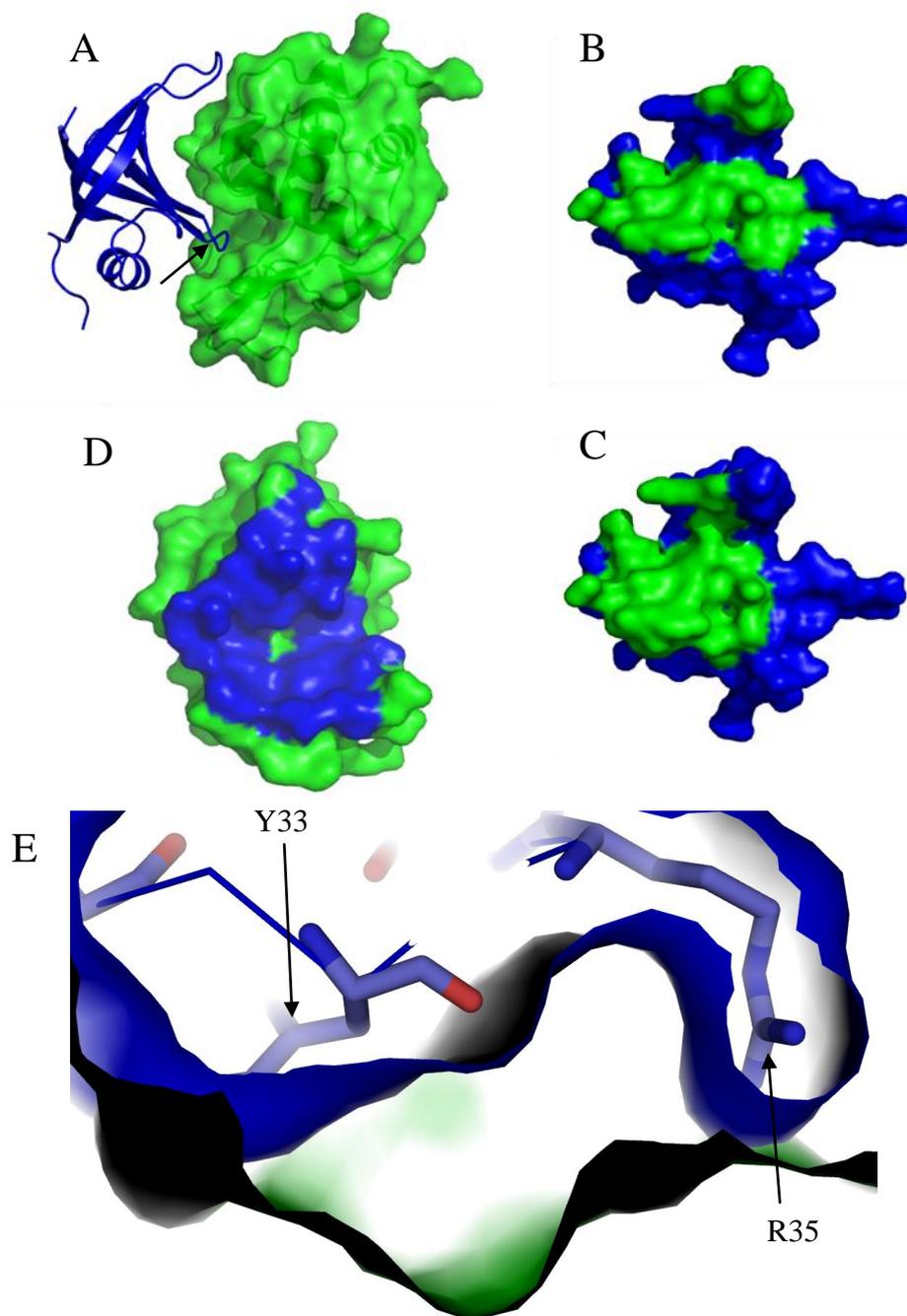


Figure 3.3 Interface Contacts vs. Library Mutations

Schematic representations of the 13mRL L8-HEWL complex interface. (A) Contact by the Obody (blue cartoon) is mediated primarily by β -strand one and two (arrow), which inserts into the HEWL (green, solvent exposed surface) substrate binding cleft. Comparison of the library-targeted Obody residues (B, highlighted green on blue surface model) with calculated contact residues (C, highlighted green) illustrates the extensive involvement of native residues in HEWL binding. The contact residues on HEWL (D, highlighted in blue) show the extent of the binding interface. (E) A cutaway diagram of surface representations of the Obody (blue) and HEWL (green) shows the position of a void at the interface between Y33 and R35.

3.2.1.2 Intermolecular polar bonds

Analysis of the two interfaces using the PDBePISA web server shows two similar but different configurations of binding (Table 3.2). In the AD interface, 10 polar bonds are predicted, involving 7 residues from the Obody (D32, Y33, R35, K40, E55, K86, E93). Amongst the polar bonds are three salt bridges (D32, R35 and E93). The salt bridge formed by R35 is of particular interest as it involves the HEWL catalytic residues E35 and D52. Residues involved in predicted polar bonds form a ring around the hydrophobic centre of the binding face (Figure 3.4). Hydrophobic interactions also contribute to the binding, especially G31, Y33, G34, V36 and I38, and to a lesser extent Y53, F67, K86 and F95. Compared to the AD interface, the BC interface contains net fewer (9) polar bonds, which are accounted for by the loss of bonds to residues K86 (due to a 180° flip of the carbonyl oxygen) and E55 but gain of a single bond to S85. Hydrophobic interactions across the BC face are identical to the AD face. Inspection of polar and hydrophobic interactions across the interface shows a single patch of hydrophobic residues at the centre, consisting of Y33, V36 and I38, surrounded by a rough circle of residues mediating hydrogen and electrostatic bonds (Figure 3.4).

Table 3.2 Intermolecular polar bonds between 13mRL L8 and HEWL

Determined by the PDBePISA server, both chains in the asymmetric unit show a common motif of polar interactions centred on 5 residues at the interface. The remaining three residues listed below (E55, S85 and K86) may represent alternative binding configurations, stabilised into two distinct complexes in the crystal structure.

Residue	Chain C	Chain D	Type†	HEWL Partner(s)
Asp 32	2	2	HS	W63, R61
Tyr 33	2	2	H	D101, N103
Arg 35	2	2	HS	E35, D52
Lys 40	1	1	H	G102‡
Glu 55		1	H	N113
Ser 85	1		H	K116
Lys 86		1	H	N113‡
Glu 93	2	2	HS	R112, K116

† H = hydrogen bond, S = salt bridge

‡ Backbone-mediated interaction

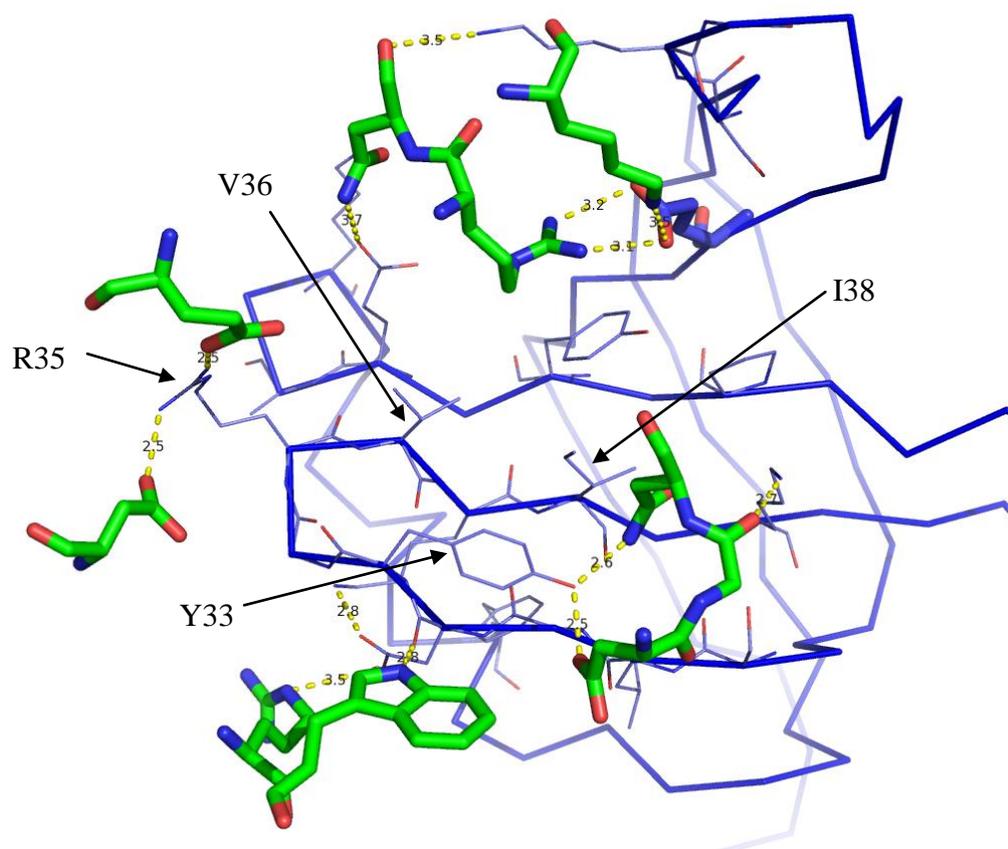


Figure 3.4 Intermolecular Polar Bonds

HEWL residues forming intermolecular polar bonds are shown as green stick diagrams, Obody as a blue $C\alpha$ trace. Obody residues with buried solvent accessible surface area at the interface are shown as blue line diagrams. Polar bonds are indicated by dotted yellow lines, labelled with the bond length. Polar bonds are arranged in roughly a circle around a hydrophobic patch on the Obody at the centre of the binding face formed by Y33, V36 and I38.

3.2.1.3 Hot spots

Analysis of the complexes for “hot spot” residues was performed using Robetta (Kim *et al.* 2004) and KFC (Darnell *et al.* 2007) web servers, which identified residues in each complex as hotspots using three metrics: The KFC server used Fast Atomic Density Evaluation (FADE) for shape complementarity and, independently, the satisfaction of biochemical contacts (Van der Waals polar/nonpolar, hydrogen bonds and electrostatics), while the Robetta server used *in silico* alanine scanning. Residues identified as a predicted hot spot are summarised in Table 3.3 and shown graphically in Figure 3.5. Y33 was identified as the single most important residue at the interface; it was the only position consistently identified as a binding hot spot across all methods used.

Surprisingly, R35 was not consistently picked, despite forming multiple hydrogen and electrostatic bonds with HEWL active site residues. It is also interesting to note that of the four residues most frequently picked as hot spots, two of them were not varied in the library (D32 and E93). D32 makes two separate contacts; one a hydrogen bond to the side chain nitrogen of HEWL W63 mediated by the D32 carbonyl oxygen, the other a salt bridge with HEWL R61. E93 also makes two contacts, to R112 and K116 on HEWL, both of which can potentially form an electrostatic or hydrogen bond.

Table 3.3 Hot spot predictions for the 13mRL L8 HEWL complex

Binding hot spots, calculated using web servers KFC and Robetta, are summarised below (A). Y33 is the only residue picked consistently as a hot spot binding residue across all three models and in both Obody-HEWL complexes. N/P indicates a residue that was not picked as an interface residue in that complex. Average $\Delta\Delta G$ across both Obody monomers as calculated by Robetta is shown for contact residues in (B). Note the negative value for K37, indicative of a negative contribution to binding.

A	Residue	Chain	K-FADE	K-CON	Robetta	B	Residue	$\Delta\Delta G$
							(kcal/mol)	
	Gly 31	C	N/P	N/P	N/P		Asp 32	0.43
		D		Yes			Tyr 33	3.76
	Asp 32	C	Yes				Arg 35	0.49
		D	Yes	Yes			Val 36	0.95
	Tyr 33	C	Yes	Yes	Yes		Lys 37	-0.37
		D	Yes	Yes	Yes		Ile 38	0.82
	Gly 34	C		Yes			Lys 40	0.85
		D	Yes	Yes			Tyr 53	1.09
	Arg 35	C	Yes				Glu 55	0.31
		D					Ser 85	-0.02
	Tyr 53	C		Yes			Lys 86	0.42
		D	Yes				Glu 93	4.21
	Glu 55	C	N/P	N/P	N/P		Phe 95	0.10
		D		Yes				
	Glu 93	C			Yes			
		D			Yes			

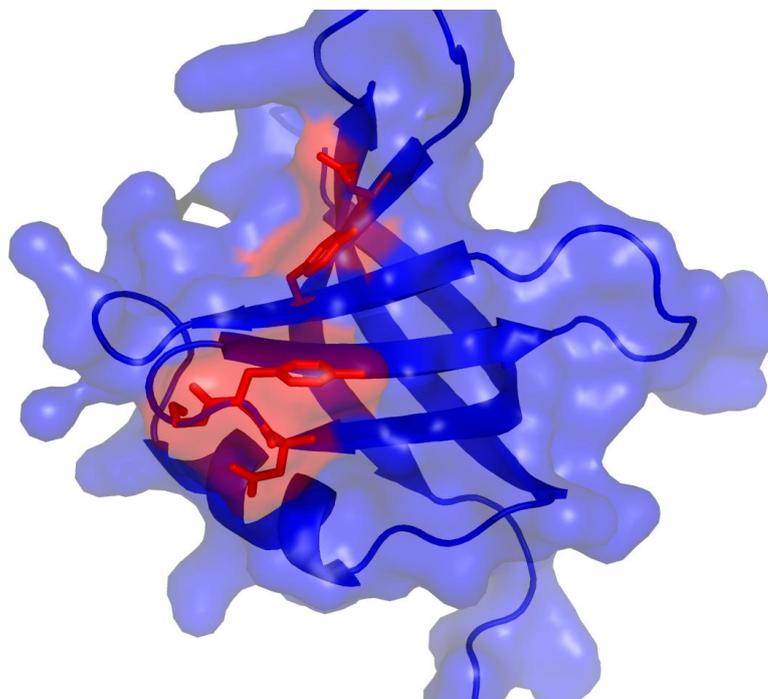


Figure 3.5 *Obody predicted hotspots*

Obody 13mRL L8 chain D, showing the location of hotspot predictions (red). Three predictions are contiguous residues on β -strand 1, which, together with R35, insert into the HEWL substrate binding cleft.

3.2.2 *The Obody-HEWL interface*

Superposition of the two Obody monomers shows little deviation ($C\alpha$ r.m.s.d. = 0.44 Å), with a similar result for the HEWLs ($C\alpha$ r.m.s.d. = 0.11 Å), which reflects the strict NCS adopted for backbone atoms during refinement. When the two Obody monomers are overlaid, the region around E55 and Y53 shows evidence for differences in side-chain conformation, and is generally less well-defined compared to the rest of the interface (Figure 3.6). Y53 shows some evidence for multiple conformations within each monomer, but these were not modelled due to the moderate resolution of the data. Due to proximity, Y53 probably influences the conformational state of E55, which is visible as two different rotamers in the two monomers, despite being very close to the interface and actually participating in complex AD by forming an intermolecular hydrogen bond. The greatest degree of backbone deviations in the Obodies maps to the loops and terminal regions, where 13 out of a total of 86 residues account for all r.m.s.d. values over 0.4 Å. These larger deviations can be at least partially accounted for by poor electron density leading to uncertain placement.

Interestingly, of the three disulfide bonds found in native HEWL, making a total of six in the asymmetric unit, only two, in HEWL chain A, were completely oxidised. Evidence for multiple conformations of the remaining cysteines indicates populations of HEWL monomers with different levels of oxidation in the crystal. It isn't clear what effect this might have on the structure, as pertaining to the Obody-HEWL complex.

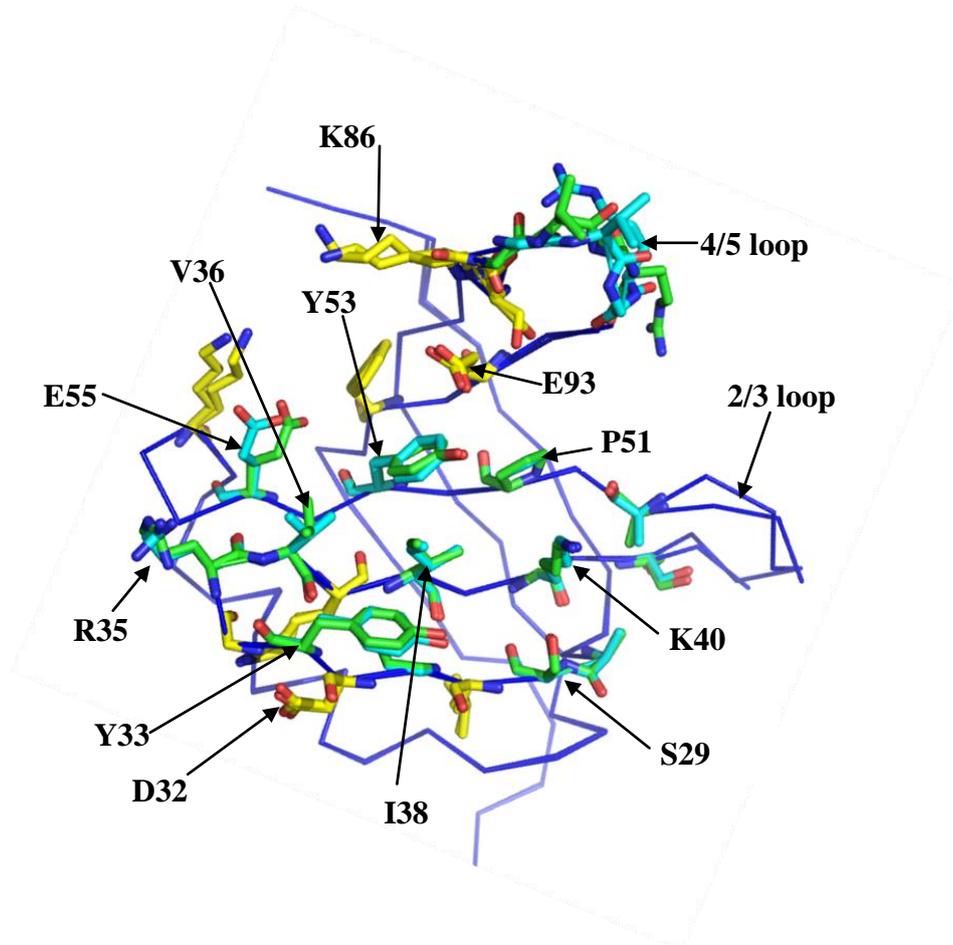


Figure 3.6 *Overlay of the Obody monomers*

The two Obody monomers are shown as blue Ca ribbons superimposed (Ca r.m.s.d. 0.44 \AA). The residues shown in blue (monomer C) and green (monomer D) are those randomised in the 13mRL library. Those shown in yellow are the ancestral sequence, but are calculated to have some solvent exposed surface area buried at the formation of the complex. The regions which contribute most to the Ca deviations are the 4/5 and 2/3 loops. The 4/5 loop was very poorly defined by electron density, suggesting that it plays no direct role in complexation with HEWL, although long-range electrostatic interactions would not necessarily be reflected by electron density. Some electron density indicated multiple conformations of Y53, but the combination of a marginal data:parameter ratio and poor signal for the second conformers lead to the decision not to model them.

3.2.3 Comparison with the initial library model

Because the original library design was based on a modelled template domain and not an experimentally determined structure, the modelled OB-fold domain was superimposed with the 13mRL L8 crystal structure, which shows a close match (Figure 3.7). The two Obody domains overlaid with the template OB-fold give an r.m.s.d. of 1.21 and 1.14 Å for chains C and D respectively. These figures were distorted upwards by large variation in loops, particularly the 4/5 loop. With the 4/5 and 2/3 loops removed, the r.m.s.d. dropped to 0.83 Å. Similarly, the refined lysozyme model compared closely with the starting model, with a r.m.s.d. of 0.2 Å. Some minor differences were observed in side chain positioning at the Obody-HEWL interface.

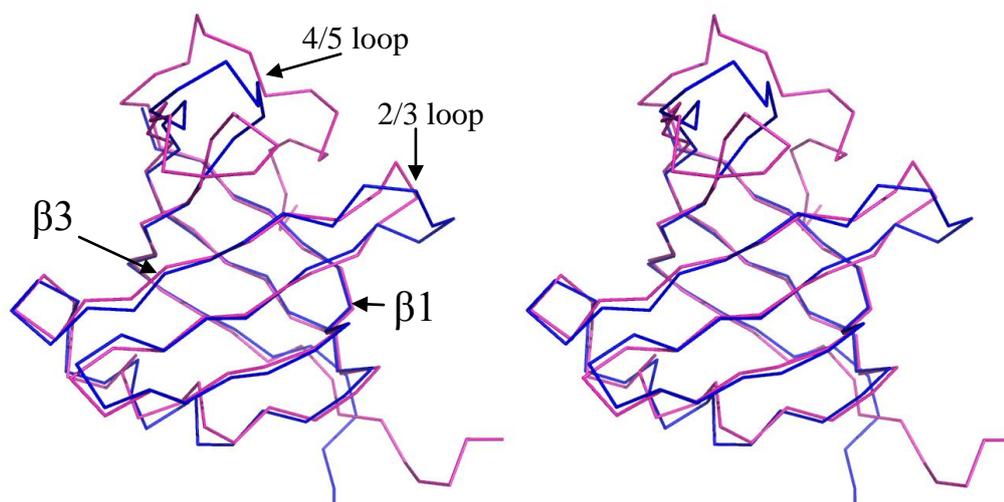


Figure 3.7 Superposition of 13mRL L8 with the library modelled domain

Shown in wall-eye stereo, the starting model of paOB3 wt (pink) was constructed by homology modeling, but overlays the crystal structure 13mRL L8 (blue, chain D) with an r.m.s.d. of 0.83 Å, excluding the 4/5 and 2/3 loops. The domains are oriented to show the three β -strand face targeted for randomisation to the front. Importantly, the residues selected for randomisation using the model were found in the correct position in the crystal structure.

3.2.4 Comparison with Other HEWL complexes

The natural bacterial inhibitor of HEWL, YkfE, binds on the opposite face of lysozyme when compared to the Obody. However, there is a surprising intersection of the inhibitory histidine from YkfE and the Obody R35 which

reaches into the active site of HEWL and bonds with the catalytic residues Glu35 and Asp52 (Figure 3.8A).

The binding of the camelid single domain antibody to HEWL also has some similarities to the Obody-HEWL complex. A long CDR3 loop from the camelid antibody binds into the active site cleft of HEWL and presents ~70% of the binding interface. This CDR3 loop has been compared to substrate binding by HEWL (Transue *et al.* 1998) and there are striking similarities between the substrate and atoms in the CDR3 loop providing clear details for the mechanism of inhibition. This is similar to the C-terminus of $\beta 1$ and the short loop between $\beta 1$ and $\beta 2$ strands of the Obody, which inserts into the active site cleft and allows R35 to reach out and H-bond with the active site acidic residues of HEWL. The relative C α positions of Y103 from the camelid antibody and Y33 from the Obody is striking (Figure 3.8B), as well as the similar hydrophobic interactions of Y103 and V36. While the Obody loop does not penetrate the active site cleft to the same degree as the camelid antibody, there are nevertheless sufficient similarities between this Obody loop and the camelid loop to suggest that efficient inhibition would be possible with higher affinity binding. Indeed, the larger Obody surface area presented to HEWL by way of the OB-fold binding face and the flanking variable loop (4/5 loop) points towards attainable tighter binding variants. Attempts to measure inhibition of HEWL by a binding Obody by digestion of killed Gram-positive bacteria (Parry Jr. *et al.* 1965) were inconclusive, probably due to the low affinity of this complex.

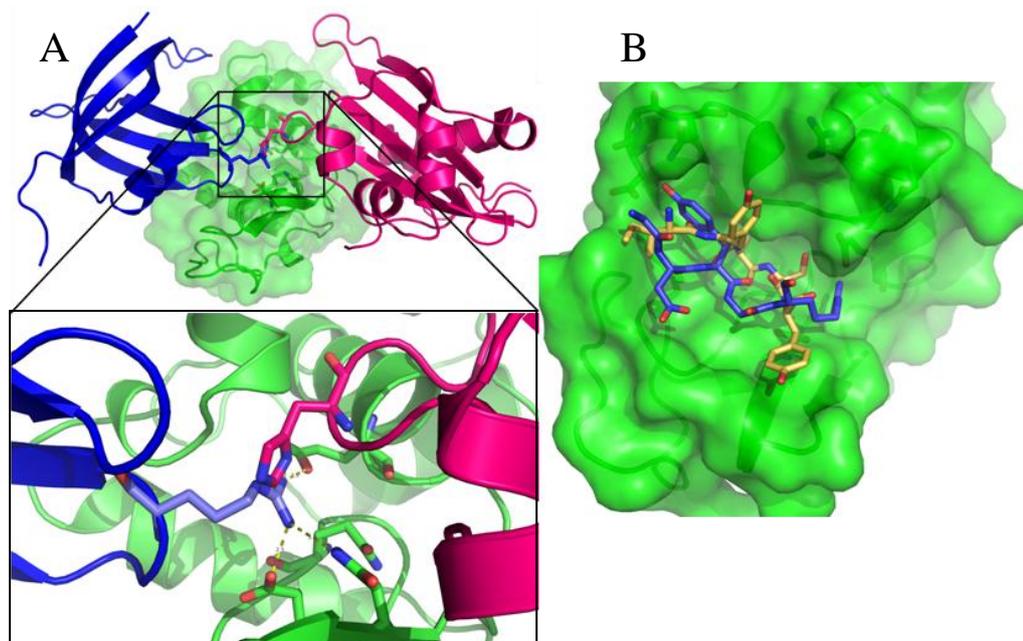


Figure 3.8 Comparison with other known HEWL binders

(A) The lysozyme inhibitor protein YkfE from *E. coli* (Monchois *et al.* 2001) is shown in pink, Obody in blue and HEWL in green. The overlay of R35 and the inhibitory histidine from YkfE is highlighted in the zoomed inset picture. (B) Overlay of the CDR3 loop (yellow) from an inhibitory camelid antibody (Transue *et al.* 1998) with the Obody peptide (blue) in the active site cleft of HEWL (green).

The Obody-HEWL interface is in line with typical antibody-antigen interfaces (in terms of buried surface area, number of H-bonds and the gap-volume index), as shown in Table 3.4. This raises the question of the modest affinity of the Obody-HEWL interaction in comparison to 3-4 orders of magnitude greater affinity of antibody-antigen interactions, despite similar properties of the interaction interface. First, the fact that interfaces with similar properties achieve much higher affinities suggests that the Obody-HEWL interface can also be modified to generate much tighter binding. Secondly, it points towards the presence of unfavourable interactions decreasing the affinity between the Obody and HEWL when compared to other lysozyme complexes. To address this, I have inspected the interface for unfavourable interactions and have identified two residues that may be responsible for lowering the affinity of the complex, as well as two other non-contacting residues which, if re-randomised and selected, may produce beneficial interactions to further stabilise the complex.

Table 3.4 Comparison of Obody-HEWL interface

These data were calculated using the PROTORP server (Reynolds *et al.* 2009). The salt bridge count was defined as the number of charged residues on the Obody with an appropriately charged HEWL residue within 4Å. Gap volume index is the ratio between buried surface area and accumulated volume of gaps at the interface, as defined by Jones and Thornton (1996).

	13mRL L8	Inhibitor YkfE, 1GPQ	Camelid Ab 1JTP	Mouse Fab 1FDH
Buried surface area*	840 Å ² (14%)	796 Å ² (11%)	800 Å ² (11%)	647 Å ² (3%)
H-bonds	10	10	8	12
Salt bridge interactions	3	2	0	0
Polar:Non-polar atoms %	43:57	47:53	28:72	51:49
Gap Volume Index[‡]	2.94	3.06	2.22	3.27
K_d	36 μM	~1 nM	50 nM	22 nM

*Average antibody/antigen buried surface area = 950 Å²

‡Average Gap Volume Index (antibody/antigen) = 3.0

3.2.5 Structure-Based Library Design

Based on the structure presented here, residues were identified and targeted for affinity maturation. The primary focus was the 4/5 loop; none of the randomised residues in the loop contact HEWL and it is poorly ordered. The four residues from the original 13mRL library were re-randomised, and two additional random codons were also added, bringing the total loop residues to six.

On the face itself, four positions were identified; S29, K37, P51 and A56 (Figure 3.9). S29 and A56 were identified as residues with small side chains close to the interface that did not make any significant contacts with HEWL. Re-selection of these positions was expected to result in bulkier residues with more bonding potential. P51 was found in the middle of the third β-strand of the binding face. The assumption was made that, as proline is a β-breaking residue (Chou and Fasman 1974), this may de-stabilise the barrel. However, the close match to the template model domain, which lacked a proline at that position, suggests that this may not be the case.

K37 was in close proximity (3 Å) to R61 from HEWL in both crystallographic complexes, although no polar or hydrophobic interactions were predicted between the two. At neutral pH (crystals were grown in pH 7.3, phage display selection was performed at pH 7.4) both lysine and arginine

would be protonated and positively charged, suggesting that K37 may be detrimental to overall binding. Examination of the structure showed the majority of the K37 side chain buried in a highly hydrophobic region at one end of the β -barrel, with only the terminal amine group solvent exposed. This suggested that some of positive effects on binding could be attributed to intramolecular interactions, by fulfilling hydrophobic contacts on the opposite side of the binding face.

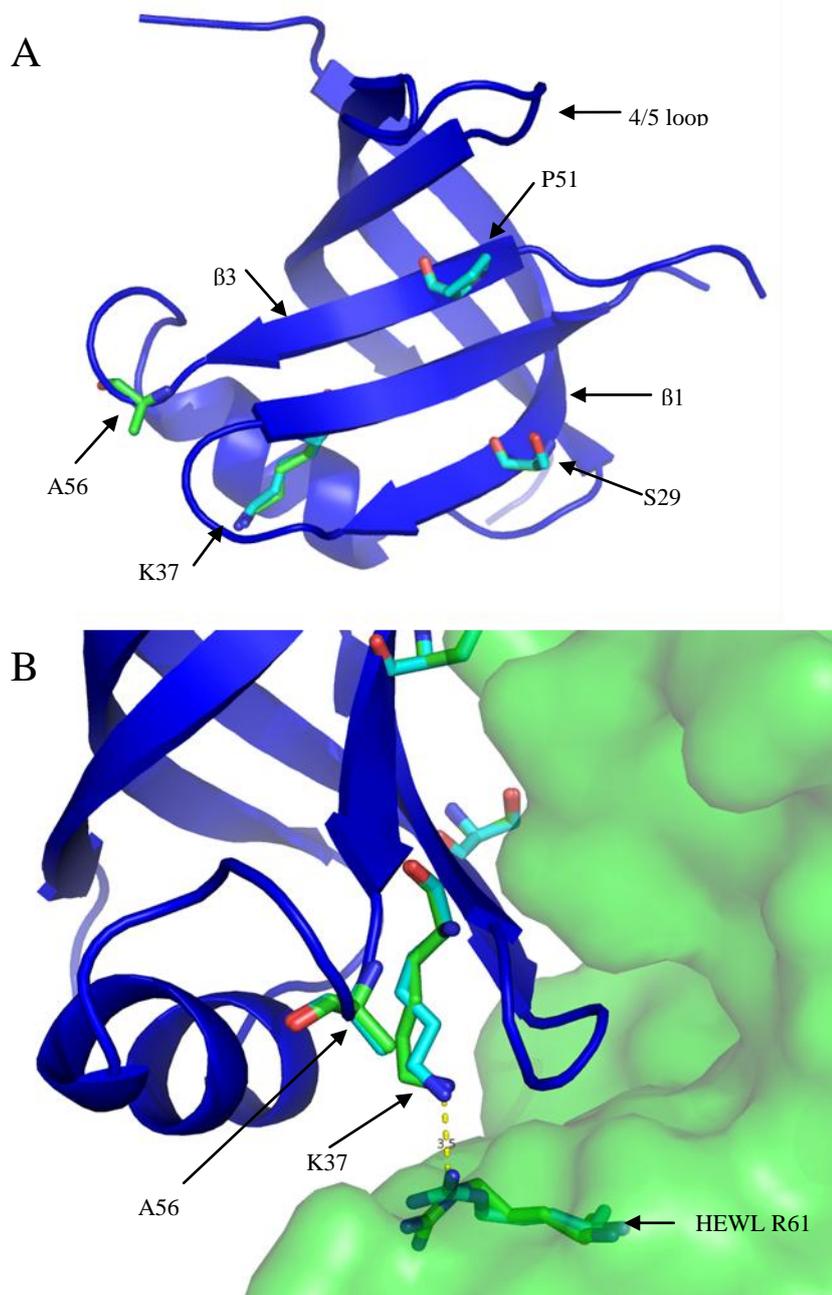


Figure 3.9 Selected face residues for the L8 10m library

Residues at the binding face identified as candidates for re-mutation for affinity maturation. A cartoon of Obody monomer chain D is shown in dark blue, the cognate HEWL is shown as a green surface. Residues

are shown as light blue stick diagrams (complex BC) or green (complex AD). (A) S29 makes no bonds to HEWL (surface representation), although a large polar patch is close. P51 was assumed to be potentially destabilising to the Obody itself as a β -breaker. (B) K37 appears to interact with R61 from HEWL (shown as green Ca trace), though no hydrogen bonds are predicted. A56 made no contacts with HEWL (surface representation) in either complex in the asymmetric unit and was not part of the original complement of mutated residues.

3.3 Discussion

3.3.1 *Structural Validation of Library Design*

The 13mRL L8 structure demonstrates without doubt that engineering of the presented Obody was successful. The domain uses the engineered face to bind to its ligand and all of the targeted residue side chains are located on the exterior of the domain, as predicted by the homology model used to design the 13mRL library. Indeed, the fact that the crystal structure Obody domain so closely resembles the homology model is testament to the high degree of structural conservation in OB-folds.

The interface itself conforms to the general description of protein-protein interfaces: a hydrophobic centre, surrounded by a series of polar interactions, salt bridge interactions for long-range orientation and a buried surface area that can be compared to interface averages (total of $\sim 1600 \text{ \AA}^2$ from both partners) from a protein-protein interaction survey (Lo Conte *et al.* 1999). Based on the presence of two non-crystallographic conformers, the interface appears to have two stable arrangements, though moderate data quality precludes certainty on this matter. Binding is focused on three major residues; Y33 is by far the most important and is the only residue to be consistently picked as vital across all hot spot analysis models employed. Surprisingly, though it makes several interactions with HEWL acidic active site residues, R35 was not picked as a hot spot consistently. This can probably be explained by the low degree of geometric complementarity it exhibits, due to both the unfilled cavity shown in Figure 3.3 and its location on the edge of the interface.

In addition to the randomised library residues, there is extensive contact with HEWL by native residues, especially by those in β -strands 1 and 2, and, to a lesser extent, at the end of β -strand 4 and beginning of β -strand 5. Indeed, two of the three residues with predicted salt bridges to HEWL are native

Obody residues. As electrostatic interactions are thought to be the first “filter” for attraction that precedes tight binding (Schreiber and Fersht 1995), it appears that selection of the mutant face enriched for a composition that relied heavily on the non-varied periphery of the interface. This is hardly surprising, given the cooperative nature of the residues on the surface of β -barrels in general, although it does raise certain issues about limitations on targets with the current face design. The salt-bridge contributions by native residues are all acidic, which makes sense considering that HEWL has a pI of ~ 9 , so at pH 7.4 possesses a net positive charge. It does however raise the possibility that, with the current engineering scheme, targets with an acidic pI may be at a disadvantage. Consequently future naïve libraries may need to consider including D32 and E93 as randomised residues to broaden the applicability of the scaffold.

3.3.2 *Affinity maturation*

The structure was used as the basis for selecting residues for further mutation in an attempt to improve binding affinity by semi-rational design. Two residues, P51 and K37, were identified as negatively impacting binding. The remaining binding face sites were selected for the straightforward reason that they don't make contacts with HEWL, but are in close enough proximity that a different residue may do so and thereby contribute to binding. The 4/5 loop in particular can potentially form a large contact area if a complementary sequence is selected, and towards this end the four residues in the loop were selected for inclusion in an affinity maturation library. Because the loop is constrained by the extensive contacts made by flanking residues K86 and E93, it may be that failure of the loop to take direct part in binding is because it simply cannot reach across the interface. Consequently, extension by two residues may promote discovery of loop residue-mediated interactions, while allowing existing contacts to persist unaltered.

At the interface there is a large cavity, presumably filled with water, between R35 and Y33. Lack of complementarity at this point probably imposes an affinity penalty, especially considering that Y33 is picked as the single most important residue at the interface. The cavity is formed because of the β -sheet nature of the Obody peptide binding into the relatively irregular HEWL

substrate binding cleft. Increasing geometric complementarity at this point has the potential of drastically increasing affinity.

3.3.2.1 Amelioration of negative contributions

K37 has a side-chain extending inwards towards the hydrophobic core and away from HEWL, but exposes its amine group to solvent adjacent to the underside of β -strand 1. It was the most obvious candidate as simple examination showed a like-charge clash with an arginine from HEWL. In support, Robetta virtual alanine scanning showed that substitution for alanine actually resulted in a reduction in free energy of binding. Significantly, K37 was not part of the original complement of library residues. Its inclusion in the HEWL contact table illustrates the nature of the selected binding; the interface doesn't use the randomised face uniformly, and makes contact with HEWL using residues from both sides of the β -sheet inserted into the substrate binding cleft.

P51 was a less certain inclusion. In principle a proline in the centre of a β -strand prevents continuous formation of β -sheet character. However, while this may actually destabilise the fold itself to some degree, its impact on binding isn't altogether clear, as the presence of a proline imposes unique conformational restrictions that may be beneficial for binding. Of course, re-randomising the residue allows for the possibility of re-selection, but since the goal of affinity maturation is optimisation of the binding face, removing the restrictions on surrounding residues imposed by P51 will potentially increase the conformational search space of the library.

The following chapter deals with affinity maturation attempts based on the structure presented here, in addition to an alternative naïve library, to improve Obody binding for the model protein target HEWL.

4 Gene Library Design and Selection

4.1 Introduction

The criteria outlined by Skerra (2000) and Binz & Pluckthun (2005) for selecting a scaffold candidate for investigation are largely met by the oligosaccharide/oligonucleotide-binding (OB) fold; a small single domain with a compact hydrophobic core, a binding area available and competent for mutation and lack of disulfides. However, to validate it as a viable protein scaffold, a candidate domain must be demonstrated as capable of tolerating the mutations necessary to develop demonstrable affinity for a model target. Previous work has shown mutational tolerance and preliminary selection for affinity to hen egg-white lysozyme (HEWL) using a phage display library based on the anti-codon recognition domain of the aspartyl-tRNA synthetase from *Pyrobaculum aerophilum* (Appendix A2). The moderate affinity achieved ($\sim 40 \mu\text{M}$) while adequate as a proof of principle, it is necessary to show tighter binding to bring the Obody scaffold up to a standard commensurate with other engineered proteins.

4.1.1 *Outline*

This chapter describes the experiments performed with the goal of producing an Obody with nM-range affinity for model target HEWL. First, the template gene was altered to facilitate more efficient combinatorial assembly. Second, two structure-based libraries were investigated, building iteratively on information gained during each step. Third, departing from structure-based library design, an affinity maturation library was investigated using randomly distributed mutations instead of site-directed mutations. Lastly, a new format of naïve library was produced. This library attempted to address the problem of the transformation bottleneck by limiting diversity at each mutant codon. The results of each of these experiments are presented and discussed below.

4.2 Results

4.2.1 *Library nomenclature*

Gene libraries were named according to the number of mutations, where in the gene the mutations are located and the template gene on which each library was based. Clones isolated during screening were suffixed with an alpha numerical code to indicate the panning stage and clone number. The general format is *Template Gene: Number of Mutations: Clone Number*. The naming scheme is summarised in Figure 4.1.

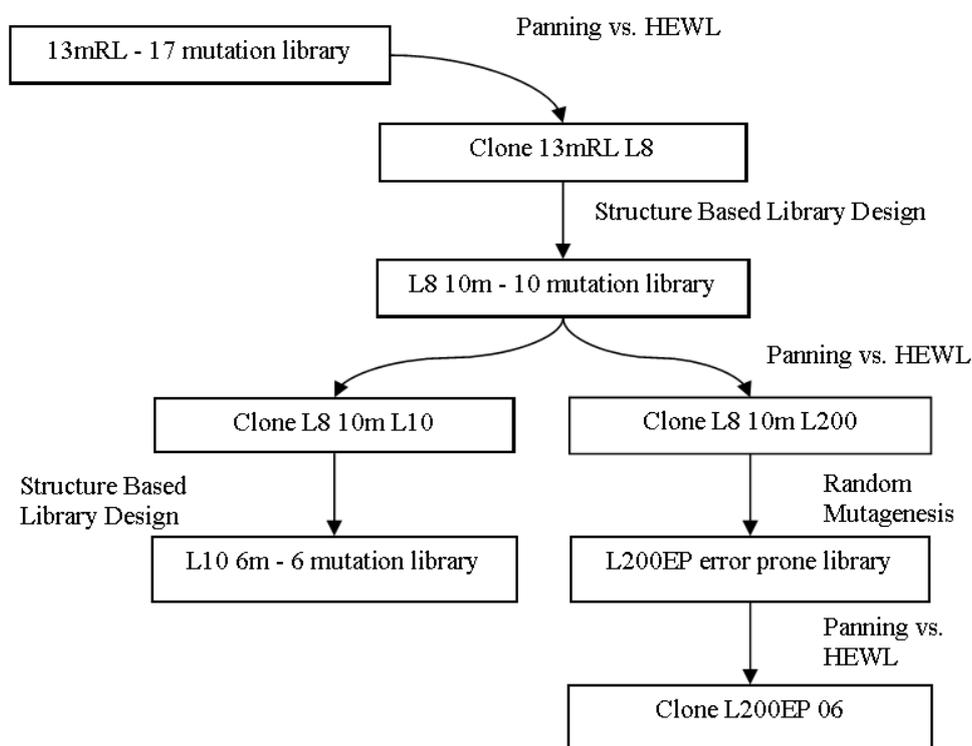


Figure 4.1 *Library nomenclature conventions*

Each library was named according to a systematic convention, as outlined above, consisting of the template gene clone number and the number of mutations. The exception is naïve libraries based on the wild-type domain from *P. aerophilum*, where the gene name has been omitted. Following screening of each library, clone numbers were assigned according to their status. Clones extracted from HEWL-panned libraries were labelled “L” followed by a number, similarly for RNase A-panned libraries with an “R”. Unselected variants examined during library quality control were prefixed with a “U” instead.

4.2.2 *Assessment of phagemid pRpsp2*

To get an estimate of the efficiency of the ligation procedure using phagemid pRpsp2, test ligations were performed. Of concern was the rate of religation of digested vector, efficiency of vector double digestion with NcoI and

NotI and efficiency of successful ligation of an Obody insert cut with the same enzymes.

4.2.2.1 Double-Digestion Efficiency

To test for the rate of vector re-ligation, eight ligations were run in parallel, as summarised in Table 4.1. These data showed that digestion of the plasmid was efficient, as no colony forming units (cfu) were observed from transformation of any of the cut vector samples that were not ligated. Where ligase was present in the reactions, there was significant vector self-ligation, as shown by the large cfu count obtained using untreated vector without the presence of insert. This effect was completely eliminated when the digested vector was dephosphorylated by SAP treatment, although it did seem to result in a reduction in efficiency when insert was present to be ligated. When taken together, these data suggested that while virtually all plasmid in the digestion was cut at least once, it seemed to be a mixed-species population of doubly- and singly-cut. Consequently, a quantitative measure of the efficiency of ligation was needed in planning the appropriate ligation scale for library production.

Table 4.1 Effect of dephosphorylated vector on transformation

1 μ L of each ligation reaction was transformed into DH5 α and the colonies counted were extrapolated in to the total number of transformable plasmids per μ L of ligation reaction. The results show that using dephosphorylated vector in ligation allows the assumption to be made that 100% of cfu contain plasmid with an insert.

	Vector	Insert	No Insert
Ligated	Dephosphorylated	2940	0
	Untreated	8120	12000
Dummy	Dephosphorylated	0	0
	Untreated	0	0

4.2.2.2 Insert ligation efficiency

Ligation efficiency was estimated by comparison of the number transformants obtained from a ligation with the maximum number possible, where the maximum assumed that all plasmid in the ligation was transformable. Cell competency was calculated in cfu per ng of plasmid by transformation of a known quantity of pRpsp2.

A standard 20 μ L ligation was constructed using pRpsp2 cut with NcoI and NotI and dephosphorylated with shrimp alkaline phosphatase, mixed with

a wt paOB insert cut with the same enzymes. The ligation contained 116 ng of digested pRpsp2 (5.8 ng/ μ L). After incubation, 1 μ L of the ligation (5.8 ng of plasmid) was used to transform an aliquot of electrocompetent DH5 α cells with a known efficiency of 600 cfu/ng of pRpsp2. This transformation yielded 1061 cfu, which equated to 1.8 ng of transformable plasmid per μ L of ligation, giving a total of 35 ng per 20 μ L ligation. Comparison with the total plasmid input gave a 31% ligation efficiency rate. As ligations using dephosphorylated plasmid but containing no insert did not yield any colonies, the assumption was made that all transformants obtained in this manner contained an insert. Also, colony PCRs using plasmid annealing primers routinely showed that an insert of the correct size was present in 100% of colonies tested.

4.2.3 *paOB3 wt template modification*

In order to facilitate the PCR assembly method chosen for generation of the gene libraries, modifications were made to the template paOB3 wt gene in two stages. A visual summary of the process is displayed in Figure 4.2. First, the wt with an extended loop (wtEL) was produced, with a 6 bp insertion (GGCGCG) between the second and third β -strands of the binding face. The insertion was introduced to allow for a sufficient overlap between the two mutational oligonucleotides used to mutate the three β -strands of the binding face. The wtEL intermediary was constructed by PCR assembly of fragments 005/056 and 068/006. Second, three silent point mutations were introduced into paOB3 wtEL to improve amplification efficiency when using overlapping primer pair 058 and 059, reducing self-annealing, hair-pins and melting temperature (T_m). The modified template was labelled paOB3 wtELMC. Subsequently the primers 058 and 059 were replaced with 168 and 169 respectively, which annealed at the same position, but incorporated the sequence changes to significantly improve gene fragment PCR. After sequencing testing for successful display as a pIII fusion, paOB3 wtELMC was used as the template for all subsequent libraries.

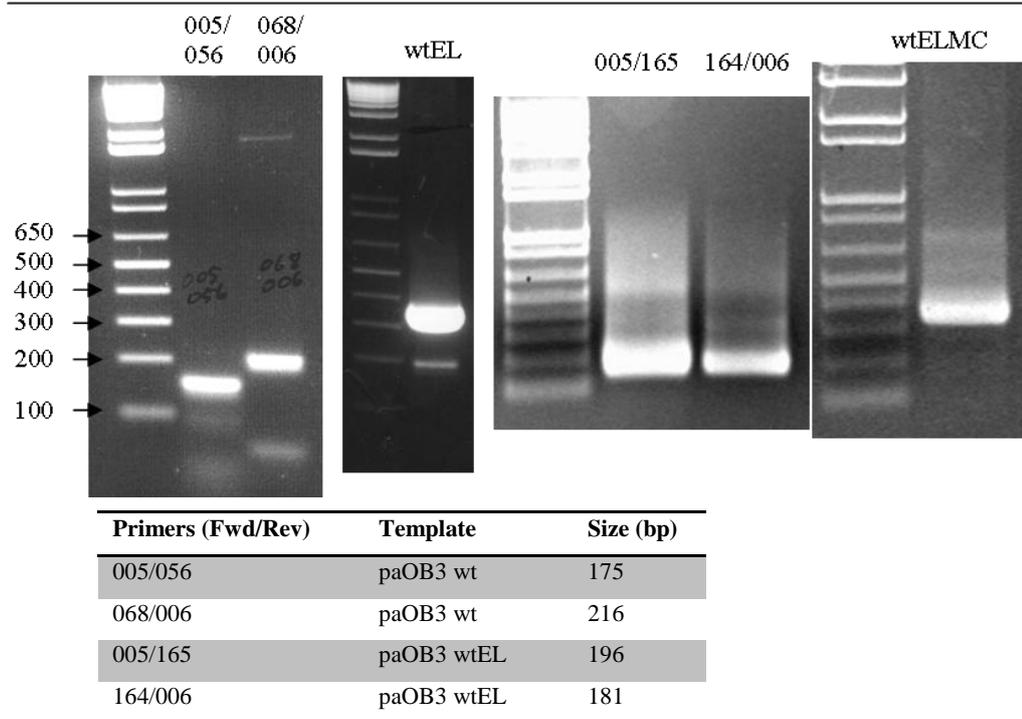
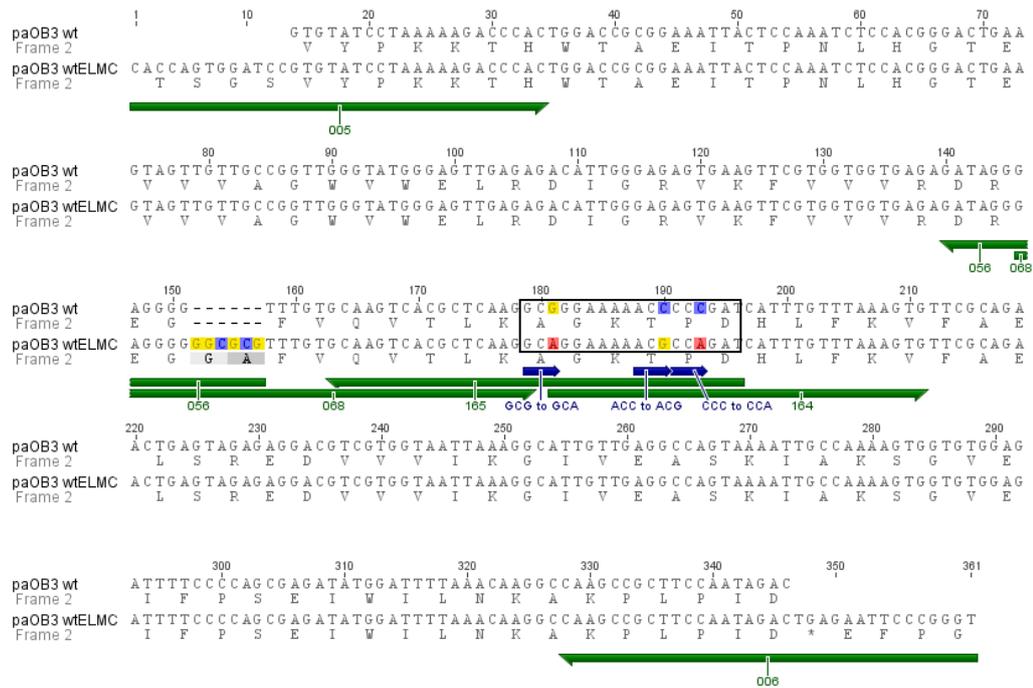


Figure 4.2 paOB3 template modification

(Top) Alignment of paOB3 wt with sequenced paOB3 wtELMC. Primers used in the modification process are shown in green, altered nucleotides are shown coloured according to the base. The annealing site for primers 058/059 and 168/169 is indicated by a black rectangle. The single nucleotide substitutions are highlighted with blue arrows and labelled with the change each represents. All three mutations, made in the third positions of three codons, are silent. (Middle) Agarose gels of the fragments generated, labelled by the primers used in the amplification, and the intact wtELMC gene insert. (Bottom) Table of gene fragments use in the modification process, and their predicted size, in bp. The intact gene has an expect size of 362 bp.

4.2.4 *Structure-based affinity maturation – L8 10m*

In order to improve affinity for HEWL, the structure of 13mRL L8 was used as the basis for selective re-randomisation of residues at the binding face and one flanking loop (section 3.2.5).

To recapitulate briefly, four residues in the binding face were selected for randomisation. Serine 29 and alanine 56 were targeted because they are both on the periphery of the face and make no contact with HEWL. While most selected positions were granted a full complement of amino acids with NNK codons, in the case of serine 29 the randomisation was limited to a subset of residues with an NRK codon (see appendix A1.7 for IUPAC abbreviations). This reduced the maximum theoretical diversity to 12 possible amino acids and placed emphasis on large, polar residues. K37, which contacts HEWL on the periphery of the interface, appeared to make an unfavourable contact and was therefore targeted. P51 was also selected on the assumption that proline is a β -sheet-breaking residue and the domain as a whole may be de-stabilised by its presence on the third β -strand of the binding face. Finally, the 13mRL L8-HEWL complex structure showed that the loop randomised in the original 17-mutation library made no contacts with HEWL and was poorly ordered. Consequently this loop was targeted for complete re-randomisation of the four original positions and extended by two additional residues to six randomised codons, making a total of 10 mutant codons over all. This new library design was therefore labelled L8 10m.

4.2.4.1 **Library design and assembly**

The mutant gene library was produced and assembled according to the general principles laid out in section 2.4.2. Long synthetic oligonucleotides 183 2m, 184 2m and 163 6m were designed to introduce the mutant triplets at the selected positions, along with the various primers needed to dissect the gene properly to incorporate the mutational oligos (Figure 4.3), based on the template gene 13mRL L8.

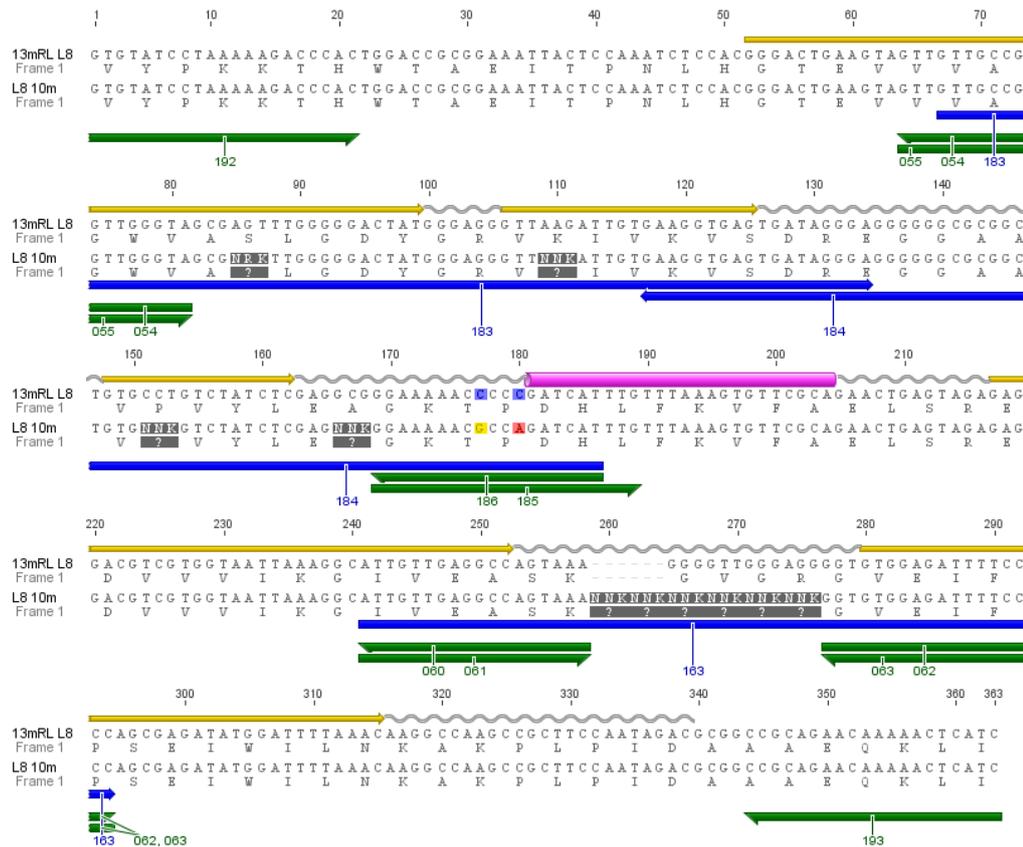


Figure 4.3 L8 10m library design

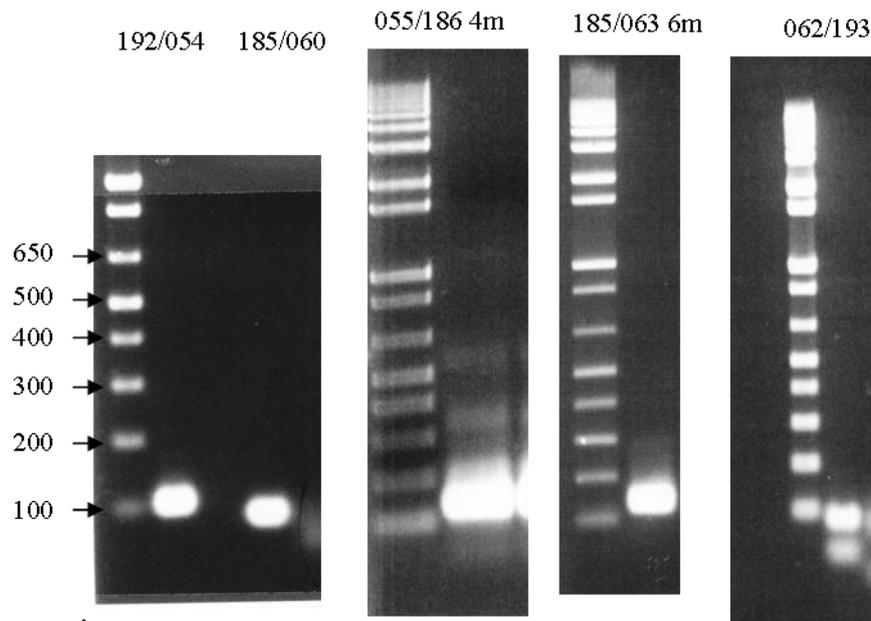
A schematic of the various primers used in PCR assembly of the L8 10m library. Short (18-21 bp) flanking primers are shown in green, long mutational oligos are shown in blue, with randomised codons highlighted in grey. Secondary structure annotations are shown in yellow (strand), pink (helix) and grey (coil). Nucleotide changes between the template gene and library plan are highlighted according to the base shown.

Single-stranded oligos 183 and 184 were combined in equimolar amounts (10 pmol each) in a 25 μ L PCR reaction and subjected to two cycles in a thermocycler, allowing each oligo to be extended by the other as a template. The cycles were as follows: 3 min at 94°C (denaturation), 1 min at 60°C (annealing), 1 min @ 68°C, 1 min at 80°C, 1 min at 60°C, 1 min at 68°C. This resulted in the generation of no more than 10 pmol of a double stranded segment incorporating all four mutant triplets, labelled 183/184 4m. The whole 25 μ L reaction was then used as a template for a larger-scale (100 μ L) reaction to amplify correctly assembled 183/184 4m using flanking primer pair 055 and 186 to yield the 055/186 4m fragment.

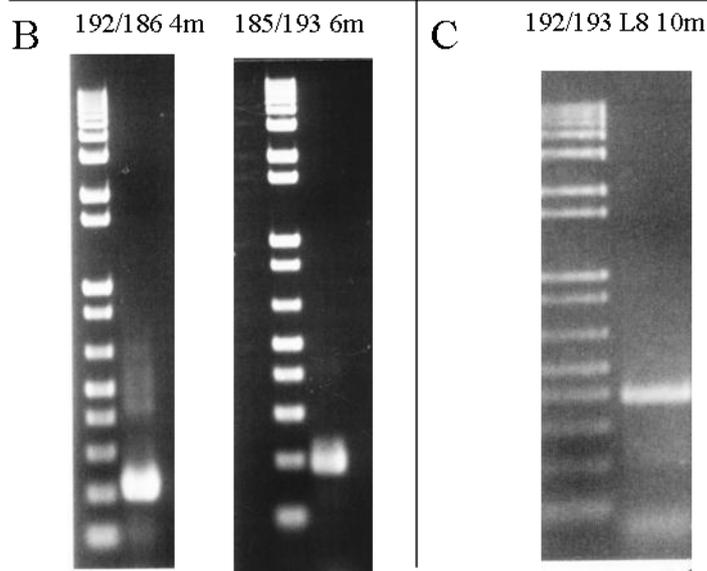
The 163 6m oligo was incorporated into the gene library by generating a complementary strand with primer 063, amplifying with primer pair 061/063 and gel purifying by freeze and squeeze.

The intact gene was assembled from PCR fragments in the following steps: 192/054 + 055/186 4m = 192/186 4m; 192/186 4m + 185/060 = 192/060 4m; 061/063 6m + 062/193 = 061/193 6m; 192/060 4m + 061/193 6m = 192/193 10m. The maximum diversity for this library was 5.6×10^{12} , therefore at each stage of the process, the theoretical minimum amount of fragment needed in each reaction, to maintain full coverage of the diversity complement, was maintained. The various fragments generated and assembled during this process are shown in Figure 4.4.

The library was assembled, gel purified, digested with restriction enzymes NcoI and NotI then test-ligated into pRpsp2. Positive clones were plasmid prepped and sequenced as the unselected representatives. A 2 mL library-scale scale ligation and transformation was performed as per section 2.4.3, with a resulting library size of 10^7 .



A



Fragment	Template	Size (bp)
192/054	paOB3 wtELMC	104
183/184 4m	-	120
055/186 4m	183/184 4m	123
185/060	paOB3 wtELMC	90
163/063 6m	-	54
061/063 6m	163 6m (ss oligo)	54
062/193	paOB3 wtELMC	87

Figure 4.4 L8 10m library gel fragments

(A-C) Agarose gels of the various gene fragments generated during the library assembly process, labelled for the flanking primers in the PCR reaction. (A) Fragments amplified directly from the template DNA, (B) partially assembled library in two fragments (209 bp and 195 bp, left to right) and (C) the final full-length library (387 bp). DNA standards, labelled in bp, are the same for all gels, with 100bp as the lowest band shown in every case. A list of the primary fragments used and their predicted sizes is also presented (D).

4.2.4.2 Panning

Phage library preparation and panning was performed as described in section 2.4.3, with helper phage Vcsm13. Ligand concentration for immunotube adsorption was varied through the rounds, beginning at 1mg/mL HEWL in rounds 1 and 2, 100 μ g/mL in rounds 3 and 4, with the lowest concentration of 10 μ g/mL in rounds 5 and 6. This allowed initial low-stringency selection from a high-diversity, low copy number population, followed by increased competition for binding sites to select for the best members of the positively selected population (Xu *et al.* 2002). Phage were eluted with HEWL at 1 mg/mL. Analysis of the panning showed positive selection for binding, as well as for displayed Obody (Figure 4.5).

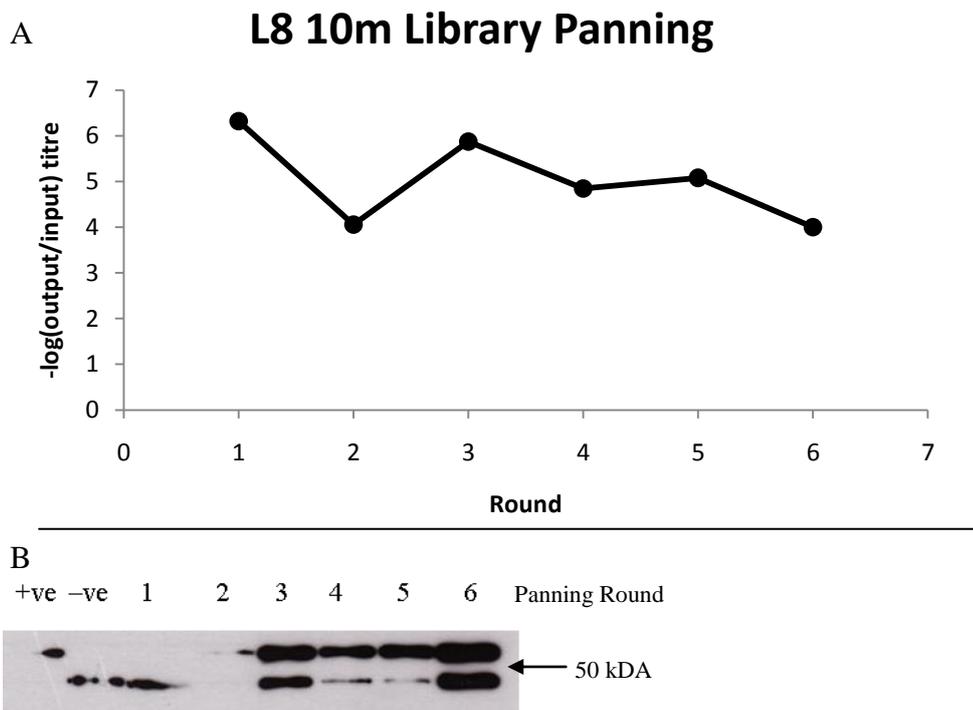


Figure 4.5 Panning results with L8 10m phage library

(A) Enrichment for binding phage is shown by the decreasing $-\log$ ratio of input:output phage titre over 6 rounds of panning. Significant enrichment is seen by round 2 because the library was based on a pre-selected lysozyme binder. The sudden increase in the ratio at round 3 reflects a reduction in immobilised ligand, with re-enrichment in the subsequent rounds of a population of tighter binding Obodies. (B) Enrichment for displayed Obody by western blot of input phage from each round. The western shows each sample as one or two bands. Controls are cMyc-tagged pIII with (+ve) or without (-ve) an N-terminal paOB3 wt fusion. Enrichment is seen in an increase in the upper band compared to the lower band, indicating a higher proportion of phage displaying intact Obody-pIII fusion product. Approximate position of a molecular weight marker is indicated by the arrow (not visible on the western).

4.2.4.3 Selected Clone Analysis

Although the library size (10^7) was adequate for a non-naïve library, sequencing of unselected variants revealed that the region of the gene covered by the mutagenic primers 183 and 184 contained numerous point mutations and deletions (see Appendix A1.4.3.1 for sequences). As a consequence, the practical library size was estimated downward to 10^6 . Sequencing of individual clones from round six gave five unique sequences in 25 clones analysed (Figure 4.6). What is most striking about this selection is the prevalence of reversions back to the ancestral residue at positions S29 and K37. Proline 51 was also re-selected in two of the six variants, and large residues (Tyr, Arg) were selected in three of six variants at position 56. There is virtually no consensus in the 4/5 loop residues, indicating that any contributions to binding at these positions were probably subordinate to changes at the binding face.

Position	29	37	51	56		4/5 loop residues					
13mRL L8	S	K	P	A	-	G	V	-	-	G	R
L05	N	K	S	Y	-	E	P	H	G	Q	D
L06	S	Q	P	L	-	W	T	D	E	L	M
L10	S	K	S	Y	-	A	A	D	M	H	N
L18	S	K	D	A	-	S	V	A	A	N	A
L200	H	M	P	R	-	Q	Y	G	S	D	T

Figure 4.6 Unique sequences from the L8 10m library

A schematic showing the mutant positions in the clones selected for affinity to lysozyme, as compared to ancestor 13mRL L8. Residues changes are highlighted, coloured according to character (yellow = non polar, blue = basic, red = acidic, green = polar).

Each of the unique genes was cloned and expressed in pProEx Htb as an N-terminal His-tag fusion protein and purified by IMAC (section 2.5). The His-tag was removed with rTEV protease and the cleaved Obody was re-purified by a two step gel-filtration process, with a preparative-grade 16/60 S75 column followed by an analytical grade 10/300 S75 column (GE). All Obodies purified showed identical purification profiles. Yield was between 10-15 mg of pure protein from 1 L of 2YT media (see Figure 5.1 for representative purification data).

Affinity was determined by SPR analysis of saturation binding, based on a 1:1 Langmuir binding model. It was found that inclusion of a linear non-specific component in the curve greatly increased the quality of the fit. It

seemed reasonable to include this factor, given that the high concentrations of Obody used in these measurements (up to 100 μM in most cases) most likely exacerbated an electrostatic non-specific effect, which would not be accounted for by subtraction of the reference flow cell. This non-specific effect can be seen in the data as an apparent failure to plateau at higher concentrations, the gradient of which was closely matched by the linear portion of the fitted model in all cases. The measured affinities ranged between 4 and 8 μM (Table 4.2, Figure 4.7 and Figure 4.8). SPR measurements were corroborated by ITC of a single variant, L8 10m L200. To elucidate the nature of the modified binding face, the crystal structure of L8 10m L10 in complex with HEWL was solved to a resolution of 1.95 \AA (section 5.2).

Table 4.2 Affinity for HEWL of L8 10m library variants

The affinity of selected variants, as determined by SPR analysis, including the linear non-specific portion, represented by the gradient of the line as calculated by GraphPad Prism with Equation 2.1

Variant	K_D (M)*	Linear NS†
L05	$6.9 (\pm 0.4) \times 10^{-6}$	6.7×10^{-6}
L06	$8.2 (\pm 0.5) \times 10^{-6}$	2.4×10^{-6}
L18	$7.1 (\pm 0.6) \times 10^{-6}$	2.2×10^{-6}
L200 (SPR)	$6.2 (\pm 2.5) \times 10^{-6}$	4.1×10^{-6}
L200 (ITC)	$4.1 (\pm 1.6) \times 10^{-6}$	-
L10	$4.6 (\pm 1.4) \times 10^{-6}$	9.8×10^{-6}
L10 (+10 mM glycerol)	$6.2 (\pm 1.9) \times 10^{-6}$	6.6×10^{-6}

*Errors given are 95% confidence intervals for the model fitting only

SPR Raw Data for L8 10m L200

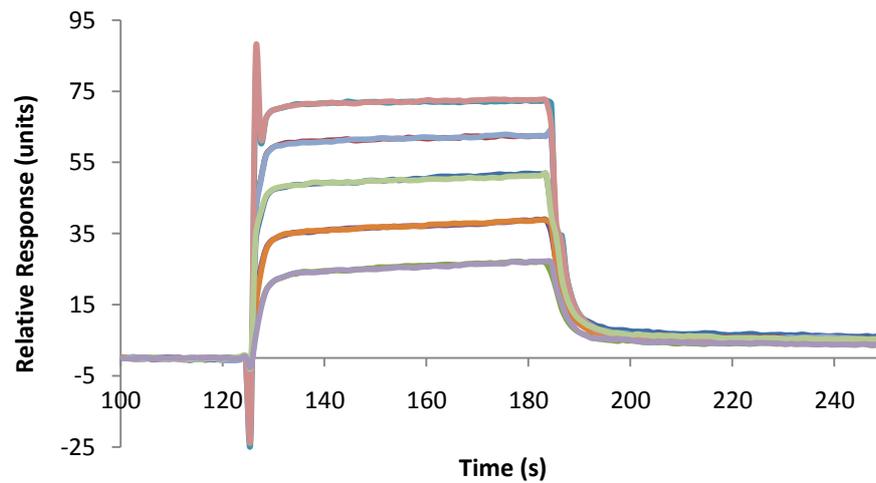


Figure 4.7 Raw SPR data

A sample of the raw output SPR data from a Biacore analysis run of L8 10m L200. The graph overlays five relative response curves at different concentrations of Obody, each sample measured in triplicate.

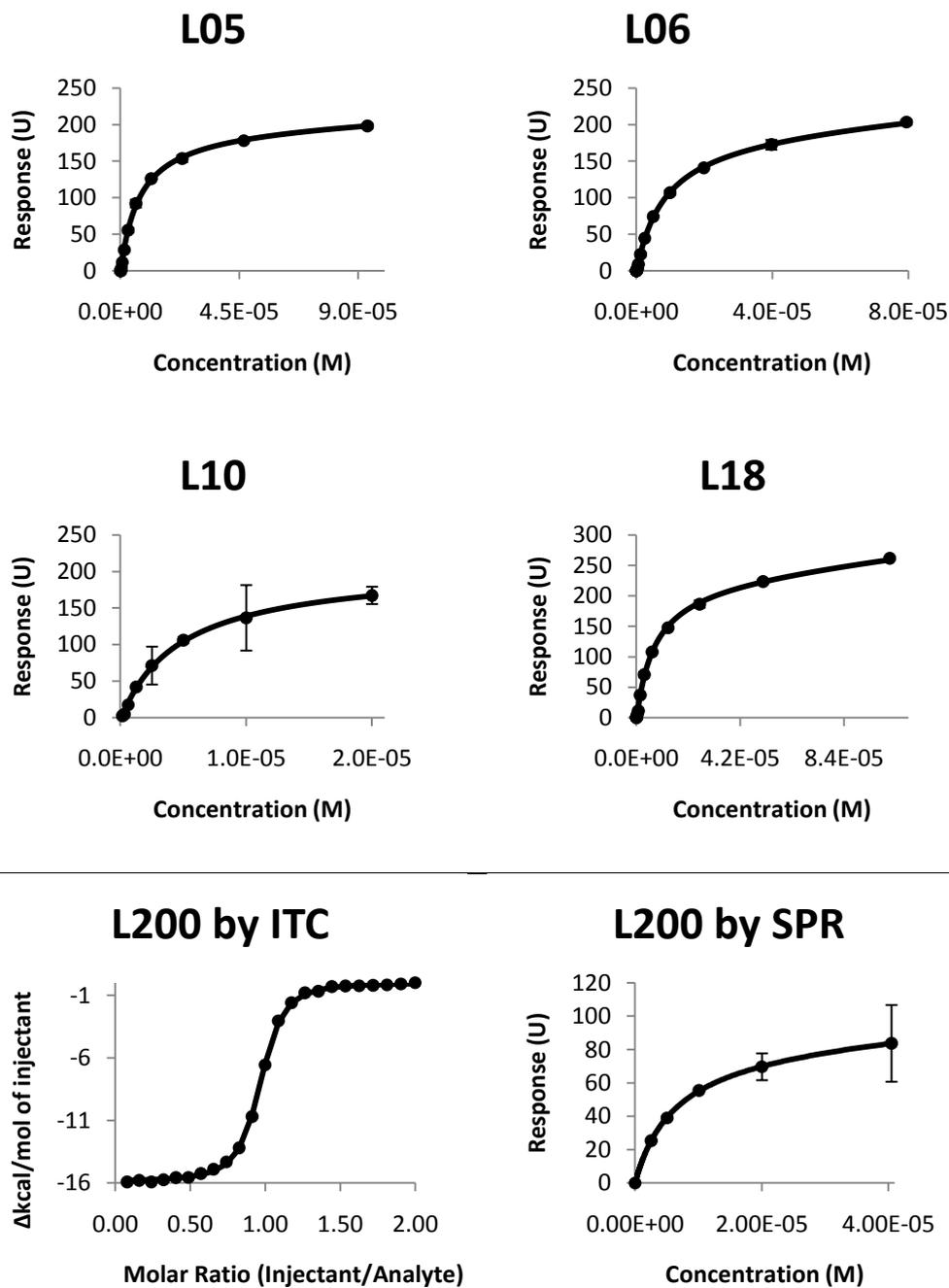


Figure 4.8 Affinity analysis of variants from the L8 10m library

Biacore analysis was performed as described in section 2.4.6.3, based on 1:1 binding stoichiometry and includes a linear non-specific component. Error bars show a 99% confidence interval for the mean of each data point. ITC analysis was performed as described in section 2.4.7, no error bars are shown as each data point is a single measurement.

4.2.4.4 Glycerol as a binding co-factor

Although the L8 10m library improved binding affinity by approximately 10-fold, persistence of a cavity at the interface not filled by the Obody was still considered a problem (Figure 3.3). This cavity was hypothesised to be a significant limitation to the specificity and affinity of this Obody lineage to HEWL, so was attempted to be filled by a small molecule. Glycerol was selected, as it possesses a combination of hydrophobic and polar atoms thought necessary to fill the cavity. The affinity of L8 10m L10 was re-measured with 10 mM glycerol included in the Obody buffer, but no significant change in affinity was observed (Table 4.2).

4.2.5 Structure-based affinity maturation – 6m library

Following from the library panned in the previous section, a new library was designed to try to fill the interface cavity. This structure-based library was designed using the crystal structure of L8 10m 10 in complex with HEWL, as presented in section 5.2. The library was conceived to attempt to fill a cavity at the interface with Obody atoms instead of waters. The loop between β -strands 1 and 2 (1/2 loop) was re-randomised. Y33, which forms the lower edge of a hydrophobic patch, was nominated the N-terminal constant boundary. The randomised region extended from G34 to I38, retaining K37 and adding two additional codons to the 1/2 loop to produce a 6 mutation library denoted L10 6m (Figure 4.9). K37 was not re-randomised in this library as it had been re-selected as lysine in the L8 10m library. It was thought that expansion of the loop and removal of the interaction made by R35 would allow a new peptide with greater complementarity to be selected.

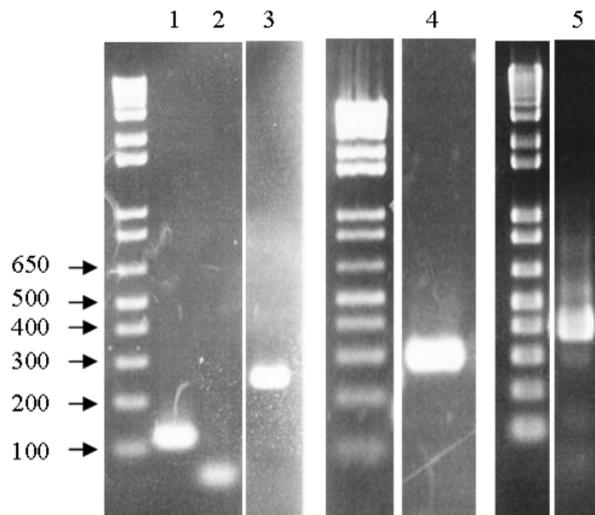


Figure 4.9 L10 6m library design

Based on the sequence of the L10 variant from the L8 10m library, this gene library was designed to alter the binding around R35, which reaches into the HEWL active site. Primer annealing sites and direction are shown in as numbered green arrows, the mutational oligo is shown in blue.

4.2.5.1 Gene Assembly

The L10 6m library was assembled by combinatorial PCR, as described in general principles in section 2.4.2, from three PCR generated fragments (Figure 4.10). Mutational oligo 211 was incorporated by double stranding with primer 210, then amplification with primer pair 207/10 to yield the fragment 207/210 6m. The assembly order was as follows: 207/210 6m +209/192 = 207/193 6m; 192/208 + 207/193 6m = 192/193 6m. Representative unselected variants are shown aligned in appendix A1.4.4.



	Fragment	Template	Size (bp)
1	192/208	L8 10m L10	140
2	207/210 6m	211 6m	57
3	209/193	L8 10m L10	249

Figure 4.10 L10 6m library assembly

The assembly process is summarised in a table (Top) outlining the primary fragments and agarose gels showing the progression of fragment assembly (bottom). The numbers in the first column of the table correspond to the numbering of the first three lanes of the gels. Lanes 4 and 5 show fragments 207/193 6m (expected size 288 bp) and the intact gene 192/193 6m respectively (expected 393 bp). Each fragment should be compared to the DNA ladder to the left. Where a space is visible between the sample lane and the ladder, the gel contained other unrelated samples which were excised from the picture.

4.2.5.2 Panning

The assembled library was cloned into pRpsp2 as per section 2.4.3 and yielded a 10^7 transformant library, compared to a theoretical diversity of 6.4×10^7 . Panning was done using a constant 100 $\mu\text{g}/\text{mL}$ HEWL in immunotubes over four rounds, as per section 2.4.4. Panning showed successful enrichment HEWL-binding phage (Figure 4.11) but sequencing of selected clones revealed that only the ancestral L8 10m L10 was present, with no mutant variants selected. This line of investigation was not pursued further.

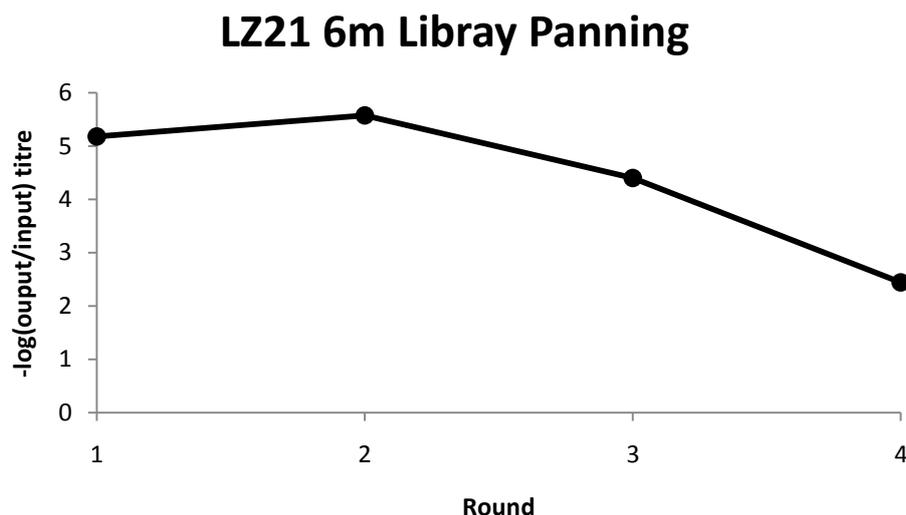


Figure 4.11 LZ21 6m library panning titre

The graph shows enrichment for positively selected HEWL-binding phage library members over 4 rounds of panning by plotting the ratio of input:output phage as a negative log. The reducing value from round to round indicates relative increase in output phage counts compared to input phage.

4.2.6 *Random Mutation Library*

To discover mutations that improve affinity without specific structural information requires a random set of mutations distributed throughout the gene. To this end, a single variant (L8 10m L200) was subjected to error-prone PCR mutation (see section 2.4.2.3 for method). Although the actual number of mutations per gene is difficult to quantify without deep sequencing, it can be estimated at an average of 1-3 substitutions per variant (Cadwell and Joyce 1992), though template dependent factors like GC content can cause this to vary widely (Pritchard *et al.* 2005). Sequencing of unselected variants showed an acceptable level of mutation in line with this estimation (Appendix A1.4.5.1). The library, named L200EP, was ligated into phagemid pRsp2 and transformed to yield a library of 10^8 transformants, then panned against HEWL.

4.2.6.1 **Panning**

Phage library panning was done in immunotubes, beginning with 100 $\mu\text{g}/\text{mL}$ HEWL in round one and dropping to 10 $\mu\text{g}/\text{mL}$ in subsequent rounds. This panning experiment was the only one to use trypsin sensitive helper phage KM13 (Kristensen and Winter 1998). Consequently, output phage samples were treated with 1 mg/mL trypsin after specific elution with 1 mg/mL

HEWL in PBS. Panning showed a 100-fold enrichment in retained phage counts after a single round. Results are summarised in Figure 4.12.

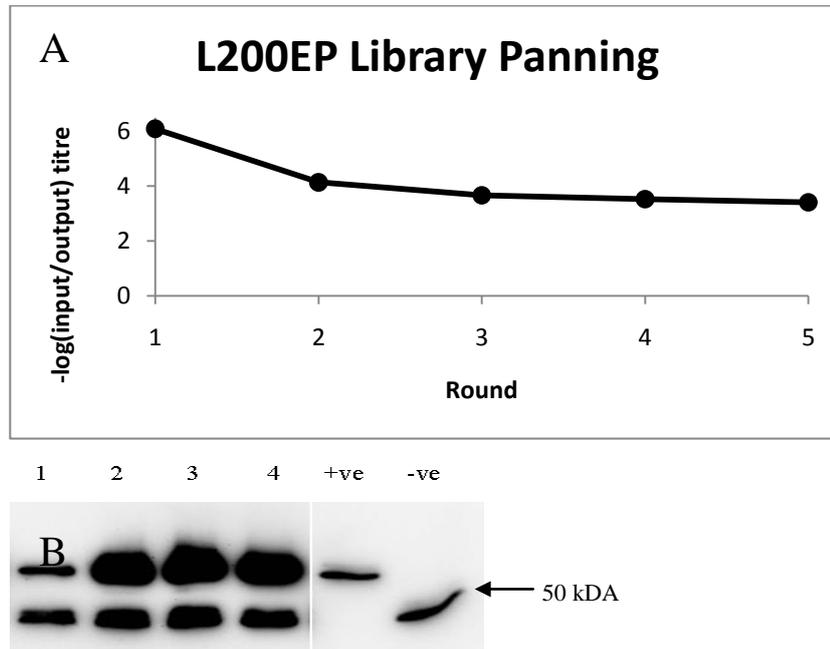


Figure 4.12 *L200EP library panning titre*

Two indications for enrichment for HEWL binding. The graph (A) shows successful positive selection for HEWL binding phage from the error-prone PCR library based on 10m L200, over 5 rounds, by visualising a reduction in the ratio between input and output phage titres. A western blot (B) shows an increase in displayed Obody-pIII fusion (upper band) compared to pIII without an Obody fusion partner (lower band) in input phage samples over four rounds of panning. Controls are cMyc-tagged pIII (-ve) and paOB3 wt-cMyc-pIII fusion (+ve). The position of a 50 kDa marker is indicated by an arrow.

4.2.6.2 Sequencing data

Individual variants were isolated from output phage samples from rounds three and five for sequence analysis (protein sequences are shown in Figure 4.13. The full gene sequences can be found in Appendix A1.4.5.2). The distribution of mutations in the selection genes appears non-random, as illustrated in Figure 4.13. One position in particular, V23F, stands out as virtually fixated. However, it is possible that this mutation is merely an artefact of the error-prone PCR process; incidence of this mutation occurs with high frequency in round three clones, but only in a single variant in round five. It isn't located on or near the binding interface, although that in itself doesn't preclude influence on binding.

Most of the mutations located directly in the binding regions of the protein can be rationalised as trivial by examination of the crystal structure. For

example, K58R is moderately conservative and would present a similar profile to ancestor type; arginine and lysine are a similar size, are both positively charged at pH 7.4, they can both mediate hydrophobic interactions and there are no aromatics available on HEWL at that position for an arginine to stack with.

Although it only occurs in less than a quarter of the sequences, substitution of M37 for lysine or arginine is interesting, considering that it was one of the original residues selected for randomisation in the L8 10m library. This mutation makes sense if the like-charge clash has been eliminated by rearrangement of the corresponding arginine, as in L8 10m L10.

The 4/5 loop sequence showed the largest cluster of mutations located in an interface region, specifically at position 86, but also 88. Given that K86 made extensive hydrophobic contacts with HEWL in the 13mRL L8 and L8 10m L10 structures, mutation of this position almost certainly has impact on binding. Six variants from round five were selected to sample the all of the selected diversity in this region to express and measure affinity by SPR.

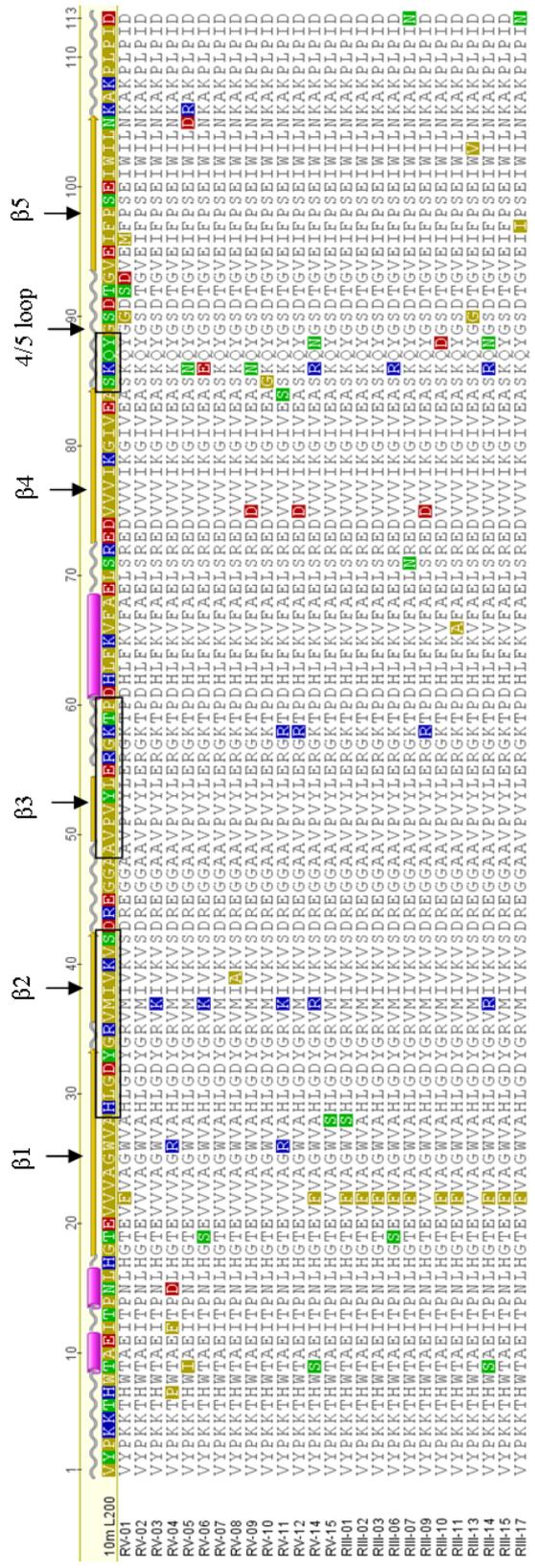


Figure 4.13 Mutation clustering in L200 EP variants

Shown above are sequences selected from the L200EP library, from round 5 (RV) or round 3 (RIII). The first sequence in the list is ancestor type L8 10m L200, which has been annotated to show the position of secondary structure elements (pink cylinders are helices, light brown arrows denote β -strands). HEWL-interface residues are outlined with black rectangles. The mutant sequences are listed below the ancestor and are highlighted according to residue changes, with reference to the ancestor. Colours are by residue character (Blue = basic, red = acidic, green = polar, yellow = non-polar).

4.2.6.3 Selected variant affinity analysis

Sequencing of individual clones from various rounds did not reveal convergence towards a particular sequence, although this was to be expected as most mutations would have little effect on the binding. Certain positions were over-represented, indicating their importance, but this may also indicate dominance of particular mutations in the unselected population from early incidence during the EP-PCR. Mutations were clustered particularly in the 4/5 loop and adjacent residues, less so in the non-binding face β -strands and very infrequently in the residues directly involved in binding (Appendix A1.4.5.2) for sequence data). Six variants from round five were selected on the basis of maximum diversity at those semi-conserved positions for expression and affinity measurement by SPR (Figure 4.14, Table 4.3), cloned into expression vector pProEx Htb and purified by IMAC and gel filtration. For a representative purification, see 5.3.1. Yield was 10-15 mg of purified protein per litre of culture for all variants except L200EP 06, which gave 45 mg per litre.

However, due to sequencing errors, one variant tested for affinity (L200EP 07) was later discovered to not contain any amino acid mutations. This discovery significantly altered interpretation of the SPR data. Because L200EP 07 was ancestor type, the measured affinity of 1.2 μ M must be considered to be within error of the previous measurements, 4.5 and 6 μ M. That the affinity of the other, phenotypically-different variants were determined to be no different requires the conclusion that, taken individually, the various mutations gave no benefit to binding.

L200EP 06 appeared to have an affinity approximately 2-fold tighter than the other 5 variants tested (Table 4.3), of K_D 612.8 nM, compared to 1.1-1.4 μ M found in contemporary variants, and 40 μ M in the 13mRL L8 ancestor. To investigate the structural basis for the apparent increase in affinity, the structure of L200EP 06 in complex with HEWL was solved to 1.86 Å. See section 5.3 for a full description and discussion of this structure.

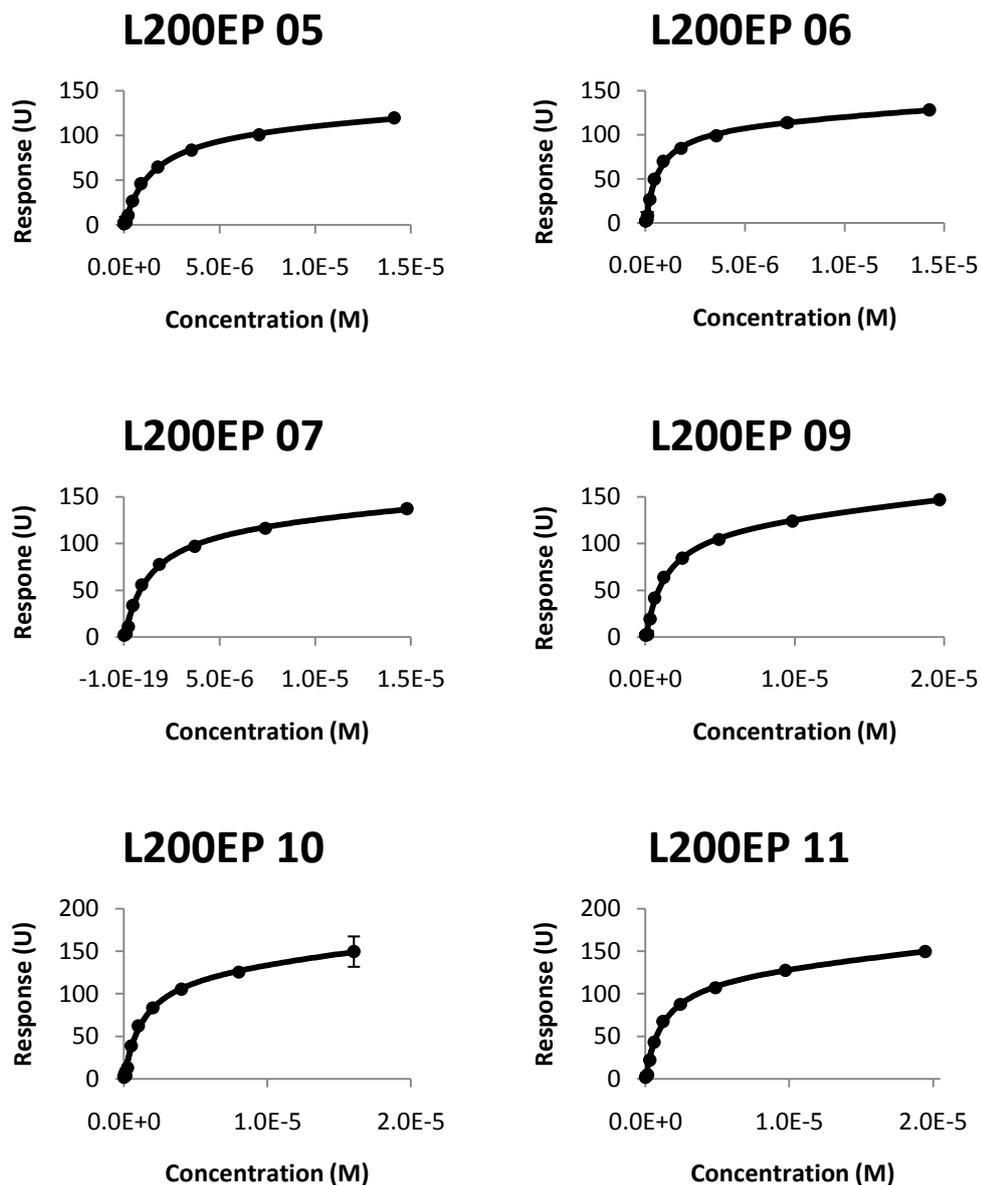


Figure 4.14 Biacore analysis of L200EP round five variants

Equilibrium affinity determination graphs, modelled as 1:1 Langmuir binding. Biacore measurements were performed as described in section 2.4.6. Data was fitted using Graphpad Prism. Error bars show 99% confidence intervals for the mean of each data point (measured in triplicate). Calculated affinities are shown in Table 4.3.

Table 4.3 Affinities from L200EP library

The affinity of six variants for HEWL as determined by SPR analysis, including the linear non-specific portion, represented by the gradient of the line as calculated by GraphPad Prism with Equation 2.1.

Variant	K_D (M)*	Linear NS gradient
L200EP 05	$1.3 \pm 0.4 \times 10^{-6}$	7.46×10^{-7}
L200EP 06	$0.6 \pm 0.2 \times 10^{-6}$	6.55×10^{-7}
L200EP 07	$1.2 \pm 0.3 \times 10^{-6}$	8.40×10^{-7}
L200EP 09	$1.2 \pm 0.1 \times 10^{-6}$	5.74×10^{-7}
L200EP 10	$1.1 \pm 0.2 \times 10^{-6}$	4.36×10^{-7}
L200EP 11	$1.2 \pm 0.3 \times 10^{-6}$	6.23×10^{-7}

*Errors given are a 95% confidence interval for the model fitting only

4.2.7 Limited Codon Library

Towards the goal of investigating the capacity of paOB3 as a scaffold for general application as a binding molecule, a new naïve library was generated and panned against HEWL and bovine pancreatic ribonuclease A (RA). The limited codon library (denoted 13mLC) was designed (Figure 4.15) as a compromise between maximising available diversity at each randomised position and increasing the fraction of theoretical diversity represented in a phage library. The position of each randomised codon wasn't altered from the original library design, but the NNK mutational codon format was replaced with DVK. This limited the number of possible residues to 12, while retaining a full complement of residue character. Other libraries have been built using limited complements, the most extreme example being only tyrosine and serine (Koide *et al.* 2007). The DVK mutational codon complement included aromatics Tyr and Trp, charged residues Lys, Arg, Asp and Glu, hydrophobic residues Ala and Gly and polar residues Asn, Thr and Ser. The residue sub-set also included Cys.

In terms of diversity, 12 possible residues at each of 13 positions gave a maximum diversity of 1.06×10^{14} , compared to the full 20 residue complement of 8.1×10^{16} . This meant that during assembly, cloning and transformation, sampling of the sequence space in this library covered a larger fraction of the maximum theoretical diversity by more than two orders of magnitude, compared to a library with the full 20-residue complement.

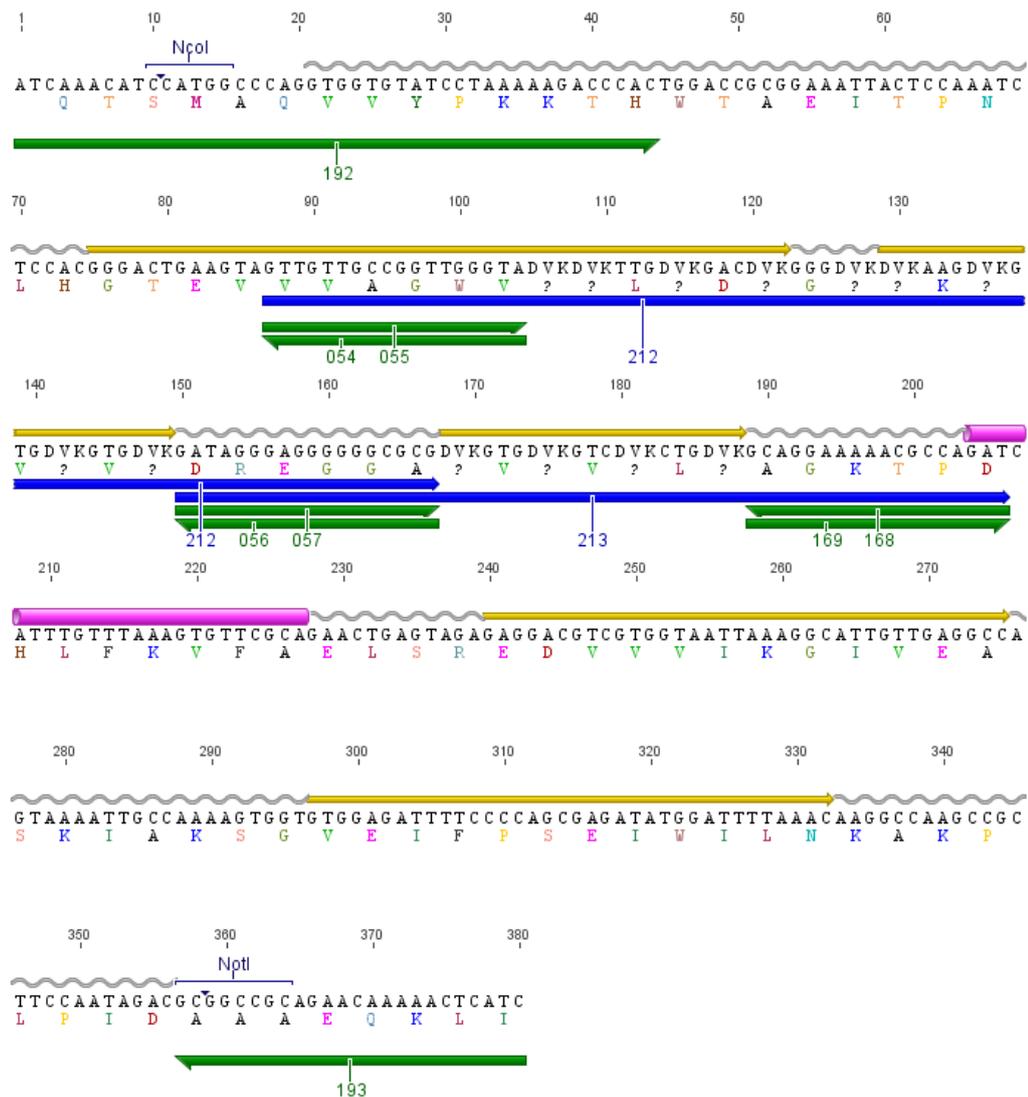


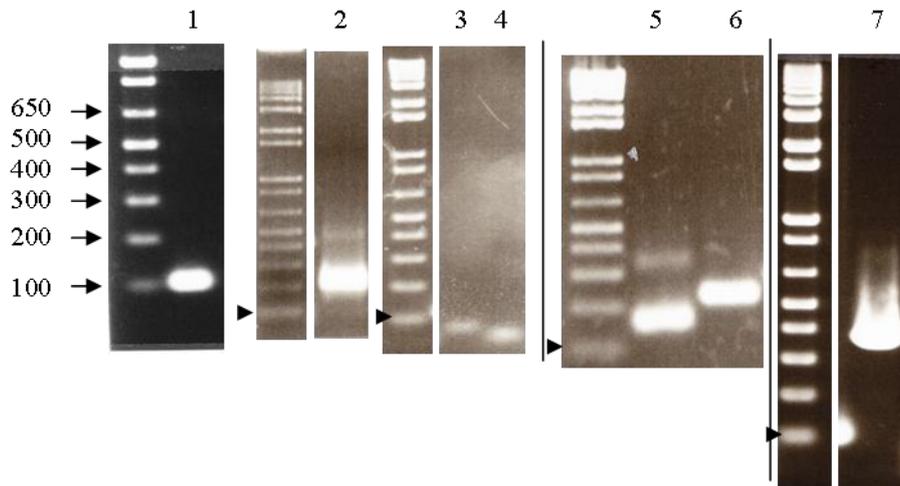
Figure 4.15 13mLC library plan

The 13mLC library was designed to incorporate mutational oligos 212 9m and 213 4m (blue) into the gene as templates for the first three β -strands (yellow, helix in pink). Flanking primer for gene fragment amplification are shown in green. Note the DVK codons at the mutated codon positions.

4.2.7.1 Construction

The limited codon library was built using 13 mutant codons on the first three β -sheets with oligos 212 9m and 213 4m (appendix A1.3). These two mutational oligos were incorporated into the gene in two separate fragments. First, each oligo was double stranded by primer extension, using the mutational oligos as the template in 100 μ L single-cycle PCR-like reactions (3 min at 94 $^{\circ}$ C, 1 min at 55 $^{\circ}$ C, 1 min at 68 $^{\circ}$ C), using reverse primers 056 (for 212 9m) and 168 (for 213 4m) in equimolar amounts (10 pmol each) to yield fragments 212/056 9m and 213/168 4m. Oligo 212 9m had a maximum theoretical

diversity of 5.16×10^9 , which was covered 1000x by 10 pmol. Assembly of the intact gene followed the principles outlined in section 2.4.2 and proceeded as follows: 192/054 + 212/056 9m = 192/056 9m; 213/168 4m + 169/193 = 057/193 4m; 192/056 9m + 057/193 4m = 192/193 13mLC. The library was ligated in pRsp2 phagemid and transformed into TG1, yielding 2×10^8 transformants. Results are summarised in Figure 4.16.



Fragment	Template	Size (bp)
192/054	wtELMC	104
169/193	wtELMC	192
212/056 9m	212 9m	81
213/168 4m	213 4m	57

Figure 4.16 13mLC assembly process

Shown here are agarose gels (top) with the various fragments generated during assembly. Fragments are labelled as follows: **1)** 192/054, **2)** 169/193, **3)** 212/056 9m, **4)** 213/168 4m, **5)** 192/056 9m, **6)** 057/193 4m **7)** 192/193 13mLC. The accompanying table lists the primary fragments, the template used in each case and their predicted size in bp. Remaining expected sizes are **5)** 167 bp, **6)** 231 bp and **7)** 380 bp

4.2.7.2 Panning

The library was panned as described in section 2.4.4 against HEWL or RA over 6 rounds, using ligand immobilised in 5mL immunotubes with helper phage Vcsm13. Ligand concentration was varied, starting at 1 mg/mL for the first three rounds, dropping to 0.1 mg/mL in subsequent rounds. Panning against both ligands showed enrichment, both in output:input phage titre ratio and an increase in displayed Obody (Figure 4.17).

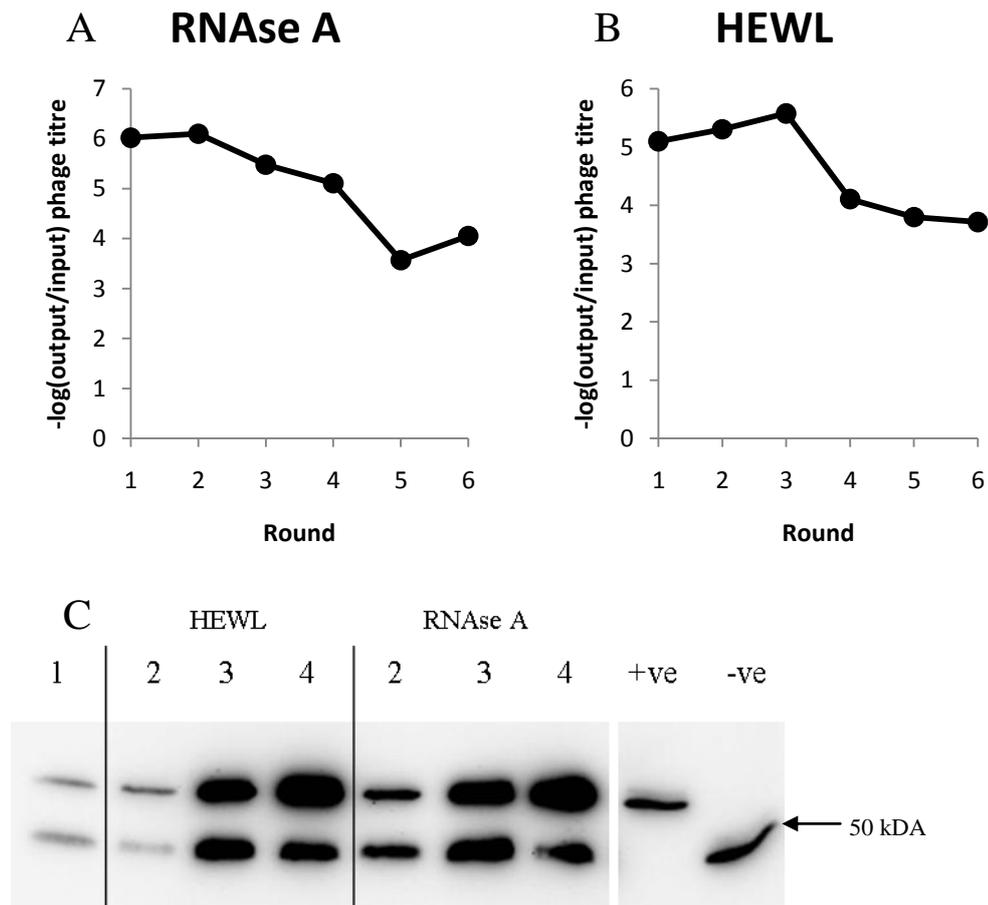


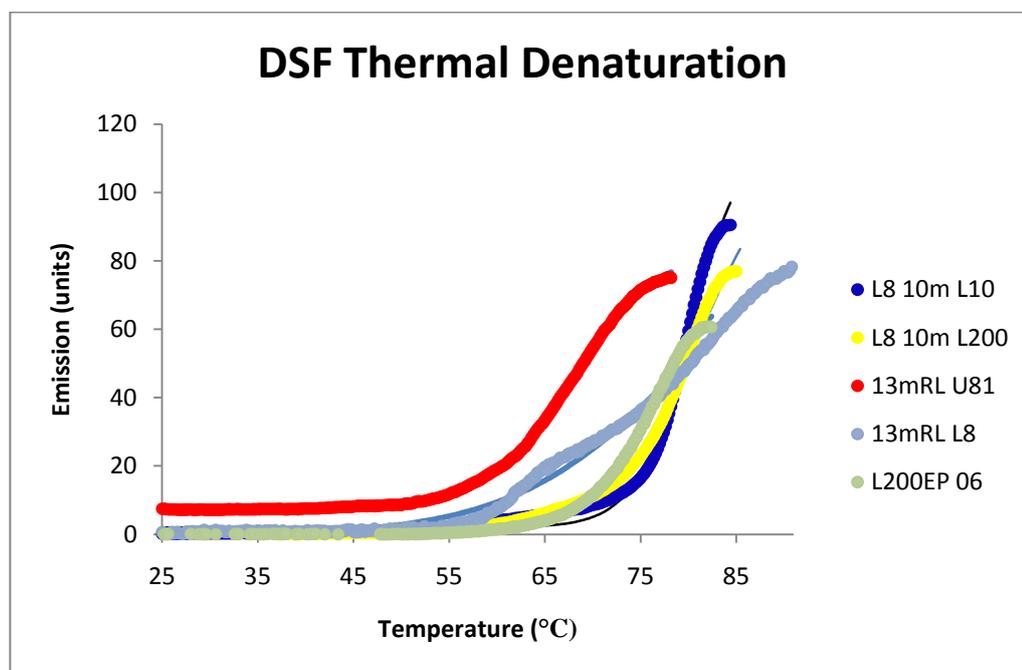
Figure 4.17 13mLC panning results

The 13mLC library showed enrichment panning against two model protein ligands. (A/B) Enrichment for phage retention with both ligands (as labelled) shown by the decreasing negative log of the ratio of eluted:input phage titres over six rounds of panning. (C) Western blot of input phage samples. Numbers indicated the panning round each sample came from. Only a single round 1 sample is included, as each ligand panning stream began from the same input phage sample. Both ligands showed enrichment of the level of displayed cMyc-tagged Obody-pIII fusion (upper band) as compared to cMyc-tagged pIII only (lower band. Controls are paOB3-pIII fusion (+ve) and pIII only (-ve), both cMyc tagged. The approximate position of the 50 kDa protein marker is indicated by an arrow.

Sequence analysis of round six HEWL-selected variants showed fixation of a cross-contaminant species identical to L8 10m L200, the presence of which is hypothesised to have disrupted proper selection from this library. One variant from round three that was clearly descended from the 13mLC library (based on the sequence and mutation pattern) was qualitatively shown to bind HEWL in a Ni²⁺ pull-down assay, but attempts at quantitative measurement did not show binding. Round six RA-selected variants showed sequence convergence, but binding to RA could not be shown by soluble protein. This library was not pursued further.

4.2.8 Thermal stability of engineered Obodies

To test the relative stability of mutant Obodies, a panel of purified variants were subjected to Differential Scanning Fluorimetry (DSF)-based temperature gradient melt assay, as described in section 2.4.8. The panel included an unselected variant from the 13mRL library, U81. This variant was included in the place of the wild-type (wt) domain, as it was found to express solubly and all 17 randomised positions were different from the wt sequence. The wt domain exhibited a toxic effect while expressing, and co-purified with quantities of RNA bound, making it problematic to handle. Thus, 13mRL U81 was substituted as a comparison for subsequently selected variants. The results are summarised in Figure 4.18.



Obody	Tm (°C)
L8 10m L10	80.2
L8 10m L200	81.2
13mRL U81 (unselected)	67.8
13mRL L8	79.6
L200EP 06	75.1

Figure 4.18 Differential scanning fluorimetry thermal stability curves

The fluorescence of five engineered Obodies denaturing along a temperature gradient is shown in the graph (top), with the Tm calculated by GraphPad Prism shown in the table below. Each curve was measured in triplicate and the average is plotted above.

Conditions for the DSF conditions were determined by screening around protein and SYPro Orange dye concentration, to determine optimal conditions for the qPCR instrument detector. Experimental runs subsequently used protein at 100-200 $\mu\text{g}/\text{mL}$ and SYPro Orange at 100x concentration.

Although the unselected (13mRL U81) variant tested showed a considerably lower T_m (67.8 $^{\circ}\text{C}$), the remaining four were very similar. At approximately the same temperature that 13mRL U81 begins to denature, all other Obodies also showed a marked increase in fluorescence which was then followed by a definite, much larger, peak. This may indicate a difference in the process of unfolding, but further experiments to produce better quality data, such as differential scanning calorimetry, are needed before this might be expanded upon.

4.3 Discussion

The over-arching goal of this part of the project was to demonstrate the ability of the selected template OB-fold domain, to be matured into a high-affinity binding domain for a model target ligand. Towards this end, two broad streams were employed. First, an existing engineered Obody with moderate affinity for HEWL (~40 μM) was targeted for stages of affinity maturation by structure-based and random library design. This can be considered as the final stages of the proof of concept and domain validation that was begun with the work presented in the manuscript found in the appendix (A2). Second, a new format of naïve library was investigated to try to improve the achievable affinity from “single pot” libraries. This is the standard achieved by the best scaffolds discovered to date and must ultimately be the goal of wider Obody development project as well.

4.3.1 *Structure-based design*

4.3.1.1 **L8 10m**

The first library investigated here was based on the structure of 13mRL L8 in complex with HEWL. It introduced four mutations on the β -sheet binding face, re-randomised four residues in the 4/5 loop and lengthened it by two additional randomised positions, giving a total of 10 mutant codons in the whole library. The residues targeted either made no contacts on HEWL or seemed to make a negative contact, as in the case of K37.

Surprisingly, this library showed a similar ratio of successfully displayed Obody to degraded pIII fusion as libraries with larger numbers of mutations, despite the fact that most of the additional mutations were located at previously randomised positions, or in the 4/5 loop, which seemed more tolerant to mutation than other sections of the domain (Appendix A2 – manuscript). While the maximum library size was limited to $\sim 10^7$ by the number of transformants, I estimated the practical library size downwards by one order of magnitude, due to the presence of extensive point mutations found in the mutational oligonucleotides (~90% of sequenced unselected cloned variants contained stop codons). Given these two large limitations, it is encouraging that even

with an estimated practical library size of 10^6 , panning still resulted in several different sequences with 10-fold improvement in affinity.

Two of the four residues at the binding were re-selected as ancestor type. P51 was seen in 2 of 5 unique sequences, and K37 was seen in 3 of 5. While the restricted numbers of sequences and the small initial diversity limit the conclusions able to be drawn over this, it may indicate that P51 played a cooperative role in binding. In the case of K37, the crystal structure of the 13mRL L8 ancestor clearly showed a like-charge clash with HEWL R61, which would remain unless the library produced significant re-arrangements in the interface. It may be that that lysine at this position contributes both to binding and stability of the Obody, given it is almost entirely buried in the hydrophobic core behind the binding face. This was supported by selection from the L200 error prone library; the library ancestor type had a methionine at position 37, and mutation to lysine was one of the substitutions seen in conjunction with a 2-fold improvement in affinity.

4.3.1.2 L10 6m Library

Following panning of the L8 10m library, a single variant (L10), selected as the variant with the best affinity, was solved in complex with HEWL (section 5.2). What this structure revealed was persistence of a cavity at the interface, filled with ordered waters. Thus the L8 10m L10 variant was used as the template for a new library to try to fill this cavity.

Hotspot analysis indicated that the primary binding residues were located along the first β -strand of the Obody binding face, including three that were not part of the original library mutational complement, and the two residue turn into the second β -strand. The peptide of Y33, G34 and R35 inserted into the substrate groove and active site of HEWL, but geometric complementarity was poor in this region, with a water-filled cavity visible between the Y33 and R35 side chains. It was hypothesised that in filling this cavity by extending the 1/2 loop an improved affinity may be produced, but the central role in binding played by this region suggested that a simple extension of the loop would not be sufficient. In addition to the β -sheet character, conformational freedom to find a better binding mode would be severely curtailed by the close proximity of critical binding residues. Therefore, the entire region was permitted to be re-

selected to accommodate any new residue combination that bound with greater complementarity to the HEWL binding groove. This rationale led to design of the L10 6m library, which mutated residues G34, R35, V36, I38 and added two random residues. The expectation was that with the extended loop, the peptide chain main-chain atoms would be free to bind closer to HEWL and residues with less bulky side chains would be selected. Unfortunately, the only sequence selected from this library was ancestor type and did not, therefore, produce variants with an improved binding affinity

The failure of this library to select any mutant variants was surprising, especially in light of the almost complete coverage of the maximum theoretical diversity. While there are many possible causes for this, the two most probable scenarios in my view are: i) that binding-promoting conformations were in direct competition with the formation of stable β -sheet and ii) that the transformation count was not actually an accurate reflection of the library diversity. The second option cannot be commented on any further here, as it can't be meaningfully validated without deep sequencing, although sequenced unselected variants did show diverse sequences (appendix A1.4.4). In addition, the presence of the ancestor gene in the library was almost certainly a contaminant, which would have inhibited selection of mutant variants with lower affinity, or those with higher affinity but poorer display profiles.

Induction of conformational changes in proteins on binding is not energetically favourable, and critical residues are thought to be pre-arranged into binding-promoting conformations in their non-bound state (Li *et al.* 2004; Gallicchio *et al.* 2010). Thus, any new arrangement of the randomised peptide to fill the aforementioned cavity must compete with the inherent stability of the existing β -sheet. This suggests that the two residues added to the chain manifested as a loop outside of the binding region targeted, rather than as a broken β -strand in the HEWL active site, as was desired.

Elimination of the arginine at position 35 removed a salt bridge from the interface. It was expected that a smaller residue would be needed to bind into the active site of HEWL, such as histidine in Ykfe (Figure 3.8). But at pH 7.4, the imidazole group on the histidine would only be partially charged, and would therefore be a less efficient mediator of long-range electrostatic

attraction. It could be that disruption of this important electrostatic interaction also contributed to the domination ancestor-type sequence over mutant library variants.

4.3.2 *Error-Prone Affinity Maturation*

During the immune response, there are two processes which serve to generate better antibodies for particular antibodies. The first is selection from a repertoire of recombined gene segments, commonly cited at 10^7 - 10^8 unique combinations (Rajewsky 1996). Virtually all of the diversity is located on the complementarity determining region (CDR) loops. The second process, known as somatic hypermutation (Griffiths *et al.* 1984; Malipiero *et al.* 1987), spreads point mutations through the variable domain to improve the affinity of a particular set of variants. This process doesn't necessarily involve mutations of the residues directly involved with binding, but rather can allow the domain scaffold residues to better accommodate the binding residues by changing their environment and thus their charge, hydration or conformational freedom (Kettleborough *et al.* 1991).

Based on this concept, a random distribution of mutations was induced in a single HEWL-binding Obody by error-prone PCR to search sequence space not explored by previous libraries. In order to improve affinity, mutant residues positively selected are required to contribute more to binding than the wt residues. This produces two competing forces, where mutations at the interface would be generally selected against, and mutations away from the face would generally be neutral.

Several rounds of selection generated a bank of evolutionary descendants with different complements of point mutations. In a natural evolutionary relationship, frequent mutation at one position would indicate a lesser importance in function than conserved residues. The sequencing data produced here reflects that, with very few mutations seen at the positions previously identified as critical for binding in 13mRL L8. However, as they are the result of positive selection, frequent mutation can also indicate a site where a substitution can more often give a net benefit to binding. To a large extent, mutations in the selected genes appear to have little effect on affinity, but there appeared to be regions which have mutations more closely clustered, though no

formal statistical treatment was undertaken. Representation of these frequently mutated regions, particularly where it was a single position, was used during selection of individual variants for affinity measurement.

4.3.2.1 **Re-selection of ancestor-type Obody**

During the clone isolation and selection process, a sequencing error resulted in the inclusion of one variant that, while it contained nucleotide substitutions, had an ancestor-type phenotype (L200EP 07). Unfortunately, the ancestor-type sequence showed an affinity that was not different from four out of five of the confirmed mutant genes screened. This resulted in a revision of the apparent improvement in affinity from 10-fold to 2-fold, with only a single variant displaying the improvement (L200EP 06). While disappointing, this result does support the notion that the current binding mode may require a prohibitively large number of mutations to improve affinity by the extent desired.

In the comparison of measured affinities across multiple libraries, it is necessary to consider the sources of experimental error in order to draw conclusions about statistically significant differences. The process of measurement of affinity using the Biacore SPR instrument involves the repeated generation of chip surfaces with covalently attached HEWL. As described in section 2.4.6, HEWL was attached via the introduction of N-hydroxysuccinimide groups on to the chip surface, which then form amide bonds with one of six surface lysines or the amino-terminus of the HEWL. Each newly generated chip represents a surface with different degrees of steric hinderance on Obody-HEWL complex formation. Thus, the major sources of experimental error during affinity calculation come primarily from batch variation of protein preparations and, more significantly, from the surface to surface variation on different chips. In the case of the error-prone library variants under discussion here, the affinities were determined in a medium-throughput fashion, with all protein purified in a short space of time, using the same batches of buffers, host cells and other reagents, and all were analysed using the same chip surface over only a few days. Thus, the experimental errors between the determined affinities can be considered to be reduced to a minimum. In contrast, the affinities determined months earlier for the ancestor

sequence are comparable with much less precision. Hence it seems reasonable to accept errors large enough to make K_D values of 1.2 and 4.6 μM not significantly different when comparing between batches, while simultaneously accepting a smaller 2-fold difference as significant when comparing within the same batch of experiments.

4.3.2.2 Structural Analysis of Affinity Improvement

Compared to the template gene, L200EP 06 it had three substitutions, two of which were considered significant; K86E, based on the 13mRL L8-HEWL complex structure, may influence the dominant conformations of the 4/5 loop, as well as critical residue E93, located on β -strand 5. This variant was the only one to show this particular substitution, which is evidence for its involvement in affinity improvement, but this can equally be taken as evidence against; a real difference as a result of this substitution should have been selected more often. The second substitution, M37K, shows a convergence to lysine that was seen in the L8 10m library and is a position that was mutated often. L200 EP06 was not the only variant with a lysine at this position; the same mutation was seen in EP11, which did not show an improvement in affinity. This leads to the conclusion that it was these mutations which, in combination, caused a 2-fold change in K_D , were not sufficient in isolation. The third substitution was T19S. Located on the opposite side of the domain from the interface, this mutation was considered to be neutral. Structural investigation of this variant to attempt to determine the basis for the affinity improvement was undertaken using x-ray crystallography, and the results are presented in Chapter 5.

The poor performance of this library in terms of improvement in affinity raises the question of what, during selection, the dominant factor was that promoted retention of phage during selection. Although affinity for the target ligand was designed to be the major filter, other selection pressures, like the propensity to be displayed on phage, or effects on bacterial growth during the amplification phase play a role as well. Two factors which may provide clues are the T_m and expression levels. Compared to the ancestor type, the T_m reduced slightly (Figure 4.18) while expression levels in *E. coli* tripled (45 mg per litre of culture compared to 15 mg for the contemporary variants purified).

Domains which are very stable may be at a disadvantage when using the *sec* secretion pathway (Steiner *et al.* 2006), though the DSF method for measuring T_m results in data that lacks the upper portion of the sigmoidal curve which introduces uncertainty into the point of inflection, making this conclusion tentative

4.3.2.3 **High tolerance to mutation**

Aside from the selection for affinity, that so many different mutations were tolerated by the Obody scaffold while maintaining binding highlights the potential as a molecular recognition reagent. From this library alone, 32 different substitutions at 25 positions are shown. These are of course in addition to the sequence space searched in the previous two libraries. These data, coupled with a more in-depth analysis of stability and survey of naturally evolved related domains, have the potential to inform the design of a new synthetic Obody scaffold with improved thermal stability parameters, as has been performed with other scaffolds (Mosavi *et al.* 2002) (Knappik *et al.* 2000).

4.3.3 *Limited Codon Library*

Any mutational library is limited by the ability of the template domain to tolerate the mutations introduced and remain stable. A phage display library is restricted by the relative level of display for each unique variant, which is influenced by domain stability and the secretion pathway used. When dealing with large numbers of mutations, sampling of the mutational space adds a further sequence bottleneck that, when combined with the already mentioned limitations, may restrict or completely prohibit discovery of specific binders to any one target. In the absence of information for site-specific mutational tolerance, it may be beneficial to reduce the available diversity at each position to reflect the bias towards certain residues seen in natural protein-protein interfaces (Lo Conte *et al.* 1999), thereby also increasing the percentage coverage of the theoretical maximum diversity. This approach has been used to generate antibody phage libraries with a two (Fellouse *et al.* 2005) and four (Fellouse *et al.* 2004) residue complement at each position, yielding μM -nM

range binders. Similar result was achieved with fibronectin type III scaffold (Koide *et al.* 2007).

The corollary of this bias is that with the more limited sequence space comes an increased possibility that the required sequence for binding is completely excluded, or that binding is inhibited by forced inclusion of sub-optimal residues. With these considerations in mind, the question becomes; does the library benefit sufficiently from the increased percentage coverage and bias towards binding-promoting residues (primarily Ser, Tyr and Arg) to make up for the diversity penalty imposed by sequence limitation? The fact that this library failed to produce binders of any significance implies that, with the twelve residue complement and the distribution of mutations used in the 13mLC library, it does not. It has been shown since that a library with the full complement of amino acids can out-compete a library with the same design but with a restricted complement (Hackel and Wittrup 2010). This suggests that while the restricted-complement approach can indeed generate binders, it seems to result in reduction of the probability of selecting a binder with high affinity.

4.3.4 *Limitations of the phage-display format*

During library generation, cloning and transformation, the diversity of each library goes through a series of bottlenecks. Because of the combinatorial PCR-based method used here for gene assemble, reaction size and quantities of DNA represent the first practical limitation, calculated at 3.011×10^{13} possible unique variants (section 2.4.2). Once the gene library is assembled, ligation into pRpsp2 again imposes a bottleneck, which I calculated at a 30% success rate (section 4.2.2). As library-scale ligations contained 3 pmol of plasmid, this equates to a maximum of ~ 1 pmol (1.8×10^{12}) ligated plasmids available for transformation. By far the most severe bottleneck is transformation. The method employed here for producing electrocompetent cells gave, at best, an efficiency of 10^9 cfu/ μg of pRpsp2, but batch variation resulted in a practical limit of $\sim 10^8$ transformants. As a result, any library with a theoretical diversity of greater than 10^9 will suffer sampling bias during the transformation step.

In addition, while it is not expected that all members of a transformed library will be produced stably, the ability to be displayed on the surface of a

phage may not necessarily overlap with domain stability. For example, large numbers of hydrophobic residues may not prevent the domain from folding, but may preclude secretion into the periplasm, where m13 phage is assembled. For future work, it may be beneficial to investigate an alternate secretion signal, such as the SRP-pathway (Steiner *et al.* 2006), to decrease the proportion of pIII with no fusion partner.

4.3.5 *Surface Plasmon Resonance*

The affinities of all Obodies were determined by SPR analysis on a Biacore 3000 instrument using an equilibrium binding model. Attempts were made to individually measure k_{on}/k_{off} directly from the raw data, however these invariably gave spurious and widely varying results. One possible explanation is that the μM affinity ranges dealt with in the course of this project lie at the edge of the Biacore's accurate range, which may cause problems during the model fitting process. In a related problem, in order to measure μM -range affinities, high concentrations of analyte are needed, resulting in a large density difference between the instrument running buffer and the protein sample. Sudden changes in density can result in large "injection spikes", in relative response curves, as can be seen in the upper most curves in the SPR raw data in Figure 4.7. These difficulties precluded the determination of k_{on}/k_{off} and left the method employed (steady-state kinetics) the most reliable at extracting a dissociation constant.

4.3.6 *Conclusions*

The above results show that affinity maturation for HEWL was successful by demonstration of sub- μM affinity for model target HEWL, though the target (<100 nM) affinities were not reached. Although affinities have remained modest, development is tracking in a similar arc to other scaffolds as they were first investigated, with the first pass libraries yielding μM -scale binding, and better affinity only achieved after further development of the scaffold domain and library design.

This allows me to state with confidence that the OB-fold is a viable scaffold for exploitation as an engineering scaffold, though more work is needed in designing better naïve libraries for new selections. Further improvement the process of selection is also needed. In particular, the

immobilisation method (adsorption on the surface of immunotubes) concentrates the ligand on a surface, potentially imposing steric restrictions. Solvent-phase panning has been described using biotinylated ligand (Barbas III *et al.* 2001) and this should be considered as an alternative method.

5 Structural Analysis of Engineered OB-folds

5.1 Introduction

This chapter presents the crystal structures of two affinity-matured Obodies in complex with model ligand HEWL, derived from libraries described in Chapter 4. In the first section, a variant from rationally designed library L8 10m is presented, with emphasis on the effect of the library mutations and their possible contributions to the observed improvement in binding affinity. The second section presents a variant from an error-prone library and attempts to rationalise the observed affinity improvement with the three point mutations present in the structure.

5.2 Results - L8 10m L10

This gene was selected for crystal trials from a panel of five unique variants among which it showed the highest affinity for HEWL (section 4.2.4.3). Compared to the ancestor gene, it contained two mutations at the binding face and six in a flanking loop.

5.2.1 *Expression and Purification*

The gene was cloned into and expressed from pRoEx Htb with an N-terminal His-tag. It was expressed in batches of 1 L, purified by hand using a HiTrap chelating column (section 2.5.3) and the His-tag removed by digestion with rTEV (section 2.5.5). The protein was concentrated using a 5,000 Da cut-off membrane ultrafiltration spin device (Vivaspin) and purified in a second step using an S75 10/300 size exclusion column, by serial injections of 250 uL. Because all Obodies thus far have purified identically, the position of the peak at ~12 mL was taken as diagnostic for a successfully purified Obody. Initial small-scale purification attempts were checked by SDS-PAGE. A typical yield gave between 10 and 15 mg of purified protein from 1 L of culture after the first step of purification. Purification data are shown in Figure 5.1.

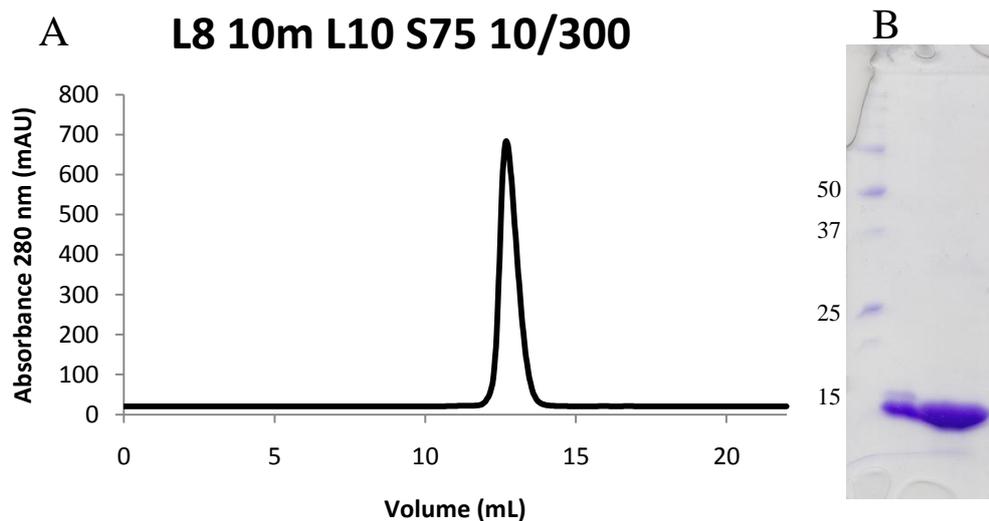


Figure 5.1 L8 10m L10 purification

(A) A representative chromatogram of elution from an S75 10/300 size exclusion column. 0.5 mL fractions were collected and pooled across the peak at ~12 mL. (B) Representative SDS-PAGE gel showing purified, His-tagged Obody from two different batches. Protein markers are labelled in kDa.

5.2.2 Crystallisation and Data Collection

L8 10mL10 was not subjected to a wide condition screen, but instead was trialled using a fine screen constructed around the condition that yielded crystals of the 13mRL L8 complex (see screen in appendix A1.6), at a similar concentration (41 mg/mL with equimolar HEWL) in a 1 μ L + 1 μ L sitting drop format. Crystals were identified in approximately 50% of the conditions trialled. A single crystal was found to diffract to resolution of at least 2 Å in condition B7 of the fine screen (precipitant: 0.2 M HEPES pH 7.4, 9% MPEG 5000) and a 360° dataset was collected with a phi oscillation of 0.5° at the Stanford Synchrotron Radiation Lightsource (SSRL) on beamline 9-1 at 100 K, to a maximum diffraction of 1.95 Å resolution (Figure 5.2).

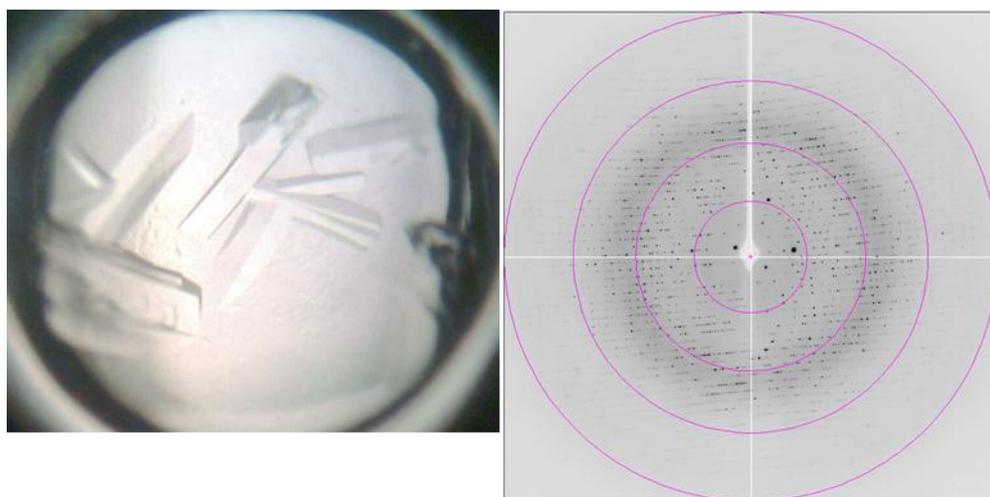


Figure 5.2 *L8 10m L10-HEWL complex crystals and diffraction*

Grown in condition 0.2 M HEPES pH 7.4, 9% MPEG 5000, these crystals (left) were carefully broken into manageable fragments for diffraction testing. The image shows a sitting drop approximately 1.5 mm in diameter. A sample diffraction image is shown on the right. Starting from the centre of the image, resolution rings mark 8.0, 4.0, 2.7 and 2.0 Å.

5.2.3 *Model building and Refinement*

While the conditions for crystallisation were very similar to that of the ancestral Obody-HEWL complex, the crystal was markedly different. The unit cell was much larger, with the symmetry of a $P2_12_12_1$ space group. Initially indexed in $p222$ using XDS with a long c axis and a mosaicity of 0.2° , molecular replacement was performed with PHASER from the CCP4 suite of programs in all related space groups, using 13mRL L8 (chain D) and HEWL (PDB accession 193L) as search models. Only a single solution in $P2_12_12_1$ was found, with 9 copies of each monomer in the asymmetric unit. Translational symmetry was evident in the asymmetric unit, and Patterson map peaks of 90 at the origin, and significant peaks at $x/y/z$ coordinates $0/0/0.33$ (52), $0/0.33/0$ (36) and $0/0.33/0.33$ (30) also showed pseudosymmetry. SCALA did not detect any twinning during scaling.

Both to improve the parameter-data ratio and reduce the delay between refinement rounds, ncs relationships were determined automatically using Phenix auto-ncs, which split the Obody chains into two groups. All nine of the HEWL molecules were restrained to chain A over residue range 15-129; Obody chains D, F and H were restrained to chain B over residue ranges 2-52, 60-90 and 99-106; Obody chains L, N, P and R were restrained to chain J over the same residue ranges, designed to exclude loops and missing segments. All

restraints were defined as “tight” for main chain atoms and “loose” for side chain atoms, assuming that the folds don’t vary much across the asymmetric unit, with most variation expected in the side chains.

The model was manually mutated to reflect the correct amino acid sequence and iteratively refined using Refmac restrained refinement with TLS (each chain was defined as an independent TLS zone) and Phenix minimisation with simulated annealing. The final R-free was 22.97%. Geometry was validated with PROCHECK, with 97.6% in preferred regions, 2.4% in allowed regions and no residues unfavourable or disallowed regions of the Ramachandran plot. Refinement statistics are summarised in Table 5.1.

Table 5.1 L8 10m L10-HEWL complex structure statistics

Those figures in brackets represent the highest resolution shell

	SSRL
Data Collection	
Space Group	P2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	
a	60.54
b	186.26
c	245.70
α, β, γ	90
Wavelength (Å)	0.95666
Resolution Limits (Å)	29.76 – 1.95 (2.05 – 1.95)
Measured Reflections	2,883,973 (345,369)
Unique Reflections	201,772 (27,775)
Multiplicity	14.3 (12.4)
Completeness (%)	99.0 (95.0)
Rmerge (%)†	6.4 (48.4)
<I/σ(I)>	24.5 (5.0)
Wilson B (Å ²)	28.19
Mosaicity (°)	0.2
Molecular Replacement‡	
Log-Likelihood Gain	17355.736
Refinement	
Resolution (Å)	29.76 – 1.95 (2.01 – 1.95)
Reflections	191,052 (12500)
R _{work} (%)*	18.80 (21.34)
R _{free} (%)*	22.97 (27.03)
Free reflections	10,136 (674)
Protein Atoms	18,959
r.m.s.d bond lengths Å	0.007
r.m.s.d bond angles (°)	1.052
<B factor> (Å ²)	30.2

† $R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - \langle I \rangle|}{\sum I_{\text{obs}}}$

‡ Ref. (McCoy *et al.* 2007)

* $R = \frac{\sum ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum |F_{\text{obs}}|}$

5.2.4 *Library Mutations*

To determine the structural basis for the 10-fold improvement in affinity, the mutations in this variant were individually examined for possible contributions (Figure 5.3). The bulk of the mutations made at the beginning of this library were located in the 4/5 loop, making up six of the total 10 mutations. Of the six mutant positions in the loop, only one (A87) was sufficiently ordered to be visible in the structure. This residue made a hydrophobic contact with HEWL in all complexes, with approximately 20-30% accessible surface area buried at the interface.

Of the four residue positions in the face subjected to mutation, two resulted in ancestor-type reversion (K37 and S29). While there appeared to be little change with S29, the distance between the *Z* nitrogen of K37 and R61 was increased from ~ 3 Å in the 13mRL L8 complex to ~ 5 Å in 10m L10, ameliorating the like-charge clash and introducing possible water-mediated interactions (Figure 5.3 B). Regarding the P51S mutation, no hydrogen bond was predicted to form between S51 and HEWL and virtual alanine scanning showed a marginally negative interaction. Comparison of the 13mRL L8 structures with L8 10mL10 showed a major change in the 2/3 loop C α trace, which could be a consequence of i) crystallographic contacts involving this loop in the L8 10mL10 structure that were not seen in 13mRL L8, ii) that the loop was poorly ordered in the 13mRL L8 structure, leading to larger errors in atom placement, and/or iii) a change in conformational restriction by the P51S substitution. As any actual direct contribution to binding by S51 seemed to be very small, it seemed likely that relaxation of conformational restrictions on the surrounding peptides was the dominant contributor of any change to affinity for HEWL stemming from this mutation.

A56Y was more easily interpretable, as the mutation increased hydrophobic interactions and introduced a hydrogen bond (Figure 5.3 C). Virtual alanine scanning calculated an average $\Delta\Delta G$ of 0.82 kcal/mol for Y56, and the interface buried an average of 25.2% of solvent accessible surface area of that residue. Relative to the Obody, this region showed amongst the greatest difference in HEWL position at the interface as compared to the 13mRL L8 structure (C α shift of 2.6 Å), although the relative shift was more pronounced distal to the interface (Figure 5.3 A). This movement appears to be a rotation of

HEWL relative to the Obody about the longitudinal axis of the complex, centred on the hydrophobic patch (Y33, V36 and I38), of approximately 10 °.

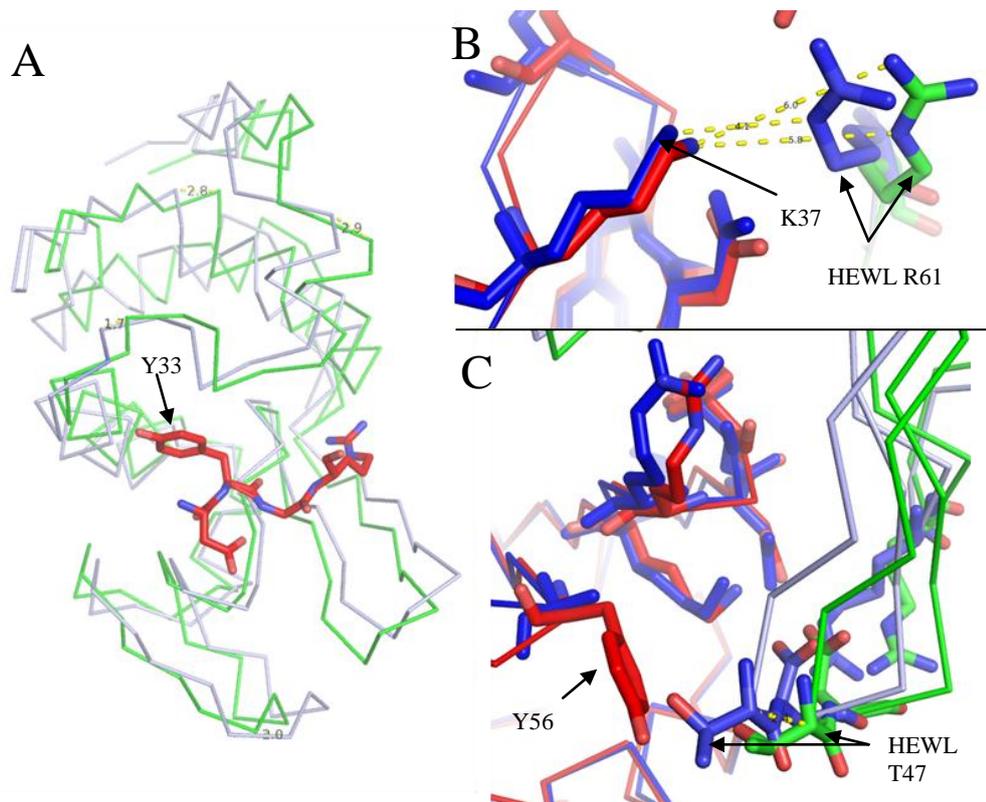


Figure 5.3 *L8 10mL10 mutations*

(A) Rotation of the interface is illustrated by superposition of Obodies and comparison of the relative positions of the HEWLs in complex with 13mRL L8 (grey Ca trace) and L8 10m L10 (green (Ca trace), with distances between representative Ca atoms labelled in Å. A four-residue peptide from L8 10m L10 is shown as a red stick diagram as a reference point. L8 10m L10 residues (red, HEWL in green) that are hypothesised to contribute to this shift are shown in B and C, superimposed with 13mRL L8 (blue, cognate HEWL also in blue). (B) Compared to 13mRL L8, the distance between K37 and HEWL R61 has increased by ~2 Å and an interaction is now mediated by an ordered water molecule (not shown). (C) A56Y, shown in red, is associated with largest structural change as illustrated by comparison of the Ca positions of HEWL T47.

5.2.5 Interface Residues

The most striking aspect of the different Obody-HEWL complexes in the asymmetric unit is illustrated in Figure 5.4; different conformers of, Y53, K58, E83, E95 and F97 were observed across the nine complexes. Y53 has the least consistent position as compared from monomer to monomer; all Obodies in the asymmetric unit exhibit evidence for multiple conformations of Y53, to varying degrees. Occupancy was reduced to 0.5 for Y53 side chain atoms where the first conformer intersected with its Fo-Fc electron density map when

contoured to 5x its sigma level. An additional conformer was added where unmodelled Fo-Fc density exceeded a sigma of three or 2Fo-Fc exceeded a sigma of one. The other four residues cluster above and behind (relative to the binding face) the third β -strand and are hypothesised to represent multiple binding modes of L8 10m L10 to HEWL.

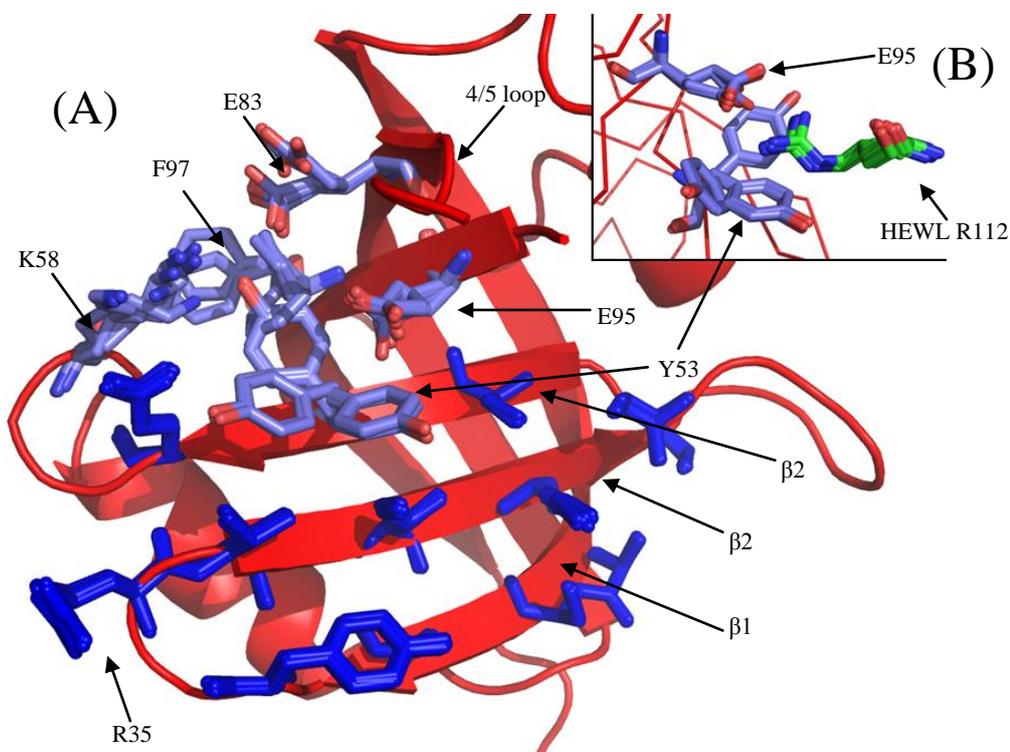


Figure 5.4 Varied modes of L8 10mL10 Obody-HEWL complexation

(A) Cartoon view of an Obody overlay (red), with the three β -strands of the binding facing to the front. The red cartoon is a single chain, whereas all nine monomers are represented by blue stick diagrams, highlighting mutant residues that overlay well. Residues which overlay poorly are shown in light blue: residues Y53, K58, E83, E95 and F97. These residues show clear conformational differences between the complexes in the asymmetric unit, particularly Y53 with evidence for multiple conformers visible on individual chains. (B) Sub-optimal aromatic stacking (approximately a 40° angle between aromatic planes) with R112 from HEWL may influence Y53 conformation, though R112 appears to primarily interact with the Obody by formation of salt bridges with E95.

In the absence of homologous structures to analyse for clues into the importance of individual residues for binding, relative stability and incidence of bond formation across the asymmetric unit may be used as a proxy. Analysis of the nine interfaces in the asymmetric unit showed variation in terms of buried surface area ($809 - 861 \text{ \AA}^2$, calculated by PDBePISA), contacting residues and bonding between the Obody and HEWL. The polar bond predictions are summarised in Table 5.2, binding hot spot analysis in Table 5.3.

Taken as an average across all complexes in the asymmetric unit, the contacts between the Obody and HEWL were similar to those seen in 13mRL L8, though interface analysis by PROTORP showed a net gain in hydrogen bonds (the output was not explicit about which residues were involved). According to PDBePISA, the same set of residues formed intermolecular hydrogen bonds (though not all bonds are found in all complexes), noting that because of a two-residue insertion into the 4/5 loop, E95 in L8 10mL10 is structurally equivalent to E93 in 13mRL L8. However, the complex lost one electrostatic bond (between D32 and HEWL R61) and R35 gained an H-bond to HEWL Q57, bringing the number of polar bonds made by this residue to three in the active site. Hot spot predictions saw a similar profile of residues selected, but showed some significant differences in calculated energy. Y33 and surrounding residues formed the nucleus of the interface, for which Robetta virtual alanine scanning calculated an average $\Delta\Delta G$ 5.6 kcal/mol, compared to 3.7 kcal/mol in the ancestor interface. Across the interface an over-all increase in $\Delta\Delta G$ was observed compared to the ancestral complex, especially R35 with a 4-fold increase (0.5 to 1.99 kcal/mol). A notable exception is E95, which was not picked as a contact point at all by KFC or Robetta-ala servers, even though PDBePISA predicted multiple polar bonds mediated by this residue. Given the close contact E95 makes with HEWL residues suitable for forming polar bonds, it seems likely that its exclusion by the hotspot detection servers reflects a limitation of their contact-determining algorithms, as opposed to a real lack of biochemical contacts across the interface.

Table 5.2 Polar bonds across the asymmetric unit

Presented here are the intermolecular polar bonds for the nine Obody-HEWL complexes in the asymmetric unit, as predicted by PDBePISA. Obody chain IDs head each column. While the numbers per residue vary slightly between interfaces, the same core six residues provide the majority of predicted polar bonds in all nine complexes.

Residue	B	D	F	H	J	L	N	P	R	Mean	Type [†]	HEWL contact
Asp 32	1	1	1	1	1	1	1	1	1	1.0	H	W63 [‡]
Tyr 33	2	2	2	2	2	2	2	2	2	2.0	H	D101, N103
Arg 35	3	2	3	3	3	3	2	3	3	2.8	HS	D52, Q57, E35
Lys 40	1	1	1	1	1	1	1	1	1	1.0	H	G102
Glu 55	1	1	1	1	1	1	1	1	1	1.0	H	N113
Tyr 56	-	-	-	-	-	1	-	-	1	0.2	H	T47
Lys 58	1	-	-	-	-	-	-	-	-	0.1	H	N113
Lys 86	-	-	-	-	-	-	-	2	1	0.3	H	R112 [‡] , N113 [‡]
Glu 95	2	2	2	2	2	2	2	2	2	2.0	HS	N113, R112
Total	11	9	10	10	10	12	9	12	12	10.56		

[†] H = hydrogen bond, S = salt bridge; [‡] mediated by Obody backbone atoms

Table 5.3 Hot Spot predictions for the L8 10mL10 interface

Using three different models (see section 2.6.4.2), the nine complexes were analysed for hot spots, giving a total of 27 predictions for each residue (A). Contacting residues with no hot spot predictions are omitted. R35 may be under-represented as the calculated $\Delta\Delta G$ from virtual alanine scanning was very close to the cut-off level (2 kcal/mol) for inclusion as a hot spot. Average $\Delta\Delta G$ values calculated by Robetta alanine scanning for interface residues are shown in (B)

A		B	
Hot Spots		$\Delta\Delta G$ (kcal/mol)	
Leu 30	1	Asp 32	0.69
Gly 31	2	Tyr 33	5.54
Asp 32	7	Arg 35	1.99
Tyr 33	27	Val 36	1.04
Gly 34	7	Ile 38	0.58
Arg 35	11	Lys 40	0.71
Ile 38	6	Ser 51	-0.02
Tyr 53	6	Tyr 53	0.79
Glu 55	6	Glu 55	0.60
		Tyr 56	0.82
		Lys 58	-0.21
		Ser 85	0.17
		Glu 86	0.32

5.2.5.1 Waters at the interface

Like 13mRL L8, a cavity at the interface is visible between R35 and Y33 (Figure 5.5). In this structure six waters were clearly defined in electron density, raising the possibility of water-mediated interactions. Polar atoms in range for hydrogen bonding were available on both molecules. Comparison with a monomeric HEWL structure (PDB accession 193L) showed that these waters roughly overlaid with those found at the interface in the absence of a binding Obody, but did not match precisely; four were within 0.6 Å and the remaining two seen in the complex replaced only one in the non-complexed HEWL structure (Figure 5.5, inset).

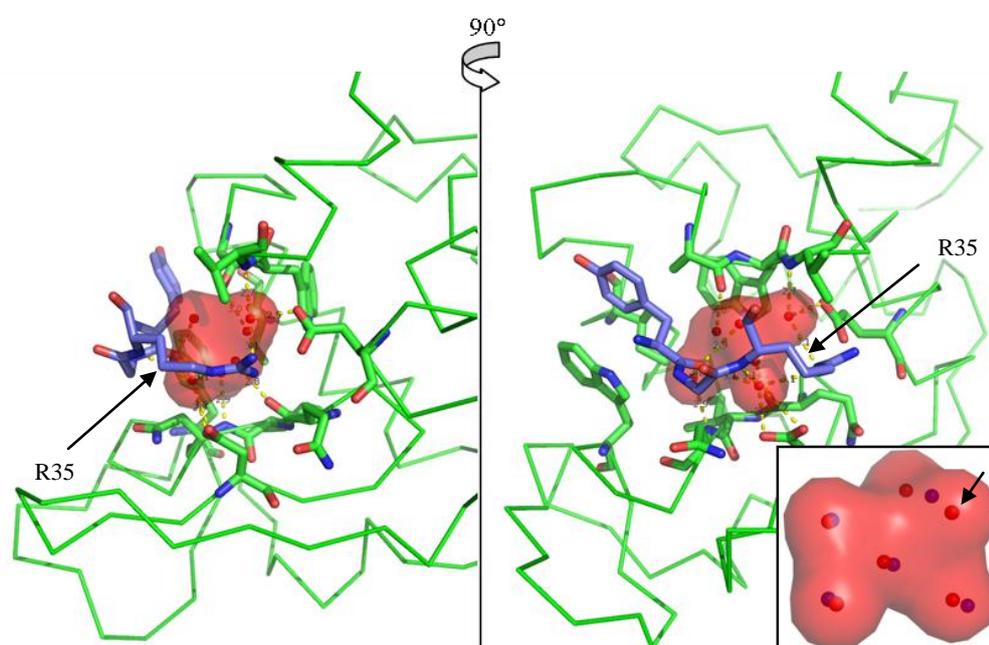


Figure 5.5 Water-filled cavity at the L8 10m L10-HEWL interface

Two views of the water-filled cavity at the Obody-HEWL interface, rotated 90°, represented by a surface model of the ordered waters (red). HEWL is shown in green as a Ca trace, residues that can form hydrogen bonds with the cavity waters are highlighted as stick diagrams. A three-residue peptide (Y33, G34, R35) from the Obody is shown in blue. The inset figure shows the waters in isolation (red surface/spheres), with an overlay of ordered water molecules from HEWL in the absence of Obody (blue spheres). Note the additional water introduced into the interface on Obody binding (arrow).

5.2.6 Summary

The L8 10m library was mutated at specific positions and re-selected for binding to HEWL, giving a ~10-fold affinity improvement. One variant was expressed and the crystal structure solved to a resolution of 1.95 Å in complex with HEWL to be used to determine the structural basis of that improvement.

The interface as whole revealed a further optimised state, evident as over-all increases in calculated $\Delta\Delta G$. In the complement of 10 mutant positions, one mutation in particular, A56Y, was identified as the most important, as it was closely associated with the largest structural change; the whole interface has rotated approximately 10° relative to the HEWL molecule in the ancestor Obody complex.

5.3 Results - L200EP 06

This gene was selected from panel of five mutant variants derived from a random mutation library on the basis improved affinity. The following section presents expression, purification, crystallisation and the structure of the protein in complex with HEWL for the purpose of determining the structural basis of an observed 2-fold affinity improvement.

5.3.1 *Expression and Purification*

Expression and purification was as described in section 2.5. Briefly, the Obody was expressed with an N-terminal His-tag from vector pProEx in cell strain DH5 α , purified on an IMAC column by hand (section 2.5.3.3), digested with rTEV (section 2.5.5) then re-purified on an S75 10/300 size exclusion column (Figure 5.6). Yield was approximately 45 mg from 1 L of culture.

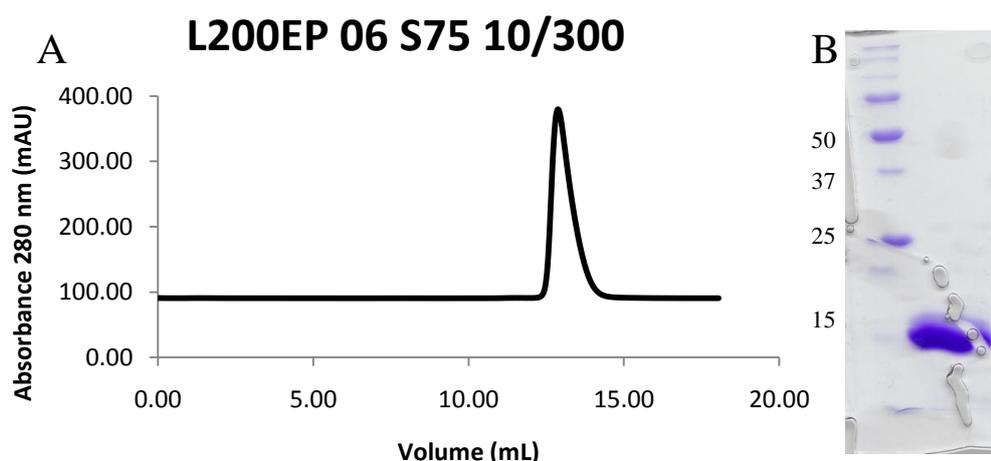


Figure 5.6 Purification of L200EP 06

(A) A representative chromatogram from the purification, showing elution from an S75 10/300 size exclusion column. Fractions were collected and pooled across the peak beginning at ~11.5 mL. (B) A representative SDS-PAGE gel showing purified His-tagged Obody, expected weight of 15,400 Da. Protein markers are labelled according to their weight, in kDa

5.3.2 *Crystallisation and Data Collection*

L200EP 06 was trialled for crystallisation in complex with HEWL using the same conditions, including concentration, as used with L8 10mL10 (section 5.2.2). Crystals were identified in approximately 50% of the conditions screened. Diffraction testing on the home source beamline at the Maurice

Wilkins centre identified two crystals diffracting to the detector limit (1.86 Å resolution at a distance of 150 mm) in condition E1 (precipitant: 0.2 M HEPES pH 7.0, 13% MPEG 5000), and one crystal diffracting to 2.21 Å resolution in condition G1 (precipitant 0.2 M HEPES pH 7.0, 17% MPEG 5000). Two datasets were collected, one from an E1 crystal (180 1° phi oscillation images), the other from the G1 crystal (Figure 5.7), with the same phi oscillation for a total of 159 images. Both datasets showed the same space group and a virtually identical unit cell (Table 5.4). Datasets were integrated with Mosflm, in spacegroup P222. The Matthews co-efficient for both datasets indicated a single complex in the asymmetric unit (2.3 Da/Å², 46% solvent).

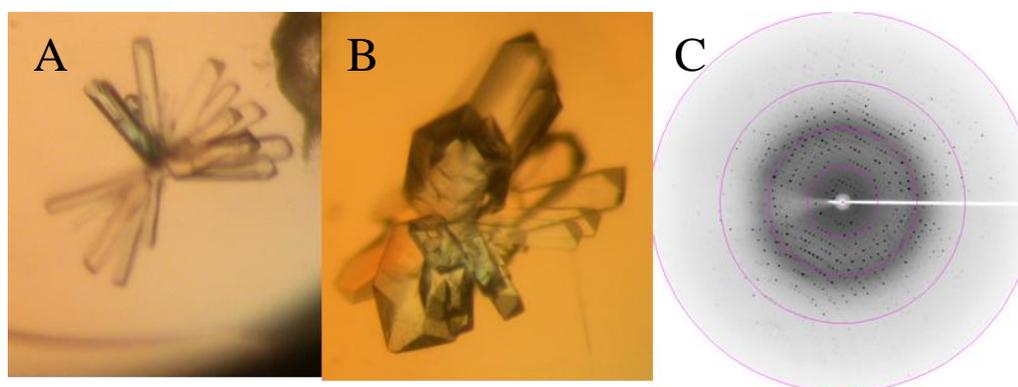


Figure 5.7 Crystals of L200EP 06 in complex with HEWL

These two clusters of crystals in conditions G1 (a) and E1 (b) were carefully broken up and fragments individually mounted for diffraction testing. Although scale was not formally measured, both clusters were approximately 0.2 mm from the centre to the end of the longest crystal. (c) A sample x-ray diffraction image from crystal 3 from condition E1. Starting from the centre of the image, circles mark 7.4, 3.7, 2.5 and 1.9 Å resolution.

5.3.3 Model building and Refinement

Dataset G1 was phased by molecular replacement using Phaser from the CCP4 suite. All possible space groups related to P222 were searched with one copy each of HEWL (PDB accession 193L) and L8 10m L10 (chain B). Only a single solution was found in $p2_12_12_1$ with one copy of each monomer in the asymmetric unit, with an overall LLG of 1612. The amino acid sequence was corrected by manual mutation and minimally refined, using Phenix minimisation with simulated annealing and Refmac, dropping the R-free to 26.83%. The dataset from crystal E1, which was found to possess the same spacegroup and unit cell, was used from that point on in the refinement. The

structure was refined using alternating rounds of Phenix minimisation with simulated annealing and TLS (TLS parameters were determined using the TLS Motion Determination server (Painter and Merritt 2006a; Painter and Merritt 2006b)) and Refmac restrained refinement, down to an R-free of 19.3%. The final model was validated with PROCHECK, which showed 96.8% of residues in favoured regions, 3.2% in allowed regions and no residues in unfavourable or disallowed regions of the Ramachandran plot. Final refinement statistics are summarised in Table 5.4.

Table 5.4 L200EP 06-HEWL complex structure statistics

Those figures in brackets represent the highest resolution shell.

	Crystal G1	Crystal E1
Data Collection		
Space Group	p212121	p212121
Unit cell (Å)		
a	50.5	50.43
b	57.86	58.33
c	82.84	81.82
α, β, γ (°)	90	90
Wavelength (Å)	1.54179	1.54179
Resolution Limits (Å)	31.79 - 2.21 (2.33 - 2.21)	34.58 - 1.86 (1.96 - 1.86)
Measured Reflections	75787 (9463)	143392 (18275)
Unique Reflections	12195 (1570)	20797 (2819)
Multiplicity	6.2 (6.0)	6.9 (6.5)
Completeness (%)	97.7 (93.0)	99.4 (94.1)
Rmerge (%)†	8.7 (44.1)	3.9 (22.6)
$\langle I/\sigma(I) \rangle$	15.9 (4.3)	29.2 (7.5)
Wilson B (Å ²)	32.7	26.4
Mosaicity (°)	0.8	0.5
Molecular Replacement‡		
Log-likelihood Gain	1611.90	-
Refinement		
Resolution (Å)	35.79 - 2.31 (2.37 - 2.31)	34.58 - 1.86 (1.95 - 1.86)
Reflections	10404 (752)	20534 (2293)
R_{work} (%)*	21.97 (23.7)	15.59 (17.3)
R_{free} (%)*	26.83 (36.1)	19.33 (22.0)
Free reflections	519 (31)	1341 (60)
Refined Atoms	1831	2156
r.m.s.d bond lengths Å	0.023	0.025
r.m.s.d bond angles (°)	1.982	1.886
$\langle B \text{ factor} \rangle$ (Å ²)	44.25	20.443

† $R_{\text{merge}} = \sum |I_{\text{obs}} - \langle I \rangle| / \sum I_{\text{obs}}$

‡ Ref. (McCoy *et al.* 2007)

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

5.3.4 *Library Mutations*

The ancestor gene of L200EP 06 (L8 10m L200) was derived from the same library that yielded the second structure presented in this chapter, L8 10m L10 (section 5.2), and while affinities are comparable (Table 4.2), residues selected from the L8 10m library were different at all 10 positions. Of the four residues targeted on the binding face, one (P51) was retained as ancestor-type. The remaining three mutations were S29H (Figure 5.8 B), K37M (reverted back to lysine in the subsequent random mutation library, Figure 5.10) and A56R (Figure 5.8 A). The six mutant residues in 4/5 loop were also completely different, but no direct structural comparison can be made as this loop was disordered in all previous structures.

In contrast to the A56Y substitution seen in L8 10m L10, arginine at this position on the L200EO 06 Obody does not make any hydrogen bonds with HEWL and Robetta alanine scanning indicates no direct impact on binding, although proximity to an acidic residue on HEWL hints at a possible electrostatic interaction (Figure 5.8). The R56 guanidine group stacks with F65, located on the Obody's only α -helix. Compared to the ancestral 13mRL L8 structure, the backbone carbonyl group of residue R56 has been flipped 180°, accompanied by a shift in the backbone of neighbouring residue G57.

S29H as seen in the present structure does not make any predicted intermolecular hydrogen bonds (Figure 5.9). However, this may be a function of the charge state of histidine; proximity (3.1 Å) to an H-bond receptor in the form of an aspartate carbonyl oxygen suggests that if it exists partially protonated, as it would at pH 7.4, an H-bond may be able to form. It would also potentially make a long-range electrostatic interaction (distance >4 Å) with the side chain of the same residue. This is of course assuming that the pKa of the imidazole group is unaffected by neighbouring charged residues. Robetta alanine scanning calculated a $\Delta\Delta G$ of 0.8 kcal/mol, indicating a positive contribution to binding.

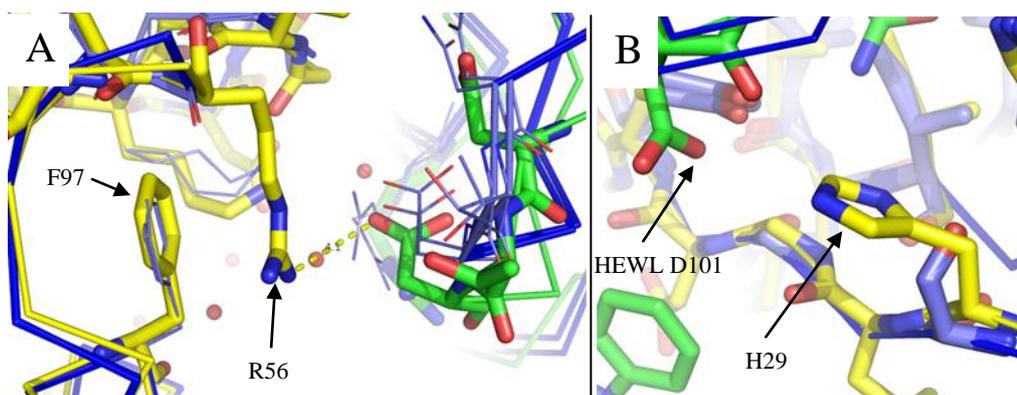


Figure 5.8 S29H and A56R

The L200EP 06 Obody is shown in yellow, its cognate HEWL in green. Overlaid for comparison is the 13mRL L8-HEWL complex, both in blue. R56 (A) stacks with F97 on the Obody's only helix capping one end of the barrel. S29H, while polar groups are close, is only partially protonated at pH 7.4, so the interaction with the pictured HEWL aspartate carbonyl is probably primarily via polar Van Der Waals contact.

5.3.4.1 Error-Prone Library Mutations

Compared to the ancestral sequence, L200EP 06 contained three residue substitutions: T19S, M37K and K86E (Figure 5.9). None of these mutations resulted in the formation of direct contacts with the HEWL ligand, so secondary and indirect contributions were specifically looked for. Consideration of the L200EP 06-HEWL complex interface for the structural basis of affinity improvement by this library was also in light of the pre-existing differences compared to the 13mRL L8 structure, which prevented complete confidence in isolation of the observed 2-fold affinity improvement to any of the three point mutations selected from the error-prone library. Indeed, such an improvement is unlikely to be represented by large structural changes and was expected to be a stabilisation or optimisation of existing interactions rather than introduction of new ones.

The first mutation was T19S, located on β -strand one, but on the opposite side of the Obody from the binding face. Interpretation of this mutation as a component of improved binding is difficult, as structural effects of the substitution are probably too subtle to be detected by x-ray crystallography. Effects on affinity, if any, are most likely to be a motion dynamics role better suited to an NMR study, therefore this mutation was disregarded.

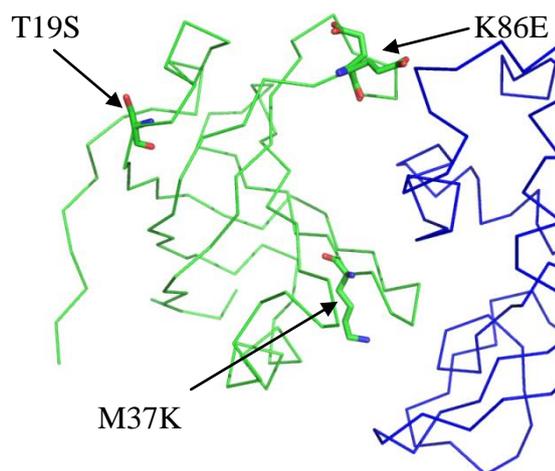


Figure 5.9 Mutations in L200EP 06

The Obody is shown as a green Ca trace in complex with HEWL, shown in blue. Mutant residues on L200EP 06 are highlighted as green stick diagrams.

5.3.4.2 M37K mutation

The second mutation in the L200EP 06 structure (M37K) represents a striking convergence to lysine (Figure 5.10) along with the sequences of both Obody structures already presented in complex with HEWL in this thesis (Figure 3.9 **Error! Reference source not found.**, Figure 5.3), reproducing the water mediated interactions found in L8 10mL10. Because the utilised models for protein-protein interface analysis exclude explicit water molecules, the impact of this interaction is under-represented in the computational interface analysis methods used here. If one disregarded this limitation of the models employed, the curious durability of this particular substitution suggests an important role in binding, over and above simply removing the like-charge clash seen in the 13mRL L8-HEWL complex.

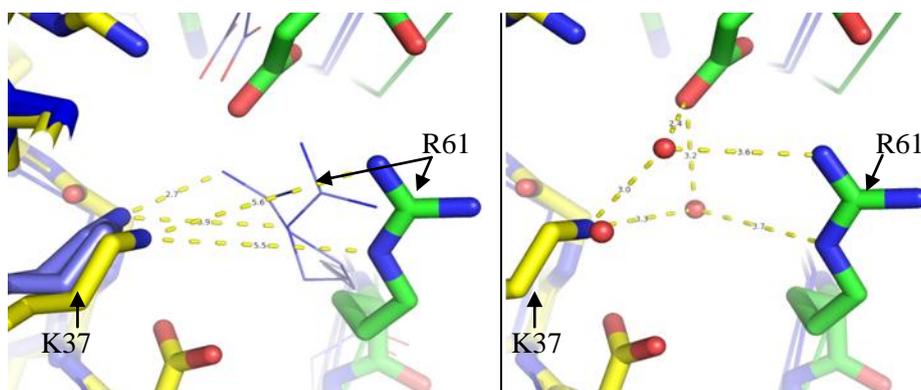


Figure 5.10 M37K Mutation

K37, originally targeted for mutation due to proximity HEWL R61, here contributes to a water-mediated interaction network. (A) comparison of the contact 13mRL L8-HEWL residues (shown in blue). The L200 EP06 structure shows that the HEWL arginine has shifted and been replaced with two ordered waters (B).

5.3.4.3 K86E Mutation

The third mutation, K86E, appeared to have the greatest potential impact (Figure 5.11). In both 13mRL L8 and L8 10mL10, K86 clearly played a role in binding, as indicated by the availability of a hydrogen bond donor on HEWL, buried surface area (82 \AA^2), hydrophobic packing with HEWL K116 and a highly ordered side-chain in the structures. Its replacement with a glutamate in L200EP 06 prevented formation of the same H-bond and hydrophobic interactions. Robetta alanine scanning indicated that the glutamate at this position is actually mildly destabilising (Table 5.6) though this is most likely an artefact of poor electron density leading to uncertain side chain placement.

K86 is one of the most commonly mutated positions positively selected from the library, with mutations to glutamate, arginine or asparagine in 6 of 23 unique sequences (Figure 4.13). It borders the 4/5 loop region, so it follows that a tight positional restraint imposed by this residue has consequences on the conformation of the loop residues, as well as those flanking the loop. Indeed, the β -strands supporting the 4/5 loop provide a major salt bridge interaction in E95, which has undergone significant reorganisation during the two rounds of affinity maturation of which this structure is a product (Figure 5.11B) culminating in a 50% increase in calculated $\Delta\Delta G$.

In addition this substitution was hypothesised to influence was the dynamics of the 4/5 loop. The most obvious difference between L200EP 06

and the two previous structures is the ordered 4/5 loop and its involvement in binding. No direct contacts with HEWL were observed by the first residue of the loop, Q87, but H-bonds to E95 and Y53 were evident. Y53 appears to have stabilised considerably, with no evidence for multiple conformers as visible in the previous two structures. As well as a hydrogen bond to Q87, aromatic stacking with HEWL 112 improved, with the angle between the two aromatic planes considerably closer to parallel (approximately 10° from parallel, compared to 40 ° in 13mRL L8). The second residue, Y88 showed an extensive burial of hydrophobic surface (140 Å² of solvent accessible surface area), binding into a depression on the surface of HEWL (Figure 5.11A). This arrangement involving E86, Q87 and Y88 would be altered significantly if the K86E substitution were reversed, via competition with the hydrophobic interaction made by K86, forcing the backbone of the residue at position 86 into a different conformation and altering the dynamics of the loop and neighbouring residues.

The importance of the aforementioned residues (E95, Y88 and Y53) in binding is highlighted in the computational interface analysis (Table 5.6). Both E95 and Y88 were picked as hot spots by Robetta virtual alanine scanning, and ranked 2nd (E95, 6 kcal/mol) and 3rd (Y88, 2.6 kcal/mol), for calculated $\Delta\Delta G$, after Y33 (7 kcal/mol).

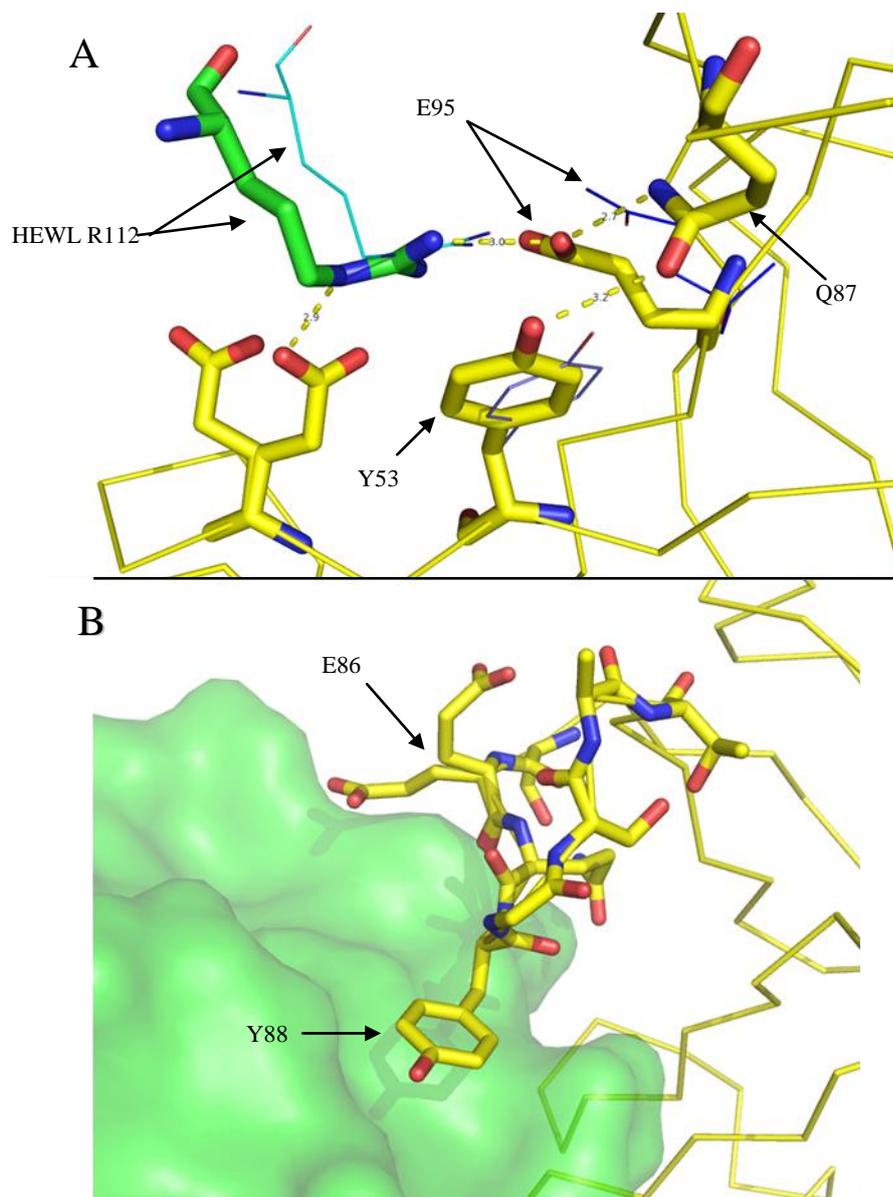


Figure 5.11 Loop mutations and their effect on the interface

The L200EP 06 complex structure, shown as a yellow Ca trace with residues highlighted as stick diagrams, with the position of selected ancestor (13mRL L8) residues shown as line diagrams, positioned by a Ca superposition of the Obodies. HEWL is represented by a green surface, or residue. (A) Loop residue Q87 protrudes down into the top of the interface, hydrogen bonding to Y53 and E95, both of which are in contact range of HEWL residue R112. The changes in residue position are obvious by comparison with the ancestor-type residues. (B) The largest increase in buried surface area can be attributed to Y88, which lies in a groove on the HEWL surface, burying 140 Å². It is proposed that the substitution K86E contributes to the positioning of both Y88 and Q87, by altering the conformational dynamics of the loop and surrounding residues. E86 is present as two equally occupied alternate conformers, suggesting there are no stable bonds formed by this residue.

5.3.5 Description of the Obody Interface.

The L200EP 06 structure shows Obody and HEWL chains that do not significantly depart from the previous structures in the core fold region; structural alignment of the L200EP 06 Obody with the nine monomers in the L8 10mL10 asymmetric unit gives r.m.s.d. values between 0.36 and 0.58 Å. Similarly, HEWL aligns with r.m.s.d. of between 0.22 and 0.25 Å. A water-filled cavity is found at the interface between Y33 and R35 which closely resembles that shown in both previous structures, down to the position of the waters (Figure 3.3 and Figure 5.5), although the gap index for this complex has dropped to 2.1 Å³, indicating a greater degrees of complementarity. As calculated by PDBePISA the complex does not contain more hydrogen bonds or salt bridges (Table 5.5), although the PROTORP server shows a gain of 5 over 13mRL L8-HEWL complex (Table 5.7). The hydrophobic patch of Y33, V36 and I38 at the centre of the interface overlays compares closely with the previous two structures, although the relative importance of V36 and I38 appear to have decreased in favour of Y33 and K40. Polar interactions are also similar, with 10 polar bonds including two salt bridges predicted by PDBePISA. The most significant difference is an increase in buried surface area up to 945 Å² on the Obody alone (total 1800 Å²), mostly attributable to Y88 in the 4/5 loop, which is ordered and contacting HEWL in this complex (Figure 5.12). The crystal form is identical to that seen in L8 10m L10, although with a much smaller unit cell; it is interesting to note here that axes b and c are almost exactly one third the length.

Table 5.5 Polar bonds at the L200EP-HEWL interface

Calculated by PISA, polar bonds are listed along with their interaction partners on HEWL.

Residue	Count	Type†	HEWL Contact
Asp 32	1	H	W63‡
Tyr 33	2	H	N103, D101
Arg 35	2	HS	E35, D52
Lys 40	1	H	G102
Glu 55	1	H	N113
Ser 85	1	H	K116‡
Glu 95	2	HS	R112

† H = hydrogen bond, S = salt bridge

‡ Mediated by Obody backbone atoms

Table 5.6 Hot spot predictions of the L200EP 06-HEWL interface

Analysis of the binding interface using the KFC and Robetta servers show predicted hot spot residues (A). The raw $\Delta\Delta G$ data from the Robetta alanine scanning server is shown in (B).

A	K-FADE	K-CON	Robetta	B	$\Delta\Delta G$ (kcal/mol)
	Gly 31		Yes		
Asp 32		Yes		Asp 32	0.63
Tyr 33	Yes	Yes	Yes	Tyr 33	7.37
Gly 34		Yes		Arg 35	1.84
Arg 35				Val 36	1.08
Val 36				Ile 38	0.63
Lys 40				Lys 40	1.32
Tyr 53		Yes		Tyr 53	1.66
Tyr 88			Yes	Glu 55	0.47
Glu 95			Yes	Arg 56	-0.01
				Ser 85	-0.05
				Glu 86	-0.09
				Tyr 88	2.60
				Glu 95	6.13
				Phe 97	0.04

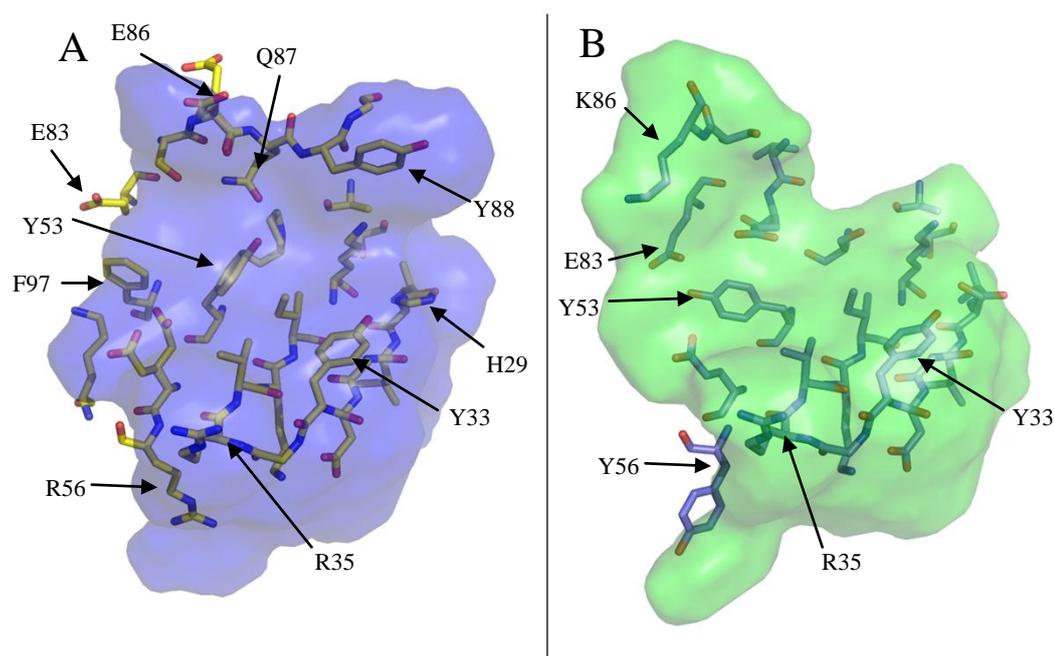


Figure 5.12 FADE surface complementarity at the interface

A side by side comparison of shape complementarity at the interfaces of L200EP 06 and L8 10mL10 (AB complex) as determined by FADE, represented by a surface model and associated Obody residues. The L200EP 06 interface contacts (A) differ from the L8 10mL10 interface (B) in two main regions. Y88 illustrates the largest difference, making extensive contacts with HEWL, increasing buried surface area of the L200EP 06 complex by 144 \AA^2 . Removal of K86 also seems to remove hydrophobic clashes with E83, potentially influencing stability of Y53 and surrounding residues.

5.3.6 *Three-structure comparison*

Superposition of the all three Obody structures and comparison of the relative position of the HEWLs in complex reveals that L200EP 06 has a different arrangement than the other two structures (Figure 5.13). Similar to the L8 10m L10 complex, L200EP 06 association with HEWL has undergone a rotation of at least 10 ° about the longitudinal axis of binding, but also a lateral shift, centred on Y33. Examination of the HEWL structures shows the place of least variation in C α position across all three structures is the pocket into which Y33 binds cementing this residue as the most important single hot spot in the interface.

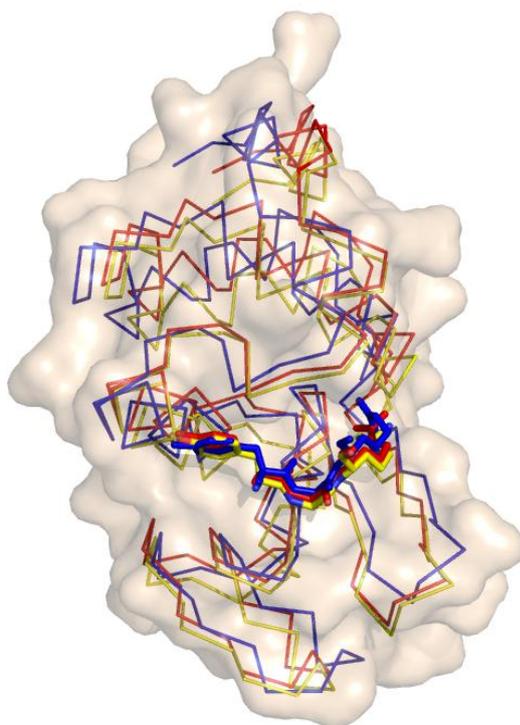


Figure 5.13 *Overlay of all three Obodies and their cognate HEWLs*

13mRL L8 (blue), L8 10m L10 (red) and L200EP 06 (yellow) in complex with HEWL are overlaid using the Obody C α trace (displaying only the three residue peptide Y33, G34, R35 from the Obodies). Cognate HEWLs (same colour key, C α trace) show the orientation differences between their associations with the Obodies.

For comparative purposes, the interface statistics calculated by PROTORP (Reynolds *et al.* 2009) are displayed in Table 5.7. The comparison highlights the iterative interface improvements, in terms of increasing numbers

of hydrogen bonds, increasing buried surface area and reducing gap index. Differences between these and the previously cited values for buried surface area and hydrogen bond quotients can be accounted for by different calculation methods used by PROTORP and PDBePISA.

Table 5.7 Comparative interface statistics

These data were calculated using the PROTORP interface comparison server

	13mRL L8	L8 10m L10	L200EP 06
Buried surface area	840 Å ² (14%)	860 Å ² (14%)	962 Å ² (16%)
H-bonds	8	11	13
Salt bridge interactions[†]	5	5	4
Polar:Non-polar atoms %	43:57	44:56	38:62
Gap Volume Index[‡]	2.94	2.50	2.10
<i>K_d</i>	36 μM	5 μM	610 nM

[†]Salt bridge count is defined as the number of charged HEWL residues within 4 Å of an appropriately charged Obody residue.

[‡]Gap Volume index is defined as the ratio between buried surface area and enclosed volume not occupied by protein, as per (Jones and Thornton 1996).

5.4 Discussion

Structural information about a potential scaffold is an invaluable source of insight when both designing a library and analysing resulting selections. This chapter presented two structures, both the product of affinity maturation libraries based on a single variant from a naïve library selected for affinity for HEWL.

5.4.1 Affinity maturation - L8 10m L10

The first structure of an Obody (13mRL L8, Chapter 3) presented as a validation for the initial library design, was used to design a new library to improve affinity. Rational selection of residues for mutation in the L8 10m library yielded several variants with diverse sequences but similar affinities, of which the one with the best affinity (L8 10m L10, K_D 5 μM) was solved by x-ray crystallography in complex with HEWL to determine the structural basis for affinity improvement over the 13mRL L8 ancestor. Immediately obvious in comparison was that the fold tolerates mutation well, and was capable of

accommodating changes to residues in both the β -strands and the loops while maintaining stability.

What seemed evident from the sequence data of the HEWL-selected L8 10m library variants was that any binding using the 4/5 loop was subordinate to selection of residues at the face; no consensus was evident in the loop residues but ample at the face positions. This was supported by the structure of variant L10 presented here (section 5.2) in which the 4/5 loop was almost completely disordered in the structure, strongly implying no contribution to binding in this complex. Instead the major differences were tracked to the periphery of the interface, where introduction of a Tyr for Ala at position 56 appears responsible for a rotational shift in HEWL relative to the Obody by approximately 10° , producing a new polar bond and some limited hydrophobic contact.

Although the re-selection of ancestor-type residues at position S29 and K37 was surprising (section 5.2.4) considering the 10-fold increase in affinity, the structure offered clues why this may be the case; both positions were altered in their interaction with HEWL by the relative shift outlined above. K37 no longer clashed with HEWL R61. S29, originally identified for replacement with a larger residue to bridge the gap between the two proteins, was brought into closer contact with HEWL. Although this residue didn't make significant intermolecular contacts, retention of S29 with its small, polar side chain may accommodate the shift imposed by Y56 on the other side of the interface (Figure 5.3). By itself, the average calculated $\Delta\Delta G$ of 0.82 kcal/mol of Y56 accounted for a four-fold increase in affinity. It is expected that the remainder of the gain can be attributed to an overall optimisation of the interactions across the face, as alluded to by a general increase in calculated $\Delta\Delta G$ across the interface residues (Table 5.6).

The presence of buried waters at the interface (Figure 5.5) provided a clue as to why the affinity of this Obody remains only moderate (5 μM) when the interface statistics compare favourably with other HEWL-binding proteins (Table 5.7, Table 3.4). Although conserved waters in protein-protein interfaces are thought to contribute positively to binding (Reichmann *et al.* 2008), perturbation of the HEWL solvation shell, as shown in Figure 5.5 by the

Obody during binding may impose an energy penalty , especially considering that the waters are in a pocket bordered by hydrophobic atoms from the critical Obody residues Y33 and R35).

5.4.1.1 **Consideration of x-ray data**

The crystal used to solve this structure showed a surprisingly long c axis in the unit cell (245 Å). Statistics presented by SCALA did not show a twinned crystal, but a Patterson map suggested the presence of pseudotranslation in one axis, which was indeed reflected in the asymmetric unit of nine complexes. The later structure of L200EP 06 in complex with HEWL exhibited very similar crystal packing in the same spacegroup, although with only one complex in the asymmetric unit. Interestingly, the unit cell parameters calculated for the L8 10m L10 complex structure were almost exactly three times those calculated for the L200EP 06-HEWL complex on two axes, strongly implying that the nine-complex asymmetric unit was a pseudotranslational expansion.

This consideration impacts on my decision to treat the variation seen across the asymmetric unit as real; crystal packing may be the cause of the disorder, rather than the other way around. However, the same interface residues showing variable conformations in this structure were observed as variable in the 13mRL L8 structure, where crystallisation conditions, space group assignment and crystal packing were different. In the same vein, the increase in affinity seen in L200EP 06 seems to be correlated with stabilisation of the aforementioned variable residues, where the crystallisation conditions are very similar and the space group is the same. Additionally, in the nine complexes there were three dominant conformations seen in the position of Y53 (Figure 5.4), with evidence for further alternative side chain conformers in all complexes. That being said, a general note for crystal structures states that the conformations seen in the asymmetric unit do not necessarily reflect the nature of the crystallised protein(s) in solution. So while each Obody-HEWL complex in the L8 10m L10 structure may not individually represent a different species in solution, it seems reasonable to conclude that the observed variation is relevant to binding.

5.4.2 *Affinity maturation – L200EP 06*

The second approach to affinity maturation was random mutagenesis. Because of the random component, the beneficial (i.e. affinity-increasing) mutations can potentially affect binding by long-range propagation of conformational changes, making it difficult to isolate binding effects as the consequence of a particular residue change. The choice of template gene for investigation in this process (L8 10mL200) prevents a direct comparison with the ancestor gene, so 13mRL L8 was used instead.

Because the direct ancestor of L200EP 06 does not have a structure available, examination of the interface avoided inferences that rely on the positions which are different between L10 and L200EP 06, but unchanged from the L200EP 06 ancestor gene. Thus, while direct comparison does give some information about the differences between the three Obodies that may account for affinity changes, it is not sufficient to say with certainty the effect of a particular mutation.

5.4.2.1 **Proposed source of affinity improvement**

L8 10m L200 differed from the L10 variant at every position that was randomised. S29 was replaced by a histidine, A56 with an arginine, K37 with a methionine, while P51 is retained as ancestor (13mRL L8)-type. Consequently, these positions, along with all of the 4/5 loop residues, were largely discounted as the causative agents for increased affinity in terms of the new contacts they introduce. Bearing that in mind, the structure suggested that a loss of one interacting residue (K86E) and the re-selection of another (M37K) are the clearest causes of affinity improvement compared to the direct ancestor, L8 10m L200.

Allowing the 4/5 loop to more easily assume a conformation favourable to binding by removing competing interactions formed by K86 may account for a portion of the improvement of affinity for HEWL. Even if no additional polar bonds are formed, direct involvement of the 4/5 loop does increase the buried surface area by more than 100 \AA^2 , primarily by hydrophobic contacts between Y88 and HEWL residues R21, Y23 and G104. This conclusion must be considered in the knowledge that the 4/5 loop was present unchanged in the ancestor of L200EP 06, so the observed binding affinity improvement cannot

be attributed simply to the presence of the loop sequence. Instead, any influence by mutation of position 86 must come instead from altered dynamics of that loop, or surrounding residues.

Aside from the loop residues themselves, the two β -strands (4 & 5) which support that loop has shifted compared to 13mRL L8, which I suggest is to accommodate the different 4/5 loop configuration. Two structurally adjacent glutamates from these strands were also involved in binding, at positions E83 and E95. E83 did not directly take part in binding, but was identified as part of the cluster of residues with varied conformations in L8 10mL10 (Figure 5.4), suggesting that it may play a co-operative role in organisation of the neighbouring interface residues. Where position 86 was a lysine, as in L8 10mL10, the carboxyl group of E83 appeared to be flattened between the hydrophobic portion of the lysine side chain and F97, which almost certainly influenced the position of structurally adjacent residues Y53 and K55. Given that residues which are important for binding tend to be in states of restricted conformational freedom (i.e. pre-organised into binding configurations, (Li *et al.* 2004)), co-operative interactions such as these may have significant impacts on binding by ameliorating the loss of degrees of freedom entropy. E95, located on the fifth β -strand immediately after the 4/5 loop, formed a salt bridge with two HEWL residues in all structures. Compared to central binding residues Y33 and R35, arrangement of the salt bridge between E95 and HEWL R112 shifted by 2 Å, which also resulted in a more favourable aromatic interaction with Y53.

5.4.2.2 Caveats

Although a general conformational stabilisation of the so-called variable cluster seen in L8 10m L10 (Figure 5.4) is hypothesised to be involved with improvement of affinity, other differences that were present before the error-prone affinity maturation library may also influence the conformational stability of the cluster as compared to L8 10m L10. Q87, the first residue in the randomised portion of the 4/5 loop, pushes down into the interface and hydrogen bonds with Y53, thereby influencing its conformational freedom. Mounted on the loop at the end of β -strand 3, R56 shows aromatic stacking with a phenylalanine, which may account for a minor shift seen in the

backbone of that loop and consequently E55 as well which bonds directly to HEWL. Because a structural comparison with the direct ancestor is not available, any improvement attributed to the K86E mutation cannot be deconvoluted from these pre-existing differences. I must therefore conclude that stabilisation of this variable cluster is the result of the combination of all of these mutations.

5.4.3 *Conclusion*

The over-all result is one of a general improvement across most interactions of their contribution to ΔG of binding, following a re-orientation of the Obody-HEWL interface. I propose that the shifts seen in both structure presented here resulted in further optimisation of the various intermolecular interactions, particularly at the critical hot spot residues, driven by changes on the periphery of the interfaces. This sort of rigid-body settling of the interface towards a more optimal binding configuration was the expected result from the two rounds of affinity maturation.

6 General Discussion

Specific, customised binding reagents are vital tools in science, for understanding biochemical processes, sensitive detection of important molecules, or as therapeutics. Before molecular techniques for the routine manipulation of biological macromolecules became available, reagents were derived from the naturally occurring reservoir of selectable diversity that was available in the vertebrate immune system. While immunisation did, and still does, provide adequate tools for many applications, what became apparent was that antibodies in their naturally occurring forms were not competent to perform all of the tasks that researchers and clinicians required. Specifically, the large size, sensitivity to redox state and reliance on glycosylation drove decades of work seeking to simplify, improve and eventually replace natural antibodies with engineered immunoglobulin domain constructs as the benchmark scaffold for molecular recognition (Better *et al.* 1988; Huston *et al.* 1988; Riechmann *et al.* 1988a; Riechmann *et al.* 1988b; Bird and Walker 1991; Buchner and Rudolph 1991; Vaughan *et al.* 1998; Knappik *et al.* 2000).

Today, using developments in gene manipulation and synthesis, application of evolutionary principles have been used not only to improve antibodies, but to investigate a wide range of different protein folds as affinity scaffolds (Skerra 2000; Binz *et al.* 2005; Binz and Pluckthun 2005; Skerra 2007a; Gebauer and Skerra 2009). Particular emphasis is often placed on therapeutic applications in the literature (Sidhu and Fellouse 2006; Tolmachev *et al.* 2007a; Zafir-Lavie *et al.* 2007; Bloom and Calabro 2009; Friedman and Stahl 2009), but these technologies have potential for use in any context which requires specific binding. While the techniques employed are in wide-spread use, the process of investigation remains time-consuming, expensive and technically challenging.

The Oligonucleotide/Oligosaccharide binding fold unifies a large collection of structural superfamilies, with no sequence conservation (Murzin 1993; Arcus 2002). It is a small, stable β -barrel fold, found performing diverse functions in evolutionarily diverse organisms and therefore meets the criteria identified as necessary to be a candidate for engineering as a scaffold (Skerra 2000; Binz and Pluckthun 2005). Previously, we selected as a template the

anticodon binding domain of aspartyl tRNA synthetase from *Pyrobaculum aerophilum* for investigation into OB-folds as molecular recognition scaffolds. Its ability to tolerate extensive mutation in an artificially randomised binding face, remain soluble and be displayed as a pIII fusion product on the surface of M13 filamentous phage was demonstrated. Using phage display and a library with 17 randomised codons, preliminary selection experiments against hen egg white lysozyme (HEWL) as a model protein target succeeded in converting the nucleic acid binding domain into a specific protein binder, with a moderate K_D of $\sim 40 \mu\text{M}$. This first-pass investigation of our candidate scaffold gave results comparable with other synthetic scaffolds from first-pass phage display libraries (Nord *et al.* 1997; Koide *et al.* 1998). However, further validation of the fold was needed to begin to bring it up to the standard achieved by other scaffolds.

6.1.1 *Structural Validation and Library Improvement*

The data presented in this thesis builds on that preliminary work, to establish OB-folds as a viable scaffold candidate for generation of customisable molecular recognition reagents. First, a major flaw in the initial selection of the domain was that no structure was available from which to make definitive decisions about which residues to target for mutation. To address this, the crystal structure of an Obody variant from the preliminary selection experiment was solved in complex with its target, hen egg white lysozyme (HEWL), to validate both the library design and investigate the structural basis for binding. The structure of 13mRL L8 in complex with HEWL showed in atomic detail the specifics of the selected binding face (section 0), revealing that the critical intermolecular interactions were indeed mediated by mutant residues, though a significant component of binding was contributed by native residues. What was surprising was that although the 13mRL L8 Obody binding face possessed all of the defining characteristics of a good interface, and that the statistics of the face compared favourably with other HEWL complexes in terms of buried surface area, gap index and complement of polar interactions, the measured affinity of the Obody was 3-4 orders of magnitude lower than the other published HEWL complexes examined (section 3.2.4). Working on the assumption that this reflected the

presence of residues preventing tighter binding, affinity maturation phage libraries were designed based on the structure.

6.1.2 *Structure-based affinity maturation for HEWL*

Two attempts were made to use structural information as the basis for affinity maturation libraries. The first was a second-generation library based on a gene extracted from a naïve library, with 10 randomised codons, named L8 10m (section 4.2.4). Panning against HEWL resulted in a 10-fold improvement in affinity. Frustratingly, this library was hampered by extensive undesired point mutations made during synthesis of the incorporated mutational oligonucleotides, so it is my belief that this library should have performed better than the 10-fold increase in affinity it produced ($K_d \sim 5 \mu\text{M}$), though it clearly still retained sufficient numbers of viable genes to produce a better binder. This problem was not re-addressed due to time and budget constraints, and because improvement was indeed shown, the newly selected variants were taken as the new benchmark for further experiments.

The crystal structure of the L10 variant from this library was solved in complex with HEWL to a resolution of 1.95 Å (section 5.2). That structure provided fairly clear evidence for the nature of the structural changes that led to affinity improvement, which was proposed as the substitution of an alanine with a tyrosine at position 56 and a rigid-body rearrangement of the complex (Figure 5.13), resulting in a closer to optimal interaction, as reflected by an average increase in calculated $\Delta\Delta G$ across the interface residues. This re-shuffling of the interface appeared to remove the like-charge interaction targeted by the library design, even though the ancestor-type Obody residue was reselected at this position (Figure 5.3). Critically, the core fold remained robust and essentially unchanged with the additional mutations introduced.

6.1.3 *Third generation structure-based affinity maturation*

The second attempt to use structural information was performed using the structure of the second generation Obody in complex with HEWL. One indication of high-quality binding is shape complementarity; the precision with which residues fill clefts in target proteins is important enough for binding that

it has been used to identify hot spots (Li *et al.* 2004). What was evident in both structures solved at the time of designing this library was a large, water-filled gap right next to two of the most critical residues for binding, namely R35 and Y33. Filling this gap would more closely match the Obody to the contours of the HEWL active site, which could in principle result in a significant improvement in affinity. The failure of this library to positively select for new variants provided an insight into the limitations of the binding mode that this particular lineage of Obody has adopted. It may be that, because of the β -sheet nature of the peptide inserted by the Obody into the HEWL active site, a more complementary shape that fills the cavity may not be energetically favourable, as it may require denaturation of the β -sheet character of certain residues, in addition to displacement of the ordered waters which would also impose an energetic penalty (Hendsch and Tidor 1994; Dong and Zhou 2006).

To illustrate this point with an analogous system of binding evolution; selection of antibodies in the vertebrate immune system goes through two stages, each resulting in a polyclonal response (Rajewsky 1996). In the first stage, selection is done from the initial complement of naïve B-cells, resulting in the amplification of the subset which express antibody genes that show binding. The second stage occurs after re-exposure to the same antigen. Those lymphocytes selected in the first round go through a second round of proliferation and mutation of their antibody genes. The critical aspect here is that the highest affinity variants that come to dominate the immune response after re-infection are not necessarily descended from those variants which were the dominant after initial exposure (Malipiero *et al.* 1987). In other words, in an energy landscape of the free energy of binding, a gene sequence in a lower local minimum may be at a disadvantage versus those with a higher binding free energy when affinity maturation is attempted. This effect may be at play here with this particular Obody library, and the energy penalties may be too great to overcome through mutation of the few residues targeted. By extension, a different variant from the original 13mRL library may perform better.

6.1.4 *Third generation random affinity maturation*

Drawing from affinity maturation processes seen naturally in antibodies (Griffiths *et al.* 1984), random mutation by error-prone PCR was used to introduce variation into a third library. Random sampling of sequence space over the whole protein is capable of finding mutations that optimise the interface which rationally-designed libraries would typically not consider. Considering the failure of the previous attempt based on the L10 variant from L8 10m (the L10 6m library) a different gene was selected; variant L200. This gene showed a similar affinity as L10 for HEWL (section 4.2.4.3), but seemed more attractive as it arose more frequently than other variants. The mutant residue complement in L200 also seemed to be more in line with the rationale of the original design of the library which produced it (L8 10m, section 3.2.5); bulky residues were selected where I expected them be, a like-charge clash was removed (replacement of K37 with methionine), and the 4/5 loop contained a tyrosine, which has been found to be over-represented in protein-protein interfaces (Lo Conte *et al.* 1999).

This effort was partially successful and produced the first sub-micromolar Obody, measured at 600 nM (L200EP 06), which, when compared to an ancestor-type variant control measured using the same chip surface, is a 2-fold improvement in affinity. This result was called a partial success as the aim was to produce a binder with a K_D of 100 nM or less, which was not achieved. That a substantially tighter binder was not produced from a library of 10^8 variants supports the, admittedly speculative, conclusion that this lineage of Obodies may occupy a local minimum in free energy of binding that is prohibitively remote from other, better minima. The scale of mutations needed, starting from the template sequence used here, may not be available by the methods employed. However, I cannot eliminate the possibility that the variants present in the library were simply an unlucky sampling of the sequence space.

6.1.4.1 **Structural Analysis**

The crystal structure of this variant in complex with HEWL was solved to high resolution (1.85 Å) and showed an interface that PDBePISA calculated as similar to the previous two, but with a large increase in buried surface area of more than 100 Å², attributed to a single tyrosine (Y88). Data from an

alternative interface server (PROTORP) showed a steady increase, with affinity maturation, in the complement of intermolecular polar bonds, increase in buried surface area and decrease in gap index. Unfortunately, because a structure of the immediate ancestor complex was not available, the impact of the three mutations found in this variant was difficult to determine precisely. Instead, the structure was compared to its ancestor, 13mRL L8, and the cumulative increase in affinity was attributed to the mutations which occurred in both maturation steps it underwent.

In a similar manner to the L8 10m L10 variant complex, the L200EP 06 complex showed a substantial over-all shift of HEWL relative to the Obody, implying that a major determinant of affinity improvement was optimisation of the interactions already present on the binding face, in addition to the introduction of buried hydrophobic surface area on Y88 (Figure 5.13). This conclusion was supported by the an average increase of calculated $\Delta\Delta G$ from this structure as compared to the others (Table 3.3, Table 5.3, Table 5.6) indicating that individual residues found more energetically favourable binding states, which can be tentatively extrapolated to the face as a whole. For example, Y33, consistently identified as the most important single residue for binding, showed a $\Delta\Delta G$ of more than 7 kcal/mol, compared to the same residue in 13mRL L8 with less than 4 kcal/mol. This can be compared to the most substantial new interaction at Y88, which is calculated at 2 kcal/mol, which may be an over-estimate given that the electron density map was less-well defined than other interface residues, resulting in greater uncertainty in the position of the side chain atoms.

Speculatively, the K86E substitution was proposed as a major source of the observed improvement from the error-prone library, by virtue of proximity to three major binding residues, but the data at hand are not sufficient to speak with certainty about the real impact that this mutation might have had. Perhaps more obviously, a convergence with the other Obody structures at position 37 to lysine was noted, which established a water-mediated interaction network. Unique amongst the structures presented thus far there was only a single conformer evident in the structure. Taken by itself, this implied that a major driver in the evolution of L200EP 06 variant was in fact stabilisation from multiple binding modes down to one, as seen in the previous two structures,

although this can partially be accounted for by uncertainty in atomic placement given the moderate resolution (2.7 Å) of the 13mRL L8 structure, and the pseudosymmetry evident in the L8 10m L10 structure.

6.1.5 *Conclusions*

This research has shown that while considerable work towards optimisation of the panning process needs to be done to improve achieved affinities, the basic premise of this thesis is correct; engineered OB-folds can be used as a scaffold for mounting customised functional residues for the purposes of specific molecular recognition.

While the complexation arrangement was clearly different in each structure, and even variable within a single complex, the conservation of the OB-fold β -barrel residue positions displayed to good effect the competency of this domain as a scaffold; where residue sequence and interactions vary, the core residues showed very little movement. This was also reflected in the T_m values of around 80°C for all selected variants (section 4.2.8), showing the retention of a thermostable fold even after mutation of 21 positions as in L200EP 06. Together, the four selected variants with a measured T_m constituted a survey of 23 positions, with 34 unique substitutions compared to the wild-type nucleic acid-binding ancestor domain. That all of these exhibited a thermostable fold is a potent indicator of the potential of this domain, and OB-folds in general, as a stable scaffold for molecular recognition work.

There are several considerations that need to be addressed when moving forward with this work. First, the matter of potential applications for any produced Obody needs to inform the choice of fold. For example, it will probably be beneficial to adopt a human-sourced OB-fold as a new scaffold if a therapeutic binder is desired. However the scaffold presented here remains a viable choice for development as a research reagent, primarily due to its thermophilic origins. Second, the structures presented here provide the only experimental structural data for an engineered Obody domain. Design of future libraries can take advantage of this information to try to provide a better interface on the protein surface, by re-consideration of the selection of mutant positions in particular, but also point mutation of residues adjacent to the randomised face. For example, D32 and E95 (E93 in 13mRL L8) both formed

the acidic half of salt bridges in the structures presented here, but were not part of the original complement of mutant residues. It may be beneficial to remove these charged residues in some circumstances, to allow electrostatic interactions to be selected from among the mutant residues, instead of having them imposed as an existing pre-condition of binding.

A major goal which was only partially fulfilled by this work was to show low nanomolar-range affinity for a model protein target, which we arbitrarily set at 100 nM or below. While significant (100-fold) improvement to binding was achieved, the best affinity measured was K_D 610 nM. Future work will examine in detail the possible causes for this. One has already been alluded to in the previous chapter; the mode in which the Obodies of this lineage filled the substrate binding cleft may have been in a local binding free energy minimum too deep to realistically escape from, thus affinity maturation was only ever going to produce incremental gains. Naïve libraries informed by the three structures should provide better coverage of sequence space, and therefore produce better binders.

A second possible factor is practical aspects of the phage display process that was used. The pRpsp2 phagemid system was adopted whole from our collaborators, who designed it for display of toxic proteins (Beekwilder *et al.* 1999). The very tight control of expression and low copy number of this plasmid may negatively impact display levels of Obodies on phage particles, and therefore the efficiency of retention of phage due to binding over background retention. A smaller plasmid than pRpsps2 would also be at an advantage during transformation due to higher efficiency of uptake by the competent cells. In the same line, phage contamination was an ongoing and severe problem. Re-addressing the panning processes to improve sterile technique and minimise cross-contamination will be necessary for any future work. Ultimately, directed evolution is an exercise in process design, and the lessons learned here about the capabilities and drawbacks of the Obodies will be of great benefit in planning future investigations into their development.

6.1.6 *Summary*

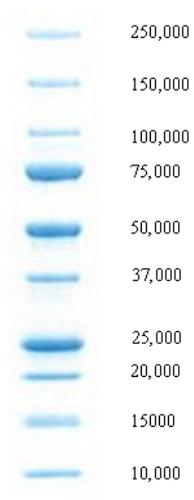
This thesis presented three crystal structures and four gene libraries investigating the OB-fold as a scaffold for use as a molecular binding reagent.

Affinity maturation of an initial binder was undertaken by a combination of rational and random mutation, which was successful, although not to the degree that was desired. Structural analysis of three Obody-HEWL complexes was instrumental in allowing the determination of the critical residues involved in binding and the structural basis for affinity maturation, and also will provide a vital resource for the design of future libraries based on this, and related folds.

A1 Appendix

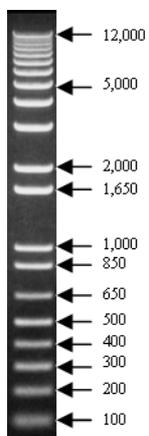
A1.1 Electrophoresis standards

A1.1.1 *Precision Plus Protein Standards*



Shown stained with coomassie, bands are labelled with molecular weight in Da.

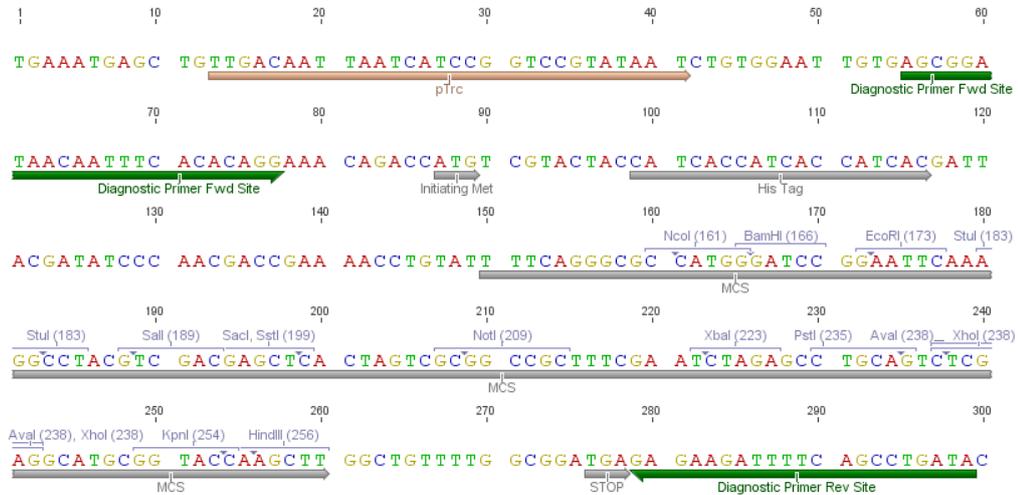
A1.1.2 *1 kb Plus DNA Standards*



Stained with ethidium bromide, bands are labelled with fragment length, in base pairs.

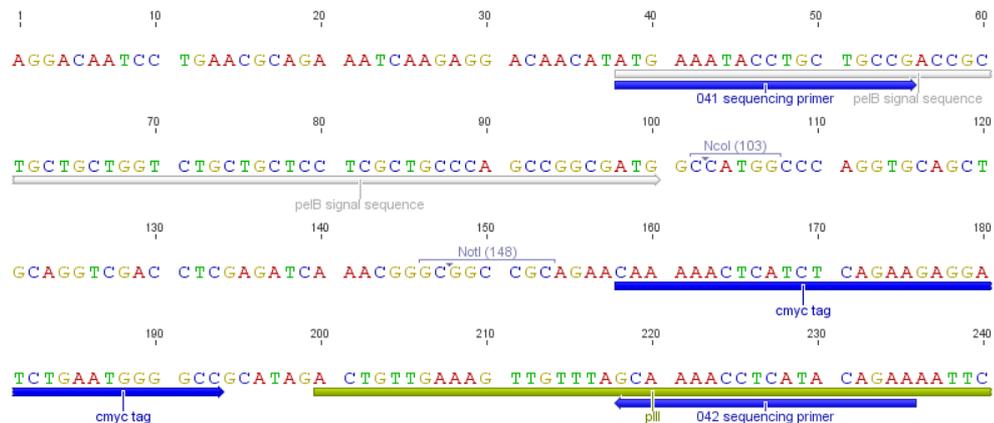
A1.2 Vectors

A1.2.1 *pProEx Htb*



The multiple cloning site (MCS) from pProEx Htb, showing the His-tag, promoter, translation initiating methionine and annealing sites for diagnostic primers.

A1.2.2 *pRpsp2*



The cloning site from pRpsp2, including signal sequence, restriction sites, cMyc tag, the beginning of the gIII ORF and annealing sites for diagnostic primers.

A1.3 Oligonucleotides

No.	Sequence (5' - 3')	Size (bp)	Description
005	CACCAGTGGATCCGTGTATCCTAAAAAGACC	31	pProEx Htb cloning (fwd)
006	ACCCGGGAATTCTCAGTCTATTGGAAGCGGCTT	33	pProEx Htb cloning (rev)
041	ATGAAATACCTGCTGCCG	18	pRsp2 sequencing (fwd)
042	TTCTGTATGAGGTTTTC	18	pRsp2 sequencing (rev)
044	GACCGACGGACTGCGGCCGCGTCTATTGGAAGCGGCTTGGCC	42	pRsp2 cloning (fwd)
051	GTTGTTGCCGGTTGGGTANNKNNKTGNNKGACNNKGGGNNKNNKAAGNNKGTGNNKGTGNNKGATAGGGAGGGGGCGCG	81	9m mutational oligo (fwd)
055	GTTGTTGCCGGTTGGGTA	18	Library construction (fwd)
056	CGCGCCCCCTCCCTATC	18	Library construction (rev)
057	GATAGGGAGGGGGCGCG	18	Library construction (fwd)
060	TTTACTGGCCTCAACAAT	18	Library construction (rev)
061	ATTGTTGAGCCAGTAAA	18	Library construction (fwd)
062	GGTGTGGAGATTTTCCC	18	Library construction (rev)
063	GGGGAAAATCTCCACACC	18	Library construction (fwd)
068	GAGGGGGCGCGTTTGTGCAAGTCACGCTCAAGG	34	Template modification (fwd)
155	GGCAGCGCGCGGTGTATCCTAAAAAGACC	30	Specific Gateway adaptor (fwd)
156	GAAAGCTGGGTGTCTCAGTCTATTGGAAGCGGCTT	33	Specific Gateway adaptor (rev)
161	ATCTGGCGTTTTTCCTGCMNNCAGMNNGACMNNCAGMNNCGCGCCCCCTCCCTATC	57	4m mutational oligo (rev)
163	ATTGTTGAGGCCAGTAAANNKNNKNNKNNKNNKNNKNNKGGTGTGGAGATTTCCC	54	4/5 loop 6m mutational oligo
164	AGGAAAAACGCCAGATCATTGTTTAAAGTG	31	Template modification (rev)
165	ATCTGGCGTTTTTCCTGCCTTGAGCGTGACTTGC	34	Template modification (fwd)
166	AGCGGATAACAATTTACACAGG	23	pProEx Htb sequencing (fwd)
167	TATCAGGCTGAAAATCTTCTC	21	pProEx Htb sequencing (rev)

168	ATCTGGCGTTTTTCCTGC	18	Library construction (rev)
169	GCAGGAAAAACGCCAGAT	18	Library construction (fwd)
183	GTTGTTGCCGTTGGGTAGCGNRKTTGGGGACTATGGGAGGGTTNNKATTGTGAAGGTGAGTGATAGGGA	71	L8 10m mutational oligo (fwd)
184	ATGATCTGGCGTTTTTCCMNNCTCGAGATAGACMNNCACAGCCGCGCCTCCCTCCCTATCACTCACCTTC	70	L8 10m mutational oligo (rev)
185	GGAAAAACGCCAGATCATTTG	21	Library construction (fwd)
186	ATGATCTGGCGTTTTTCC	18	Library construction (rev)
192	ATCAAACAT CCATGG CCAGGTG GTGTATCCTAAAAAGACCCAC	44	pRsp2 cloning (fwd)
193	GATGAGTTTTTGTCTGCGG	20	pRsp2 cloning (rev)
207	GCGAGTTTGGGGACTAT	18	Library construction (fwd)
208	ATAGTCCCCAAACTCGC	18	Library construction (rev)
209	GTGAAGGTGAGTGATAGG	18	Library construction (fwd)
210	CCTATCACTCACCTTCAC	18	Library construction (rev)
211	GCGAGTTTGGGGACTATNNKNNKNNKNNKNNKAAGNNKGTGAAGGTGAGTGATAGG	57	L10 6m mutational oligo (fwd)
212	GTTGTTGCCGTTGGGTADVVDVKTGDKVKGACDKVGGDVKDKAAGDVKGTGDKVKGATAGGGAGGGGGCGCG	81	13mLC mutational oligo (fwd)
213	GATAGGGAGGGGGCGCDVKGTGDKVKTCDVKCTGDKVKGKAGGAAAAACGCCAGAT	57	13mLC mutational oligo (rev)
-	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAAACCTGTATTTTCAGGGCAGCGGCGCA	61	Generic gateway adaptor (fwd)
-	GGGGACCACTTTGTACAAGAAAGCTGGGTG	29	Generic gateway adaptor (rev)

A1.4 Sequencing Data

A1.4.1 *OB-fold domain from aspartyl tRNA-synthetase*

The construct used as the wild-type template gene was taken from the OB-fold anticodon-binding domain from the above-mentioned gene from *Pyrobacculum aerophilum*. GenBank accession number is AE009441, gene name is PAE0703.

```
5` GTGTATCCTA  AAAAGACCCA  CTGGACCGCG  GAAATTACTC  CAAATCTCCA50  CGGGACTGAA  GTAGTTGTTG  CCGGTGGGGT
    ATGGGAGTTG  AGAGACATTG100  GGAGAGTGAA  GTTCGTGGTG200  GTGAGAGATA  GGGAGGGGTT  TGTGCAAGTC150  ACGCTCAAGG
    CGGGAAAAAC  CCCC250GATCAT  TTGTTTAAAG  TGTTTCGCAGA  ACTGAGTAGA  GAGGACGTCG300  TGGTAATTAA  AGGCATTGTT
    GAGGCCAGTA1  AAATTGCCAA  AAGTGGTGTG  GAGATTTTCC  CCAGCGAGAT  ATGGATTTTA1  AACAAGGCCA  AGCCGCTTCC
    AATAGAC

N- VYPKKTHWTA  EITPNLHGTE  VVAGVWVWEL  RDIGRVKFVV  VRDREGFVQV  TLKAGKTPDH  LFKVFAELSR  EDVVVIKIV
    EASKIAKSGV  EIFPSEIWIL  NKAKPLPID
```

A1.4.2 Unselected Variant 13mRL U81

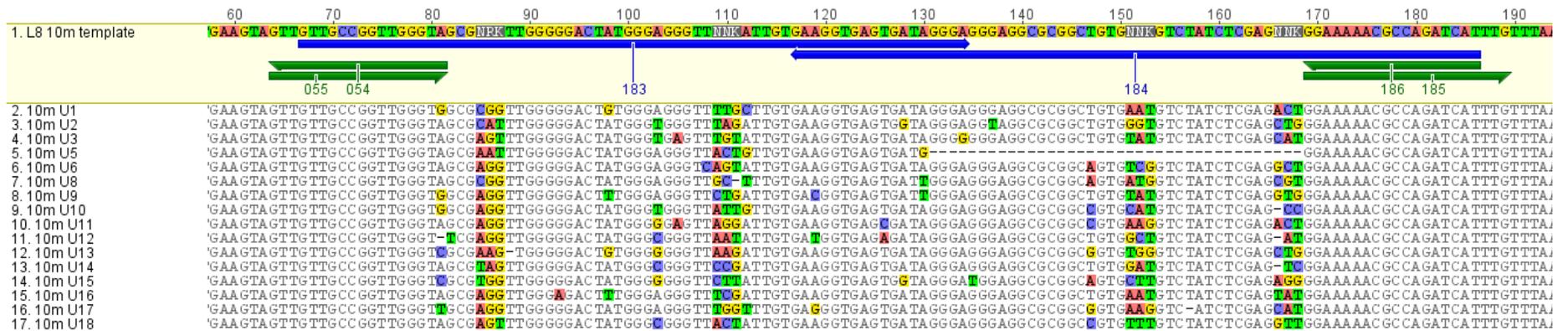
```
5` GTGTATCCTA AAAAGACCCA CTGGACCGCG GAAATTACTC CAAATCTCCA CGGGACTGAA GTAGTTGTTG CCGGTTGGGT
AAAGGGTTTG GTTGACATGG GGTTGTTGAA GGGGGTGACG GTGGGTGATA GGGAGGGGGG CGCGAGTGTG CTTGTCCGGC
TCACTGCGGG AAAAACCCCC GATCATTTGT TTAAAGTGTT CGCAGAAGTG AGTAGAGAGG ACGTCGTGGT AATTAAAGGC
ATTGTTGAGG CCAGTAAACT TGTGCCGCAG GGTGTGGAGA TTTTCCCAG CGAGATATGG ATTTTAAACA AGGCCAAGCCG
CTTCCAATAG AC

N- VYPKKTHWTA EITPNLHGTE VVAGWVKGL VDMGLLKGVV VGDREGGASV LVRLTAGKTP DHLFKVFAEL SREDVVVIKG
IVEASKLVPQ GVEIFPSEIW ILNKAKPLPI D
```

A1.4.3 L8 10m Library

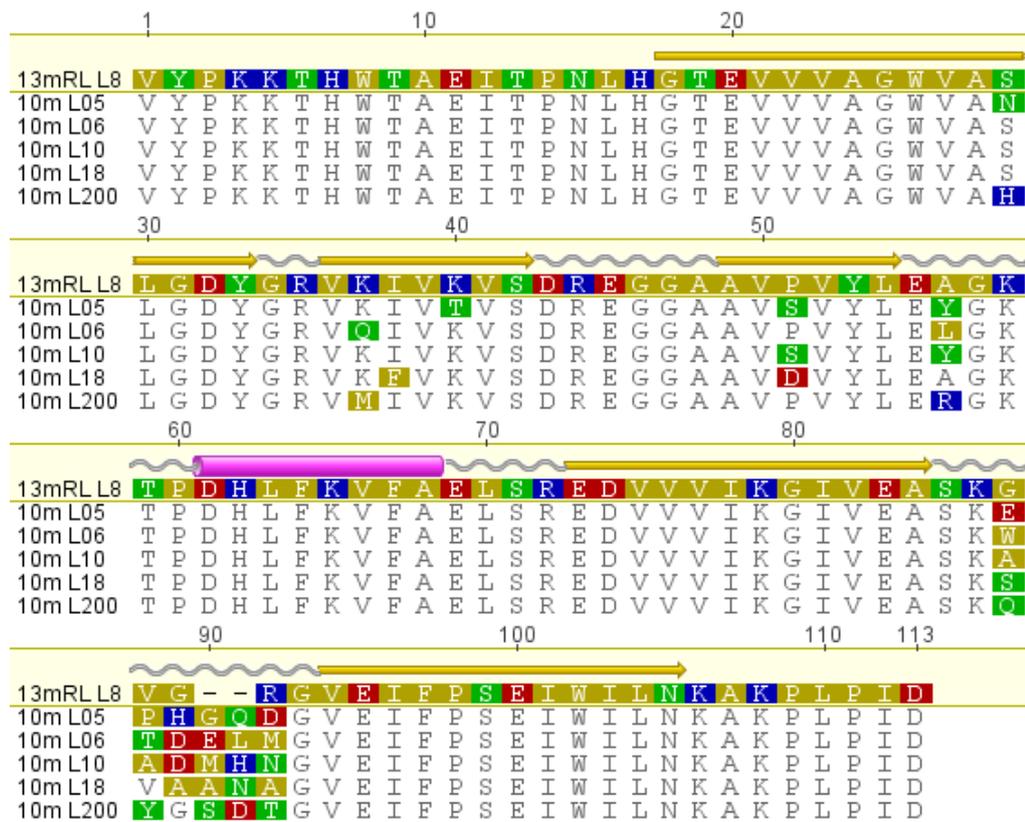
A1.4.3.1 Unselected

Shown here is the section of the L18 10m library where point mutations were introduced during incorporation of the 183/184 mutational oligonucleotides, along with deletions. Based on this data, the library was estimated to have 90% of variants as truncated Obodies.



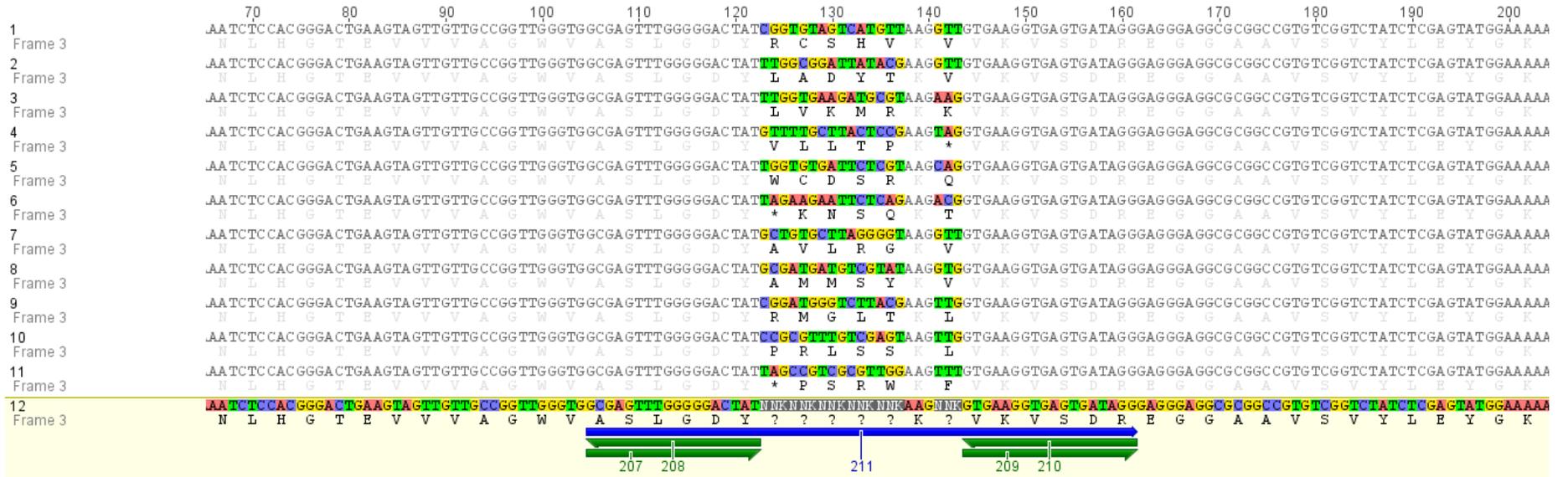
A1.4.3.2 Unique variants full sequences

The five unique variants pulled from the L8 10m library panning against HEWL, full protein sequences. The β -strands are shown as yellow arrows, helix as a pink cylinder. Residues are highlighted by disagreement with ancestral sequence 13mRL L8, coloured according to character (yellow = nonpolar, green = polar, red = acidic, blue = basic).



A1.4.4 L10 6m library

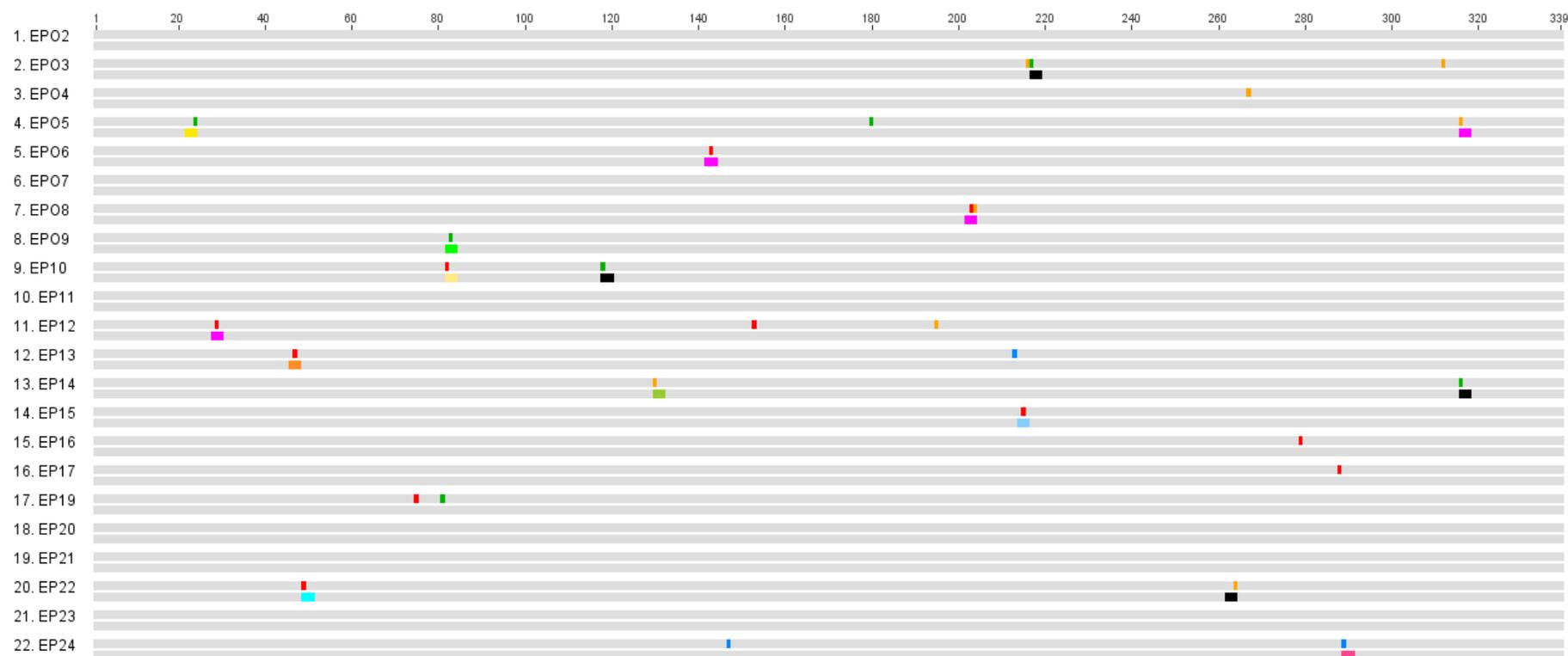
A close up on the mutated region from the L10 6m library, showing 11 unselected variants (1-11) and the template gene with primer used during assembly. Note that where the codon is TAG, the sequencing program has interpreted this as a STOP codon. However, because the TG1 *E. coli* strain has a supE genotype, this should instead be read as a glutamine.



A1.4.5 Error-prone library

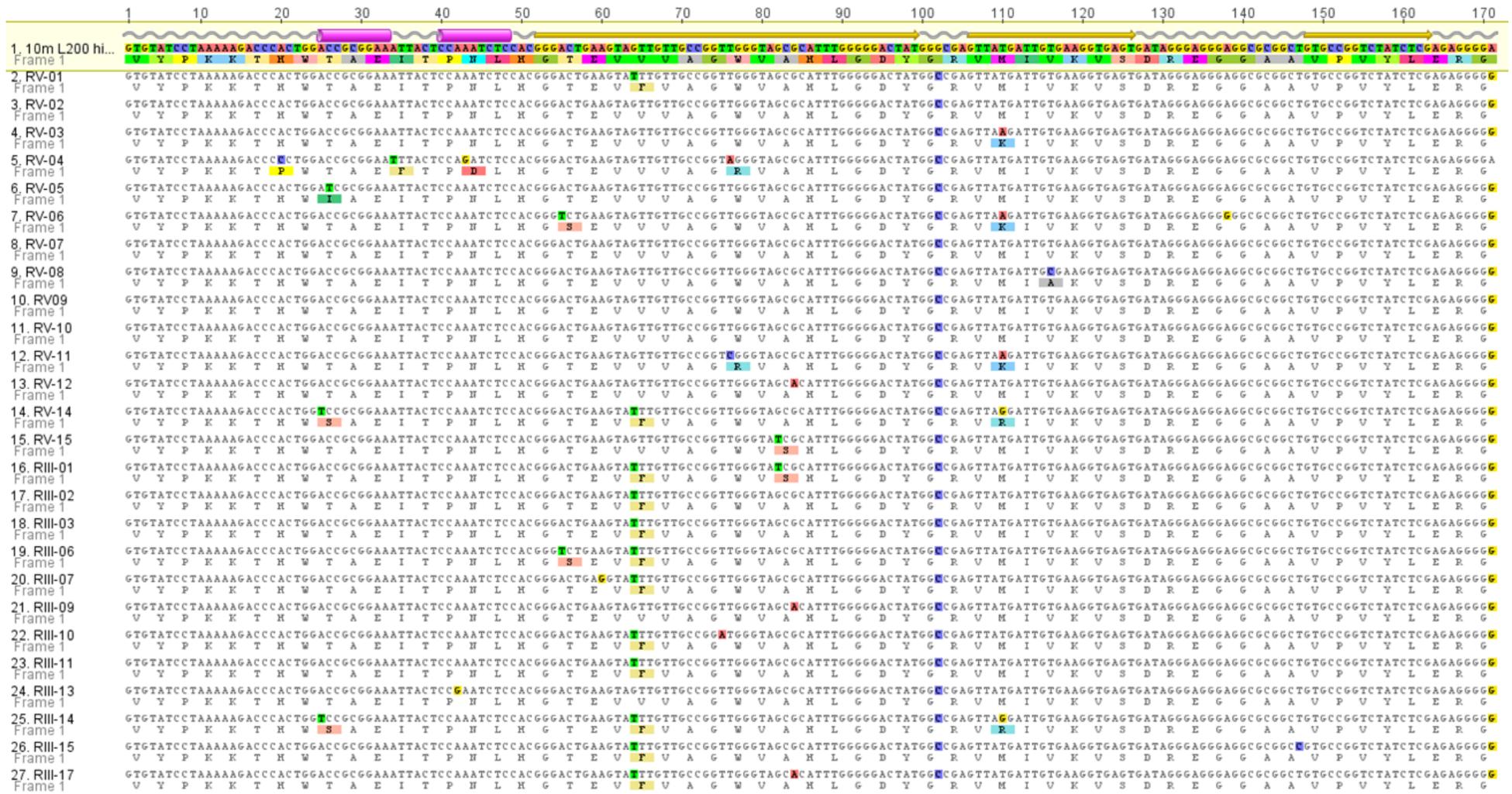
A1.4.5.1 Unselected Variants

Each sequence is represented by two lines. The first highlights nucleotide changes, the second show any resulting amino acid changes.



A1.4.5.2 Selected Sequences

Shown on the following two pages, error-prone sequences from rounds 5 (prefixed V) and 3 (prefixed III) panning against HEWL, aligned with the ancestral sequence, L8 10m L200. Each residue highlighted differs from the ancestral sequences and is coloured according to the residue name.



A1.5 rTEV Protease

Recombinant tobacco etch virus (rTEV) protease was produced from expression plasmid pMTHdelta238, using *E. coli* strain Rosetta2(DE3) (Blommel and Fox 2007). The construct, expressed as a self-cleaving MBP-fusion protein, was engineered for increased stability and expression. The expression clone was acquired from the Arizona State University Biodesign Institute DNASU Plasmid Repository. rTEV recognises a canonical peptide, ENLYFQ(G/S), and cleaves between the Q and G/S positions.

Expression was by a modified method, courtesy of James Busby, Richard Bunker and Ghader Bashiri from the Maurice Wilkins Centre, School of Biological Sciences, University of Auckland.

A1.5.1 Reagents

A1.5.1.1 Phosphate Mix

0.17 M KH_2PO_4

0.72 M K_2HPO_4

Filter to 0.2 μm to sterilise.

A1.5.1.2 Autoinduction expression media (Studier 2005)

12 g Tryptone

24 g Yeast extract

8 ml Glycerol

5 g Lactose

0.15 g glucose

Make up to 883 ml with water and autoclave. When cool, add:

100 ml phosphate mix

2 ml sterile 1M MgSO_4

15 ml 25% (w/v) Aspartic acid

2 ml 50 mg/ml kanamycin

1 ml 25 mg/ml chloramphenicol

A1.5.1.3 **Lysis Buffer**

50mM Tris/HCl pH 7.5

500 mM NaCl

1 mM tris(2-carboxyethyl)phosphine (TCEP)

20% (v/v) Ethylene Glycol

A1.5.1.4 **Binding Buffer**

50 mM Tris/HCl pH 7.5

500 mM NaCl

1 mM β -mercaptoethanol

A1.5.1.5 **Elution Buffer**

50 mM Tris/HCl pH 7.5

500 mM NaCl

500 mM Imidazole pH 7.5

1 mM β -mercaptoethanol

A1.5.1.6 **Dialysis Buffer**

20 mM Tris/HCl pH 7.5

150 mM NaCl

0.5 mM TCEP

A1.5.2 Expression Method

An LB/Agar plate supplemented with 25 μ g/mL chloramphenicol and 50 μ g/mL kanamycin was streaked from a glycerol stock of the expression clone and grown at 37°C overnight. Two colonies were picked and used to inoculate two 25 mL cultures in non-inducing media MDAP(Studier 2005) which were grown overnight with constant shaking in 100 mL conical flasks at 37°C, with the same antibiotics as above. Two 2 L baffled conical flasks with

500 mL Autoinduction expression media in each (and antibiotics as above) were inoculated with the two 25 mL overnight cultures and grown overnight with constant shaking at 37°C.

The following day the cells were harvested at 8,000 g for 15 min, the supernatant discarded and the pellet resuspended in lysis buffer. Lysis was by sonication. The soluble fraction was separated by centrifugation at 15,000 g for 20 min, filtered to 0.2 µm by syringe and loaded on to a Ni²⁺ charged HisTrap FF 5 mL column (GE Healthcare). Using an FPLC, the column was washed with five column volumes of binding buffer, then five column volumes of 85% binding buffer/15% elution buffer (75 mM imidazole). Bound protein was eluted with an elution buffer gradient, from 15-100% over 50 mL, collecting 2 mL fractions.

Fractions located in the protein peak (as determined by inline UV₂₈₀ absorbance) were pooled and dialysed overnight in 6-8,000 Da dialysis tubing into 2 L dialysis buffer. The dialysed protein was quantified by UV₂₈₀ absorbance, adjusted to 2 mg/mL with dialysis buffer then to 1 mg/mL with 100% glycerol and snap frozen in liquid nitrogen as 1 mL aliquots. Stored at -80°C.

A1.6 Crystal Screen

Condition	MPEG 5K (%)	Buffer (0.2 M)	pH
1	5	HEPES	7.0
2	7	HEPES	7.0
3	9	HEPES	7.0
4	11	HEPES	7.0
5	13	HEPES	7.0
6	15	HEPES	7.0
7	17	HEPES	7.0
8	19	HEPES	7.0
9	21	HEPES	7.0
10	23	HEPES	7.0
11	24	HEPES	7.0
12	25	HEPES	7.0
13	5	HEPES	7.4
14	7	HEPES	7.4

15	9	HEPES	7.4
16	11	HEPES	7.4
17	13	HEPES	7.4
18	15	HEPES	7.4
19	17	HEPES	7.4
20	19	HEPES	7.4
21	21	HEPES	7.4
22	23	HEPES	7.4
23	24	HEPES	7.4
24	25	HEPES	7.4
25	5	HEPES	7.8
26	7	HEPES	7.8
27	9	HEPES	7.8
28	11	HEPES	7.8
29	13	HEPES	7.8
30	15	HEPES	7.8
31	17	HEPES	7.8
32	19	HEPES	7.8
33	21	HEPES	7.8
34	23	HEPES	7.8
35	24	HEPES	7.8
36	25	HEPES	7.8
37	5	TAPS	8.0
38	7	TAPS	8.0
39	9	TAPS	8.0
40	11	TAPS	8.0
41	13	TAPS	8.0
42	15	TAPS	8.0
43	17	TAPS	8.0
44	19	TAPS	8.0
45	21	TAPS	8.0
46	23	TAPS	8.0
47	24	TAPS	8.0
48	25	TAPS	8.0
49	5	TAPS	8.4
50	7	TAPS	8.4
51	9	TAPS	8.4
52	11	TAPS	8.4
53	13	TAPS	8.4
54	15	TAPS	8.4
55	17	TAPS	8.4
56	19	TAPS	8.4
57	21	TAPS	8.4
58	23	TAPS	8.4
59	24	TAPS	8.4
60	25	TAPS	8.4
61	5	TAPS	8.6
62	7	TAPS	8.6
63	9	TAPS	8.6
64	11	TAPS	8.6

65	13	TAPS	8.6
66	15	TAPS	8.6
67	17	TAPS	8.6
68	19	TAPS	8.6
69	21	TAPS	8.6
70	23	TAPS	8.6
71	24	TAPS	8.6
72	25	TAPS	8.6

A1.7 IUPAC Nucleotide codes

Code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

A2 Appendix – digital media

The accompanying CD contains data files for the three structures, and the unpublished manuscript detailing the initial work on Obodies.

References

- Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). Phenix: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **66**: 213.
- Agterberg, M., H. Adriaanse and J. Tommassen (1987). Use of outer-membrane protein phoe as a carrier for the transport of a foreign antigenic determinant to the cell-surface of *escherichia coli* k-12 *Gene* **59**(1): 145.
- Akamatsu, Y., M. S. Cole, J. Y. Tso and N. Tsurushita (1993). Construction of a human ig combinatorial library from genomic-v segments and synthetic cdr3 fragments. *Journal of Immunology* **151**(9): 4651.
- Allazikani, B., A. M. Lesk and C. Chothia (1997). Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology* **273**(4): 927.
- Almagro, J. C., V. Quintero-Hernandez, M. Ortiz-Leon, A. Velandia, S. L. Smith and B. Becerril (2006). Design and validation of a synthetic v-h repertoire with tailored diversity for protein recognition. *Journal of Molecular Recognition* **19**(5): 413.
- Amstutz, P., H. K. Binz, P. Parizek, M. T. Stumpp, A. Kohl, M. G. Grutter, P. Forrer and C. Pluckthun (2005). Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *Journal of Biological Chemistry* **280**(26): 24715.
- Andreeva, A., D. Howorth, S. E. Brenner, T. J. P. Hubbard, C. Chothia and A. G. Murzin (2004). Scop database in 2004: Refinements integrate structure and sequence family data. *Nucleic Acids Research* **32**: D226.
- Andreeva, A., D. Howorth, J. M. Chandonia, S. E. Brenner, T. J. P. Hubbard, C. Chothia and A. G. Murzin (2008). Data growth and its impact on the scop database: New developments. *Nucleic Acids Research* **36**: D419.
- Arcus, V. (2002). Ob-fold domains: A snapshot of the evolution of sequence, structure and function. *Current Opinion in Structural Biology* **12**(6): 794.
- Arcus, V., R. Langley, T. Proft, J. D. Fraser and E. N. Baker (2002). The three-dimensional structure of a superantigen-like protein, set3, from a pathogenicity island of the staphylococcus aureus genome. *Journal of Biological Chemistry* **277**(35): 32274.
- Arcus, V., T. Proft, J. A. Sigrell, H. M. Baker, J. D. Fraser and E. N. Baker (2000). Conservation and variation in superantigen structure and activity highlighted by the three-dimensional structures of two new superantigens from streptococcus pyogenes. *Journal of Molecular Biology* **299**(1): 157.
- Armstrong, J., R. N. Perham and J. E. Walker (1981). Domain structure of bacteriophage fd adsorption protein. *FEBS Lett* **135**(1): 167.
- Arnold, F. H. and A. A. Volkov (1999). Directed evolution of biocatalysts. *Current Opinion in Chemical Biology* **3**(1): 54.

- Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). The swiss-model workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**(2): 195.
- Bailey, S. (1994). The ccp4 suite - programs for protein crystallography. *Acta Crystallographica Section D-Biological Crystallography* **50**: 760.
- Ban, Y. E. A., H. Edelsbrunner and J. Rudolph (2006). Interface surfaces for protein-protein complexes. *Journal of the Association for Computing Machinery* **53**(3): 361.
- Bandeiras, T. M., R. C. Hillig, P. M. Matias, U. Eberspaecher, J. Fanghanel, M. Thomaz, S. Miranda, K. Crusius, V. Puetter, P. Amstutz, M. Gulotti-Georgieva, H. K. Binz, C. Holz, A. A. P. Schmitz, C. Lang, P. Donner, U. Egner, M. A. Carrondo and B. Muller-Tiemann (2008). Structure of wild-type plk-1 kinase domain in complex with a selective darpin. *Acta Crystallographica Section D-Biological Crystallography* **64**: 339.
- Barbas III, C. F., D. R. Burton, J. K. Scott and G. J. Silverman (2001). Phage display; a laboratory manual, **Cold Spring Harbour Laboratory Press**.
- Baron, M., D. G. Norman and I. D. Campbell (1991). Protein modules. *Trends in Biochemical Sciences* **16**(1): 13.
- Batori, V., A. Koide and S. Koide (2002). Exploring the potential of the monobody scaffold: Effects of loop elongation on the stability of a fibronectin type iii domain. *Protein Engineering* **15**(12): 1015.
- Beekwilder, J., J. Rakonjac, M. Jongsma and D. Bosch (1999). A phagemid vector using the e-coli phage shock promoter facilitates phage display of toxic proteins. *Gene* **228**(1-2): 23.
- Bertschinger, J. and D. Neri (2004). Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Protein Engineering Design & Selection* **17**(9): 699.
- Beste, G., F. S. Schmidt, T. Stibora and A. Skerra (1999). Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proceedings of the National Academy of Sciences of the United States of America* **96**(5): 1898.
- Better, M., C. P. Chang, R. R. Robinson and A. H. Horwitz (1988). *Escherichia coli* secretion of an active chimeric antibody fragment *Science* **240**(4855): 1041.
- Binz, H. K., P. Amstutz, A. Kohl, M. T. Stumpp, C. Briand, P. Forrer, M. G. Grutter and A. Pluckthun (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnology* **22**(5): 575.
- Binz, H. K., P. Amstutz and A. Pluckthun (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nature Biotechnology* **23**(10): 1257.
- Binz, H. K. and A. Pluckthun (2005). Engineered proteins as specific binding reagents. *Current Opinion in Biotechnology* **16**(4): 459.
- Binz, H. K., M. T. Stumpp, P. Forrer, P. Amstutz and A. Pluckthun (2003). Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of Molecular Biology* **332**(2): 489.
- Bird, R. E. and B. W. Walker (1991). Single chain antibody variable regions. *Trends in Biotechnology* **9**(4): 132.
- Blommel, P. G. and B. G. Fox (2007). A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expression and Purification* **55**: 53.

- Bloom, L. and V. Calabro (2009). Fn3: A new protein scaffold reaches the clinic. *Drug Discovery Today* **14**(19-20): 949.
- Boder, E. T. and K. D. Wittrup (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* **15**(6): 553.
- Bork, P., L. Holm and C. Sander (1994). The immunoglobulin fold - structural classification, sequence patterns and common core. *Journal of Molecular Biology* **242**(4): 309.
- Borrebaeck, C. A. K., L. Danielsson and S. A. Moller (1988). Human monoclonal-antibodies produced by primary *in vitro* immunization of peripheral blood lymphocytes
Proceedings of the National Academy of Sciences of the United States of America **85**(11): 3995.
- Braisted, A. C. and J. A. Wells (1996). Minimizing a binding domain from protein a. *Proceedings of the National Academy of Sciences of the United States of America* **93**(12): 5688.
- Buchner, J. and R. Rudolph (1991). Renaturation, purification and characterization of recombinant fab fragments produced in *escherichia coli*. *Bio-Technology* **9**(2): 157.
- Burritt, J. B., C. W. Bond, K. W. Doss and A. J. Jesaitis (1996). Filamentous phage display of oligopeptide libraries. *Analytical Biochemistry* **238**(1): 1.
- Cadwell, R. C. and G. F. Joyce (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl* **2**(1): 28.
- Castro, M. J. M. and S. Anderson (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: Effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin. *Biochemistry* **35**(35): 11435.
- Chakrabarti, P. and J. Janin (2002). Dissecting protein-protein recognition sites. *Proteins-Structure Function and Genetics* **47**(3): 334.
- Chaput, J. C. and J. W. Szostak (2004). Evolutionary optimization of a nonbiological ATP binding protein for improved folding stability. *Chemistry & Biology* **11**(6): 865.
- Charbit, A., J. C. Boulain, A. Ryter and M. Hofnung (1986). Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope; expression at the cell surface. *Embo Journal* **5**(11): 3029.
- Charbit, A., A. Molla, W. Saurin and M. Hofnung (1988). Versatility of a vector for expressing foreign polypeptides at the surface of gram-negative bacteria. *Gene* **70**(1): 181.
- Chou, P. Y. and G. D. Fasman (1974). Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* **13**(2): 211.
- Christ, D., K. Famm and G. Winter (2007). Repertoires of aggregation-resistant human antibody domains. *Protein Engineering Design & Selection* **20**(8): 413.
- Clackson, T., H. R. Hoogenboom, A. D. Griffiths and G. Winter (1991). Making antibody fragments using phage display libraries. *Nature* **352**(6336): 624.
- Clackson, T., M. H. Ultsch, J. A. Wells and A. M. de Vos (1998). Structural and functional analysis of the 1 : 1 growth hormone : Receptor complex reveals the molecular basis for receptor affinity. *Journal of Molecular Biology* **277**(5): 1111.

- Coco, W. M., W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires and D. J. Monticello (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology* **19**(4): 354.
- Commans, S., P. Plateau, S. Blanquet and F. Dardel (1995). Solution structure of the anticodon-binding domain of *escherichia coli* lysyl-transfer-rna synthetase and studies of its interaction with trna(lys). *Journal of Molecular Biology* **253**(1): 100.
- Cortese, R., F. Felici, G. Galfre, A. Luzzago, P. Monaci and A. Nicosia (1994). Epitope discovery using peptide libraries displayed on phage. *Trends in Biotechnology* **12**(7): 262.
- Cowan, S. W., M. E. Newcomer and T. A. Jones (1990). Crystallographic refinement of human serum retinol binding-protein at 2 Å resolution. *Proteins-Structure Function and Genetics* **8**(1): 44.
- Cwirla, S. E., E. A. Peters, R. W. Barrett and W. J. Dower (1990). Peptides on phage - a vast library of peptides for identifying ligands. *Proceedings of the National Academy of Sciences of the United States of America* **87**(16): 6378.
- Dai, M., J. Temirov, E. Pesavento, C. Kiss, N. Velappan, P. Pavlik, J. H. Werner and A. R. M. Bradbury (2008). Using t7 phage display to select gfp-based binders. *Protein Engineering Design & Selection* **21**(7): 413.
- Darnell, S. J., D. Page and J. C. Mitchell (2007). An automated decision-tree approach to predicting protein interaction hot spots. *Proteins-Structure Function and Bioinformatics* **68**(4): 813.
- Dennis, M. S. and R. A. Lazarus (1994a). Kunitz domain inhibitors of tissue factor-factor viia .1. Potent inhibitors selected from libraries by phage display. *Journal of Biological Chemistry* **269**(35): 22129.
- Dennis, M. S. and R. A. Lazarus (1994b). Kunitz domain inhibitors of tissue factor-factor viia .2. Potent and specific inhibitors by competitive phage selection. *Journal of Biological Chemistry* **269**(35): 22137.
- Derbyshire, K. M., J. J. Salvo and N. D. Grindley (1986). A simple and efficient procedure for saturation mutagenesis using mixed oligodeoxynucleotides. *Gene* **46**(2-3): 145.
- Doi, N. and H. Yanagawa (1999). Stable: Protein-DNA fusion system for screening of combinatorial protein libraries in vitro. *Febs Letters* **457**(2): 227.
- Dong, F. and H. X. Zhou (2006). Electrostatic contribution to the binding stability of protein-protein complexes. *Proteins-Structure Function and Bioinformatics* **65**(1): 87.
- Duan, J., J. Wu, C. A. Valencia and R. Liu (2007). Fibronectin type iii domain based monobody with high avidity. *Biochemistry* **46**(44): 12656.
- Eigenbrot, C., M. Ultsch, A. Dubnovitsky, L. Abrahmsen and T. Hard (2010). Structural basis for high-affinity her2 receptor binding by an engineered protein. *Proceedings of the National Academy of Sciences of the United States of America* **107**(34): 15039.
- Eiler, S., A. C. Dock-Bregeon, L. Moulinier, J. C. Thierry and D. Moras (1999). Synthesis of aspartyl-trna(asp) in *escherichia coli* - a snapshot of the second step. *Embo Journal* **18**(22): 6532.
- Eklund, M., L. Axelsson, M. Uhlen and P. A. Nygren (2002). Anti-idiotypic protein domains selected from protein a-based affibody libraries. *Proteins-Structure Function and Genetics* **48**(3): 454.

- Eklund, M., K. Sandstrom, T. T. Teeri and P. A. Nygren (2004). Site-specific and reversible anchoring of active proteins on to cellulose using a cellulosome-like complex. *Journal of Biotechnology* **109**(3): 277.
- Emsley, P. and K. Cowtan (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **60**: 2126.
- Engfeldt, T., A. Orlova, T. Tran, A. Bruskin, C. Widstrom, A. E. Karlstrom and V. Tolmachev (2007). Imaging of her2-expressing tumours using a synthetic affibody molecule containing the tc-99m-chelating mercaptoacetyl-glycyl-glycyl-glycyl (mag3) sequence. *European Journal of Nuclear Medicine and Molecular Imaging* **34**(5): 722.
- Ericsson, U. B., B. M. Hallberg, G. T. DeTitta, N. Dekker and P. Nordlund (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry* **357**(2): 289.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography* **62**: 72.
- Feldhaus, M. J., R. W. Siegel, L. K. Opresko, J. R. Coleman, J. M. W. Feldhaus, Y. A. Yeung, J. R. Cochran, P. Heinzelman, D. Colby, J. Swers, C. Graff, H. S. Wiley and K. D. Wittrup (2003). Flow-cytometric isolation of human antibodies from a nonimmune *saccharomyces cerevisiae* surface display library. *Nature Biotechnology* **21**(2): 163.
- Fellouse, F. A., B. Li, D. M. Compaan, A. A. Peden, S. G. Hymowitz and S. S. Sidhu (2005). Molecular recognition by a binary code. *Journal of Molecular Biology* **348**(5): 1153.
- Fellouse, F. A., C. Wiesmann and S. S. Sidhu (2004). Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition. *Proceedings of the National Academy of Sciences of the United States of America* **101**(34): 12467.
- Figini, M., J. D. Marks, G. Winter and A. D. Griffiths (1994). *In vitro* assembly of repertoires of antibody chains on the surface of phage by renaturation. *Journal of Molecular Biology* **239**(1): 68.
- Flower, D. R. (1996). The lipocalin protein family: Structure and function. *Biochemical Journal* **318**: 1.
- Foote, J. and G. Winter (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *Journal of Molecular Biology* **224**(2): 487.
- Forrer, P., M. T. Stumpp, H. K. Binz and A. Pluckthun (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *Febs Letters* **539**(1-3): 2.
- Francisco, J. A., R. Campbell, B. L. Iverson and G. Georgiou (1993a). Production and fluorescence-activated cell sorting of *escherichia coli* expressing a functional antibody fragment on the external surface. *Proceedings of the National Academy of Sciences of the United States of America* **90**(22): 10444.
- Francisco, J. A., C. Stathopoulos, R. A. J. Warren, D. G. Kilburn and G. Georgiou (1993b). Specific adhesion and hydrolysis of cellulose by intact *escherichia coli* expressing surface anchored cellulase or cellulose binding domains. *Bio-Technology* **11**(4): 491.
- Fraser, M. E., M. M. Chernaia, Y. V. Kozlov and M. N. G. James (1994). Crystal structure of the holotoxin from *shigella dysenteriae* at 2.5-angstrom resolution. *Nature Structural Biology* **1**(1): 59.

- Fraser, M. E., M. Fujinaga, M. M. Cherney, A. R. Melton-Celsa, E. M. Twiddy, A. D. O'Brien and M. N. G. James (2004). Structure of shiga toxin type 2 (stx2) from *escherichia coli* o157 : H7. *Journal of Biological Chemistry* **279**(26): 27511.
- French, D. L., R. Laskov and M. D. Scharff (1989). The role of somatic hypermutation in the generation of antibody diversity. *Science* **244**(4909): 1152.
- French, S. and K. S. Wilson (1978). On the treatment of negative intensity observations. *Acta Crystallographica* **A34**(4): 517.
- Freudl, R. (1989). Insertion of peptides into cell-surface-exposed areas of the *escherichia coli* ompa protein does not interfere with export and membrane assembly. *Gene* **82**(2): 229.
- Friedman, M. and S. Stahl (2009). Engineered affinity proteins for tumour-targeting applications. *Biotechnology and Applied Biochemistry* **53**(1): 1.
- Friesen, W. J. and M. K. Darby (1998). Specific rna binding proteins constructed from zinc fingers. *Nature Structural Biology* **5**(7): 543.
- Friesen, W. J. and M. K. Darby (2001). Specific rna binding by a single c2h2 zinc finger. *Journal of Biological Chemistry* **276**(3): 1968.
- Fuchs, P., F. Breitling, S. Dubel, T. Seehaus and M. Little (1991). Targeting recombinant antibodies to the surface of *escherichia coli* - fusion to a peptidoglycan associated lipoprotein. *Bio-Technology* **9**(12): 1369.
- Gallicchio, E., M. Lapelosa and R. M. Levy (2010). Binding energy distribution analysis method (bedam) for estimation of protein ligand binding affinities. *Journal of Chemical Theory and Computation* **6**(9): 2961.
- Garman, E. F. and T. R. Schneider (1997). Macromolecular cryocrystallography. *Journal of Applied Crystallography* **30**(3): 211.
- Gebauer, M. and A. Skerra (2009). Engineered protein scaffolds as next-generation antibody therapeutics. *Current Opinion in Chemical Biology* **13**(3): 245.
- Georgiou, G., C. Stathopoulos, P. S. Daugherty, A. R. Nayak, B. L. Iverson and R. Curtiss (1997). Display of heterologous proteins on the surface of microorganisms: From the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnology* **15**(1): 29.
- Ghadessy, F. J., J. L. Ong and P. Holliger (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* **98**(8): 4552.
- Ghahroudi, M. A., A. Desmyter, L. Wyns, R. Hamers and S. Muyldermans (1997). Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *Febs Letters* **414**(3): 521.
- Gibbs, M. D., K. M. H. Nevalainen and P. L. Bergquist (2001). Degenerate oligonucleotide gene shuffling (dogs): A method for enhancing the frequency of recombination with family shuffling. *Gene* **271**(1): 13.
- Glockshuber, R., M. Malia, I. Pfitzinger and A. Pluckthun (1990). A comparison of strategies to stabilize immunoglobulin fv-fragments. *Biochemistry* **29**(6): 1362.
- Goletz, S., P. A. Christensen, P. Kristensen, D. Blohm, I. Tomlinson, G. Winter and U. Karsten (2002). Selection of large diversities of

- antiidiotypic antibody fragments by phage display. *Journal of Molecular Biology* **315**(5): 1087.
- Gram, H., L. A. Marconi, C. F. Barbas, T. A. Collet, R. A. Lerner and A. S. Kang (1992). *In vitro* selection and affinity maturation of antibodies from a naïve combinatorial immunoglobulin library. *Proceedings of the National Academy of Sciences of the United States of America* **89**(8): 3576.
- Grant, R. A., T. C. Lin, W. Konigsberg and R. E. Webster (1981). Structure of the filamentous bacteriophage fl. Location of the a, c, and d minor coat proteins. *J Biol Chem* **256**(1): 539.
- Griffiths, A. D., S. C. Williams, O. Hartley, I. M. Tomlinson, P. Waterhouse, W. L. Crosby, R. E. Kontermann, P. T. Jones, N. M. Low, T. J. Allison, T. D. Prospero, H. R. Hoogenboom, A. Nissim, J. P. L. Cox, J. L. Harrison, M. Zaccolo, E. Gherardi and G. Winter (1994). Isolation of high-affinity human-antibodies directly from large synthetic repertoires. *Embo Journal* **13**(14): 3245.
- Griffiths, G. M., C. Berek, M. Kaartinen and C. Milstein (1984). Somatic mutation and the maturation of immune response to 2-phenyl oxazolone. *Nature* **312**(5991): 271.
- Grubisha, O., M. Kaminska, S. Duquerroy, E. Fontan, F. Cordier, A. Haouz, B. Raynal, J. Chiaravalli, M. Delepierre, A. Israel, M. Veron and F. Agou (2010). Darpin-assisted crystallography of the cc2-lz domain of nemo reveals a coupling between dimerization and ubiquitin binding. *Journal of Molecular Biology* **395**(1): 89.
- Gunnarsson, L. C., L. Dexlin, E. N. Karlsson, O. Holst and M. Ohlin (2006a). Evolution of a carbohydrate binding module into a protein-specific binder. *Biomolecular Engineering* **23**(2-3): 111.
- Gunnarsson, L. C., E. N. Karlsson, A. S. Albrekt, M. Andersson, O. Holst and M. Ohlin (2004). A carbohydrate binding module as a diversity-carrying scaffold. *Protein Engineering Design & Selection* **17**(3): 213.
- Gunnarsson, L. C., Q. Zhou, C. Montanier, E. N. Karlsson, H. Brumer and M. Ohlin (2006b). Engineered xyloglucan specificity in a carbohydrate-binding module. *Glycobiology* **16**(12): 1171.
- Gunneriusson, E., K. Nord, M. Uhlen and P. A. Nygren (1999). Affinity maturation of a taq DNA polymerase specific affibody by helix shuffling. *Protein Engineering* **12**(10): 873.
- Gunneriusson, E., P. Samuelson, M. Uhlen, P. A. Nygren and S. Stahl (1996). Surface display of a functional single-chain fv antibody on staphylococci. *Journal of Bacteriology* **178**(5): 1341.
- Hackel, B. J., A. Kapila and K. D. Wittrup (2008). Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *Journal of Molecular Biology* **381**(5): 1238.
- Hackel, B. J. and K. D. Wittrup (2010). The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin g binders from the fibronectin scaffold. *Protein Engineering Design & Selection* **23**(4): 211.
- Hamerscasterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman and R. Hamers (1993). Naturally occurring antibodies devoid of light-chains. *Nature* **363**(6428): 446.
- Hanes, J. and A. Pluckthun (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the*

National Academy of Sciences of the United States of America **94**(10): 4937.

- Hartley, J. L., G. F. Temple and M. A. Brasch (2000). DNA cloning using in vitro site-specific recombination. *Genome Research* **10**(11): 1788.
- Hawkins, R. E., S. J. Russell and G. Winter (1992). Selection of phage antibodies by binding-affinity - mimicking affinity maturation. *Journal of Molecular Biology* **226**(3): 889.
- Hendsch, Z. S. and B. Tidor (1994). Do salt bridges stabilize proteins - a continuum electrostatic analysis. *Protein Science* **3**(2): 211.
- Heyd, B., F. Pecorari, B. Collinet, E. Adjadj, M. Desmadril and P. Minard (2003). In vitro evolution of the binding specificity of neocarzinostatin, an enediyne-binding chromoprotein. *Biochemistry* **42**(19): 5674.
- Hibbert, E. G., F. Baganz, H. C. Hailes, J. M. Ward, G. J. Lye, J. M. Woodley and P. A. Dalby (2005). Directed evolution of biocatalytic processes. *Biomolecular Engineering* **22**(1-3): 11.
- Hogbom, M., M. Eklund, P. A. Nygren and P. Nordlund (2003). Structural basis for recognition by an in vitro evolved affibody. *Proceedings of the National Academy of Sciences of the United States of America* **100**(6): 3191.
- Holler, P. D., A. R. Lim, B. K. Cho, L. A. Rund and D. M. Kranz (2001). Cd8(-) t cell transfectants that express a high affinity t cell receptor exhibit enhanced peptide-dependent activation. *Journal of Experimental Medicine* **194**(8): 1043.
- Hoogenboom, H. R. and G. Winter (1992). Bypassing immunization - human-antibodies from synthetic repertoires of germline vh gene segments rearranged *in vitro*. *Journal of Molecular Biology* **227**(2): 381.
- Hudson, P. J. and C. Souriau (2003). Engineered antibodies. *Nature Medicine* **9**(1): 129.
- Hufton, S. E., N. van Neer, T. van den Beuken, J. Desmet, E. Sablon and H. R. Hoogenboom (2000). Development and application of cytotoxic t lymphocyte-associated antigen 4 as a protein scaffold for the generation of novel binding ligands. *Febs Letters* **475**(3): 225.
- Huse, W. D., L. Sastry, S. A. Iverson, A. S. Kang, M. Altingmees, D. R. Burton, S. J. Benkovic and R. A. Lerner (1989). Generation of a large combinatorial library of the immunoglobulin repertoire in phage-lambda. *Science* **246**(4935): 1275.
- Huston, J. S., D. Levinson, M. Mudgetthunter, M. S. Tai, J. Novotny, M. N. Margolies, R. J. Ridge, R. E. Brucoleri, E. Haber, R. Crea and H. Oppermann (1988). Protein engineering of antibody-binding sites - recovery of specific activity in an anti-digoxin single-chain fv analog produced in *escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **85**(16): 5879.
- Hynes, T. R. and R. O. Fox (1991). The crystal-structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins-Structure Function and Genetics* **10**(2): 92.
- Inokuchi, H., F. Yamao, H. Sakano and H. Ozeki (1979). Identification of transfer rna suppressors in *escherichia coli*. I. Amber suppressor su+2, an anticodon mutant of trna2gln. *J Mol Biol* **132**(4): 649.
- Iyer, S., S. Wei, K. Brew and K. R. Acharya (2007). Crystal structure of the catalytic domain of matrix metalloproteinase-1 in complex with the inhibitory domain of tissue inhibitor of metalloproteinase-1. *Journal of Biological Chemistry* **282**(1): 364.

- Jacobs, G. H. (1992). Determination of the base recognition positions of zinc fingers from sequence-analysis. *Embo Journal* **11**(12): 4507.
- Jones, P. T., P. H. Dear, J. Foote, M. S. Neuberger and G. Winter (1986). Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* **321**(6069): 522.
- Jones, S. and J. M. Thornton (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **93**(1): 13.
- Jones, S. and J. M. Thornton (1997). Prediction of protein-protein interaction sites using patch analysis. *Journal of Molecular Biology* **272**(1): 133.
- Jostock, T. and S. Dubel (2005). Screening of molecular repertoires by microbial surface display. *Combinatorial Chemistry & High Throughput Screening* **8**(2): 127.
- Kabsch, W. (2010). Xds. *Acta Crystallographica Section D-Biological Crystallography* **66**: 125.
- Keefe, A. D. and J. W. Szostak (2001). Functional proteins from a random-sequence library. *Nature* **410**(6829): 715.
- Kettleborough, C. A., J. Saldanha, V. J. Heath, C. J. Morrison and M. M. Bendig (1991). Humanization of a mouse monoclonal-antibody by cdr-grafting - the importance of framework residues on loop conformation. *Protein Engineering* **4**(7): 773.
- Kiefer, F., K. Arnold, M. Kunzli, L. Bordoli and T. Schwede (2009). The swiss-model repository and associated resources. *Nucleic Acids Research* **37**: D387.
- Kim, D. E., D. Chivian and D. Baker (2004). Protein structure prediction and analysis using the rosetta server. *Nucleic Acids Research* **32**(Suppl. 2): W526.
- Knappik, A., L. M. Ge, A. Honegger, P. Pack, M. Fischer, G. Wellenhofer, A. Hoess, J. Wolle, A. Pluckthun and B. Virnekas (2000). Fully synthetic human combinatorial antibody libraries (hucal) based on modular consensus frameworks and cdrs randomized with trinucleotides. *Journal of Molecular Biology* **296**(1): 57.
- Kohl, A., P. Amstutz, P. Parizek, H. K. Binz, C. Briand, G. Capitani, P. Forrer, A. Pluckthun and M. G. Grutter (2005). Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure* **13**(8): 1131.
- Kohl, A., H. K. Binz, P. Forrer, M. T. Stumpp, A. Pluckthun and M. G. Grutter (2003). Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proceedings of the National Academy of Sciences of the United States of America* **100**(4): 1700.
- Kohler, G. and C. Milstein (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**(5517): 495.
- Koide, A., S. Abbatiello, L. Rothgery and S. Koide (2002). Probing protein conformational changes in living cells by using designer binding proteins: Application to the estrogen receptor. *Proceedings of the National Academy of Sciences of the United States of America* **99**(3): 1253.
- Koide, A., C. W. Bailey, X. L. Huang and S. Koide (1998). The fibronectin type iii domain as a scaffold for novel binding proteins. *Journal of Molecular Biology* **284**(4): 1141.
- Koide, A., R. N. Gilbreth, K. Esaki, V. Tereshko and S. Koide (2007). High-affinity single-domain binding proteins with a binary-code interface.

- Proceedings of the National Academy of Sciences of the United States of America* **104**(16): 6632.
- Koide, A., M. R. Jordan, S. R. Horner, V. Batori and S. Koide (2001). Stabilization of a fibronectin type iii domain by the removal of unfavorable electrostatic interactions on the protein surface. *Biochemistry* **40**(34): 10326.
- Korn, A. P. and R. M. Burnett (1991). Distribution and complementarity of hydrophathy in multisubunit proteins. *Proteins-Structure Function and Genetics* **9**(1): 37.
- Korndorfer, I. P., G. Beste and A. Skerra (2003). Crystallographic analysis of an "Anticalin" With tailored specificity for fluorescein reveals high structural plasticity of the lipocalin loop region. *Proteins-Structure Function and Genetics* **53**(1): 121.
- Krehenbrink, M., M. Chami, I. Guilvout, P. M. Alzari, F. Pecorari and A. P. Pugsley (2008). Artificial binding proteins (affitins) as probes for conformational changes in secretin puld. *Journal of Molecular Biology* **383**(5): 1058.
- Krissinel, E. and K. Henrick (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**(3): 774.
- Kristensen, P. and G. Winter (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Folding & Design* **3**(5): 321.
- Ku, J. and P. G. Schultz (1995). Alternate protein frameworks for molecular recognition. *Proceedings of the National Academy of Sciences of the United States of America* **92**(14): 6552.
- Kuus-Reichel, K., L. S. Grauer, L. M. Karavodin, C. Knott, M. Krusemeier and N. E. Kay (1994). Will immunogenicity limit the use, efficacy, and future development of therapeutic monoclonal antibodies? *Clin Diagn Lab Immunol* **1**(4): 365.
- Leslie, A. G. W. (1992). Recent changes to the mosflm package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**.
- Li, X., O. Keskin, B. Y. Ma, R. Nussinov and J. Liang (2004). Protein-protein interactions: Hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: Implications for docking. *Journal of Molecular Biology* **344**(3): 781.
- Li, Z. H. and J. Y. Li (2010). Geometrically centered region: A "Wet" Model of protein binding hot spots not excluding water molecules. *Proteins-Structure Function and Bioinformatics* **78**(16): 3304.
- Liao, H. I., C. A. Olson, S. M. Hwang, H. Y. Deng, E. Wong, R. S. Baric, R. W. Roberts and R. Sun (2009). Mrna display design of fibronectin-based intrabodies that detect and inhibit severe acute respiratory syndrome coronavirus nucleocapsid protein. *Journal of Biological Chemistry* **284**(26): 17512.
- Lichtarge, O., H. R. Bourne and F. E. Cohen (1996). An evolutionary trace method defines binding surfaces common to protein families. *Journal of Molecular Biology* **257**(2): 342.
- Lichtarge, O. and M. E. Sowa (2002). Evolutionary predictions of binding surfaces and interactions. *Current Opinion in Structural Biology* **12**(1): 21.
- Lin, S. L., C. J. Tsai and R. Nussinov (1995). A study of 4-helix bundles - investigating protein-folding via similar architectural motifs in protein

- cores and in subunit interfaces. *Journal of Molecular Biology* **248**(1): 151.
- Lo Conte, L., B. Ailey, T. J. P. Hubbard, S. E. Brenner, A. G. Murzin and C. Chothia (2000). Scop: A structural classification of proteins database. *Nucleic Acids Research* **28**(1): 257.
- Lo Conte, L., S. E. Brenner, T. J. P. Hubbard, C. Chothia and A. G. Murzin (2002). Scop database in 2002: Refinements accommodate structural genomics. *Nucleic Acids Research* **30**(1): 264.
- Lo Conte, L., C. Chothia and J. Janin (1999). The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* **285**(5): 2177.
- Lowman, H. B., S. H. Bass, N. Simpson and J. A. Wells (1991). Selecting high-affinity binding-proteins by monovalent phage display. *Biochemistry* **30**(45): 10832.
- Ma, B. Y., T. Elkayam, H. Wolfson and R. Nussinov (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences of the United States of America* **100**(10): 5772.
- Magnusson, M. K., P. Henning, S. Myhre, M. Wikman, T. G. Uil, M. Friedman, K. M. E. Andersson, S. S. Hong, R. C. Hoeben, N. A. Habib, S. Stahl, P. Boulanger and L. Lindholm (2007). Adenovirus 5 vector genetically re-targeted by an affibody molecule with specificity for tumor antigen her2/neu. *Cancer Gene Therapy* **14**(5): 468.
- Malipiero, U. V., N. S. Levy and P. J. Gearhart (1987). Somatic mutation in antiphosphorylcholine antibodies
Immunological Reviews **96**: 59.
- Markland, W., A. C. Ley and R. C. Ladner (1996a). Iterative optimization of high-affinity protease inhibitors using phage display .2. Plasma kallikrein and thrombin. *Biochemistry* **35**(24): 8058.
- Markland, W., A. C. Ley, S. W. Lee and R. C. Ladner (1996b). Iterative optimization of high-affinity protease inhibitors using phage display .1. Plasmin. *Biochemistry* **35**(24): 8045.
- Markland, W., B. L. Roberts, M. J. Saxena, S. K. Guterman and R. C. Ladner (1991). Design, construction and function of a multicopy display vector using fusions to the major coat protein of bacteriophage-m13. *Gene* **109**(1): 13.
- Marks, J. D., H. R. Hoogenboom, T. P. Bonnert, J. McCafferty, A. D. Griffiths and G. Winter (1991). Bypassing immunization - human-antibodies from v-gene libraries displayed on phage. *Journal of Molecular Biology* **222**(3): 581.
- Marquart, M., J. Walter, J. Deisenhofer, W. Bode and R. Huber (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallographica Section B Structural Science* **39**(4): 480.
- Massova, I. and P. A. Kollman (1999). Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *Journal of the American Chemical Society* **121**(36): 8133.
- Mattheakis, L. C., R. R. Bhatt and W. J. Dower (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proceedings of the National Academy of Sciences of the United States of America* **91**(19): 9022.

- McConnell, S. J. and R. H. Hoess (1995). Tendamistat as a scaffold for conformationally constrained phage peptide libraries. *Journal of Molecular Biology* **250**(4): 460.
- McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). Phaser crystallographic software. *Journal of Applied Crystallography* **40**: 658.
- Mitchell, J. C., R. Kerr and L. F. Ten Eyck (2001). Rapid atomic density methods for molecular shape characterization. *Journal of Molecular Graphics & Modelling* **19**(3-4): 325.
- Mitton-Fry, R. M., E. M. Anderson, T. R. Hughes, V. Lundblad and D. S. Wuttke (2002). Conserved structure for single-stranded telomeric DNA recognition. *Science* **296**(5565): 145.
- Mitton-Fry, R. M., E. M. Anderson, D. L. Theobald, L. W. Glustrom and D. S. Wuttke (2004). Structural basis for telomeric single-stranded DNA recognition by yeast cdc13. *Journal of Molecular Biology* **338**(2): 241.
- Miyachi, Y., N. Shimizu, C. Ogino and A. Kondo (2010). Selection of DNA aptamers using atomic force microscopy. *Nucleic Acids Research* **38**(4): 8.
- Monchois, V., C. Abergel, J. Sturgis, S. Jeudy and J. M. Claverie (2001). Escherichia coli ykfe orfan gene encodes a potent inhibitor of c-type lysozyme. *Journal of Biological Chemistry* **276**(21): 18437.
- Moore, G. L. and C. D. Maranas (2000). Modeling DNA mutation and recombination for directed evolution experiments. *Journal of Theoretical Biology* **205**(3): 483.
- Moreland, N., R. Ashton, H. M. Baker, I. Ivanovic, S. Patterson, V. L. Arcus, E. N. Baker and J. S. Lott (2005). A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallographica Section D-Biological Crystallography* **61**: 1378.
- Mosavi, L. K., D. L. Minor and Z. Y. Peng (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proceedings of the National Academy of Sciences of the United States of America* **99**(25): 16029.
- Moulinier, L., S. Eiler, G. Eriani, J. Gangloff, J. C. Thierry, K. Gabriel, W. H. McClain and D. Moras (2001). The structure of an asprs-trna(asp) complex reveals a trna-dependent control mechanism. *Embo Journal* **20**(18): 5290.
- Mouratou, B., F. Schaeffer, I. Guilvout, D. Tello-Manigne, A. P. Pugsley, P. M. Alzari and F. Pecorari (2007). Remodeling a DNA-binding protein as a specific in vivo inhibitor of bacterial secretin puld. *Proceedings of the National Academy of Sciences of the United States of America* **104**(46): 17983.
- Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* **53**: 240.
- Murzin, A. G. (1993). Ob(oligonucleotide oligosaccharide binding)-fold - common structural and functional solution for nonhomologous sequences. *Embo Journal* **12**(3): 861.
- Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia (1995). Scop - a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology* **247**(4): 536.
- Muyldermans, S. (2001). Single domain camel antibodies: Current status. *Reviews in Molecular Biotechnology* **74**(4): 277.

- Nagi, A. D. and L. Regan (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. *Folding & Design* **2**(1): 67.
- Nemoto, N., E. MiyamotoSato, Y. Husimi and H. Yanagawa (1997). In vitro virus: Bonding of mrna bearing puromycin at the 3'-terminal end to the c-terminal end of its encoded protein on the ribosome in vitro. *Febs Letters* **414**(2): 405.
- Nguyen, T. N., M. H. Gourdon, M. Hansson, A. Robert, P. Samuelson, C. Libon, C. Andreoni, P. A. Nygren, H. Binz, M. Uhlen and et al. (1995). Hydrophobicity engineering to facilitate surface display of heterologous gene products on *staphylococcus xylosus*. *J Biotechnol* **42**(3): 207.
- Nguyen, T. N., M. Hansson, S. Stahl, T. Bachi, A. Robert, W. Domzig, H. Binz and M. Uhlen (1993). Cell-surface display of heterologous epitopes on staphylococcus-xylosus as a potential delivery system for oral vaccination. *Gene* **128**(1): 89.
- Niesen, F. H., H. Berglund and M. Vedadi (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2**(9): 2212.
- Nissim, A., H. R. Hoogenboom, I. M. Tomlinson, G. Flynn, C. Midgley, D. Lane and G. Winter (1994). Antibody fragments from a single pot phage display library as immunochemical reagents. *Journal of Cellular Biochemistry* **56**(Suppl. 18D): 203.
- Nord, K., E. Gunneriusson, J. Ringdahl, S. Stahl, M. Uhlen and P. A. Nygren (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nature Biotechnology* **15**(8): 772.
- Nord, K., E. Gunneriusson, M. Uhlen and P. A. Nygren (2000). Ligands selected from combinatorial libraries of protein a for use in affinity capture of apolipoprotein a-1(m) and taq DNA polymerase. *Journal of Biotechnology* **80**(1): 45.
- Nord, K., J. Nilsson, B. Nilsson, M. Uhlen and P. A. Nygren (1995). A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Engineering* **8**(6): 601.
- Norel, R., S. L. Lin, H. J. Wolfson and R. Nussinov (1995). Molecular-surface complementarity at protein-protein interfaces - the critical role played by surface normals at well placed, sparse, points in docking. *Journal of Molecular Biology* **252**(2): 263.
- Norel, R., D. Petrey, H. J. Wolfson and R. Nussinov (1999). Examination of shape complementarity in docking of unbound proteins. *Proteins-Structure Function and Genetics* **36**(3): 307.
- Norman, T. C., D. L. Smith, P. K. Sorger, B. L. Drees, S. M. O'Rourke, T. R. Hughes, C. J. Roberts, S. H. Friend, S. Fields and A. W. Murray (1999). Genetic selection of peptide inhibitors of biological pathways. *Science* **285**(5427): 591.
- Novotny, J. and K. Sharp (1992). Electrostatic fields in antibodies and antibody antigen complexes. *Progress in Biophysics & Molecular Biology* **58**(3): 203.
- Nuttall, S. D., M. J. M. Rousch, R. A. Irving, S. E. Hufton, H. R. Hoogenboom and P. J. Hudson (1999). Design and expression of soluble ctla-4 variable domain as a scaffold for the display of functional polypeptides. *Proteins-Structure Function and Genetics* **36**(2): 217.
- Ochoa-Leyva, A., X. Soberon, F. Sanchez, M. Arguello, G. Montero-Moran and G. Saab-Rincon (2009). Protein design through systematic catalytic

- loop exchange in the (beta/alpha)₈ fold. *Journal of Molecular Biology* **387**(4): 949.
- Olson, C. A., H. I. Liao, R. Sun and R. W. Roberts (2008). Mrna display selection of a high-affinity, modification-specific phospho- κ b alpha-binding fibronectin. *Acs Chemical Biology* **3**(8): 480.
- Oneil, K. T., R. H. Hoess, S. A. Jackson, N. S. Ramachandran, S. A. Mousa and W. F. Degrado (1992). Identification of novel peptide antagonists for gpIIB/IIIa from a conformationally constrained phage peptide library. *Proteins-Structure Function and Genetics* **14**(4): 509.
- Orlova, A., M. Magnusson, T. L. J. Eriksson, M. Nilsson, B. Larsson, I. Hoiden-Guthenherg, C. Widstrom, J. Carlsson, V. Tolmachev, S. Stahl and F. Y. Nilsson (2006). Tumor imaging using a picomolar affinity her2 binding affibody molecule. *Cancer Research* **66**(8): 4339.
- Ortlund, E., C. L. Parker, S. F. Schreck, S. Ginell, W. Minor, J. M. Sodetz and L. Lebioda (2002). Crystal structure of human complement protein c8 gamma at 1.2 angstrom resolution reveals a lipocalin fold and a distinct ligand binding site. *Biochemistry* **41**(22): 7030.
- Painter, J. and E. A. Merritt (2006a). Optimal description of a protein structure in terms of multiple groups undergoing tls motion. *Acta Crystallographica* **D62**: 439.
- Painter, J. and E. A. Merritt (2006b). Tlsmd web serve for the generation of multi-group tls models. *Journal of Applied Crystallography* **39**: 109.
- Park, H. S., S. H. Nam, J. K. Lee, C. N. Yoon, B. Mannervik, S. J. Benkovic and H. S. Kim (2006). Design and evolution of new catalytic activity with an existing protein scaffold. *Science* **311**(5760): 535.
- Parmeggiani, F., R. Pellarin, A. P. Larsen, G. Varadamsetty, M. T. Stump, O. Zerbe, A. Caflisch and A. Pluckthun (2008). Designed armadillo repeat proteins as general peptide-binding scaffolds: Consensus design and computational optimization of the hydrophobic core. *Journal of Molecular Biology* **376**(5): 1282.
- Parry Jr., R. M., R. C. Chandan and K. M. Shahani (1965). A rapid and sensitive assay of muramindase. *Proceedings of the Society of Experimental Biology and Medicine* **119**: 384.
- Peitsch, M. C. (1995). Protein modelling by email. *Bio-Technology* **13**(8): 723.
- Predki, P. F., V. Agrawal, A. T. Brunger and L. Regan (1996). Amino-acid substitutions in a surface turn modulate protein stability. *Nature Structural Biology* **3**(1): 54.
- Pritchard, L., D. Corne, D. Kell, J. Rowland and M. Winson (2005). A general model of error-prone pcr. *Journal of Theoretical Biology* **234**(4): 497.
- Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* **381**(6585): 751.
- Rakonjac, J., G. Jovanovic and P. Model (1997). Filamentous phage infection-mediated gene expression: Construction and propagation of the gIII deletion mutant helper phage r408d3. *Gene* **198**(1-2): 99.
- Rebar, E. J. and C. O. Pabo (1994). Zinc-finger phage - affinity selection of fingers with new DNA-binding specificities. *Science* **263**(5147): 671.
- Rees, B., G. Webster, M. Delarue, M. Boeglin and D. Moras (2000). Aspartyl trna-synthetase from escherichia coli: Flexibility and adaptability to the substrates. *Journal of Molecular Biology* **299**(5): 1157.
- Reichmann, D., Y. Phillip, A. Carmi and G. Schreiber (2008). On the contribution of water-mediated interactions to protein-complex stability. *Biochemistry* **47**(3): 1051.

- Reiter, Y., U. Brinkmann, R. J. Kreitman, S. H. Jung, B. Lee and I. Pastan (1994a). Stabilization of the fv fragments in recombinant immunotoxins by disulfide bonds engineered into conserved framework regions. *Biochemistry* **33**(18): 5451.
- Reiter, Y., U. Brinkmann, K. O. Webber, S. H. Jung, B. K. Lee and I. Pastan (1994b). Engineering interchain disulfide bonds into conserved framework regions of fv fragments - improved biochemical characteristics of recombinant immunotoxins containing disulfide-stabilized fv. *Protein Engineering* **7**(5): 697.
- Renberg, B., J. Nordin, A. Merca, M. Uhlen, J. Feldwisch, P. A. Nygren and A. E. Karlstrom (2007). Affibody molecules in protein capture microarrays: Evaluation of multidomain ligands and different detection formats. *Journal of Proteome Research* **6**(1): 171.
- Reynolds, C., D. Damerell and S. Jones (2009). Protorp: A protein-protein interaction analysis server. *Bioinformatics* **25**(3): 413.
- Riechmann, L., M. Clark, H. Waldmann and G. Winter (1988a). Reshaping human antibodies for therapy. *Nature* **332**(6162): 323.
- Riechmann, L., J. Foote and G. Winter (1988b). Expression of an antibody fv fragment in myeloma cells. *Journal of Molecular Biology* **203**(3): 825.
- Roberts, B. L., W. Markland, A. C. Ley, R. B. Kent, D. W. White, S. K. Guterman and R. C. Ladner (1992a). Directed evolution of a protein: Selection of potent neutrophil elastase inhibitors displayed on m13 fusion phage. *Proceedings of the National Academy of Sciences of the United States of America* **89**(6): 2429.
- Roberts, B. L., W. Markland, K. Siranosian, M. J. Saxena, S. K. Guterman and R. C. Ladner (1992b). Protease inhibitor display m13-phage - selection of high-affinity neutrophil elastase inhibitors. *Gene* **121**(1): 9.
- Roberts, R. W. and J. W. Szostak (1997). Rna-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America* **94**(23): 12297.
- Rodier, F., R. P. Bahadur, P. Chakrabarti and J. Janin (2005). Hydration of protein-protein interfaces. *Proteins-Structure Function and Bioinformatics* **60**(1): 36.
- Rondot, S., J. Koch, F. Breitling and S. Dubel (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnology* **19**(1): 75.
- Rothe, A., R. J. Hosse, A. Nathanielsz, F. Oberhaeuser, E. P. von Strandmann, A. Engert, P. J. Hudson and B. E. Power (2006). Selection of human anti-cd22 scfvs from an acute lymphocytic leukemia related scfv library with ribosome display. *Blood: 48th Annual Meeting of the American-Society-of-Hematology*, Orlando, FL.
- Rottgen, P. and J. Collins (1995). A human pancreatic secretory trypsin-inhibitor presenting a hypervariable highly constrained epitope via monovalent phagemid display. *Gene* **164**(2): 243.
- Roussel, A., B. F. Anderson, H. M. Baker, J. D. Fraser and E. N. Baker (1997). Crystal structure of the streptococcal superantigen spe-c: Dimerization and zinc binding suggest a novel mode of interaction with mhc class ii molecules. *Nature Structural Biology* **4**(8): 635.
- Sambrook, J. and D. Russel (2001). Molecular cloning: A laboratory manual, **CSHL Press**.

- Schaffitzel, C., J. Hanes, L. Jermutus and A. Pluckthun (1999). Ribosome display: An in vitro method for selection and evolution of antibodies from libraries. *Journal of Immunological Methods* **231**(1-2): 119.
- Schlehuber, S., G. Beste and A. Skerra (2000). A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *Journal of Molecular Biology* **297**(5): 1105.
- Schlehuber, S. and A. Skerra (2001). Duocalins: Engineered ligand-binding proteins with dual specificity derived from the lipocalin fold. *Biological Chemistry* **382**(9): 1335.
- Schlehuber, S. and A. Skerra (2005). Anticalins in drug development. *Biodrugs* **19**(5): 279.
- Schmitt, E., L. Moulinier, S. Fujiwara, T. Imanaka, J. C. Thierry and D. Moras (1998). Crystal structure of aspartyl-trna synthetase from pyrococcus kodakaraensis kod: Archaeon specificity and catalytic mechanism of adenylate formation. *Embo Journal* **17**(17): 5227.
- Schonfeld, D., G. Matschiner, L. Chatwell, S. Trentmann, H. Gille, M. Hulsmeyer, N. Brown, P. M. Kaye, S. Schlehuber, A. M. Hohlbaum and A. Skerra (2009). An engineered lipocalin specific for ctla-4 reveals a combining site with structural and conformational features similar to antibodies. *Proceedings of the National Academy of Sciences of the United States of America* **106**(20): 8198.
- Schreiber, G. and A. R. Fersht (1995). Energetics of protein-protein interactions - analysis of the barnase-barstar interface by single mutations and double mutant cycles. *Journal of Molecular Biology* **248**(2): 478.
- Schweizer, A., H. Roschitzki-Voser, P. Amstutz, C. Briand, M. Gulotti-Georgieva, E. Prenosil, H. K. Binz, G. Capitani, A. Baici, A. Pluckthun and M. G. Grutter (2007). Inhibition of caspase-2 by a designed ankyrin repeat protein: Specificity, structure, and inhibition mechanism. *Structure* **15**(5): 625.
- Scott, J. K. and G. P. Smith (1990). Searching for peptide ligands with an epitope library. *Science* **249**(4967): 386.
- Seelig, B. and J. W. Szostak (2007). Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448**(7155): 828.
- Segal, D. J., J. W. Crotty, M. S. Bhakta, C. F. Barbas and N. C. Horton (2006). Structure of aart, a designed six-finger zinc finger peptide, bound to DNA. *Journal of Molecular Biology* **363**(2): 405.
- Shanahan, H. P., M. A. Garcia, S. Jones and J. M. Thornton (2004). Identifying DNA-binding proteins using structural motifs and the electrostatic potential. *Nucleic Acids Research* **32**(16): 4732.
- Sheinerman, F. B. and B. Honig (2002). On the role of electrostatic interactions in the design of protein-protein interfaces. *Journal of Molecular Biology* **318**(1): 161.
- Sidhu, S. S. (2000). Phage display in pharmaceutical biotechnology. *Current Opinion in Biotechnology* **11**(6): 610.
- Sidhu, S. S. and F. A. Fellouse (2006). Synthetic therapeutic antibodies. *Nature Chemical Biology* **2**(12): 682.
- Sidhu, S. S., G. A. Weiss and J. A. Wells (2000). High copy display of large proteins on phage for functional selections. *Journal of Molecular Biology* **296**(2): 487.

- Sieber, V., A. Pluckthun and F. X. Schmid (1998). Selecting proteins with improved stability by a phage-based method. *Nature Biotechnology* **16**(10): 955.
- Skerra, A. (1993). Bacterial expression of immunoglobulin fragments. *Current Opinion in Immunology* **5**(2): 256.
- Skerra, A. (2000). Engineered protein scaffolds for molecular recognition. *Journal of Molecular Recognition* **13**(4): 167.
- Skerra, A. (2001). 'anticalins': A new class of engineered ligand-binding proteins with antibody-like properties. *Reviews in Molecular Biotechnology* **74**(4): 257.
- Skerra, A. (2007a). Alternative non-antibody scaffolds for molecular recognition. *Current Opinion in Biotechnology* **18**(4): 295.
- Skerra, A. (2007b). Anticalins as alternative binding proteins for therapeutic use. *Current Opinion in Molecular Therapeutics* **9**(4): 336.
- Smith, G. P. (1985). Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* **228**(4705): 1315.
- Spada, S., C. Krebber and A. Pluckthun (1997). Selectively infective phages (sip). *Biological Chemistry* **378**(6): 445.
- Srivastava, A. K., Y. Sharma and K. V. R. Chary (2010). A natively unfolded beta gamma-crystallin domain from hahella chejuensis. *Biochemistry* **49**(45): 9746.
- Steffen, A. C., A. Orlova, M. Wikman, F. Y. Nilsson, S. Stahl, G. P. Adams, V. Tolmachev and J. Carlsson (2006). Affibody-mediated tumour targeting of her-2 expressing xenografts in mice. *European Journal of Nuclear Medicine and Molecular Imaging* **33**(6): 631.
- Steffen, A. C., M. Wikman, V. Tolmachev, G. P. Adams, F. Y. Nilsson, S. Stahl and J. Carlsson (2005). In vitro characterization of a bivalent anti-her-2 affibody with potential for radionuclide-based diagnostics. *Cancer Biotherapy and Radiopharmaceuticals* **20**(3): 239.
- Stein, P. E., A. Boodhoo, G. D. Armstrong, S. A. Cockle, M. H. Klein and R. J. Read (1994). The crystal structure of pertussis toxin. *Structure* **2**(1): 45.
- Steiner, D., P. Forrer, M. T. Stumpp and A. Pluckthun (2006). Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nature Biotechnology* **24**(7): 823.
- Stemmer, W. P. C. (1994). DNA shuffling by random fragmentation and reassembly - *in vitro* recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**(22): 10747.
- Stemmer, W. P. C. (2001). Molecular breeding of genes, pathways and genomes by DNA shuffling. *5th International Symposium on Biocatalysis and Biotransformations*, Darmstadt, Germany.
- Stocks, M. R. (2004). Intrabodies: Production and promise. *Drug Discovery Today* **9**(22): 960.
- Stover, C. K., G. P. Bansal, M. S. Hanson, J. E. Burlein, S. R. Palaszynski, J. F. Young, S. Koenig, D. B. Young, A. Sadziene and A. G. Barbour (1993). Protective immunity elicited by recombinant *bacille calmette-guerin* (bcg) expressing outer surface protein-a (ospa) lipoprotein - a candidate lyme-disease vaccine. *Journal of Experimental Medicine* **178**(1): 197.
- Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**(1): 207.

- Stumpp, M. T., P. Forrer, H. K. Binz and A. Pluckthun (2003). Designing repeat proteins: Modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *Journal of Molecular Biology* **332**(2): 471.
- Swaminathan, S., W. Furey, J. Pletcher and M. Sax (1992). Crystal-structure of staphylococcal enterotoxin-b, a superantigen. *Nature* **359**(6398): 801.
- Tabuchi, I., S. Soramoto, N. Nemoto and Y. Husimi (2001). An in vitro DNA virus for in vitro protein evolution. *Febs Letters* **508**(3): 309.
- Tawfik, D. S. and A. D. Griffiths (1998). Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* **16**(7): 652.
- Teplyakov, A., G. Obmolova, K. S. Wilson, K. Ishii, H. Kaji, T. Samejima and I. Kuranova (1994). Crystal structure of inorganic pyrophosphatase from *thermus thermophilus*. *Protein Science* **3**(7): 1098.
- Theobald, D. L., R. M. Mitton-Fry and D. S. Wuttke (2003). Nucleic acid recognition by ob-fold proteins. *Annual Review of Biophysics and Biomolecular Structure* **32**: 115.
- Tolmachev, V., A. Orlova, F. Y. Nilsson, J. Feldwisch, A. Wennborg and L. Abrahmsen (2007a). Affibody molecules: Potential for in vivo imaging of molecular targets for cancer therapy. *Expert Opinion on Biological Therapy* **7**(4): 555.
- Tolmachev, V., A. Orlova, R. Pehrson, J. Galli, B. Baastrup, K. Andersson, M. Sandstrom, D. Rosik, J. Carlsson, H. Lundqvist, A. Wennborg and F. Y. Nilsson (2007b). Radionuclide therapy of her2-positive microxenografts using a lu-177-labeled her2-specific affibody molecule. *Cancer Research* **67**(6): 2773.
- Transue, T. R., E. De Genst, M. A. Ghahroudi, L. Wyns and S. Muyldermans (1998). Camel single-domain antibody inhibits enzyme by mimicking carbohydrate substrate. *Proteins-Structure Function and Bioinformatics* **32**(4): 515.
- Tuncbag, N., A. GURSOY and O. Keskin (2009). Identification of computational hot spots in protein interfaces: Combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics* **25**(12): 1513.
- Tuncbag, N., O. Keskin and A. GURSOY (2010). Hotpoint: Hot spot prediction server for protein interfaces. *Nucleic Acids Research* **38**: W402.
- Ueda, H., P. Kristensen and G. Winter (2004). Stabilization of antibody v-h-domains by proteolytic selection. *Journal of Molecular Catalysis B-Enzymatic* **28**(4-6): 173.
- Vagin, A. and A. Teplyakov (1997). Molrep: An automated program for molecular replacement. *Journal of Applied Crystallography* **30**(6): 1022.
- Vaughan, T. J., J. K. Osbourn and P. R. Tempest (1998). Human antibodies by design. *Nature Biotechnology* **16**(6): 535.
- Vaughan, T. J., A. J. Williams, K. Pritchard, J. K. Osbourn, A. R. Pope, J. C. Earnshaw, J. McCafferty, R. A. Hodits, J. Wilton and K. S. Johnson (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnology* **14**(3): 309.
- Verma, R., E. Boleti and A. J. T. George (1998). Antibody engineering: Comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of Immunological Methods* **216**(1-2): 165.

- Vernet, E., A. Konrad, E. Lundberg, P. A. Nygren and T. Graslund (2008). Affinity-based entrapment of the her2 receptor in the endoplasmic reticulum using an affibody molecule. *Journal of Immunological Methods* **338**(1-2): 1.
- Vogt, M. and A. Skerra (2004). Construction of an artificial receptor protein ("Anticalin") based on the human apolipoprotein d. *Chembiochem* **5**(2): 191.
- Wagner, U. G., E. Stupperich and C. Kratky (2000). Structure of the molybdate/tungstate binding protein mop from *sporomusa ovata*. *Structure* **8**(11): 1127.
- Wang, C. I., Q. Yang and C. S. Craik (1995). Isolation of a high-affinity inhibitor of urokinase-type plasminogen-activator by phage display of ecotin. *Journal of Biological Chemistry* **270**(20): 12250.
- Ward, E. S., D. Gussow, A. D. Griffiths, P. T. Jones and G. Winter (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *escherichia-coli*. *Nature* **341**(6242): 544.
- Warke, A. and C. Momany (2007). Addressing the protein crystallization bottleneck by cocrystallization. *Crystal Growth & Design* **7**(11): 2219.
- Watson, E., W. M. Matousek, E. L. Irimies and A. T. Alexandrescu (2007). Partially folded states of staphylococcal nuclease highlight the conserved structural hierarchy of ob-fold proteins. *Biochemistry* **46**(33): 9484.
- Weiner, L. M. and P. Carter (2005). Tunable antibodies. *Nature Biotechnology* **23**(5): 556.
- Wilson, D. S., A. D. Keefe and J. W. Szostak (2001). The use of mrna display to select high-affinity protein-binding peptides. *Proceedings of the National Academy of Sciences of the United States of America* **98**(7): 3750.
- Winter, G., A. D. Griffiths, R. E. Hawkins and H. R. Hoogenboom (1994). Making antibodies by phage display technology. *Annual Review of Immunology* **12**: 433.
- Worn, A., A. der Maur, D. Escher, A. Honegger, A. Barberis and A. Pluckthun (2000). Correlation between in vitro stability and in vivo performance of anti-gcn4 intrabodies as cytoplasmic inhibitors. *Journal of Biological Chemistry* **275**(4): 2795.
- Worn, A. and A. Pluckthun (2001). Stability engineering of antibody single-chain fv fragments. *Journal of Molecular Biology* **305**(5): 989.
- Wuttke, D. S., E. M. Anderson, W. A. Halsey, V. Lundblad, R. M. Mitton-Fry and D. L. Theobald (2004). Recognition of single-stranded DNA at telomeres by the essential protein cdc13. *18th Symposium of the Protein-Society, San Diego, CA*.
- Xu, D., S. L. Lin and R. Nussinov (1997a). Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *Journal of Molecular Biology* **265**(1): 68.
- Xu, D., C. J. Tsai and R. Nussinov (1997b). Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering* **10**(9): 999.
- Xu, L. H., P. Aha, K. Gu, R. G. Kuimelis, M. Kurz, T. Lam, A. C. Lim, H. X. Liu, P. A. Lohse, L. Sun, S. Weng, R. W. Wagner and D. Lipovsek (2002). Directed evolution of high-affinity antibody mimics using mrna display. *Chemistry & Biology* **9**(8): 933.

- Yan, C., F. Wu, R. L. Jernigan, D. Dobbs and V. Honavar (2008). Characterization of protein-protein interfaces. *Protein Journal* **27**(1): 59.
- Yanez, J., M. Arguello, J. Osuna, X. Soberon and P. Gaytan (2004). Combinatorial codon-based amino acid substitutions. *Nucleic Acids Research* **32**(20): 10.
- Yang, H. J., P. D. Jeffrey, J. Miller, E. Kinnucan, Y. T. Sun, N. H. Thoma, N. Zheng, P. L. Chen, W. H. Lee and N. P. Pavletich (2002). Brca2 function in DNA binding and recombination from a brca2-dss1-ssdna structure. *Science* **297**(5588): 1837.
- Yonezawa, M., N. Doi, Y. Kawahashi, T. Higashinakagawa and H. Yanagawa (2003). DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Research* **31**(19).
- Young, A. C. M., G. Scapin, A. Kromminga, S. B. Patel, J. H. Veerkamp and J. C. Sacchettini (1994a). Structural studies on human muscle fatty-acid-binding protein at 1.4-angstrom resolution - binding interactions with three c18 fatty-acids. *Structure* **2**(6): 523.
- Young, L., R. L. Jernigan and D. G. Covell (1994b). A role for surface hydrophobicity in protein-protein recognition. *Protein Science* **3**(5): 717.
- Zafir-Lavie, I., Y. Michaeli and Y. Reiter (2007). Novel antibodies as anticancer agents. *Oncogene* **26**(25): 3714.
- Zahnd, C., E. Wyler, J. M. Schwenk, D. Steiner, M. C. Lawrence, N. M. McKern, F. Pecorari, C. W. Ward, T. O. Joos and A. Pluckthun (2007). A designed ankyrin repeat protein evolved to picomolar affinity to her2. *Journal of Molecular Biology* **369**(4): 1015.
- Zhang, R. G., D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley and E. M. Westbrook (1995). The 3-dimensional crystal-structure of cholera-toxin. *Journal of Molecular Biology* **251**(4): 563.
- Zhao, H. M., L. Giver, Z. X. Shao, J. A. Affholter and F. H. Arnold (1998). Molecular evolution by staggered extension process (step) in vitro recombination. *Nature Biotechnology* **16**(3): 258.
- Zhou, H. X. and Y. B. Shan (2001). Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins-Structure Function and Genetics* **44**(3): 336.