Isolating novel antibodies against conformation specific epitopes and a hapten using Next Generation Antibody Phage Display

Rose H. Reader

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, May 2019.

Primary supervisors: Dr. Kevin Gough
Secondary Supervisor: Professor Richard Emes

Abstract

Phage display bypasses the immune response by bringing the entire immune repertoire to a target antigen, *in vitro*. The coupling of antibody phage display and next generation sequencing (NGS) has the potential to revolutionise the discovery of novel antibodies. In this study, next generation antibody phage display (NGAPD) strategies were devised to identify novel antibodies against conformational epitopes on the enzyme phytase, and against the hapten phytic acid. These were notably difficult targets and previous immunisation attempts for phytic acid have not demonstrated any immune response for hybridoma or the production of immune phage display libraries. This study used naïve human scFv and camelid VHH phage display libraries and it was therefore more difficult to identify potential binders due to a lack of bias towards the targets within a high level of antibody diversity. Following phage panning, the read depth provided by NGS facilitates the identification of novel ligands from within a diverse sub-library, and thus the study also aimed to optimise the use of NGS with naïve phage display libraries. The NGS approach was used in comparison with conventional colony picking. The colony picking approach successfully identified phage-fusion binders against the three targets, however when expressed as antibody-MBP fusions the binding activities were lost.

The NGS data analysis involved ranking candidates using a Z score followed by further, highly stringent sorting analyses designed to remove nonspecific ligands. Major challenges involved rescuing scFv candidates from their sub-libraries; this was later resolved with the use of the camelid VHH library that uses a single antibody framework. The NGAPD concluded with a range of NGS candidates synthesised from the final VHH experiment. The candidates were cloned in a single step into a pMAL expression system using inverse PCR. Finally, the NGS candidates were expressed as MBP fusions; however no binding activity was detected.

The final section of the study considered alternative bioinformatics analysis strategies for NGS data that have been used in the literature for both antibodies and peptides and also suggested novel enrichment profiles as a means of identifying selection dependant ligands. The VHH NGS data was re-assessed based solely on candidate frequency and enrichment throughout

panning. A selection of antibody candidates were identified by both bioinformatics methods for phytase, suggesting that the original NGS analysis strategy was successfully identifying promising candidates that were more likely to be conformation specific.

Acknowledgements

Dr Kevin Gough and Dr Ben Maddison have been inspirations throughout the study and will continue to be so. Their knowledge, support and their characters were irreplaceable. I am very grateful to them both and feel privileged to have worked beside them.

I give special thanks to the members of ADAS; Jon, Keith, Claire and Helen, and also Dr Alison Gray and my fellow postgrads Rob, Matt, Kate, Mary, Maria and Juan. Their guidance and friendships have been invaluable, and they made the whole experience so much more enjoyable.

I would also like to thank Dr Mike Bedford for his kind encouragement and advice during the project and for the invitation to present my findings.

My most heartfelt thanks go to Thomas Kirkland. His unwavering love and support throughout the study were essential to bring it to completion. Despite being on the receiving end during most of my overwhelming moments, he would always make everything right with the world.

I also thank my friends and family, most notably my sister Hazel and dear friend Christine. Thank you for always being there for me and ensuring my wine glass was always full!

Last but not least, I give thanks to ADAS, AB-Vista and the University of Nottingham for funding this PhD and for providing an excellent support network and guidance throughout the project.

Contents

Abstract 1
Acknowledgements 3
Chapter 1: Introduction 15
Aims of the study 15
1.1 Introduction to antibodies 15
1.2 Monoclonal antibody isolation 17
1.3 Antibody Phage Display 19
1.4 Study targets 31
1.5 Phytase & phytic acid 32
Chapter 2: Materials and Methods 39
2.1 Materials 39
2.2 SDS PAGE of protein samples 40
2.3 Western Blot 41
2.4 Determining the protein concentration 42
2.5 Indirect measurement of Phytase activity 42
2.6 Removal of N-Linked Glycans using PNGase F 44
2.7 Standard recipes used for phage display 45
2.8 Growing the Griffin scFv and VHH phagemid libraries 46
2.9 Phage Precipitation 49
2.10 Antigen preparation 49
2.11 Phytase immobilisation 50
2.12 Phytic acid immobilisation 50
2.13 Phytic acid staining (Toluidine Blue) 51
2.14 Panning Rounds 52
2.15 Random colony picking 54
2.16 Preparation of DNA for Ion Torrent Sequencing 54
2.17 dsDNA quantification using Qubit kit 60
2.18 NGS Data Analysis 60
2.19 scFv Rescue 73
2.20 VHH synthesis strategy 76
2.21 Making electro-competent cells 78
2.22 Colony PCR 79
2.23 Sanger sequencing to confirm the expected HCDR3 region 80
2.24 Phage ELISA 80
2.25 pMAL expression system 81
2.26 Soluble antibodies 84

Chapter 3: Identifying antibodies (scFv) against the active phytase conformation using next generation sequencing 87
3.1 Determining the size of the Quantum Blue and Megazyme phytase 87
3.2 Determining the concentration of Quantum Blue & Megazyme phytase 89
3.3 Establishing conditions to produce heat-denatured or native phytase prior to panning 90
3.3 Activity assessment for phytase preparations used for panning 97
3.4 Panning rounds 1 and 2 98
3.5 Preparation of DNA for Ion Torrent sequencing 99
3.6 Data analysis and scFv rescue 99
3.7 Testing the specificity of the scFv rescues via phage ELISA 103
3.8 Discussion 106
Chapter 4: Identifying antibodies (scFv) against active and denatured phytase conformations and phytic acid using traditional and next generation phage display 111
4.1 Determination of purity of Quantum Blue phytase 111
4.2 Antigen preparation: Phytases 114
4.3 Optimising the immobilising phytic acid 115
4.4 Staining 120
4.5 Panning Rounds 1—4 and screening 121
4.6 Preparation of DNA for Ion Torrent sequencing 123
4.7 Colony picks 123
4.8 Sub cloning scFv candidates 131
4.9 Sanger sequencing high signal clones from monoclonal picks 133
4.10 ScFv gene amplification 135
4.11 Confirmation of antibody display 136
4.12 Sub cloning scFv: Continued 141
4.13 NGPD data analysis and scFv rescue 147
4.14 Testing the specificity of the scFv rescues via colony PCR and phage ELISA 151
4.15 Inverse PCR primer specificity 155
4.16 Discussion 157
Chapter 5: Identifying single domain antibodies from a naïve VHH library using traditional and next generation phage display 165
5.1 Comparison of antibody expression levels between three phage display libraries 165
5.2 NGAPD using the camelid VHH library 170
5.3 Preparation for NGS: Phage reinfection and DNA extractions 173

5.4 Colony picking 176
5.5 Phytic Acid panning using Covalink support 193
5.6 NGS Analysis 197
5.7 VHH synthesis 211
5.8 Discussion 233
Chapter 6: Assessment and optimisation of bioinformatics analysis methods 241
6.1 scFv flanking region changes 241
6.2 VHH & Griffin library diversity 248
6.3 Copy number as an estimate of enrichment through panning 248
6.4 Reanalysing the VHH data using alternative published methods 252
6.5 Growth and selection patterns 261
6.6 Revisiting the colony picks; comparing unrelated datasets 276
6.7 Discussion 278
Chapter 7: Conclusions 294
Chapter 8: Future Approaches 303
References 304

List of Tables

Table 1.1: Examples of Helper Phage in the Literature 23

Table 2.1: Antibodies used throughout the study 39

Table 2.2: Bacterial strains used within the study 39

Table 2.3: The tag orientation in each vector and antibody combination used within the study. 40

Table 2.4: Round 1 PCR primers for NGS preparation 56

Table 2.5: Ion Torrent Barcode Primers – Round 2 PCR for NGS 57

Table 2.6: The vectors in which the anti-β-lactamase VHH clone was sub-cloned into to generate a template. 77

Table 3.1: Phytase activity assay 1 - Megazyme total phosphorous kit 91

Table 3.2: Phytase activity assay 2 - Megazyme total phosphorous kit 92

Table 3.3: Denaturing phytase assay 1 - Megazyme total phosphorous kit 94

Table 3.4: The 24 scFvs selected for phagemid rescues after panning against active phytase 101

Table 3.5: Inverse PCR primer pairs for phagemid rescues for scFvs selected against active phytase. 102

Table 3.6: Rescued scFv transformants: Sanger sequencing results 105

Table 4.1: Expected amino acid sequence coverage for gel extracts 114

Table 4.2: Summary of samples prepared for next generation sequencing for active and denatured phytase and phytic acid 124

Table 4.3: Sanger sequencing HCDR3 amino acid sequence results for the four high signal clones for denatured phytase and phytic acid. 134

Table 4.4: Sanger sequencing LCDR3 amino acid sequence results for the four monoclonal picks for denatured phytase and phytic acid. 134

Table 4.5: Troubleshooting summary for the subcloning of monoclonal picks after panning against denatured phytase. 139

Table 4.6: Troubleshooting summary for the subcloning of monoclonal picks after panning against phytic acid 140

Table 4.7: The final list of intended scFv rescues for active phytase and denatured phytase post-NGS analysis. 148

Table 4.8: Inverse PCR primer pairs for phagemid rescues for scFvs selected against active and denatured phytase. 150

Table 4.9: Sanger Sequencing for Inverse PCR transformant colonies 154

Table 5.1: Summary of samples prepared for next generation sequencing for active and denatured phytase and phytic acid 174

Table 5.2: Summary of samples prepared for next generation sequencing for active and denatured phytase and phytic acid 190

Table 5.3: The top 10 NGS candidates for active phytase, denatured phytase and phytic acid. 201

Table 5.4: The most common motifs identified by the MEME suite for all three target antigens. 207

Table 5.5: The major motifs for each target, as identified by MEME suite, and their selected contributing sequence. 210

Table 6.1: Original list of rescue candidates for panning against active and denatured phytase and phytic acid based on 1 change allowed within each flanking sequence compared to 0 or 2 changes. 243

Table 6.2: Total reads in each round of panning, for each target antigen, after 0—2 flanking region changes 246

Table 6.3: Percentage increase in total reads for each round of panning after comparison to 0 flanking region changes 246

Table 6.4: Percentage of Griffin library HCDR3 sequences captured after allowing 0—2 changes in the flanking regions 247

Table 6.5: Sequence (HCDR3) copy number and a percentage of the total reads in the 12mer VHH library. 249

Table 6.6: Sequence (HCDR3) copy number and a percentage of the total reads in the 16mer VHH library. 249

Table 6.7: Sequence (HCDR3) copy number and a percentage of the total reads in the 12mer VHH library. 250

Table 6.8: The top 20 frequency candidates for the active phytase VHH dataset. 253

Table 6.9: The top 20 frequency candidates for the denatured phytase VHH dataset. 254

Table 6.10: The top 20 frequency candidates for the phytic acid VHH dataset. 255

Table 6.11: The frequencies of the top 20 VHH candidates (frequency analysis) in each active phytase, positive replicate barcode. 256

Table 6.12: The frequencies of the top 20 VHH candidates (frequency analysis) in each denatured phytase, positive replicate barcode. 257

Table 6.13: The frequencies of the top 20 VHH candidates (frequency analysis) in each phytic acid, positive replicate barcode. 258

Table 6.14: The top 20 frequency candidates for the denatured phytase VHH dataset. 261

Table 7.1: The most promising VHH candidates identified during the study. 297

List of Figures

Figure 1.1: M13 phage structure and protein function. 20

Figure 1.2: Nomenclature of phage systems using modified pIII. 22

Figure 1.3: Antibody formats used in phage display. 24

Figure 1.4: Antibody phage display panning procedure. 28

Figure 1.5: The structure of phytic acid and dephosphorylation sites 33

Figure 2.1: Griffin scFv phagemid vector pHEN2. 47

Figure 2.2: VHH phagemid vector pSD3. 48

Figure 2.3: Covalink plate linker joining to a phosphoramidate bond 51

Figure 2.4. NGS Coverage and PCR schematic for the Griffin and VHH vectors. 55

Figure 2.5. NGS Coverage and PCR schematic for the Griffin and VHH vectors. 68

Figure 2.6. ScFv Rescue Strategy: Inverse PCR. 74

Figure 2.7. VHH synthesis strategy: Inverse PCR. 76

Figure 3.1. Quantum Blue phytase and Megazyme phytase molecular weight analysis. 88

Figure 3.2. Quantum Blue phytase and Megazyme phytase concentration assay. 89

Figure 3.3: A standard curve of Quantum Blue phytase activity. 93

Figure 3.4: Quantum Blue phytase denaturation assay. 95

Figure 3.5: Partial denaturation assay for Quantum Blue phytase. 96

Figure 3.6: Quantum Blue standard activity curve. 97

Figure 3.7: Phage-scFv ELISA against partially treated, active phytase for 22 NGS rescue candidates. 104

Figure 4.1. PNGase F treated and untreated Quantum Blue phytase molecular weight analysis on a native and denaturing protein gel. 112

Figure 4.2. PNGase F treated Quantum Blue phytase samples selected for LC-MS. 113

Figure 4.3: pH-shift Quantum Blue phytase activity assay. 115

Figure 4.4: Digestion of phytic acid from 10mg of Titansphere TiO_2 beads. 116

Figure 4.5: Digestion of immobilised phytic acid from TiO_2 or ZiO_2 coated plates. 119

Figure 4.6: Stained phytic acid bound to TiO2 and ZiO2 supports. 120

Figure 4.7: Diagrammatic representation of the panning strategy used with the Griffin scFv library for all three target antigens. 122

Figure 4.8: Colony picking phage-ELISA. 126

Figure 4.9: Repeat phage-scFv ELISA using the four monoclonals picked from round 4 panning against denatured phytase. 127

Figure 4.10: Immobilised phytic Acid, ATP, PolyP 25 and free phosphate stained with toluidine blue. 128

Figure 4.11: Phage-scFv ELISA using four monoclonal picks from round 4 panning against phytic acid and other phosphate sources immobilised on TiO_2 and ZiO_2. 129

Figure 4.12: Double digested pMal P5X vector. 131

Figure 4.13: Double digested high signal clones to denatured phytase (C7, D8, E4, F7). 132

Figure 4.14: Double digested high signal clones to phytic acid (E6, F6, G6, G11). 133

Figure 4.15: Denatured phytase and phytic acid monoclonal picks – amplified scFv genes (C7, D8, E4, F7, E6, F6, G6, G11). 135

Figure 4.16: Western blot analysis for denatured phytase and phytic acid clones with anti-pIII antibody. 137

Figure 4.17: Amplification of 'C7's and 'E6's scFv gene to add an NCOI restriction site. 141

Figure 4.18: Double digested pCR™4-TOPO vector containing C7 and E6 scFv genes. 142

Figure 4.19: Double digested P5X vector containing C7 and E6 scFv genes. 143

Figure 4.20: Western blot analysis for the expression of three anti-phytic acid and denatured phytase P5X:E6 clones with anti-MBP and anti-mouse-HRP. 144

Figure 4.21: Immobilised soluble MBP-scFv ELISA using P5X:E6 (anti-phytic acid) and P5X:C7 (anti-denatured phytase). 145

Figure 4.22: Soluble MBP-scFv ELISA using P5X:E6 (anti-phytic acid) and P5X:C7 (anti-denatured phytase). 146

Figure 4.23: Amplification of the scFv genes present in the Inverse PCR transformants for clones isolated against Active and Denatured Phytase and Phytic Acid. 152

Figure 4.24: Phage-scFv ELISA for the denatured phytase inverse PCR rescues. 153

Figure 4.25: Amplification of the pHEN2 phagemid by Inverse PCR using library and target template. 156

Figure 5.1: Western blot analysis of soluble extracts from whole cell extracts. 167

Figure 5.2: Western blot analysis of soluble extracts from the periplasm. 169

Figure 5.3: Phytase activity assay after thermal cycler denaturation. 170

Figure 5.4: Diagrammatic representation of the panning strategy used with the VHH library for all three target antigens. 172

Figure 5.5: Polyclonal phage ELISA for the active and denaturd phytases. 177
Figure 5.6: Polyclonal phage ELISA for phytic acid. 178
Figure 5.7: Polyclonal phage ELISA for active and denaturd phytase. 179
Figure 5.8: Polyclonal phage ELISA for phytic acid. 180
Figure 5.9: Colony picking phage-ELISA. 183
Figure 5.10: Repeat phage ELISA for the phytase VHH colony picks. 184
Figure 5.11: Colony picking phage-ELISAs: set 2. 187
Figure 5.12: Active and Denatured phytase colony picking phage-ELISA: set 2 repeat. 188
Figure 5.13: Repeat phage ELISA for the VHH clones derived from the colony picking approach for active and denatured phytase. 192
Figure 5.14: Phytic acid and ATP immobilised on Covalink wells. 194
Figure 5.15: Phytic acid colony picking phage-ELISA – picks 1—120 immobilised. 196
Figure 5.16: The number of reads, as a percentage of the total, within each panning round for the picked clones. 199
Figure 5.17: Common motif analysis. 206
Figure 5.18: Anti-bLAC VHH gene amplicon. 213
Figure 5.19: Anti-bLAC VHH digests from the P5X and PET22B vectors. 214
Figure 5.20: Western blot analysis of the anti-bLAC soluble VHH. 215
Figure 5.21: Anti-Elastin VHH ELISA: phage and soluble VHH formats. 217
Figure 5.22: Western blot analysis and Soluble VHH-MBP ELISA for the 15 clones raised against denatured phytase: whole cell extracts. 221
Figure 5.23: Western blot analysis and Soluble VHH-MBP ELISA for the 11 clones raised against active phytase: whole cell extracts. 222
Figure 5.24: Western blot analysis and soluble VHH-MBP ELISA for the 12 clones raised against phytic acid: whole cell extracts. 224
Figure 5.25: Western blot analysis of the anti-bLAC VHH clone. 226
Figure 5.26: Soluble VHH-MBP ELISAs for anti-bLAC clone and no insert P5X control. 227
Figure 5.27: Western blot analysis of the active and denatured phytase and phytic acid VHH candidates extracted from the periplasm. 230
Figure 5.28: Soluble VHH-MBP ELISA for the clones raised against active and denatured phytase: periplasmic extracts. 231
Figure 6.1: The number of distinct HCDR3 sequences captured during each panning round for active and denatured phytase and phytic acid. 252
Figure 6.2: The frequency of the top 20 active and denatured phytase and phytic acid candidates during panning (frequency analysis). 261

Figure 6.3: Growth and selection phase frequencies for top 7 VHH candidates. 266

Figure 6.4: Growth and selection phase frequencies for 3 known elastin binders. 269

Figure 6.5: Growth and selection phase frequencies for the top 10 anti-elastin clones (Frequency Analysis). 275

Figure 7.1 A summary of phage display approaches relevant to the present study 300

Abbreviations

AOAC	Association of Official Analytical Chemists
ATP	Adenine Tri-phosphate
BSA	Bovine serum albumin
CDR	Complementary determining region
CFU	Colony forming units
CH	Constant region
CV	Coefficient of variation
dsDNA	Double stranded deoxyribonucleic acid
ECL	Enhanced chemiluminescence
EDC	1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Fab	Antigen binding fragment
Fc	fragment crystallisable region
FW	Framework region
HCDR3	Heavy chain complementary determining region 3
HCl	Hydrochloric acid
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
Ig(X)	Immunoglobulin (**X** = **G** gamma, **M** mu, **A** alpha, **D** delta or **E** epsilon)
IP6	MyoInositol Hexaphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KLH	Keyhole limpet haemocyanin
LCDR3	Light chain complementary determining region 3
LC-MS	Liquid chromatography and mass spectrometry
LOD	Limit of detection
MBP	Maltose binding protein
MOI	Multiplicity of infection
MOPS	3-(N-morpholino)propanesulfonic acid)
NGAPD	Next generation antibody phage display
NGS	Next generation sequencing
NIR	Near-infrared spectrometry

OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
SMRT	Single molecule real-time sequencing
PVDF	Polyvinylidene difluoride
PVP	Poly 4-vinyl pyridine
SAROTUP	Scanner and reporter of target-unrelated peptides
scFv	Small chain variable fragment
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
ssDNA	Single stranded deoxyribonucleic acid
SUMO	Small ubiquitin-like modifier
TAE	Tris-acetate-EDTA
TiO_2	Titanium Dioxide
Tm	Melting temperature
TMB	3,3',5,5'-Tetramethylbenzidine
TUP	Target unrelated peptides
V_H	Variable heavy chain
VL	Variable light chain
ZiO_2	Zirconium Dioxide

Chapter 1: Introduction

Aims of the study

The study aimed to develop a phage display strategy that would facilitate the isolation of novel antibody fragments with specific binding properties using Next Generation Antibody Phage Display (NGAPD). The novel antibodies would distinguish between conformational epitopes on the enzyme phytase and bind a hapten target known as phytic acid. The applications for such antibodies would include the development of novel immunoassays that could rapidly determine the phytase functional state, or accurately quantify phytic acid.

1.1 Introduction to antibodies

Antibodies are specialised proteins produced in animals in response to foreign antigens. There are five major types of antibodies in mammals (IgG, IgM, IgA, IgE and IgD), which are classed based on the differences within their heavy chains.

- Gamma (IgG) antibodies are the smallest and most common format.
- Mu (IgM) are the largest antibody format, are usually restricted to the bloodstream. They are the first response antibodies and indicate recent exposure to their target antigens.
- Alpha (IgA) are usually found in externally secreted fluids such as saliva, sweat and tears or in the mucous membranes lining the respiratory and gastrointestinal tract.
- Epsilon (IgE) are often responsible for allergic reactions due to their role in histamine release.
- Delta (IgD) are a poorly understood format; mostly bound to B cells.

Each antibody possesses a highly specific binding site enabling them to bind to their target with considerable specificity.

The binding site, or paratope, of the antibody consists of two variable heavy (VH) and two variable light chains (VL) attached to more structurally conserved constant (C) regions (CH1, CH2 and CH3).

Light chains on antibodies are classed as either Lambda or Kappa depending on a small number of amino acid changes.

The variable chains contain three highly variable complementarity determining regions (CDRs) surrounded by four more conserved framework sequences. The variable domains differ between antibodies due to the random rearrangement of variable (V), joining (J) and in the heavy chains case, diversity (D) gene segments.

The constant regions CH2 and CH3 form the stable fragment crystallisable (fc) region of an antibody. These regions differ between the antibody classes and largely determine how antigens are destroyed; furthermore, they can also extend the half-lives of their respective antibodies (Muller et al., 2012). The fc regions are also commonly targeted by generic species-specific antibodies in diagnostic assays.

In vitro, antibodies can be used to accurately detect even trace amounts of their target antigen due to their high sensitivity and specificity. They can be ideal reagents for developing accurate, high throughput diagnostic assays and therapeutics. Antibodies used for diagnostic applications are usually polyclonal or monoclonal antibodies, the latter produced by hybridoma technology. Polyclonal antibodies are secreted by many B cells of different lineages, so although they bind to the same antigen, they will each bind to different epitopes. In contrast, monoclonal antibodies have the same B cell lineage and will only interact with one epitope.

Polyclonal antibodies are easier and less expensive to obtain than monoclonals and are traditionally purified from immunised animal sera. Although polyclonal antibodies can be more versatile and bind to a number of epitopes, they are also more likely to cross react with other, unintended antigens with similar structures.

Therapeutic antibodies are given directly to patients to help combat a range of ailments such as infectious diseases, auto-immune diseases and cancer. They are especially beneficial for immunocompromised individuals who might not be able to produce enough antibodies themselves. Therapeutic antibodies provide a form of passive immunity which aims to instruct the

immune system into attacking the antibodies target. Producing therapeutic antibodies for use in humans is difficult as they are usually produced by other species and must be re-engineered as chimeric (e.g murine variable domain with human constant regions) or humanised antibodies (e.g murine CDRs with human constant and variable frameworks). Polyclonal antibodies are at a disadvantage because of the increased risk of cross-reactivity and *in vivo* can mistakenly bind to a 'self' antigen and mount a response. Even in circumstances where foreign antibodies do not lead to exaggerated immune responses, their *in vivo* half-life is still short. Generally, same species antibodies are safer and make therapeutics that are more effective however, that is a particularly difficult goal to achieve when discovering new antibodies for human therapeutics.

Monoclonal antibodies are generally more useful as both diagnostics and therapeutics due to the high specificity and low chance of cross reaction with similar antigens. Developing immunoassays using high affinity, monoclonal antibodies is ideal and also helps maintain consistency between batches of antibodies. Unfortunately, monoclonal antibodies are not only more expensive to produce in comparison to polyclonals, but they are also more technically challenging to isolate.

1.2 Monoclonal antibody isolation

The gold standard technique for isolating monoclonal antibodies involves a technique known as hybridoma technology, developed in 1975 by Kohler and Milstein (Kohler and Milstein, 1975). The method begins by exposing an animal's immune system to the target antigen to provoke an immune response. Binding antibodies are eventually secreted by B cells and they can then be purified from the animal's sera as a polyclonal reagent. To achieve monoclonal antibodies, the Hybridoma technique involves isolating individual B cells from immune-cell rich regions such as the spleen. The cells are then fused with immortal cell lines, a myeloma, to produce a hybrid cell or hybridoma by electrofusion or using chemicals such as polyethylene glycol (PEG). The hybrid's monoclonal antibodies are then screened individually to check for specificity for the target antigen, a procedure usually involving an enzyme-linked immunosorbent assay (ELISA).

The hybridoma method is a well-developed technique for isolating novel, monoclonal antibodies but it is a very expensive method and the antibodies require 'humanizing' to be of use as human therapeutics. Furthermore, the immortal cell lines used to secrete the monoclonals are often unstable due to the inbalance of genetic material derived from both the B cell and myeloma. The method cannot be used for any target antigens that are either toxic to the host, or are non-immunogenic, for instance if they are regarded as self and thus do not cause an immune response.

Since the hybridoma technology's advancement a range of other techniques have been developed such as antibody and animal humanization, VL and VH repertoire data mining (Reddy et al., 2010), glycoengineering and antibody display techniques. The hybridoma method and mining immunised antibody-sequence data both require animal immunisations as a pre-requisite before antibodies can be isolated/identified. The use of animals raises ethical concerns and is a costly process limited to non-toxic and non-self antigens.

Antibody display systems are an increasingly common alternative to hybridoma production and can offer the flexibility of isolating antibodies against more difficult targets such as toxins or non-immunogens. One major advantage of antibody display systems is that they can theoretically be used to generate antibody fragments against a target within only a few weeks. Bacteriophage (phage) display systems are the most common format however other display systems have also been used i.e ribosomal (Hanes and Pluckthun, 1997), yeast (Boder and Wittrup, 1997) and mammalian display (King et al., 2014) systems. Bacteriophage were discovered in 1915 by William Twort, and two years later Felix d'Herelle demonstrated they could kill bacteria (Sulakvelidze et al., 2001). Phage became important research tools, helping to determine the basis of genetic material via the Hershey and Chase experiment, that 3 nucleotides encoded an amino acid (Crick et al., 1961) and also enabled the identification of restriction enzymes (Luria and Human, 1952, Bertani and Weigle, 1953). Bacteriophage can have lytic or lysogenic (non-lytic) life cycles, but temperate bacteriophage will have both. Lytic lifecycles involve infecting the host and rapidly killing the cells to release new virions. The lysogenic lifecycle enables phage to integrate their genomes into the hosts, residing in the form of plasmids that are stable for many

generations. Non-lytic phage are shed from the host cell at a cost to the host's metabolic rate but with no obvious cell death.

The Ff class of filamentous (rod shaped) bacteriophage include M13, fd and f1 phage and are a family of single stranded DNA (ssDNA) viruses that target gram negative bacteria. The three Ff classes of phage each possess the same surface proteins but were isolated in different parts of the world.

Antibody display systems are limited to phage for the purpose of this study.

1.3 Antibody Phage Display

Antibody phage display is a highly adaptable methodology that facilitates the identification of antibodies with affinity for any given target. The original phage display experiments by Smith (1985) demonstrated that filamentous bacteriophage could be genetically modified to display an EcoRI endonuclease, via their PIII surface protein, and could then be significantly enriched over phage that did not possess the fusion by immobilising them with polyclonal antibodies. The display of functional antibody fragments was demonstrated later when the gene encoding an anti-lysozyme antibody was inserted at the N-terminal of the pIII gene in a phage vector. The antibody chains were displayed with the VH and VL chains joined together by a linker (McCafferty et al., 1990).

Early phage display experiments employed lytic lambda phage with limited success (Huse et al., 1989). However, non-lytic phage such as M13 are assembled in the periplasm of the bacteria, enabling phage production without killing the host and the correct folding of antibody fragments due to the oxidising environment (Skerra and Pluckthun, 1988). Thus the use of non-lytic Ff class bacteriophage proved to be much more successful for screening antibodies and is typically the class of choice for phage display systems.

M13 phage are made up of 11 different proteins (figure 1.1).

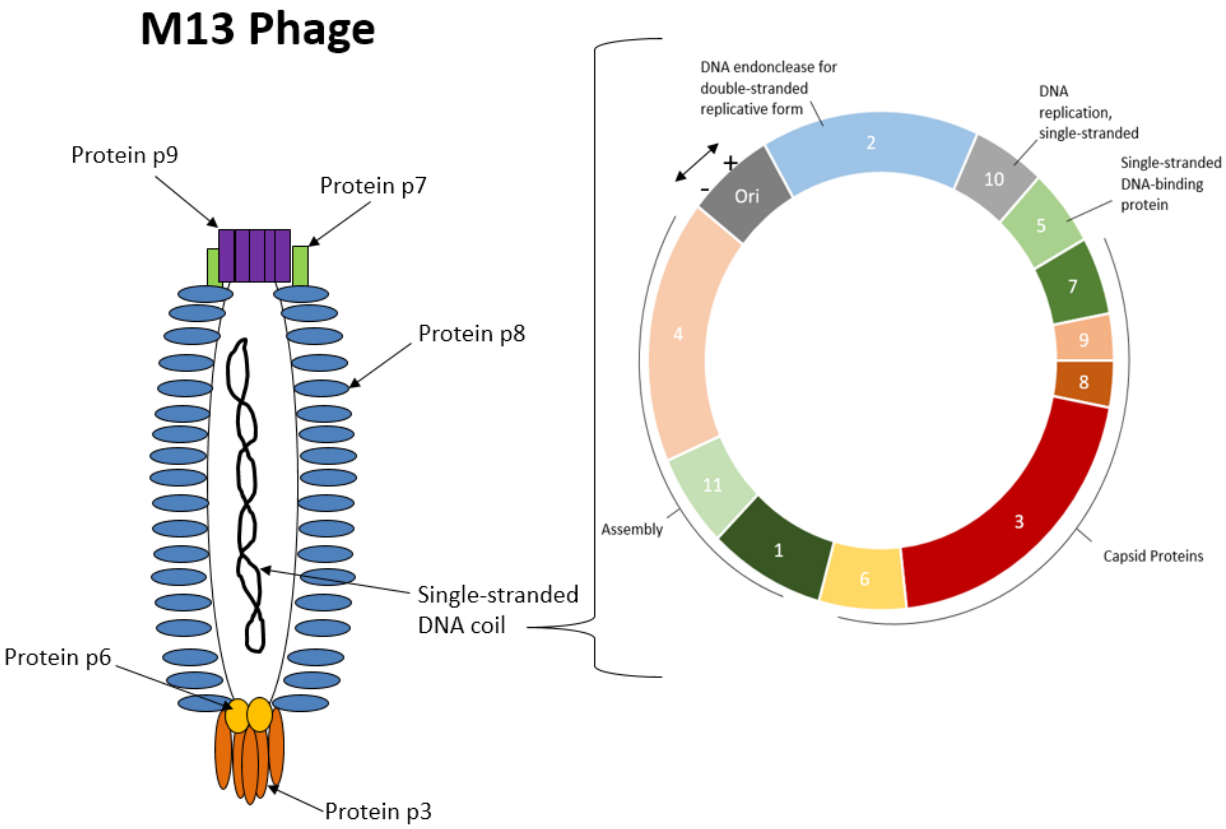


Figure 1.1: M13 phage structure and protein function.
M13 phage possess 11 proteins and collectively they are responsible for all of the basic phage functions such as DNA replication, infection and viral assembly and packaging. Adapted from Sachdev S. Sidhu (2017).

Phage display libraries usually use either pIII (a minor structural protein involved in infectivity) or pVIII (a major structural protein) surface proteins for display purposes. Any protein can be displayed on the surface of phage but care must be taken not to display too large a protein, as it may impair the phage's normal function. The pIII surface protein can be used to display another protein even larger than itself provided that it is monovalent, otherwise the individual pIII protein can no longer interact with the F pilus expressed by bacterial cells and the virus is no longer infectious. Multivalent display via pIII is possible but infectivity is severely reduced with each additional display. Alternatively, there are 2,700 of copies of pVIII per phage and although the potential for multivalent display is much greater than for pIII, the tolerated display protein size is much smaller.

The display of antibody fragments is almost always on pIII and the resulting phage displays both pIII-antibody fusion and wild type pIII. The structure of pIII has been well documented and it plays an essential role in

the phage's ability to infect the bacterial host. The pIII structure interacts with male *Escherichia coli* (*E. coli*) (cells expressing the F' pilus) to both infect and release new virions (Holliger and Riechmann, 1997, Riechmann and Holliger, 1997). The pIII protein facilitates infection by interacting with the periplasmic TolA protein (part of the TolQRA complex). The pIII N2 domain is thought to interact with TolA and bring the pIII N1 domain within reach of TolA. The mechanism is poorly understood beyond this interaction, however the pIII C domain is responsible for finalising the packaging of new virions and their subsequent release from the cell.

Phage display libraries can use either true phage or phagemids. True phage are functional virions that contain all of their necessary genes and are capable of infecting a host and replicating. Phagemids are plasmids that contain only a select number of phage genes, most importantly the phage origin of replication and packaging signalling sequences, and therefore require a helper phage to produce new virions as they do not contain all of the genes necessary to produce and package recombinant phage. The nomenclature of phage vector systems when there are multiple sources of the surface protein gene are shown by figure 1.2. The Griffin scFv and camelid VHH libraries used within this study are examples of '3+3' phagemid systems.

Phagemids have been used extensively for the purpose of displaying antibody fragments and are also more efficient for ligation purposes and library production (Hoogenboom et al., 1991). Phagemids carry two copies of the origin of replication gene, one to facilitate the production of phagemid dsDNA and the other for ssDNA and the packaging of phage particles. Phagemids will also have a morphogenetic signalling sequence for assembly of the virions (Breitling et al., 1991). Once the bacterial host is superinfected with both the phagemid and the helper phage it will use both genomes as templates for phage production. Preventative measures are required to stop the phagemids from expressing pIII before superinfection with the helper phage can occur. Even small amounts of phage proteins have been known to prevent other filamentous phage from infecting the same cell. The presence of pIII has been shown to reduce the stability of the E. coli membrane and disrupts expression of the F' pilus and thereby preventing new infections (Boeke et al., 1982). The Lac promotor gene is commonly used to control pIII

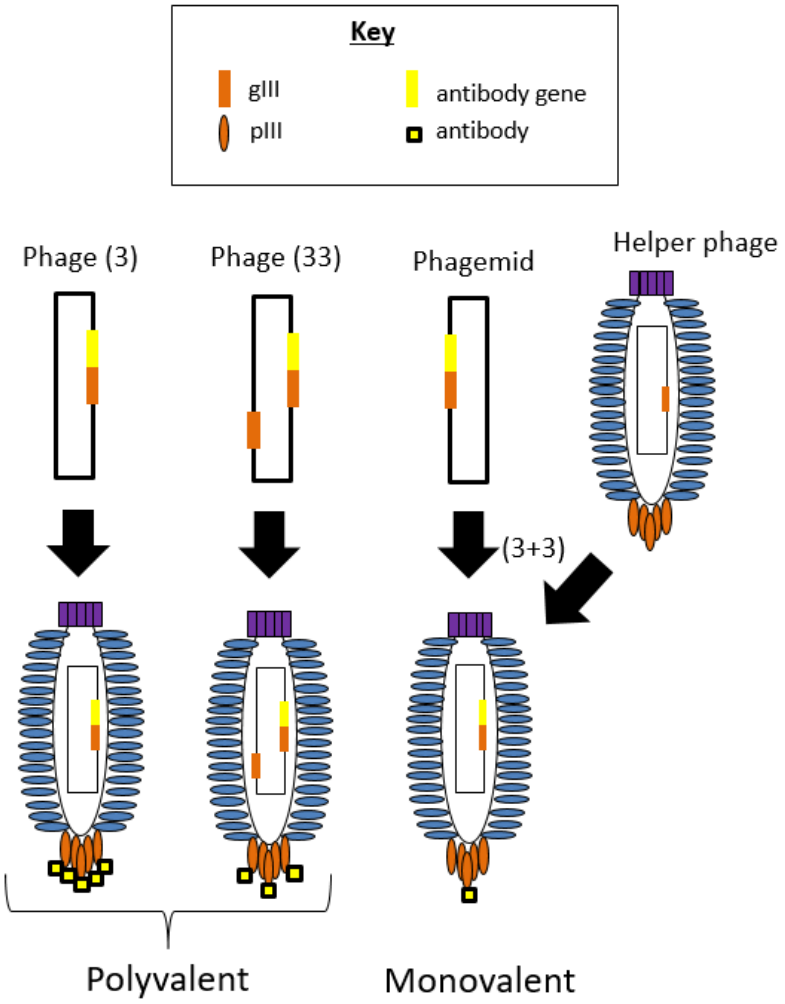


Figure 1.2: Nomenclature of phage systems using modified pIII.

When a phage system uses a vector that only has one copy of gIII, and it is modified to include the antibody gene, it is known as a '3' system. The '33' system uses a vector that possesses both a wild type gIII and a modified gIII. The phagemid vector format in combination with a helper phage is known as a '3+3' system because the two copies of gIII are from separate sources. The '3' system can only express one type of pIII whereas the '33' and '3+3' produce mosaics. Adapted from Sachdev S. Sidhu (2017).

expression for many phagemid vectors because its function is easily supressed by glucose.

Helper phage are generally modified to be less efficient at both replicating themselves and packaging to favour the incorporation of pIII-fusions and phagemid DNA into the recombinant phage. The recombinant phage is assembled more frequently around ss phagemid DNA rather than the helper phage but new virions become a mosaic of proteins acquired from both sources, the helper phage and the phagemid. This ensures that all of the

necessary genes required to produce phage are present, but also that both fusion and wild type pIII are incorporated into new phage and therefore remain infectious. Some helper phage DNA will be incorporated into recombinant phage despite the inefficient packaging. These particles are not usually selected because the incorrect resistance gene is conveyed, and any infected host cells will not be selected. A number of different helper phage have been developed for use with phage as described by table 1.1.

Table 1.1: Examples of Helper Phage in the Literature

Publication [1]	**Helper Phage** [2]
Vieira and Messing (1987a)	**M13KO7.** Contains a kanamycin resistance gene inserted in the domain B of the origin of replication. The antibiotic resistance gene enables the selection of phage infected *E. coli*.
Kramer et al. (2003)	**CT helper phage**. Prevents the amplification of phage that do not display an scFv-pIII fusion. The helper phage lacked infectivity domains N1 & N2 of PIII so only phage with PIII-scFv would be infectious.
Rondot et al. (2001)	**Hyperphage**. gIII deletion so recombinant phage could only possess scFv-pIII fusions.
Baek et al. (2002)	**Exphage.** Mutated M13KO7 derived helper phage with two amber stop codons (TAG) at the 5′ region of gIII. Reduces the expression rate of wild-type pIII thus increasing the production rate of recombinant phage displaying the protein of interest without sacrificing the incorporation of wild-type pIII.
Soltes et al. (2003)	**Pharberge**. 1 amber stop codon at 3′ end of gIII. Dramatic increase in expression ratio of Fab-pIII fusions vs wild type.
Shi et al. (2007)	The helper phage genome was incorporated into the F+*Escherichia* cells bearing a gene-III-lacking helper phage genome (VCSM13d3). The modified cells were named Hpd3cells. This enabled a high phage yield and a reduction in pIII resistance.

[1] The publication in which the helper phage was first described.
[2] The name and the advantages of each helper phage described within the cited publication.

The Ff class of M13 phage are assembled in the oxidising environment of the bacterial periplasm, facilitating the formation of disulphide bonds to

produce stable antibody structures. The virions are then extruded from the bacterial cell.

A naive antibody phage display library consists of a very large diversity (typically 10^7-10^{11}) of genetically modified bacteriophages, or phagemids, that each result in the display of a distinct antibody fragment such as a Fab, single chain variable fragment (scFv), or a single variable domain (nanobody/VHH/vNAR) (Figure 1.3) via their PIII surface protein.

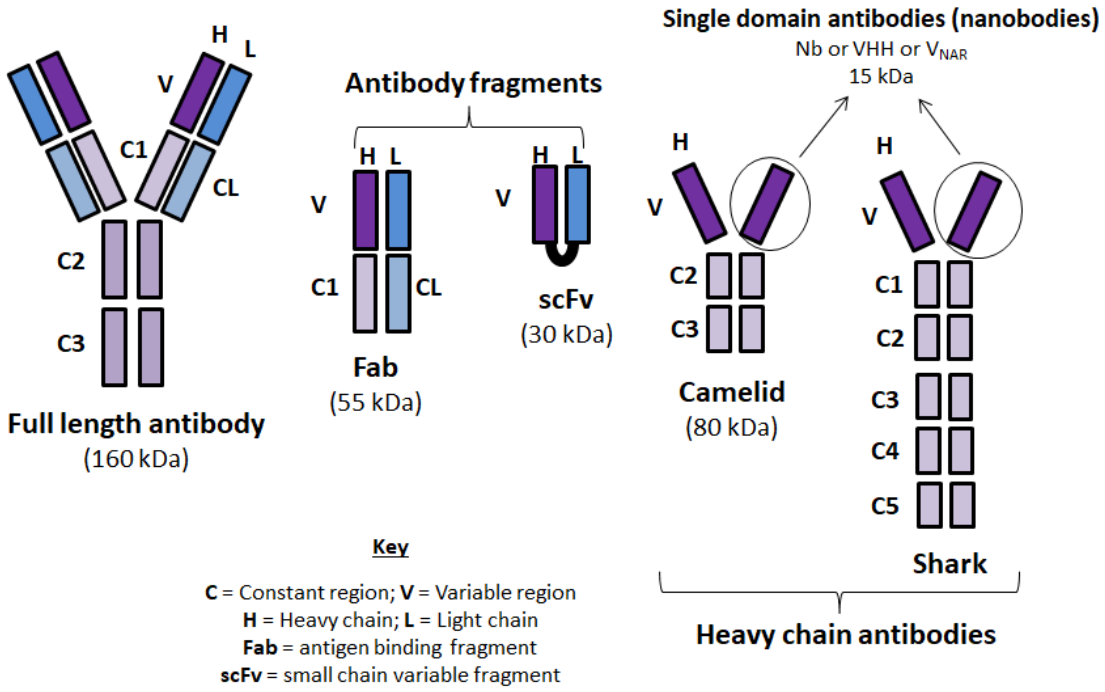


Figure 1.3: Antibody formats used in phage display.
Antibody fragments and domains commonly used in phage display. The full length antibodies are too large to display via bacteriophage and instead the antibody fragment (Fab), single chain variable fragment (scFv) or single domain (VHH/vNAR) formats are used. Their smaller sizes reduce the interference to the phage host. H= heavy, L = light, V= variable region, C = constant region. Adapted from Hammers and Stanley (2014).

The antibody fragment known as Fab consists of one constant and one variable domain of each of the heavy and light chains whereas the smaller antibody fragment, known as scFv, only consists of the heavy and light variable chains, commonly joined by a glycine-serine peptide linker (Hammers and Stanley, 2014). For scFv the variable regions can be connected in either the VH-linker-VL or VL-linker-VH orientation and these orientations can affect expression efficiency, stability, and antigen binding activity (Tsumoto et al., 1994). Although Fabs are more stable than scFv and

tend not to dimerise, their size often results in lower expression rates and a greater toxicity to the cell (Arndt et al., 2001).

Alternatively, the single domain antibodies (nanobodies) only have variable heavy chains and are typically derived from either sharks or camelids where the antigen-binding sites are known as vNAR or VHH respectively (Kovalenko et al., 2013). The variable new antigen receptor (vNAR) and VHH fragments are smaller (11—15 kDa) than both the Fabs and scFv but can still bind to a wide range of target antigens as effectively as typical antibodies and their fragmented formats, despite having fewer CDRs. Conventional antibodies possess 6 CDRs evenly distributed between the VH and VL chains, whereas VHH only have the 3 present on the VH and vNARs only possess 2. Although vNAR have similar functions to that of the other antibodies, they are not derived from the same lineages as typical antibodies, instead they are structurally similar to T-cell receptors and cell adhesion molecules suggesting a possible common ancestor molecule before becoming part of the adaptive response (Kovaleva et al., 2014). Thus, in contrast to the camelid VHH, the vNAR never had or lost a light chain but still play a vital role in the shark immune response (Kovaleva et al., 2017). The camelid VHH likely originated from a standard IgG that underwent some sort of selective pressure where a dual domain antibody was insufficient. This selection process likely resulted in the loss of the light chain (Muyldermans, 2013).

Each antibody format has been used extensively for phage display purposes with examples as cited (Wu et al., 2001, Ravn et al., 2004, Muller et al., 2012, Miyazaki et al., 2015, Lovgren et al., 2016, Turner et al., 2016, Kovaleva et al., 2017, Ubah et al., 2017, Yang et al., 2017).

Antibody genes are inserted into the phage vector or phagemid sequence at the 5’ end, ensuring they are transcribed as a fusion to the surface protein. The vector is packaged into resulting phage particles that display the antibody fragment. By providing recombinant phage with the antibody gene, this technique effectively links the antibody’s genotype to its phenotype, or binding properties. Recombinant phage will typically have one copy of the PIII-antibody fusion and 4 wild-type PIII. Multiple PIII-fusion displays are rare and a complete absence of wild-type PIII would prevent the M13-derived phage from being infectious due to the inability to interact with the bacteria’s F pilus.

Antibody phage display libraries can be naïve, immune, semi-synthetic and synthetic. Naïve and immunised libraries are produced from naturally occurring antibody gene diversity from non-immunised or immunised individuals, respectively. Alternatively, libraries either combine naturally occurring diversity with CDR sequences designed *in silico* (semi-synthetic) or are designed to be entirely synthetic CDR regions that are placed within limited numbers of "scaffold" sequences. Therefore, synthetic libraries have no naturally occurring CDR diversity within the antibody genes. The antibody diversity of a phage display library is therefore not limited by design, but instead by transformation efficiency and genetic mutations such as frameshifts during library production.

A typical phage display library will involve billions of different antibody genes at first (McCafferty et al., 1990) (Winter et al., 1994) (Griffiths and Duncan, 1998), although immune libraries can often be several fold smaller. The antibody genes, for a naïve library, are derived from an entire antibody repertoire of an immunised or naïve host, and this can include humans. Hybridoma techniques can only generate a very small number of binders against each target in comparison to phage display, and it is difficult to incorporate enrichment steps used to filter antibodies for specific functionality e.g conformation specificity. The phage can also be used to rapidly propagate the antibodies of interest, offering a considerably cheaper amplification method compared to hybridoma technology.

The phage display technique involves incubating the phage antibody library with an immobilised target, and any antibody fragments with affinity for the target can bind and any non-binders can be washed away (figure 1.4). This process is termed biopanning or panning. Immobilisation can involve adsorbing the target to solid supports such as polystyrene 96-well plates via Van-der-Waals and electrostatic interactions, or covalently, for example using maleic anhydride plates to couple via primary amines. Alternatively, phage display can be conducted in solution by adding biotin tags to the target antigens and isolating them and their phage binders using streptavidin coated magnetic beads (Zhuang et al., 2001). Naïve phage libraries can be incubated against multiple, unrelated targets and yield binders against each one independently; this is a considerable advantage over hybridoma technology

and immune display libraries. Furthermore, the targets can be antigens that would usually be toxic to animals or completely non-immunogenic.

During the panning process the removal of unbound phage antibodies is far from complete. Once the majority of non-binders have been washed away, there will still be too many different phage-antibodies that are selected but are not specificially bound to the target ligand in order to identify ligand-specific antibodies. In order to isolate specific antibody fragments, the bound phage must be eluted and re-propagated before being incubated with the target ligand again. These incubation and enrichment steps are known as panning rounds (figure 1.4). Many studies have conducted up to 4—6 rounds to enrich for high-affinity phage-antibodies that are specific for the target ligand. In theory, the antibody fragment with the greatest affinity and specificity for the target will bind to the immobilised target more frequently than those with lower specificity/affinity and thus gradually become more enriched throughout the panning rounds.

Phage display techniques are not without their pitfalls. The size of the ligand being displayed will have an impact on how effectively it is displayed, and some are simply too large to display without hindering the phage's life cycle. The rate at which recombinant phage are produced and repeatedly propagated through the panning process means that they can be prone to unintended mutations. Mutated phage can be selected for and begin to compete with other phage in the library. Phage that lack a pIII fusion, due to antibody genes not being expressed, for example those containing a stop codon, will have a growth advantage in bacteria over antibody-producing clones. In addition, antibody sequences will have a range of toxicities to the bacterial host leading to growth advantages for some phage clones and thus a reduction in overall antibody diversity despite antibody binding properties. Phage with growth advantages can become enriched over the course of multiple panning rounds and are known as non-specific or parasitic phage.

Another type of parasitic phage are those that bind to the blocking agent or the solid support used during panning, again leading to their enrichment. Phage display is therefore burdened by multiple laborious, and time-consuming, panning rounds with the possibility that non-specific phage fusions can become enriched throughout the rounds and bias the enriched sublibraries.

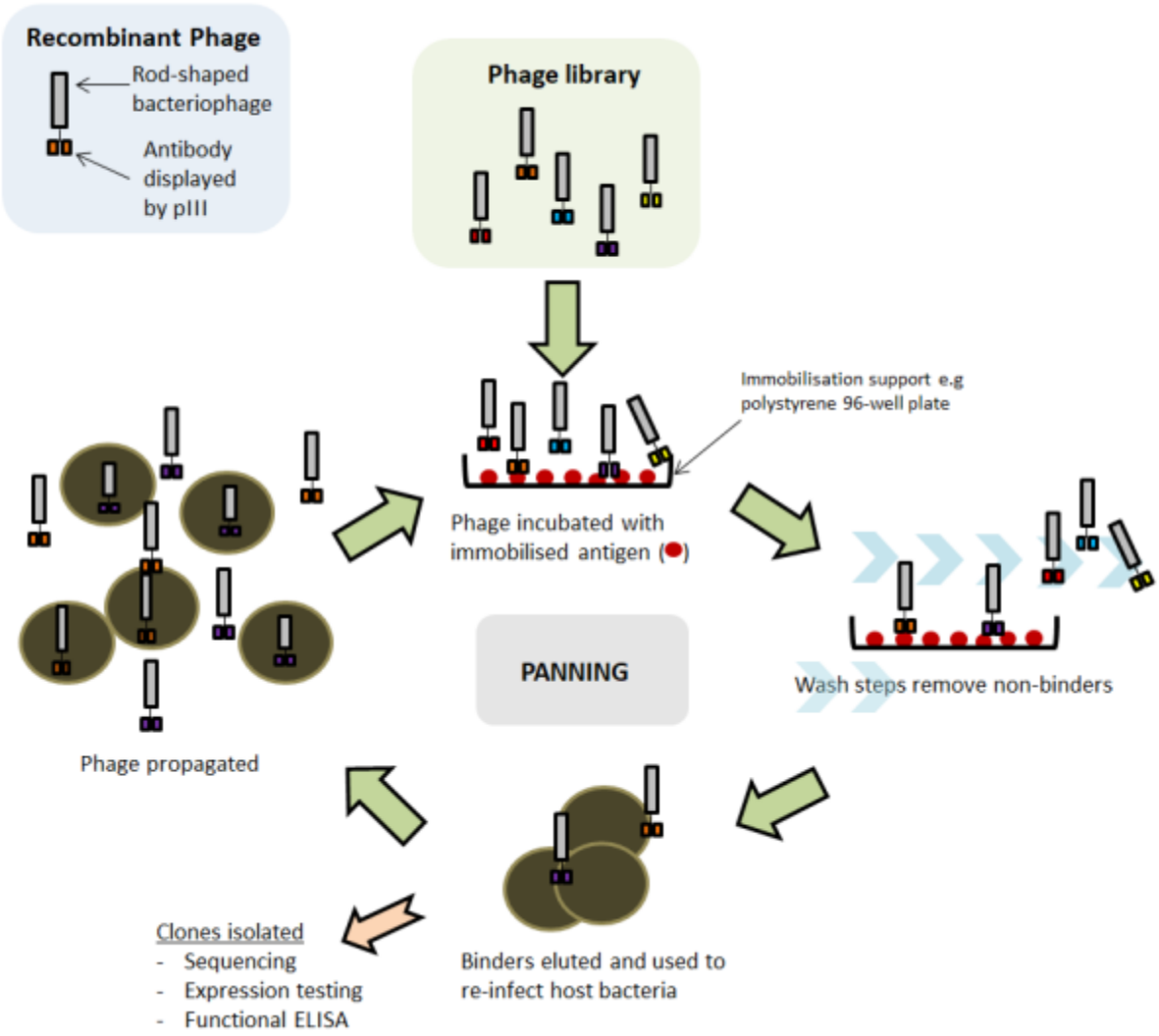


Figure 1.4: Antibody phage display panning procedure.

The antibody library (naïve or immunised) is used to produce recombinant phage. The phage are incubated with an immobilised target and non-binders are washed away after a prolonged incubation step. Bound phage are eluted (usually by a pH shift) and used to re-infect a bacterial host where they are used to create a bacterial glycerol stock and then a sub-library of phage. The new glycerol stocks serve as an enriched sub library full of potential target binders. The sub-library of phage can be used for future panning rounds, gradually producing less diverse yet more target-specific sub-libraries. Adapted from Yun et al. (2009).

Following the panning rounds, hundreds of single colonies are usually picked and propagated to produce monoclonal phage, which are then tested for specificity against the target e.g. in an ELISA. Although phage display can be useful to identify novel antibodies, those with high specificity can be rare and difficult to identify amongst a collection of parasitic clones, this can often

be the reason that panning experiments fail. The ability to identify enriched, target-specific clones in as few rounds as possible would be highly advantageous as at lower rounds the diversity of antibodies being selected is high and unrelated phage selection is unlikely to be prevalent. Random colony picking after 3—5 rounds of panning can be time consuming and laborious and such an approach is very unlikely to select specific binding clones within earlier rounds of panning. Therefore, much effort has been made to make the screening process more efficient.

High-throughput NGS has the potential to revolutionise phage display by significantly reducing the number of panning rounds required to detect enriched target specific peptides (’t Hoen et al., 2012, Ngubane et al., 2013, Rebollo et al., 2014, Naqid et al., 2016) and antibodies (Zhang et al., 2011, D'Angelo et al., 2014, Miyazaki et al., 2015). This emerging technique exploits the link between the genotype of the antibody with its phenotypic specificity by identifying those phage fusions that have become enriched beyond background levels. With NGAPD, the enriched antibody-phage fusions are identified by sequencing the heavy chain complementarity determining region 3 (HCDR3). The NGS method can then identify, in parallel, the presence and frequency of thousands to millions of antibody fragments at each round of panning. The greater sensitivity and read depth of NGS technology means that even after only one or two panning rounds highly specific phage fusions can be enriched enough within the sample to be identified. Furthermore, because the antibodies are still encoded by their bacteriophage host the clones of interest can be rescued, the antibodies expressed and tested for specificity.

The use of NGAPD is ideal for the identification of antibody fragments because it readily allows targeting and mass production of individual antibodies following data analysis.

The challenge is how to identify which antibodies are specific for the target and then how they might be rescued. Although NGAPD will allow us to sort through enriched sequence data and removed sequences with low frequency or those with stop codons, it will not allow us to sort between the enriched non-specific clones and those that are target specific (Ngubane et al., 2013, ’t Hoen et al., 2012). However, one method to try and identify target-specific phage above parasitic phage is to bind phage-antibodies to the target and a

control in parallel and then to express the relative levels of each HCDR3 as a Z score (Zhang et al., 2011). The Z score ranks each antibody sequence based on its frequency within the collection of HCDR3 sequences within the target sample, but also considers its frequency in both the target samples and the negative control and the depth of sequencing. A high Z score will be generated in the instances where an antibody-fusion has bound frequently to the immobilised target molecule, i.e. it has become enriched within the phage pool, but in the control sample, the antibody fragment will have either bound in low frequency or not at all.

Next generation sequencing platforms have been used in conjunction with several peptide phage display experiments (Matochko et al., 2012, 't Hoen et al., 2012, Ngubane et al., 2013) but not very many antibody display libraries (Zhang et al., 2011, Yang et al., 2017) unless they were for technique improvements such as clone recovery (Spiliotopoulos et al., 2015) or library quality control purposes (Glanville et al., 2009, Ravn et al., 2010b, Ravn et al., 2013).

Yang et al. (2017) compared the success of colony picking techniques to NGS analysis for the identification of potential scFv binders to prostate specific antigen. Binders were identified by both methods, but considerably more clones were identified by NGS than through colony picking methods. Although, the study did make use of pre-immunised libraries thereby, biasing the overall diversity of the scFv libraries (6.09×10^{10}, 3.64×10^{10} and 5.16×10^{10}) towards the target antigen.

It is important to bear in mind that not all antibodies are displayed or expressed well as a phage fusion. This can be a limiting factor for phage display as demonstrated by Saggy et al. (2012) who analysed an immunised murine VH VL library. In this study, the authors compared antibody identification strategies involving both a phage display format and a V gene repertoire, data mining approach. Both experiments identified target specific antibodies; however, the clones identified by data mining and phage display were poorly represented in each other's datasets. This suggests that not all antibodies seemingly enriched by the immunisation process were displayed well enough by phage to become enriched further during panning. Likewise, not all target-specific antibodies successfully expressed by the phage system were produced as high antibody titres *in vivo*. Another possibility is that

antibodies expressed by phage may have originated by PCR errors during the library construction process, rather than being originally present within the antibody repertoire and/or their specificity is generated by random heavy and light chain combinations not seen *in vivo*. As mentioned previously, phage display is known to introduce bias during the screening process, favouring non-toxic ligands that interfere less with the typical phage lifecycle, and can be expressed easily by the host bacterium. Nevertheless, in combination with an NGS approach, the phage display technology can be made even more efficient by filtering non-specific binders from the data, and identifying enriched clones that would otherwise be difficult to isolate by screening limited numbers of random clones.

1.4 Study targets

The primary target antigens of this study were the plant phosphorous storage molecule phytic acid, and the enzyme phytase. The aim was to isolate antibody fragments that would bind specifically to phytic acid, and distinguish between the active and denatured conformations of phytase.

This study used naive phage display libraries and therefore, it is more difficult to identify potential binders from the library due to the high level of antibody diversity. Immune phage antibody libraries or hybridoma technology are not suitable for identifying binders against phytic acid as it is a small (660 daltons), non-immunogenic hapten that is not readily amenable to conjugation to a carrier protein. Haptens are small molecules, generally <1000 Da, and do not elicit an immune response by themselves.

For the enzyme phytase, the objective was to identify conformation-specific antibodies and only monoclonals would be capable of consistently, binding to epitopes that were only available when the enzyme was in an active or denatured state. Polyclonal antibodies would be more likely to include antibodies that bind to common epitopes available in either state and would therefore be unable to distinguish between conformations. Again, using hybridoma technology would only produce a limited number of antibody clones and it would be difficult to screen for conformation specific binders. The antibodies that distinguish between the closely related active and denatured forms of phytase would be rare. The immune phage display

libraries would be a more suitable alternative however considerable effort and expense is required to create a bespoke immune library. Immune libraries would encounter similar problems to hybridoma when looking at conformation specific antibodies against phytase, and for phytic acid an immune library would not be possible. A naïve antibody library should be able to provide specific antibodies to the hapten phytic acid and also the two conformations of phytase.

Thus optimising the phage display NGS method using a non-immunised library was used for this study. Methodologies to identify and rescue binders easily from a non-biased library by NGAPD would be a considerable development as it would save time and resources associated with the lengthy process of conventional phage display.

1.5 Phytase & phytic acid

The enzyme phytase is responsible for the hydrolysis of phytic acid (Vats and Banerjee, 2004), or MyoInositol Hexaphosphate ($C_6H_{18}O_{24}P_6$) (IP6), an indigestible cyclic acid containing six phosphate groups surrounding an inositol ring. Phytases have been found in animals such as rats (Patwardhan, 1937), calves (Bitar and Reinhold, 1972), chickens (Cowieson et al., 2006) and pigs. They have also been found in plants, bacteria including *E. coli* (Greiner et al., 1993), *Pseudomonas* (Cosgrove et al., 1970), *I3 Bacillus* (Powar and Jagannathan, 1982, Shimizu, 1992) and *Enterobacter* (Yoon et al., 1996). Phytases are commonly found in fungi (Mullaney and Ullah, 2003) notably *Aspergillus* variants such as *A. terreus* (Yamada et al., 1968), *A. ficuum* (Gibson, 1987) and *A. niger* (Shieh and Ware, 1968).

Phytases can be classified into two groups, 3-phytases and 6-phytases, based on where within the inositol ring they initiate dephosphorylation (Figure 1.5). The 3-phytases initiate dephosphorylation at the 3rd carbon position whereas the 6-phytases target the 6th carbon position. At present, there is no evidence to suggest any difference in enzyme efficiency between these two groups of enzymes.

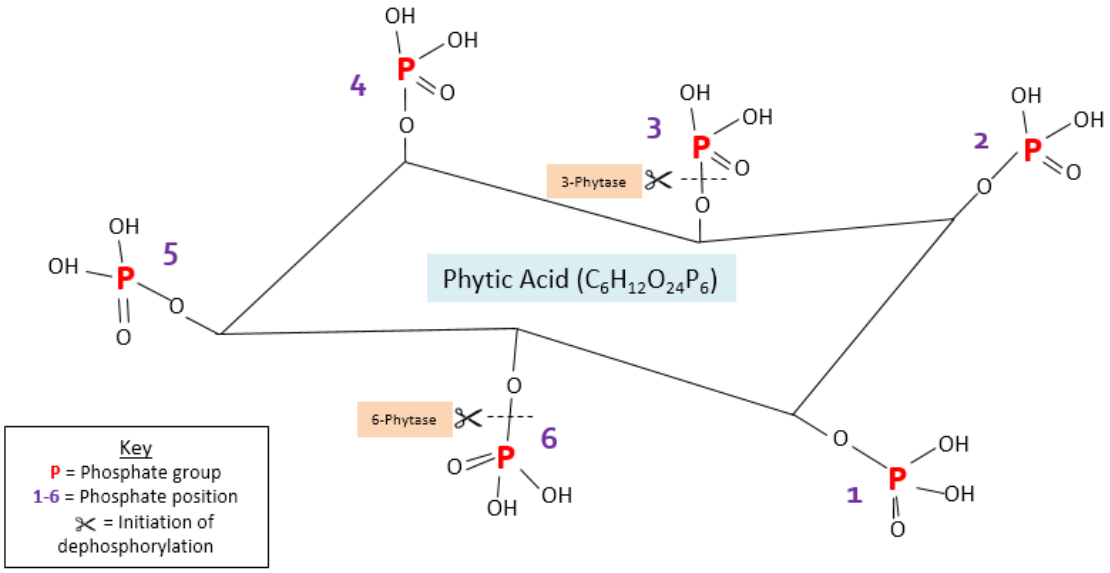


Figure 1.5: The structure of phytic acid and dephosphorylation sites
Phytic acid has 6 phosphate groups but dephosphorylation is only initiated at one of two sites depending on the class of phytase. 3-phytase and 6-phytase dephosphorylation sites are shown. Adapted from DSM.com article 'E006A', accessed April 2019.

Phytic acid is abundant in plants and serves as the primary storage form of phosphorous. Phytic acid in plants tends to exist in its salt form where it can account for up to 60—90% of the total phosphorous content (Lolas and Markakis, 1975). The salt form of phytic acid is known as phytate, and in grains, it is found most commonly as phytin, a mixed calcium-magnesium phytate. Phytin not only serves as the primary storage form of phosphorous in cereals and legumes (Cosgrove, 1980), but also acts as a source of early nourishment for developing seedlings (Loewus and Loewus, 1983) and as a means of controlling Ca2+ ion channels in guard cells (Lemtiri-Chlieh et al., 2003). In animals, phytic acid is used more commonly for trans-membrane signalling, the mobilisation of calcium and also as an extracellular neurotransmitter.

Phosphorous is an essential nutrient for all living organisms and Inositol, a carbocyclic sugar, is also beneficial as a structural component for various messenger molecules and lipids (Croze and Soulage, 2013). Phosphorous can also be supplemented in the feeds of livestock but at considerable expense, being the third most expensive feed component after energy and amino acids (Potchanakorn and Potter, 1987, Erpel et al., 2016b). Phytic acid can be a

natural source of phosphates and inositol in the presence of phytase, but in its absence the hydrolysis of phytic acid is significantly reduced. The supplementation of feed with the enzyme phytase which removes, or hydrolyses, the phosphate groups from the phytic acid molecule is much more efficient and cost effective. Feed is one of the most expensive aspects of producing livestock but the mass production of enzymes such as phytase has become an invaluable tool for maximising the utilisation of nutrients and reducing overall costs (Rathnayake et al., 2018). The use of phytase as a feed additive has a market volume of over €150 million and is one of the most commonly used enzymes in diets manufactured for monogastircs such as pigs and poultry (Powell et al., 2011, Bhavsar and Khire, 2014, Kirk et al., 2002). By maximising the availability of nutrients that are naturally present within the grains, farmers and feed manufacturers can then also reduce the supplementation of phosphorous and calcium in feeds (Selle and Ravindran, 2007).

In addition, if phytic acid remains whole it acts as a mineral sink, binding to any available cations within the digesta and making them insoluble. This mineral sink effect, in combination with the loss of a natural source of phosphates, can lead to mineral deficiencies in the animal and contribute to eutrophication in the environment.

The hydrolysis of phosphate groups from phytic acid (IP6) results in smaller catabolites of phytic acid including IP5, IP4, IP3, IP2, and IP1 until only the cyclic inositol remains with six free phosphates. The final phosphate group is particularly resistant to phytase activity and will usually only be released after prolonged incubation with phytase at a high concentration (Cosgrove et al., 1970) or in the presence of an alkaline phosphatase.

It has been demonstrated that bacterial phytase will prefer to hydrolyse the phytic acid forms with the greatest number of phosphate groups such as IP6 or IP5 (Wyss et al., 1999, Greiner and Farouk, 2007). However, the subsequent accumulation of smaller phytic acid esters can also become problematic if left undigested, as one anti-nutrient is simply replaced by another. Albeit lower molecular weight esters will chelate with available cations less readily and with a lower capacity than the larger esters (Cowieson et al., 2011).

Unlike monogastric animals, ruminants possess gut bacteria that synthesise phytase and secrete the enzyme into the rumen. The secreted phytase is then responsible for the efficient processing of phytic acid.

Monogastrics, such as pigs and poultry, can produce their own phytase but it is simply too inefficient to catabolise the vast amounts of phytic acid within their diet, nor do they possess gut bacteria that will do it for them.

Intact phytic acid has a negative, knock-on effect on other proteins and nutrients (Cowieson et al., 2004). The acidity of phytic acid can lower the pH in the stomach, effecting the activation of the protease pepsin. This causes a chain reaction during which the proteases are less effective in the stomach, which may lead to a lower liberation of polypeptides and a greater quantity of undigested proteins. Subsequently, this leads to additional secretions of hydrochloric acid (HCl) and pepsin's precursor, which contributes towards an increased amount of undigested proteins within the stomach and an increasingly acidic digesta. Upon reaching the intestines, a greater volume of sodium bicarbonate is then required to return the pH to neutral which then negatively impacts the absorption of amino acids (Ravindran et al., 2008, Cowieson et al., 2008).

Phytic acid is a negatively charged molecule (Costello, 1976, Cosgrove, 1966), and as it moves from the acidity of the stomach to the near neutral conditions of the intestines, it becomes increasingly negatively charged. As mentioned previously, it can act as a chelating agent, binding to any available cations such as zinc, iron, calcium and magnesium to form phytates. The phytic acid will also bind to positively charged proteins, rendering them insoluble and resistant to enzyme activity (Dersjant-Li et al., 2015). This effect includes those proteins that are secreted into the intestines, and because a significant proportion of available nitrogen is invested in an animal's own protein secretions it is essential that they are reabsorbed post-function (Cowieson et al., 2009).

Attempts to reduce phytate content of food has involved soaking, germination/malting, cooking, fermentation, phytase supplementation (Kumar et al., 2010) and genetically altering crops and even livestock in order to offset the mineral sink effects (Cerino Badone et al., 2012). A transgenic algae (*Chlamydomonas reinhardtii*) has also been developed that expresses a phytase suitable for the gastrointestinal environment/pH and retains

stability during the pelleting process of producing feeds. It is cheap to produce and can be used as a feed additive without the need for protein purification (Erpel et al., 2016a). Transgenic pigs have also been produced that secrete phytase in their saliva. The study suggested that the transgenic animals required little to no inorganic phosphate supplementation and excreted up to 75% less faecal phosphorous than non-transgenic pigs (Golovan et al., 2001).

It is unsurprising then that the cheap supplementation of phytase in feedstuffs with a high phytic acid content is becoming more common. It is crucial that enough phytase is provided to ensure that all of the phytic acid is dephosphorylated to the less malignant, smaller molecular weight esters within the stomach or gizzard. Phytic acid digestibility in the absence of phytase will depend on the concentrations of other phosphatases and luminal cations, but feed supplemented phytase allows the amount of available phosphorous and inositol to be maximised (Cowieson et al., 2011).

One unit of phytase activity (FTU) is defined as 'an amount that will hydrolyse one micromole of inorganic phosphorous per minute from an excess of sodium phytate at 37°C at a pH of 5.5'. Feed formulations for poultry or swine that include phytases often contain around 500 FTU/kg of the enzyme (Lei et al., 2013), however much larger doses of phytase have also been tested in an effort to break down all of the phytic acid present within the feed and increase phosphate availability. A number of studies have investigated the effects of superdosing phytase, ranging from 2,500 — 12,000 FTU/kg. Each study has consistently demonstrated a significant improvement in the hydrolysis of phytate in both swine and poultry (Simons et al., 1990, Zhang et al., 2000, Shirley and Edwards, 2003, Augspurger and Baker, 2004, Cowieson et al., 2006, Kies et al., 2006, Zeng et al., 2014). Significant efforts have been made to maximise the efficiency of phytase production (Erpel et al., 2016b) and stability of phytases. Efforts have been made to determine optimal growth media for phytase-producing microorganisms (Sasirekha et al., 2012), characterising phytases (Rodriguez et al., 1999, Igbasan et al., 2000, Shim and Oh, 2012, Kalsi et al., 2016) and isolating novel phytases with more desirable properties (Kim et al., 1998, Shi et al., 2008) such as thermostability, protease resistance and wider optimal pH spectrums. At present, when ~500 FTU/kg of microbial phytase is supplemented, over 50% of the phytate will be hydrolysed once the digesta

reaches the end of the digestive tract. A phytase that can retain activity under true *in vivo* conditions would be able to hydrolyse phytate much faster. Although the inclusion of wheat in a feed sample can provide a natural source of phytase, an *E. coli* derived phytase is often supplemented instead because it is much more efficient (Phillippy, 1999, Rapp et al., 2001).

Feed compositions can vary considerably depending on the manufacturer and purpose of the feed for example maintainence, growth, meat production or for pregnant and lactating animals. It is therefore important to know how much phytic acid there is within a feed sample so that an adequate amount of phytase can be supplemented. A cheap assay to detect and quantify phytic acid would be invaluable to the animal feed industry. The phytic acid content also changes significantly between feed compositions so the ability to quickly and accurately measure the phytic acid content of numerous feed samples would be of significant value. Existing quantification methods such as high performance liquid chromatography (HPLC) can be sensitive, adaptable and be used to determine the amount of phytic acid in a variety of feedstuffs (Cosgrove, 1980, Guse et al., 1995) (Talamond et al., 2000, Chen and Li, 2003, Dost and Tokul, 2006). Unfortunately, HPLC analysis is also time consuming, expensive and ill-suited for the high-throughput requirements of the feed industry. More recently, Near Infrared Spectrometry (NIR) has been used to rapidly predict phytate levels in feed compositions (Tahir et al., 2012, Pande and Mishra, 2015). Although NIR is fast and amenable to high-throughput it still relies on calibrations that can require frequent updates and the resulting peaks can be widely variable and sometimes overlap, making them difficult to assign to specific feed components.

It is also important that the phytase that is supplemented not only remains active after a feed has been heat-treated as part of the pelleting process, but also while it is exposed to the hostile environment of the stomach, maintaining resistance to both the acidic pH and also the resident proteases such as pepsin. A number of phytases have been identified or modified that possess more stable characteristics making them suitable as feed additives (Kim et al., 1998, Rodriguez et al., 1999, Igbasan et al., 2000, Shim and Oh, 2012). Although a range of phytases are available it can be time consuming to test the activity of each enzyme under different conditions, for example

those mimicking gastric environments. It can become tedious and time consuming to constantly establish an activity curve for each phytase.

Determining whether phytase is active would usually involve a colormetric assay that detects the release of phosphates, a frequently modified method based on the original experiments by Fiske and Subbarow (1925). Furthermore, the Association of Official Analytical Chemists (AOAC), or AOAC International, method for measuring phytase activity is also conducted at pH 5.5, which does not mimic the pH of the gastric environment. Care must be taken when selecting a commercially available phytase to ensure that it will retain activity *in vivo*. In addition, activity assessments such as these can often be inaccurate due to the presence of contaminating phosphates, phosphatases and competing smaller phytate esters.

A rapid assay that can determine whether a phytase is functional would therefore be of great value to the manufacturer, and industry if it were to bind to multiple phytases. Again the existing methods are less than ideal; they rely on substrate hydrolysis as an indicator for enzyme activity and can often be inaccurate due to the matrix effects of the feed. The ability to determine whether a phytase is active or inactive could be achieved by using antibodies that will only bind to the active form of phytase, for example that bind against a conformational epitope only present when the enzyme is active and in its native conformation.

In conclusion, the present study aimed to develop next generation antibody phage display (NGAPD) strategies for the isolation of target specific antibodies. As model targets, this study uses the hapten phytic acid and the protein phytase. In doing so, the study aims to isolate antibodies that would have the potential to be developed into novel immunoassays to quantify phytic acid in feeds, and distinguish between the active and denatured conformations of phytase (AB Vista's Quantum Blue).

Chapter 2: Materials and Methods

2.1 Materials

Table 2.1: Antibodies used throughout the study

Antibody [1]	**Dilution Factor** [2]
Anti-X6HIS-HRP (Product code: MA1-135-HRP, Thermo Scientific)	1:500
Anti-TetraHis (Product code: 34670, Qiagen)	1:2000
Anti-Mouse-HRP (Product code: A16072, Invitrogen)	1:2000
Anti-Mouse-HRP (Product code: P0447, Dako)	1:2000
Anti-PIII (Product Code: PSKAN3, 2B Scientific)	1:4000
Anti-PIII (Product Code: E8033S, New England Biolabs)	1:1000
Anti-M13-HRP (Product code: 27-9421-01, GE Healthcare)	1:5000
Anti-MBP (Product code: E8032S, New England Biolabs)	1:10000

[1] List of antibodies and their manufacturers/product codes used during the study.
[2] Dilution factor used for each antibody to produce a working solution. Antibodies were diluted into 3% (w/v) skimmed milk powder/Marvel dissolved in 1X PBS or 1M Tris Buffer (pH 7.5).

Table 2.2: Bacterial strains used within the study

***E. coli* Strain** [1]	**Genotype** [2]	**Used for** [3]
TG1	supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5 (rK – mK –) [F´ traD36 proAB lacIq ZΔM15].	phagemid propagation/PIII-fusion expression
HB2151	K12, ara, (lac-pro), thi/F' proA+B+, laclq lacZM15)	soluble antibody expression
ER2523	fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10-- TetS) endA1 (mcrC-mrr)114::IS10	recombinant protein expression
T-Vector: Ecoli Top10 cells	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG	TOPO® TA Cloning® Kit

[1] List of *E. coli* strains used throughout the study.
[2] Genotype of each *E. coli* strain concerning protease gene omissions or amber codon interactions.
[3] What each *E. coli* strain was used for throughout the study.

Amber suppressor strains (supE genotype) such as TG1 will prevent amber stop codons (TAG) expression, and instead substitute a glutamine (G) to allow read through of the affected gene. In this study, the amber suppressor strains allowed the PIII protein to be synthesised along with the N-terminal scFv/VHH as a fusion protein because the stop codon before PIII is substituted with a Q.

In non-suppressor strains such as HB2151 the amber stop codon terminates translation as normal. This causes the scFv/VHH proteins to be synthesised without the pIII.

A range of different vectors were used throughout the study to express antibody fragments. Each vector had different fusions and tag combinations and orientations as shown by table 2.3.

Table 2.3: The tag orientation in each vector and antibody combination used within the study.

Vector & Antibody [1]	6x HIS tag orientation [2]
pHEN2 scFv	scFv-x6HIS-Myc-(pIII)
pSD3 VHH	VHH-StreptaggII-6xHIS-(pIII)
pET22B VHH	VHH-Strep-taggII-6xHIS
p5X VHH	MBP-VHH-StreptaggII-6xHIS
p5X scFv	MBP-scFv (digested before 6x HIS)

[1] The antibody format and vector combination. Shaded cells are for phage display fusions. Non-shaded for 'soluble' antibody expression.
[2] The orientation of the 6X HIS tag from the N terminus to C terminus (left to right). The pIII protein was optional depending on whether the antibody was expressed in an amber suppressor strain of *E. coli* or non suppressor. Depending on the vector or design of the antibody insert, various tags (Myc, StreptaggII) or protein fusions (Maltose Binding Protein 'MBP') were included in addition to the x6 HIS tag. scFv from the pHEN2 vector were restriction-digested upstream from the 6x HIS tag before being inserted into the p5X vector.

2.2 SDS PAGE of protein samples

Various protein samples throughout the study were investigated by running them on a denaturing SDS-PAGE gel. These experiments facilitated the analysis of the protein's molecular weights, level of degredation and downstream applications such as western blotting. Protein samples were mixed with 2X LDS buffer/β-Mercaptoethanol (Aldrich M6250) before being loaded onto a pre-cast NuPAGE® Novex® 12% Bis-Tris protein gel (Fisher NP0342). Protein standards were also included on the gel such as 5µl of SeeBlue Plus2 (Invitrogen; LC5925) which was compatible with both Bis-Tris and Tris-Glycine gels. The SeeBlue Plus2 contained 10 pre-reduced polypeptides in the range of 4–250 kDa depending upon the buffer system.

The protein gel wells were pre-washed with 1X MOPs running buffer (Invitrogen NP0001-02) before the samples were loaded and the same buffer

was used during electrophoresis. The protein gel was run at 200 volts for 45 minutes.

The gel was then stained with Commassie blue 250-R (Fisher BP101-25) (45% methanol, 45% dH_2O, 10% glacial acetic acid) overnight on a gently rocking platform. The following day, the gels were de-stained (40% methanol, 50% dH_2O, 10% glacial acetic acid v/v) repeatedly and visualised using a white light trans-illuminator.

2.3 Western Blot

Proteins of interest were detected by western blot throughout the study. Protein samples were loaded onto a SDS-PAGE gel as described above. The protein marker Magic Mark XP (Invitrogen LC5603) was used as a standard, and contained recombinant proteins with IgG binding sites ranging between 20-220 kDa in weight. The marker was therefore compatible with any primary or secondary antibody used to detect the target. After the protein gel had finished running, the gel was separated from the plastic casing and gently rinsed with 1X transfer buffer (Life Tech. NP0006). The gel was then sandwiched between two pieces of filter paper and a PVDF transfer membrane. The PVDF transfer membrane was pre-treated with 100% methanol and soaked in transfer buffer alongside the two pieces of filter paper before coming into contact with the protein gel.

The membrane and filter papers were wedged between sponges soaked in 1X transfer buffer and loaded into a western blot tank. The compartment with the sponges and filter paper was filled with 1X transfer buffer and the surrounding compartment in dH_2O to act as a heat sink. The transfer was run at 30 volts for 75 minutes.

After the transfer was complete, the PVDF membrane was rinsed with sterile dH_2O and blocked for 1 hour in 3% (w/v) skimmed milk powder (Marvel) in 1X PBS or 1M Tris Buffer (pH 7.5) on a gently rocking platform.

Antibodies were diluted to produce 5—40ml of a working concentration using 3% (w/v) Marvel in 1X PBS or 1M Tris Buffer (pH 7.5). The blocked PVDF membrane was sealed in plastic with 5 ml of the working antibody solution or incubated with 40 ml in a square petri dish. The membrane was incubated for 1 hour with agitation.

After each step of probing the blot with antibodies, the membrane was removed from the sealed plastic or petri dish and submerged in 1X PBS or 1M Tris Buffer (pH 7.5) with 3% (w/v) Marvel and containing 0.1% (v/v) Tween 20, refreshing the buffer every 5 minutes for 15 minutes.

Following incubation with secondary-HRP antibody and additional washing, the membrane was rinsed with dH_2O before being incubated with the HRP substrate enhanced chemiluminescence (ECL) (Thermo Scientific Pierce; 32106) The substrate was prepared by following the manufacturers instructions (1.5 ml). The 1.5ml of ECL substrate was pipetted over the membrane repeatedly for 2 minutes before being re-sealed in plastic and visualised using BioRad Chemi Doc.

2.4 Determining the protein concentration

Protein concentrations were predicted using a Bradford Assay Kit (Bio-Rad 5000202). The kit provided a 1X dye reagent and seven pre-made bovine serum albumin (BSA) standards at 0.125, 0.25, 0.5, 0.75, 1.0, 1.5 and 2mg/ml. Absorbance values were determined for each BSA standard at 595nm using a Biochrom Spectrophotometer (Libra S22; 190-1100nm), 1.5ml cuvettes and a 1cm light path to generate a standard curve. The unknown concentrations of the protein samples were predicted based on the BSA standard curve using neat and diluted samples.

2.5 Indirect measurement of Phytase activity

A number of phytase activity assays were conducted using a phytic acid Total Phosphorous kit (Megazyme Product No. K-PHYT). Briefly, the kit provided it's own phytase enzyme, an alkaline phosphatase enzyme, free phosphate standards, two buffers (acidic and alkaline) and an oat powder to be used as a source of phytic acid.

From the oat powder, the phytic acid was extracted by incubating 1g of oat powder with 20 ml of hydrochloric acid (0.66M) overnight, mixing vigorously at room temperature. The following day, 1ml of the extract was transferred into a 1.5 ml eppendorf tube and centrifuged at 17,000 xg for 10 minutes. Half of the supernatant was transferred to a new eppendorf tube and neutralised with 0.5ml of sodium hydroxide (0.75M).

For each source and weight of phytic acid, a control sample was included which was not incubated with phytase or alkaline phosphatase. These samples reflected the background level of free phosphorous present in every sample that used the same phytic acid source and weight.

For each sample, 50 µl of the phytic acid source was incubated with 20 µl of phytase, 200 µl of the acidic buffer and 600 µl of dH_2O for 10 minutes at 40°C. During the incubation period, the phytase was expected to release phosphates from the phytic acid molecule, thereby increasing the amount of free phosphorous within the solution in comparison to a no-enzyme control (free phosphate negative control). For free phosphate control samples, the enzyme volumes were replaced by additional dH_2O.

After incubation, 200 µl of the alkaline buffer was added to all of the samples and either 20 µl of water (for no-enzyme controls) or alkaline phosphatase solution (samples). All samples were incubated for a further 15 minutes at 40°C. The reaction was stopped with the addition of 300 µl of 50% trichloroacetic acid. Samples were centrifuged for 10 minutes at 17,000 xg and 1 ml of the supernatant was incubated with 500 µl of colour reagent (ascorbic acid, Sigma 95210; ammonium molybdate Sigma 31402) at 40°C for 1 hour. Free phosphate groups reacted with ammonium molybdate to become molybdenum blue, the intensity of which is directly proportional to the amount of free phosphates within the sample.

After 1 hour, the absorbance of each sample was measured at 655 nm using a Biochrom Spectrophotometer (Libra S22; 190—1100nm). For every sample, 1ml of solution was loaded into a 1.5ml cuvette and measured with a 1cm light path. The absorbance values were used as a measure of enzyme activity.

Throughout the activity assays there were a number of modifications to the standard protocol.

- All of the colour reagents and phytic acid sources were replenished daily.
- In the intermediate and later activity assays, the enzyme alkaline phosphatase was removed from the assay and was replaced with water due to its contribution towards free phosphates.

- Although the kit provided its own source of phytase and phytic acid, as the activity assays progressed these were replaced with the same volumes of AB Vista's Quantum Blue phytase and Sigma's phytic acid sodium salt (sodium phytate) in order to optimise the assay for the intended panning reagents.
- Quantum Blue phytase was used either neat or diluted with dH_2O.
- Sigma's sodium phytate was prepared similarly to the original protocol by dissolving 1g in 20 ml dH_2O, transferring 1 ml into a 1.5 ml eppendorf tube and centrifuging at 17,000xg for 10 minutes. Half of the supernatant was transferred into a new eppendorf tube and diluted with 500 µl of dH_2O.

2.6 Removal of N-Linked Glycans using PNGase F

Proteins can undergo hundreds of different post-translational modifications causing molecular weight shifts and alterations to the protein's physical properties.

In eukaryotes, Glycans are commonly attached to nitrogen atoms of an asparagine (Asn) side chain when part of a particular amino acid sequence e: Asn-X-Ser, Asn-X-Thr and sometimes Asn-X-Cys, where X can be any other amino acid except for Proline. Most of the proteins secreted by fungi are thought to be glycosylated so the Quantum Blue phytase amino acid sequence (data not shown) was analysed, and was found to have only two such combinations that were potentially glycosylated.

Those sequences enable Asn to be accessible within the three dimensional structure of the protein so that N-linked glycosylation can be initiated. Furthermore, the addition of glycans can also interfere with sequencing or mass spectrometry procedures and need to be removed.

De-glycosylation can be conducted under native or denaturing conditions using a PNGase F amidase.

For denaturing conditions, 10 µg of the potentially glycosylated Quantum Blue phytase protein was mixed with 1 µl of glycoprotein denaturing buffer (10X) (NEB; B1704S) and distilled water to make a total volume of 10 µl. The mixed sample was heated to 100°C for 10 minutes and then chilled on ice. To remove glycans, 6 µl of distilled water, 2 µl of GlycoBuffer2 (10X) (NEB;

B3704S) and 2 µl of 10% NP-40 (NEB; B2704S) was mixed with the denatured sample to make the total sample volume 20 µl. The enzyme PNGase F (NEB; P07055), 1 µl or 500U, was mixed with the sample and incubated for 1 hour at 37°C.

De-glycosylation can also be conducted under native conditions by combining 10 µg of the potentially glycosylated Quantum Blue phytase protein with 2 µl of glycoBuffer2 (10X) and distilled water to make a total volume of 20 µl. After mixing, 2—5 µl or 1,000—2,500U of the enzyme PNGase F was added with the sample and incubated for 1 hour to overnight at 37°C. After de-glycosylation, the protein's molecular weight may visibly shift when re-analysed on a protein gel.

Samples were analysed using a protein gel and commassie blue 250-R.

2.7 Standard recipes used for phage display

Minimal agar

Minimal agar plates were used to select for TG1 or HB2151 colonies that were expressing their F' pilus to facilitate infection by bacteriophage. The F pilus is a special kind of tunnel protein with the diameter roughly the size of dsDNA. When a bacterial cell with an F pilus encounters another cell, the pilus is used to form a bridge between the two cells for conjugation purposes. Bacteriophage also use this structure to inject their DNA into the new host. The *E. coli* strains TG1 and HB2151 used in this study are thiamine deficient and in the absence of an F' pilus, the strains are also proline deficient. The TG1 and HB2151 F' episome contains the proA and proB genes that encode enzymes responsible for proline synthesis. By adding only thiamine to the minimal agar, the lack of proline in the media ensured that only cells expressing their F' pilus would produce colonies.

- 1.5g Agar powder was mixed into 75ml H_2O and sterilised by autoclaving
- The solution was allowed to cool to 50^{O}C before adding:
    - 25 ml 5x M9 salts
    - 1.6 ml 25% glucose
    - 100 µl l $MgCl_2$
    - 50 ul Thiamine Hydrochloride

5X M9 salts

A primary component for minimal agar plates used to select for bacteria expressing an F' pilus.

- Di-sodium Hydrogen Phosphate (Anhydrous) 15 g Na_2HPO_4 (Fisher 5/4520/53)
- 7.5 g KH2PO4 Potassium Phosphate Monobasic (Sigma P9791-500g)
- 2.5 g NH4Cl Ammonium Chloride (Fisher A/3880/53)
- 1.25 g NaCl Sodium Chloride (Fisher 5/3160/60)
- Distilled water to 500ml
- Autoclaved to sterilise

2YT (500ml)

Rich media used to culture bacterial stocks throughout the study, as recommended by the phagemid library manuals.

- 8 g Tryptone
- 5 g Yeast
- 2.5 g Sodium chloride (NaCl)
- Distilled water to 500ml
- Autoclaved to sterilise

2YT agar was made by adding 7.5g of agar powder.

2.8 Growing the Griffin scFv and VHH phagemid libraries

Phage display was conducted using either the Griffin phagemid (pHEN2) library (semi-synthetic human VH + VL scFv library) (Griffiths et al., 1994), or an in-house, naïve and synthetic camelid VHH library (Yan et al., 2014) using a single-framework, pSD3 phagemid vector.

The Griffin scFv and VHH libraries both displayed their antibody fragments via the N-terminal of the pIII surface protein. The VHH library had a synthetically randomised HCDR3, using NNK codons, that was created with three different lengths (12mer, 16mer or 21mer) producing three VHH-pIII sub-libraries. Both phagemid vectors (figure 2.1 and figure 2.2) contained an amber stop codon in between the antibody gene and gIII to allow soluble antibody expression in non-supressor cell lines. Furthermore, the Griffin and VHH library phagemids both possessed an ampR gene encoding β-lactamase, conferring ampicillin resistance to the bacterial host. The phagemid pelB leader sequences facilitated the export of the antibody-PIII fusions to the periplasm, where they were folded and incorporated into recombinant virions.

In addition, both vectors contained a ribosome binding site, an M13 origin of replication gene and a phagemid origin of replication gene (ColE1 in Griffin pHEN2 and p15A in VHH pSD3). Griffin scFv antibody fragments are produced with MYC (EQKLISEEDL) and 6X HIS (HHHHHH) tags and a full length copy of the PIII gene is used. The camelid VHH antibody fragments are produced with a Strep-Tag II (WSHPQFEK) and 6X HIS (HHHHHH) tag and used a truncated PIII (30 kDa) format.

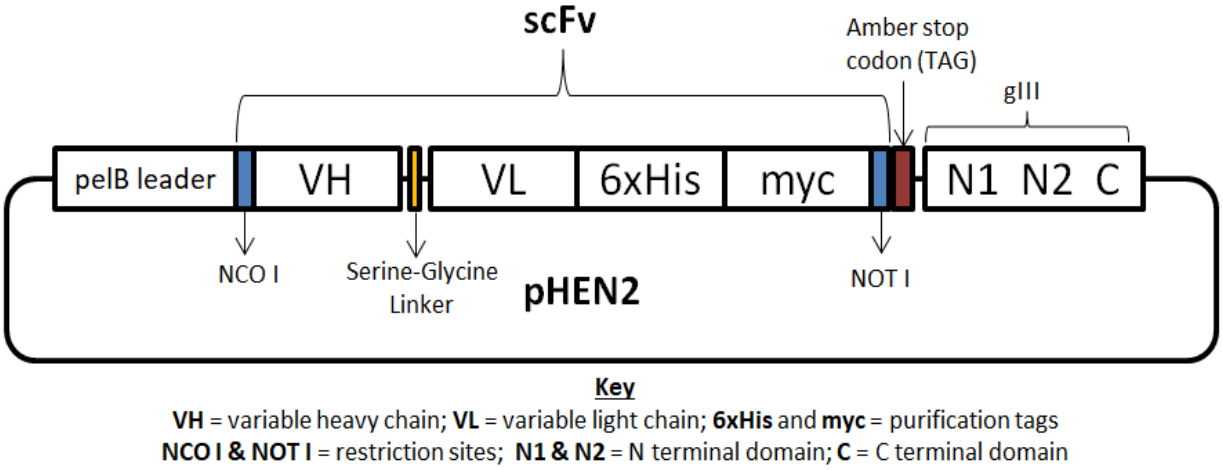


Figure 2.1: Griffin scFv phagemid vector pHEN2. The Griffin human scFv library used the pHEN2 phagemid vector which conveyed ampicillin resistance to the host. The scFv gene was downstream of the pelB leader signalling sequence and flanked by NCO1 and NOT1 restriction sites. The variable heavy (VH) and variable light (VH) chains were separated by a serine-glycine linker sequence. Downstream of the scFv gene was a 6X HIS and myc tag for purification and detection purposes followed by an amber stop codon (TAG) and a full length copy of the bacteriophage pIII minor coat protein. Adapted from Griffiths et al. (1994).

Phage were produced using the helper phage known as Exphage (Baek et al., 2002). Exphage are an M13KO7 derived helper phage, and therefore possess a mutated M13 origin of replication (Vieira and Messing, 1987b).

The helper phage provided its hosts with kanamycin resistance and all of the necessary genes to produce new virions; replication proteins pII and pX, single stranded DNA binding protein pV, major structural protein pVIII, minor structural proteins pIII, pVI, pVII, pIX and morphogenic proteins pI, pIV, pXI.

Exphage was engineered to possess two amber stop codons (TAG) at the 5’ end of the gIII substituting two glutamic acid residues. The mutated origin gene and presence of two stop codons within the helper phage’s gIII gene mean that helper phage package less efficiently overall due to its mutated ori

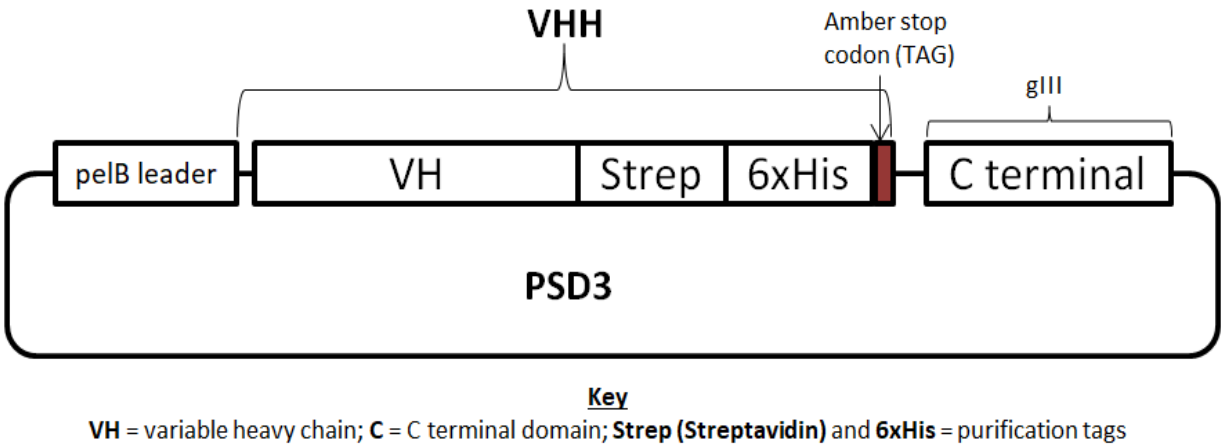


Figure 2.2: VHH phagemid vector pSD3. The camelid VHH library was constructed in the pSD3 phagemid vector with a synthetically randomised HCDR3; 12, 16 or 21 amino acids long. The vector conveyed ampicillin resistance to the host. The VHH gene was downstream of the pelB leader signalling sequence and at the 3' end of the VHH gene was a Strep-tag II and 6X HIS tag for purification and identification purposes. Just after the 6X HIS tag was an amber stop codon (TAG) followed by a truncated copy of the bacteriophage minor coat protein pIII. Adapted from Griffiths et al. (1994).

gene, and the expression of the wild type gIII protein is also lower because it is reliant on an infrequent read-through error, therefore producing a bias towards scFv/VHH-gIII fusion display from the phagemid. Recombinant phage were therefore more likely to display N-terminal scFv or VHH fusions via their pIII, and this reduces the occurrence of 'bald phage'. Superinfection of the host bacterium was mediated with the use of glucose, a LAC repressor, which prevents the expression of phagemid gIII-scFv/VHH fusion to allow superinfection by the helper phage (Krebber et al., 1996), and consequently the provision of kanamycin resistance.

The phage display libraries were initially grown in 500ml 2YT cultures containing 150 µg/ml ampicillin and 1% (w/v) glucose in a shaking incubator at 37°C. Once the optical density (OD_{600nm}) of the culture reached an absorbance value from 0.4—0.6, indicating ~$4x10^8$ cells, the helper phage Exphage was added at a multiplicity of infection (MOI) of 10. The helper phage was left to infect the culture for a total of 60 minutes, being static for 30 minutes then shaking for the remaining 30 minutes at 37°C. The culture was centrifuged for 10 minutes at 3000xg and the pellet was re-suspended in 500 ml 2YT containing 150 µg/ml ampicillin and 50 µg/ml kanamycin. The culture was left overnight shaking at 30°C to produce recombinant phagemid particles.

2.9 Phage Precipitation

Initial library cultures were acid purified to assist with purification of the recombinant phage and minimise unwanted cell debris within subsequent panning experiments.

The overnight culture was centrifuged for 20 minutes at 3000xg and the supernatant containing the phage was collected. The phage were acid precipitated by lowering the pH to 4.2 using HCl. The phage supernatant was left on ice for 1 hour before being centrifuged for 30 minutes at 8000xg. The pellet was re-suspended in 10 ml of 1X PBS and the pH was returned to 7—7.4 using sodium hydroxide (2 M). The resulting phage suspension was then transferred into sterile eppendorf tubes and centrifuged for 2 minutes at 12,000xg to remove bacterial debris. The supernatant was transferred to a sterile container and a small aliquot was used to titre the phage and obtain an approximate phage count using serial dilutions of the phage and colony counts. The phage were stored at 4°C and used for NGAPD experiments within a week of precipitation. For the camelid VHH library, the three sublibraries (12mer, 16mer and 21mer) were grown separately and the phage were acid precipitated. The three libraries were titred and then combined in equal concentrations and volumes before panning.

2.10 Antigen preparation

The study focused on three target antigens; the active conformation of the enzyme phytase, the denatured conformation of the enzyme phytase and the hapten, phytic acid.

An aliquot of Quantum Blue phytase was kindly supplied by AB Vista. This antigen is an *E. coli-derived*, highly thermostable and commercially available phytase produced in a *Trichoderma* fungal system for use as a feed additive. All subsequent dilutions of the Quantum Blue phytase were done using 1X PBS buffer unless the phytase was intended for use in Megazyme's enzyme activity assays, in which case it was diluted with dH_2O.

Phytic acid was purchased from Sigma as a sodium salt derived from rice (Sigma; 8810). The phytic acid solutions (0.125M pH 1 or 1.25M pH 7) were freshly prepared on the day of use by dissolving the powder in distilled water

and adjusting the pH to 1 using 1M Perchloric acid. The solution was then stored at 4°C until use.

2.11 Phytase immobilisation

During chapters 3 and 4, Quantum Blue phytase was immobilised overnight at 18°C on Nunc MaxiSorp® plates (Thermo. Scientific Product No. 430341) (binding capacity: 0.6—0.65 µg), a polystyrene support commonly used in ELISA assays. The phytase was expected to bind to the 96-well plates by passive adsorption of the hydrophilic molecules in preparation for phage display experiments. For chapter 3, treated-active Quantum Blue phytase and treated-denatured Quantum Blue phytase was immobilised at 20 µg/ml for both panning rounds. For chapter 4, the active and denatured Quantum Blue phytase aliquots were immobilised at 20 µg/ml for round 1 and 2 panning. For round 3 and 4 panning and screening rounds, the phytases were immobilised at 4 µg/ml and 0.8 µg/ml respectively.

During chapter 5 Quantum Blue phytase was immobilised overnight at 18°C on maleic anhydride plates (Product no. 15100, Thermo Scientific) (binding capacity: ~125pmol biotin pentylamine (328Da)/well). Prior to immobilisation, the maleic anhydride plates were rinsed 3X with PBST. The phytases were diluted using 1X PBS before immobilising 100 µl aliquots of phytase (10 µg/ml) overnight at 18°C. The phytases were expected to bind to the 96-well plates by forming an amide bond between the primary amines within the phytase polypeptide structure and the maleic anhydride

For chapter 5, the active and denatured Quantum Blue phytase aliquots were immobilised at 10 µg/ml for round 1 and 2 panning. For round 3 and 4 panning and screening rounds, the phytases were immobilised at 4 µg/ml and 0.8 µg/ml respectively.

2.12 Phytic acid immobilisation

Phytic acid was immobilised on TiO_2 (Fisher; 60110-207) or ZiO_2 (Fisher; 60110-206) coated plates or beads (Titansphere; 5020-75000). The supports (TiO_2 or ZiO_2) were washed once with dH_2O followed by 1M perchloric acid. The phytic acid solution was then incubated with the supports for at least 15

minutes, however for panning and ELISA purposes the binding reaction was always allowed to occur overnight at 4°C.

The beads were pelleted in-between wash steps by centrifuging at 3,500 xg for 5 minutes and the supernatants were removed. During binding steps, the beads were vortexed periodically.

For chapter 4 and 5, the phytic acid aliquots were immobilised at 0.125M for all 4 panning rounds and screening rounds.

For chapter 5, phytic acid and ATP were also immobilised on Covalink plates (product no. 478042; Thermo Scientific)

The CovaLink plates had a specially designed surface that would enable them to interact with molecules bearing phosphate groups. The plates formed a phosphoramidate bond (NR_2), allowing them to couple with peptides, haptens and exclusively to the 5' end of DNA. Between the phosphoramidate bond and the base of the Covalink plate was a linker 2nm in size, enabling greater access to the immobilised molecule (Figure 2.3).

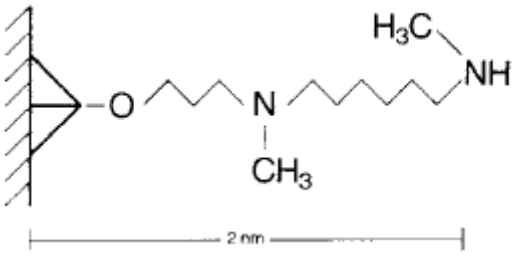


Figure 2.3: Covalink plate linker joining to a phosphoramidate bond *(© 2010 Thermo Fisher Scientific Inc., accessed 10/2017)*

CovaLink wells were pre-washed with dH_2O and 1M perchloric acid in preparation for immobilising phytic acid and ATP. To bind the phytic acid and ATP, 100 µl of a 0.1M (pH 1) solution was incubated in the pre-treated CovaLink wells overnight at 4°C.

2.13 Phytic acid staining (Toluidine Blue)

To confirm that the phytic acid molecules had been immobilised on the metal supports, the wells (TiO_2 & ZiO_2) and beads (TiO_2) were stained with

Toluidine Blue (Sigma; T3260), a dye that binds to phosphate groups. This technique would not distinguish between lower phytic acid species or free phosphates, but as the only source of phosphates on the well, would determine whether the phytic acid was present after washing and therefore was immobilised on the support as intended. After binding phytic acid to the pre-washed, metal-coated wells overnight, the phytic acid solution was removed from the well and the wells were then washed in 1M Tris buffer (pH 7.5). To stain the bound phosphate groups, 400 µl of Toluidine Blue (0.05% toluidine blue powder, 20% methanol, 79.95% dH2O) was incubated within the wells for 60 minutes followed by multiple de-staining washes (30% methanol, 70% dH2O) for 2 hours.

2.14 Panning Rounds

Phage display experiments typically involved 2—5 rounds of panning in order to enrich binders against any given target. Each antigen was immobilised overnight in preparation for panning.

The solid supports were then washed once with 300 µl 1X PBS (phytases) or 1M Tris Buffer (pH 7.5) (phytic acid) and were then blocked, along with aliquots of the overnight phage suspension, with 3% (w/v) Marvel and 1X PBS or 1M Tris Buffer (pH 7.5) for one hour. The supports were then washed with 300 µl 1X PBS or 1M Tris Buffer (pH 7.5) and 300 µl of blocked phage suspension was added to each of the supports and allowed to bind for one hour. The supports were then washed ten times with either PBST (1X PBS containing 0.1% (v/v) Tween 20) or 1M Tris buffer (pH 7.5) containing 0.1% (v/v) Tween 20 and then an additional ten times with either PBS or 1M Tris buffer (pH 7.5), and finally rinsed with dH_2O.

Phage were then eluted from the plates competitively using a 10X molar excess of the antigen by incubation for 30 minutes or acid eluted using 100 µl of 0.2M Glycine HCl (pH 2.6) buffer per well which was neutralised with 100 µl of 1M Tris buffer (pH 7.5) after 10 minutes. Alternatively, the phage were eluted by direct addition of the *E. coli* strain TG1 which were allowed to incubate on the supports for 30 minutes at 37°C static. The eluates were pooled and half of each elution pool were transferred to sterile eppendorf tubes and used to titre the enriched phage. The remaining elution was

incubated with 1 ml of TG1 if necessary and left for 30 minutes at 37°C static to allow infection. If the antigens were eluted from the supports using TG1, the samples were then pelleted at 3000 xg for 10 minutes and re-suspended in 1 ml of 2YT. The 1 ml TG1 solutions were then spread on bioassay plates (245 x 245 x 25mm) containing 2YT agar and 150 µg/ml ampicillin, 1% (w/v) glucose, and grown at 37°C overnight. The bacterial colonies were scraped from the bioassay plates and suspended in 2 ml 2YT 15% glycerol and stored at -80°C.

For subsequent panning rounds, 300 µl of the previous panning round's glycerol stock were used to inoculate 50 ml 2YT containing 150 µg/ml ampicillin and 1% (w/v) glucose and were allowed to grow, shaking at 37°C.

At an $O.D_{600nm}$ of 0.5—1, helper phage was added to each culture as detailed previously and allowed to incubate for 30 minutes static and 30 minutes shaking. Cultures were spun for 10 minutes at 3000xg and the pellets were re-suspended in 50ml 2YT containing 150 µg/ml ampicillin and 50 µg/ml kanamycin. They were then allowed to grow overnight shaking at 30°C. The following day, all cultures were centrifuged at 3,000xg and the supernatants were transferred to sterile containers and titred, ready to be used for further panning experiments.

In chapters 4 and 5, subtraction steps were conducted before or after panning. These additional subtraction steps would ideally remove non-specific binders from the pool of phage-scFv before they were incubated with the primary antigen. The phytase sub-libraries (300 µl aliquots of phage supernatant) were subtracted against their negative control antigen (opposite conformation) in 21 wells (x3 sets of 7). The phytic acid sub-library (300 µl aliquots of phage supernatant) was subtracted against 90mg of TiO2 beads (x15 sets of 6mg) with (chapter 5) or without (chapter 4) a negative control antigen (ATP) bound. In chapter 5, the Covalink panning's were subjected to subtraction steps against 21 Covalink wells containing ATP (x3 sets of 7) wells instead of TiO_2 beads. After 20 minutes with each set, the phage aliquots were transferred to a new set until completed. The remaining phage supernatants were used for panning.

2.15 Random colony picking

The colony picking approach was also used to identify binding antibodies. Random colonies were individually picked and grown in deep-96 well plates to produce monoclonal phage. The binding specificity of the phage preparations were tested by phage ELISA.

The colony picks were grown in 200 µl 2YT cultures in deep-96 well plates with 1% glucose and 150 µg/ml ampicillin. At an $O.D_{600nm}$ of 0.5 the cultures were superinfected with exphage and then the pellets re-suspended in 400 µl 2YT cultures with 150 µg/ml ampicillin and 50 µg/ml kanamycin. Phage were produced overnight and 1 µl of phage supernatant was used to determine the titres.

2.16 Preparation of DNA for Ion Torrent Sequencing

Plasmid dsDNA or phage ssDNA was isolated from the panning round glycerol stocks or eluted phage, respectively, replicates using a Qiagen Spin Mini Prep Kit (Qiagen Product No. 27104) or Qiagen QiaPrep Spin M13 Kit (Qiagen Product No. 27704) following the manufacturer's instructions. The eluates were quantified using a Nanodrop spectrophotometer and working DNA stocks of 10 ng/µl were prepared.

Two rounds of PCR were conducted to generated amplicons with linker and barcode sequences compatible with the NGS platform (Figure 2.4).

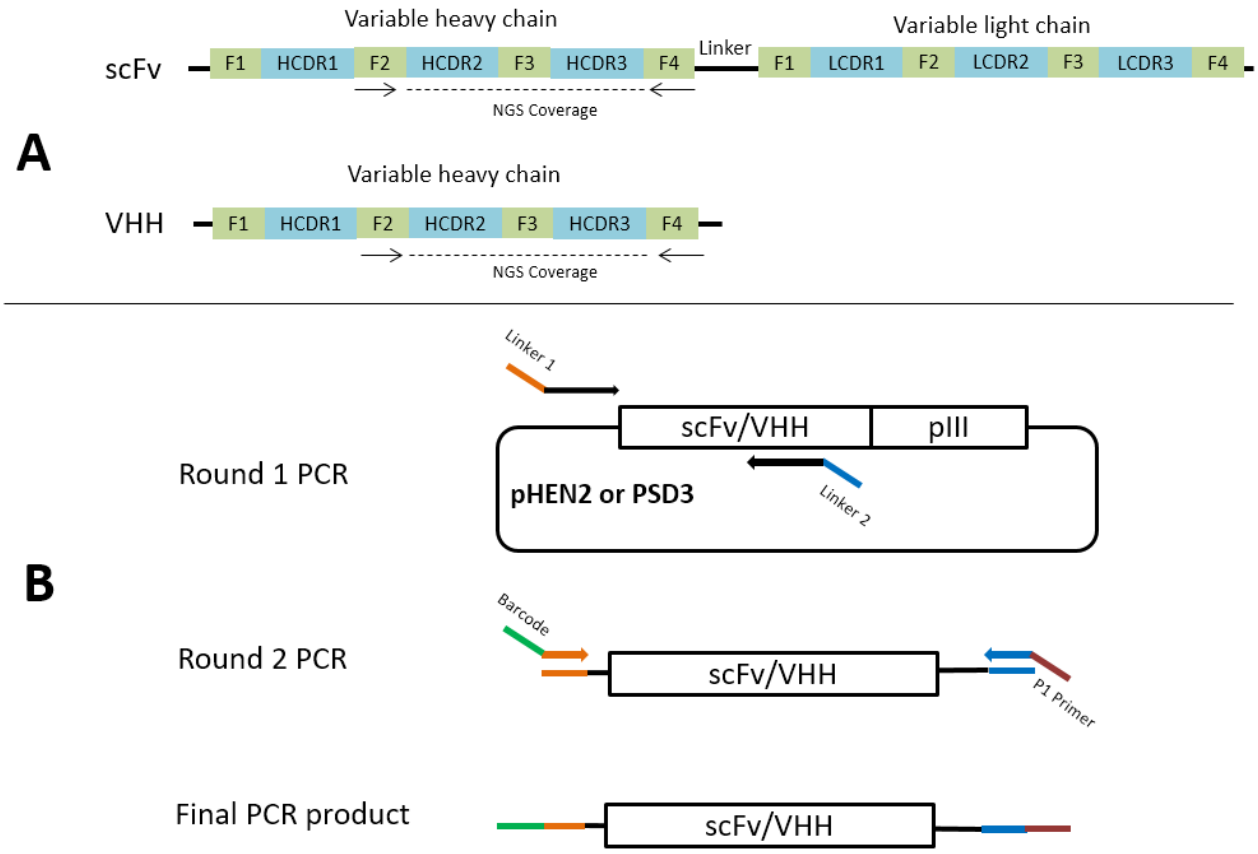


Figure 2.4. NGS Coverage and PCR schematic for the Griffin and VHH vectors. (A) The NGS strategy produces a 300 bp amplicon in round 1, amplifying the most variable HCDR3 region of the antibody genes. Vector-specific primers anneal to the framework (F1—4) regions 2 and 4 on the heavy chain. (B) Two rounds of PCR are conducted in preparation for NGS in order to provide linker and barcode sequences for each sample. Barcode sequences are later used to identify each sample during data analysis. Adapted from Yang et al. (2017).

The first round of PCR was conducted using a high fidelity Platinum Taq Polymerase (Invitrogen Product No. 11304-011) and primers complementary to the framework region 2 and 4 (table 2.4). The primer pair for round 1 PCR provided a linker sequence, serving as the template for the addition of barcode sequences in the second round of PCR. For the Griffin pHEN2 vector, the primers were degenerate in order to account for the framework variation within the vector between clones. The VHH phagemid vector had the same framework sequences between clones

apart from the randomised HCDR3, thus the primers were fixed for this library.

Table 2.4: Round 1 PCR primers for NGS preparation

Library[1]	**Primer Sequence 5' → 3'**
Griffin scFv VHFR2UniNew [2]	**CTAGAACATTTCACTTAC**GGVMARGGNCTKGASTGG
Griffin scFv JHFR4RevNew [3]	**GTAATCCTTGTGGTATCG**TGARGAGACRGTGACCRKKGT
VHH PCR 1 link2 f [2]	**CTAGAACATTTCACTTA**GATTCGGTAAAGGGCCGC
VHH PCR 1 link2 r [3]	**GTAATCCTTGTGGTATCG**CTAGAGGAAACTGTCACCTG

[1] Antibody phage display library used for panning experiments
[2] The forward primer compatible with each phage display library. Used for the purpose of adding a **linker sequence** to the round 1 PCR products in preparation for NGS.
[3] The reverse primer compatible with each phage display library. Used for the purpose of adding a **linker sequence** to the round 1 PCR products in preparation for NGS.

The final concentrations of the PCR components within each of the 50 µl reactions were as follows:

- dNTPs: 10 µM
- Primers: 0.5 µM
- DNA template: 10 ng
- $MgSO_4$: 2 mM
- Platinum Buffer: 1X
- Platinum *Taq* Polymerase: 1 Unit

Round 1 PCR was optimised and begun with an initial denaturing step at 95°C for 3 minutes. The initial two cycles were carried out with a denaturing step conducted at 95°C for 30 seconds followed by an annealing temperature of 55°C for 30 seconds and an extension time of 30 seconds at 68°C. A further 28 cycles were then conducted using a denaturing step at 95°C for 30 seconds, an annealing temperature of 60°C held for 30 seconds and an extension time of 30 seconds at 68°C. The size of round 1 PCR products were confirmed at 300bp by running 10 µl of PCR reaction on a 2% (w/v) agarose/TAE gel, stained with Nancy-520 (Sigma Product No. 01494), at 100 volts for one hour and compared against a low range ladder (LR Gene Ladder; Thermo Scientific Product No. SM1193).

For ssDNA templates the round 1 PCR protocol was altered slightly to increase the number of amplification rounds due to less genetic material originating from the phage ssDNA extracts. The number of cycles in the first round was increased from 2, to 8 cycles.

The PCR products were purified using a Nucleospin Gel and PCR purification kit (Machery-Nagel Product No. 740609) before conducting the second round of PCR. Round 2 PCR was conducted using a high fidelity Platinum Taq Polymerase and a second set of primers intended to provide unique barcode sequences for the purpose of sample identification and specific sequences (so called p1 and Akey sequences) that are compatible with the Ion Torrent sequencing platform (table 2.5).

Table 2.5: Ion Torrent Barcode Primers – Round 2 PCR for NGS

Barcode Primer [1]	**Sequence 5′ → 3′** [2]
P1prim-linker2	CCTCTCTATGGGCAGTCGGTGATCTAGAACATTTCACTTAC
Akey-BC1-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGTAATCCTTGTGGTATCG
Akey-BC2-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGTAATCCTTGTGGTATCG
Akey-BC3-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGTAATCCTTGTGGTATCG
Akey-BC4-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGTAATCCTTGTGGTATCG
Akey-BC5-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGTAATCCTTGTGGTATCG
Akey-BC6-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGTAATCCTTGTGGTATCG
Akey-BC7-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGTAATCCTTGTGGTATCG
Akey-BC8-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGTAATCCTTGTGGTATCG
Akey-BC9-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGTAATCCTTGTGGTATCG
Akey-BC10-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGTAATCCTTGTGGTATCG
Akey-BC11-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGTAATCCTTGTGGTATCG
Akey-BC12-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGTAATCCTTGTGGTATCG
Akey-BC13-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGTAATCCTTGTGGTATCG
Akey-BC14-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCGTAATCCTTGTGGTATCG
Akey-BC15-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCGTAATCCTTGTGGTATCG
Akey-BC16-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACGTAATCCTTGTGGTATCG
Akey-BC17-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTATTCGTCGTAATCCTTGTGGTATCG
Akey-BC18-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCAATTGCGTAATCCTTGTGGTATCG
Akey-BC19-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAGTCGGACGTAATCCTTGTGGTATCG
Akey-BC20-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGATCCATCGTAATCCTTGTGGTATCG
Akey-BC21-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCAATTACGTAATCCTTGTGGTATCG
Akey-BC22-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGAGACGCGTAATCCTTGTGGTATCG
Akey-BC23-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCCACGAACGTAATCCTTGTGGTATCG
Akey-BC24-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCTCATTCGTAATCCTTGTGGTATCG
Akey-BC25-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATACGTAATCCTTGTGGTATCG
Akey-BC26-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACAACCTCGTAATCCTTGTGGTATCG

Akey-BC27-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCGTAATCCTTGTGGTATCG
Akey-BC28-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGGAATCGTAATCCTTGTGGTATCG
Akey-BC29-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGACCACTCGTAATCCTTGTGGTATCG
Akey-BC30-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGGTTATCGTAATCCTTGTGGTATCG
Akey-BC31-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCAAGCTGCGTAATCCTTGTGGTATCG
Akey-BC32-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTTACACACGTAATCCTTGTGGTATCG
Akey-BC33-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTCATTGAACGTAATCCTTGTGGTATCG
Akey-BC34-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATCGTTCGTAATCCTTGTGGTATCG
Akey-BC35-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGCCATTGTCGTAATCCTTGTGGTATCG
Akey-BC36-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGGAATCGTCGTAATCCTTGTGGTATCG
Akey-BC37-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGAGAATGTCGTAATCCTTGTGGTATCG
Akey-BC38-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGGACGGACGTAATCCTTGTGGTATCG
Akey-BC39-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAACAATCGGCGTAATCCTTGTGGTATCG
Akey-BC40-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACATAATCGTAATCCTTGTGGTATCG
Akey-BC41-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCACTTCGCGTAATCCTTGTGGTATCG
Akey-BC42-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCACGAATCGTAATCCTTGTGGTATCG
Akey-BC43-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGACACCGCGTAATCCTTGTGGTATCG
Akey-BC44-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGGCCAGCGTAATCCTTGTGGTATCG
Akey-BC45-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGCTTCCTCGTAATCCTTGTGGTATCG
Akey-BC46-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGTCCGAACGTAATCCTTGTGGTATCG
Akey-BC47-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGCAACCACGTAATCCTTGTGGTATCG
Akey-BC48-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTAAGAGACGTAATCCTTGTGGTATCG
Akey-BC49-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAACATAACGTAATCCTTGTGGTATCG
Akey-BC50-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAATGGCGTAATCCTTGTGGTATCG
Akey-BC51-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTCGTAATCCTTGTGGTATCG
Akey-BC52-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCATGGAACGTAATCCTTGTGGTATCG
Akey-BC53-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGGCAATCCTCGTAATCCTTGTGGTATCG
Akey-BC54-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGGAGAATCGCGTAATCCTTGTGGTATCG
Akey-BC55-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCACCTCCTCGTAATCCTTGTGGTATCG
Akey-BC56-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCATTAATTCGTAATCCTTGTGGTATCG
Akey-BC57-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGCAACGGCGTAATCCTTGTGGTATCG
Akey-BC58-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAGAACACGTAATCCTTGTGGTATCG
Akey-BC59-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTTGATGTTCGTAATCCTTGTGGTATCG
Akey-BC60-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGCTCTTCGTAATCCTTGTGGTATCG
Akey-BC61-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTCGGATCGTAATCCTTGTGGTATCG
Akey-BC62-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTGCTTCACGTAATCCTTGTGGTATCG
Akey-BC63-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTTAGAGTTCGTAATCCTTGTGGTATCG
Akey-BC64-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGAGTTCCGACGTAATCCTTGTGGTATCG
Akey-BC65-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGGCACATCGTAATCCTTGTGGTATCG
Akey-BC66-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCAATCATCGTAATCCTTGTGGTATCG
Akey-BC67-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTACCAGTCGTAATCCTTGTGGTATCG
Akey-BC68-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAAGAAGTTCGTAATCCTTGTGGTATCG
Akey-BC69-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCAATTGGCGTAATCCTTGTGGTATCG
Akey-BC70-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTACTGGTCGTAATCCTTGTGGTATCG
Akey-BC71-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGGCTCCGACGTAATCCTTGTGGTATCG

Akey-BC72-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGGCCACACGTAATCCTTGTGGTATCG
Akey-BC73-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGCCTGTCGTAATCCTTGTGGTATCG
Akey-BC74-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATCGGTTCGTAATCCTTGTGGTATCG
Akey-BC75-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGGAATACGTAATCCTTGTGGTATCG
Akey-BC76-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGAACCTCGTAATCCTTGTGGTATCG
Akey-BC77-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGCGATTCGTAATCCTTGTGGTATCG
Akey-BC78-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCCAATTCTCGTAATCCTTGTGGTATCG
Akey-BC79-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGGTTGTCGTAATCCTTGTGGTATCG
Akey-BC80-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGAAGGCAGGCGTAATCCTTGTGGTATCG
Akey-BC81-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGCCATTCGCGTAATCCTTGTGGTATCG
Akey-BC82-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCATCTCGTAATCCTTGTGGTATCG
Akey-BC83-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGACATTCGTAATCCTTGTGGTATCG
Akey-BC84-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTCCATAACGTAATCCTTGTGGTATCG
Akey-BC85-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCAGCCTCAACGTAATCCTTGTGGTATCG
Akey-BC86-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGGTTATTCGTAATCCTTGTGGTATCG
Akey-BC87-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGCTGGACGTAATCCTTGTGGTATCG
Akey-BC88-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGAACACTTCGTAATCCTTGTGGTATCG
Akey-BC89-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTGAATCTCGTAATCCTTGTGGTATCG
Akey-BC90-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAACCACGGCGTAATCCTTGTGGTATCG
Akey-BC91-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGGATGCGTAATCCTTGTGGTATCG
Akey-BC92-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAGGAACCGCGTAATCCTTGTGGTATCG
Akey-BC93-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGTCCAATCGTAATCCTTGTGGTATCG
Akey-BC94-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCGACAAGCGTAATCCTTGTGGTATCG
Akey-BC95-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAGATCGTAATCCTTGTGGTATCG
Akey-BC96-lnk1	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAAGCGGTCGTAATCCTTGTGGTATCG

[1] Round 2 barcode (BC) primer compatible with Ion Torrent sequencing platform
[2] The barcode primer sequence with the A-key highlighted in yellow, the barcode section highlighted in red and finally the shorter linker section of the primer in black. P1prim-linker2 was the reverse primer used in combination with one of the other barcoding primers.

Round 2 PCR products had total volumes of 50 µl and used the same reagents and concentrations as round 1. The template DNA for round 2 PCR was 5 µl of the purified round 1 PCR products. For round 2 PCR an initial denaturing step was conducted at 95°C for 3 minutes. Subsequently, 25 cycles involving a denaturing step at 95°C for 30 seconds, an annealing time of 30 seconds at 63°C and an extension time of 30 seconds at 68°C were conducted. Amplified round 2 products were resolved on a 2% (w/v) agarose/TAE gel, stained with Nancy-520, at 100 volts for one hour to confirm the expected amplicon size of ~350bp.

For ssDNA templates the round 2 PCR program was altered slightly; the number of cycles was increased from 25 to 35 cycles.

Round 2 PCR products were quantified using Qubit florescence and then pooled, ensuring an equal contribution from each sample towards a total pool of 1000 ng. The pooled DNA sample was then purified using Agencourt AMPure XP kit (Beckman-Coulter A63880) and re-quantified before being sent for sequencing at the University of Pennsylvania using the high-throughput Ion Torrent platform with a 318 chip.

2.17 dsDNA quantification using Qubit kit

Qubit Broad Range (BR) or High Sensitivity (HS) kits used a fluorophore specific for dsDNA.

A master mix was made as follows:

- No. of samples x 199 µl = total buffer
- No. of samples x 1 µl = total fluorophore

For each standard, 190 µl of the master mix was combined with 10 µl of the standard. For each test sample, 198 µl of the master mix was combined with 2 µl of the test sample. The samples were vortexed and incubated at room temperature for 5 minutes. The samples were then vortexed again and read immediately using a Qubit flurometer.

2.18 NGS Data Analysis

2.18.1 Analysis pipeline

Deep sequencing data was received from the University of Pennsylvania in a compressed, FASTQ file format (.fastaq.tgz). Perl scripts in Ubuntu, a Linux operating system, were used to analyse the NGS data via the script-processing program Perl and the terminal window display. The user specified input (purple text) and output (blue text) file names were provided along with a list of the Ion Torrent barcodes used to identify individual samples. Once de-multiplexed, individual files could then be compared to identify enriched heavy chain complementary determine region 3 (HCDR3), sequences against the panning target as detailed below.

The first Perl script converted the decompressed, '.fastq' file from the University of Pennsylvania to a '.fasta' file by running the command:

seqtk seq -a XXX.fastq >XXX.fasta

Individual samples were then de-multiplexed and split into several files based on their barcode identifiers. Barcode identifiers were provided by the ion.barcodes.96.fasta file containing all 96 barcode sequences. The following command was used to de-multiplex and split the samples:

flexbar -o -n 8 -bt 0 -b ion.barcodes.96.fasta -t split -r XXX.fasta

Output files were labelled according to the barcode sequence e.g split_barcode_BC01.fasta, split_barcode_BC02.fasta, split_barcode_BC03.fasta.

The following script accounted for the Griffin/VHH library being sequenced in the reverse direction, and thereby reverse complemented the existing sequences to produce new output files (barcodes 1—5 examples).

seqtk seq -r split_barcode_BC01.fasta > split_barcode_BC01.rev.fasta

seqtk seq -r split_barcode_BC02.fasta > split_barcode_BC02.rev.fasta

seqtk seq -r split_barcode_BC03.fasta > split_barcode_BC03.rev.fasta

seqtk seq -r split_barcode_BC04.fasta > split_barcode_BC04.rev.fasta

seqtk seq -r split_barcode_BC05.fasta > split_barcode_BC05.rev.fasta

The reverse complemented output files were then used as the input files for the following translation script (barcodes 1 and 2 examples). The text within each output file was converted to all three reading frames (-f 1, 2 or 3) and ochre (TAA), opal (TGA) and amber (TAG) stop codons were replaced with o, p or q respectively using the –s opq setting.

perl translate.pl -i split_barcode_BC01.rev.fasta -s opq -f 1 > BC01.frame1.fasta

perl translate.pl -i split_barcode_BC01.rev.fasta -s opq -f 2 > BC01.frame2.fasta

perl translate.pl -i split_barcode_BC01.rev.fasta -s opq -f 3 > BC01.frame3.fasta

perl translate.pl -i split_barcode_BC02.rev.fasta -s opq -f 1 > BC02.frame1.fasta

perl translate.pl -i split_barcode_BC02.rev.fasta -s opq -f 2 > BC02.frame2.fasta

perl translate.pl -i split_barcode_BC02.rev.fasta -s opq -f 3 > BC02.frame3.fasta

The three frame files for each barcode (barcode 1 & 2 examples) were combined into one .frames file using the cat command.

cat BC01.frame1.fasta BC01.frame2.fasta BC01.frame3.fasta > BC01.frames.fasta

cat BC02.frame1.fasta BC02.frame2.fasta BC02.frame3.fasta > BC02.frames.fasta

The combined reading frame files were then used as the input files for the subsequent script.

The next script searched for the Griffin/VHH library's flanking motifs surrounding the HCDR3. The HCDR3 sequences are identified within the NGS data by finding the less variable flanking sequences at either side. The flanking sequences used in this study were 5 amino acids long.

The flanking motifs were fixed in the VHH vector so the left flanking sequence text file contained 'YYCAA' and the right text file contained 'WGQGT'.

The flanking motifs were not fixed in the Griffin vector, thus the script permitted single amino acid changes [^amino_acid] in the expected sequences. Only the heavy chains were amplified for NGS analysis so no contaminating sequences from the LCDR's were expected.

The left flanking sequences text file contained:

VYYCAR
[^V]YYCAR
V[^Y]YCAR
VY[^Y]CAR
VYY[^C]AR
VYYC[^A]R
VYYCA[^R]

The right flanking sequences text file contained:

WGQGT
[^W]GQGT
W[^G]QGT

WG[^Q]GT

WGQ[^G]T

WGQG[^T]

The perl script iterate_motifs.pl required two additional text files containing all possible combinations for the left and right flanking sequences.

A minimum number of amino acids (4) making up the HCDR3 were allowed between the flanking motifs to minimise the number of likely non-functional antibody sequences returned (barcodes 1—5 examples). The following command was then used to run the iterate.pl script.

perl iterate_motifs.pl BC01.frames.fasta 4

perl iterate_motifs.pl BC02.frames.fasta 4

perl iterate_motifs.pl BC03.frames.fasta 4

perl iterate_motifs.pl BC04.frames.fasta 4

perl iterate_motifs.pl BC05.frames.fasta 4

Three output files were generated for each barcode (barcode 1 example) - BC01.frames.L.fa, BC01.frames.LR.fa and BC01.frames.R.fa. The .L.fa and .R.fa files only contained HCDR3 sequences with either the left or right motifs only. The .LR.fa files contained HCDR3 sequences with both the expected left and right flanking sequences and were use as the final output files for comparison purposes.

The remaining part of the NGS analysis involved comparing the barcoded files against each other as necessary to identify enriched antibodies against the panning target e.g barcodes representing positive samples individually compared against the pool of negative control samples barcodes.

Barcode files were pooled into one output file as required using the following script (barcodes 61—72 examples):

cat BC61.frames.fasta.LR.fa BC62.frames.fasta.LR.fa BC63.frames.fasta.LR.fa BC64.frames.fasta.LR.fa BC65.frames.fasta.LR.fa BC66.frames.fasta.LR.fa BC67.frames.fasta.LR.fa BC68.frames.fasta.LR.fa BC70.frames.fasta.LR.fa BC71.frames.fasta.LR.fa BC72.frames.fasta.LR.fa >negative.pool.LR.fa

A comparison script was used compare the HCDR3 sequences present in the first input file (positive sample) and second input file (negative pool). For each HCDR3 sequence, the total number of reads (frequency) were calculated as well as the ratio (number of reads in the positive sample vs the number of

reads in the negative pool). The frequency and ratio were used for a two proportion Z test to calculate a Z score upon which each HCDR3 sequence could be ranked (Zhang et al., 2011). The Z score equation is shown below, where n1 = first input file read count, n2 = second input file read count, p1 = proportion of total reads 1 and p2 = proportion of total reads 2.

$$\left(z = \frac{p1 - p2}{\sqrt{\underset{n1}{p1(1-p1)} \; \underset{n2}{p2\,(1-p2)}}} \right)$$

The Z score ranking system took in to account the total frequency of each HCDR3 but also its presence in the experimental and control replicates, thus providing a system to rank each candidate based on its likely specificity to the target antigen.

The command used to compare an individual positive sample's barcode (barcode examples 49—51) against the pool of negative controls was:

perl compare.collapsed.sequences_2a.Zscore.pl BC49.frames.fasta.LR.fa negative.pool.LR.fa >49vspool.table

perl compare.collapsed.sequences_2a.Zscore.pl BC50.frames.fasta.LR.fa negative.pool.LR.fa >50vspool.table

perl compare.collapsed.sequences_2a.Zscore.pl BC51.frames.fasta.LR.fa negative.pool.LR.fa >51vspool.table

The output table files that were generated detailed each HCDR3 sequence that was present within the first input file (positive sample), the number of reads present in the first input file, the number of reads present in the second input file (negative control pool) and the final Z score for that HCDR3 sequence.

The table files were then used as input files for the following perl scripts that aimed to produce new tables files with Z score thresholds of ≥2, ≥5 or ≥10. The Z score thresholds implied a strong likelihood that the candidate had become enriched against the target due to a binding event and therefore was not simply enriched by chance.

The table files generated could not only show HCDR3 sequences with Z scores of 2, 5, 10 or greater but also the number of times the HCDR3 was

present in all of the individual positive barcodes rather than just on an individual basis.

Firstly, the Z score threshold tables were generated using the following script (barcode examples 49—51; Z score threshold of 2 as an example). The $4 setting searched for the Z score value in the table files 4th column. If it had a value of 2 or more, it was copied to the .Z2 output file.

awk ' $4>=2 { print $0 } ' <49vspool.table >49vspool.table.Z2

awk ' $4>=2 { print $0 } ' <50vspool.table >50vspool.table.Z2

awk ' $4>=2 { print $0 } ' <51vspool.table >51vspool.table.Z2

All the individual positive samples were then combined into one table file using the command below, to show how many positive samples each HCDR3 had been seen in with a Z score of 2 or more.

awk '{ print $1 }' <49vspool.table.Z2>>Z2_list

awk '{ print $1 }' <50vspool.table.Z2>>Z2_list

awk '{ print $1 }' <51vspool.table.Z2>>Z2_list

Enriched scFv sequences were then chosen by sorting through the table files manually in Microsoft Excel and selecting sequences based on their presence in the positive replicates and the Z scores as summarised in figure 2.5.

For each screening round the positive, target antigen replicates were individually compared the round's pool of negative control replicates. This comparison generated Z scores for each HCDR3 based on their total frequency and counts in the positive replicate and negative pool. High Z scores were generated for HCDR3 when they were seen in multiple positive replicates and only a few, if any, negative replicates.

Score thresholds of Z ≥ 10, 5 or 2 were applied, generating 3 datasets for each screening round against each target.

Duplicate HCDR3 sequences were removed between Z score threshold datasets against the same target, favouring the threshold entry with the highest number of replicates seen in followed by the best Z score threshold. Duplicates were also removed when seen in multiple screening rounds against the same target, following the same criteria as the Z score thresholds.

For the Griffin scFv analysis, HCDR3 containing stop codons were removed from the data (ochre (TAA), opal (TGA) and amber (TAG) stop codons were replaced with o, p or q respectively when translated). This was not the case

for the VHH library as the framework sequences including the HCDR3 were converted to optimal codon combinations for expression in *E. coli*, and this automatically removed stops.

The remaining candidates that were seen in at least 4 positive replicates were retained for dataset comparisons.

For the selection of potential binders to phytase, in addition to the selection criteria above, the data analysis procedure also removed duplicates seen in both phytase conformation datasets. Any clone appearing in both data sets would have likely been binding to both phytase conformations, or have been parasitic phage with growth advantages, with no specificity for the targets. The remaining candidates were then compared against the phytic acid dataset and duplicates were removed from both sets, as again they suggested non-specific binders.

In chapters 4 and 5, the top 20 candidates were subjected to a 'reversed analysis'. The negative replicates were individually compared to the same screening round's pool of positive replicates, generating new Z scores for each HCDR3 sequence. The new Z scores for the top 20 candidates were identified and any potential candidates with two or more negative replicates, with Z scores of 2 or greater, were excluded from further analyses.

ScFv clones were rescued from samples with the highest Z score for that particular HCDR3 sequence. The sample was identified using the following script (Using example HCDR3 sequence 'AFFIQTPD':

grep -B 1 AFFIQTPD BC60.frames.fasta.LR.fa | grep -v AFFIQTPD | grep -v '\-\-' >AFFIQTPD.fasta.list

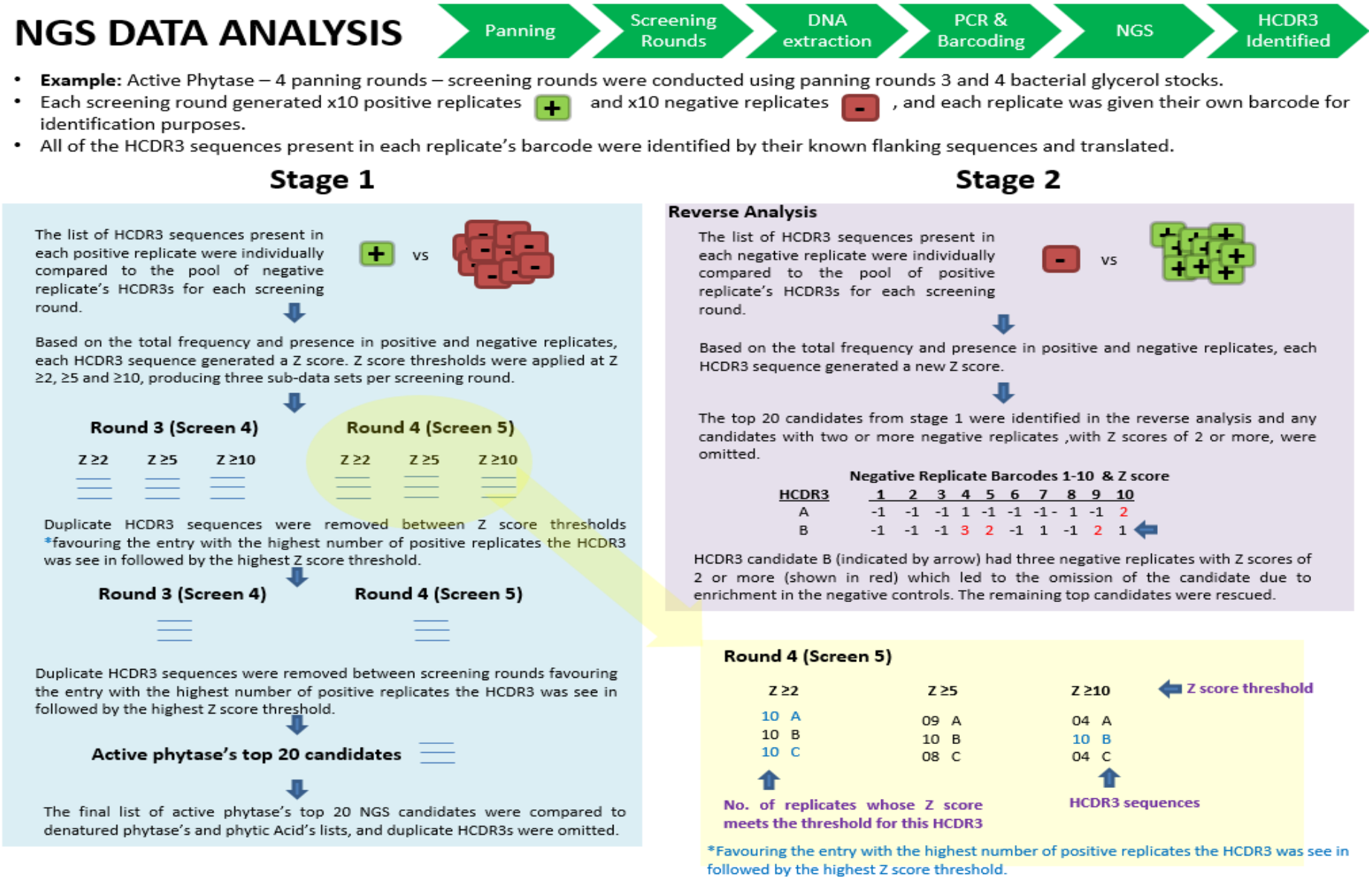
grep -A 1 -f AFFIQTPD.fasta.list split_barcode_BC60.fasta | head -n 30 | grep -v '\-\-' >AFFIQTPD.dna.list

The first line of the script looked for the example HCDR3 sequence 'AFFIQTPD' in the translated barcode file with the highest Z score, and then output its identifier codes into a new .fasta.list file. The fasta.list file formed the input file for the second line of the script, which used the identifier codes to find 'AFFIQTPD's original DNA sequence reads from the de-multiplexed DNA files. The DNA sequence information was required to rescue the clone.

The scFv candidates were rescued from their enriched sub libraries using an inverse PCR method. This was not required for VHH cloning as there was no associated VL region and also no variance in the other VH regions outside

of the CDR3. VHH clones were therefore synthesised directly into the VHH template within an expression vector.

Figure 2.5. NGS Coverage and PCR schematic for the Griffin and VHH vectors.



2.18.2 Flanking region changes

Based on the kabat numbering system, the human HCDR3 always begins 33 amino acids after the end of the HCDR2 and is usually 3-25 residues long. The typical flanking regions around the Griffin scFvs HCDR3 are 'VYYCAR' at the 5' end and 'WGQGT' at the 3' end. The residues before the HCDR3 are always cysteine (C) followed by two more amino acids, usually Alanine (A) and Arginine (R). The 3' flanking region is always Tryptophan (W), Glycine (G), any amino acid but usually Glutamine (Q), then always Glycine (G).

The flanking sequences around the LCDR3 are very similar to that of the HCDR3 and small changes can cause one to be mistaken for the other. The LCDR3 always begins 33 amino acids after the end of the LCDR2 and is usually 7-11 residues long. The typical flanking regions around the LCDR3 are VYYC at the 5' end and FGQGT at the 3' end. The 5' residue before the LCDR3 begins is always cysteine (C). The 3' flanking region always has the phenylalanine (F), Glycine (G) and last Glycine (G).

For chapter 3 and 4, the analysis pipeline permitted single amino acid changes [^amino_acid] in the expected HCDR3 flanking sequences. Only the heavy chain was amplified in preparation for the NGS analysis and further degeneracy could be permitted within the flanking regions without contaminating the data with light chain sequences. For chapter 6, the Griffin bioinformatics data analysis permitted zero and two changes in the flanking regions and the left and right flanking sequences text files were changed accordingly. The VHH bioinformatics data analysis also permitted up to 1 or 2 changes within the flanking sequences and the left and right flanking sequence text files were changed as required.

Zero changes permitted (Griffin Library)

The left flanking sequences text file contained:

- VYYCAR

The right flanking sequences text file contained:

- WGQGT

Up to two changes permitted (Griffin Library)

The left flanking sequences text file contained:

- VYYCAR
- [^V]YYCAR
- V[^Y]YCAR
- VY[^Y]CAR
- VYY[^C]AR
- VYYC[^A]R
- VYYCA[^R]
- [^V][^Y]YCAR
- [^V]Y[^Y]CAR
- [^V]YY[^C]AR
- [^V]YYC[^A]R
- [^V]YYCA[^R]
- V[^Y][^Y]CAR
- V[^Y]Y[^C]AR
- V[^Y]YC[^A]R
- V[^Y]YCA[^R]
- VY[^Y][^C]AR
- VY[^Y]C[^A]R
- VY[^Y]CA[^R]
- VYY[^C][^A]R
- VYY[^C]A[^R]
- VYYC[^A][^R]

The right flanking sequences text file contained:

- WGQGT
- [^W]GQGT
- W[^G]QGT
- WG[^Q]GT
- WGQ[^G]T
- WGQG[^T]
- [^W][^G]QGT
- [^W]G[^Q]GT
- [^W]GQ[^G]T
- [^W]GQG[^T]
- W[^G][^Q]GT
- W[^G]Q[^G]T
- W[^G]QG[^T]
- WG[^Q][^G]T
- WG[^Q]G[^T]
- WGQ[^G][^T]

One change permitted (VHH Library)

The left flanking sequences text file contained:

- YYCAA
- [^Y]YCAA
- Y[^Y]CAA
- YY[^C]AA
- YYC[^A]A
- YYCA[^A]

The right flanking sequences text file contained:

- WGQGT
- [^W]GQGT
- W[^G]QGT
- WG[^Q]GT
- WGQ[^G]T
- WGQG[^T]

Up to two changes permitted (VHH Library)

The left flanking sequences text file contained:

- YYCAA
- [^Y]YCAA
- Y[^Y]CAA
- YY[^C]AA
- YYC[^A]A
- YYCA[^A]
- [^Y][^Y]CAA
- [^Y]Y[^C]AA
- [^Y]YC[^A]A
- [^Y]YCA[^A]
- Y[^Y][^C]AA
- Y[^Y]C[^A]A
- Y[^Y]CA[^A]
- YY[^C][^A]A
- YY[^C]A[^A]
- YYC[^A][^A]

The right flanking sequences text file contained:

- WGQGT
- [^W]GQGT
- W[^G]QGT
- WG[^Q]GT
- WGQ[^G]T
- WGQG[^T]
- [^W][^G]QGT
- [^W]G[^Q]GT
- [^W]GQ[^G]T

- [^W]GQG[^T]
- W[^G][^Q]GT
- W[^G]Q[^G]T
- W[^G]QG[^T]
- WG[^Q][^G]T
- WG[^Q]G[^T]
- WGQ[^G][^T]

2.18.3 NGS diversity script

The diversity of the phage display libraries were predicted using a perl script based on the Lincoln Index statistical measure (R. E. Southwood and Henderson, 2000). The Lincoln Index was used to predict the number of antibody sequences that had not been observed, based on what had been observed by two separate sequencing attempts of the inital phage display libraries. In preparation for this analysis, the bacterial glycerol stocks for each library were miniprepped twice and both extracts were prepared for NGS as previously described. Each miniprep was allocated a barcode during round 2 PCR and used to identify the sample further downstream. The data was processed as before until each barcode was split into frame files (BC01.frames.L.fa, BC01.frames.LR.fa and BC01.frames.R.fa).

The diversity command required two lists of sequences, or frames.LR.fa files to compare against one another and looked at the total number of sequences in each file. The total number of sequences in each file were multiplied by each other and then divided by the total number of sequences that were common to both files, as described for the Lincoln Index measure. The script was run using the following command in Linux terminal:

perl fasta_diversity_estimate_v1.1.pl file1.frames.LR.fa file2.frames.LR.fa >output.txt

2.19 scFv Rescue

2.19.1 scFv rescue: Inverse PCR

Primers intended to rescue specific scFv were designed based on a degenerate, universal framework 4 sequence and the unique HCDR3 sequence (Spiliotopoulos et al., 2015) of enriched scFv (Figure 2.6). Working DNA solutions were prepared (10 ng/µl) for phage sub-libraries from the panning experiments that had the highest Z score for the HCDR3 of the scFv being rescued. The framework 4 region and HCDR3 sequences were amplified and extended to include the entire Griffin pHEN2 vector in an inverse PCR reaction. This used a high fidelity Q5 DNA Polymerase (BioLabs Product No. M0491S). The inverse PCR reaction involved final sample volumes of 50 µl with the components at the following concentrations: Q5 DNA Polymerase: 0.02 U/µl, dNTP: 400 µM, GC Enhancer Buffer (1X), Q5 reaction buffer (1X), primers (0.5 µM) and DNA Template (10—50 ng). The inverse PCR was started with an initial denaturing reaction conducted at 95°C for 3 minutes followed by 10-15 cycles of 95°C for 30 seconds, 50°C for 50 seconds and an extension time of 5 minutes at 72°C. The following 20-25 cycles involved a denaturing time of 30 seconds at 95°C, 50 seconds annealing time at 55°C and an extension time of 5 minutes at 72°C. Annealing temperatures were adjusted to suit the melting temperature (TM) of each HCDR3 primer as required.

For optimisation, annealing temperatures were reduced or additional template material or magnesium was included in the reaction. If the reactions continued to fail, the first group of cycles annealing temperature was reduced to 53°C and only the HCDR3 primer was added to the master mix. After the first set of cycles finished, the FRW4 primer was then added to the master mix as the second group of cycles began. The second set of cycles had an annealing temperature of 65°C. This alternative method was used to potentially solve any significant difference in primer TM – failing this method, the FRW4 primer was redesigned to be specific to the individual scFv's framework.

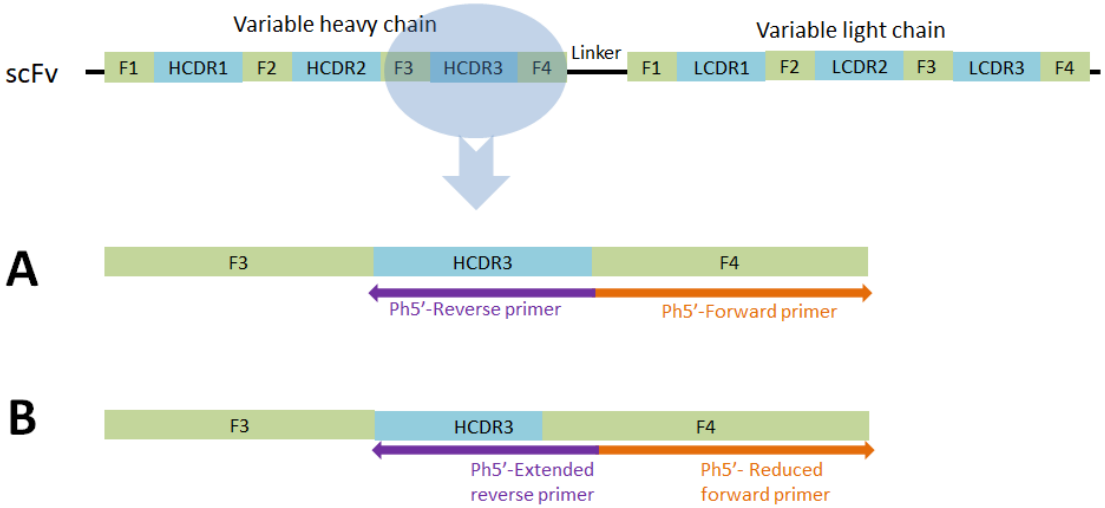


Figure 2.6. ScFv Rescue Strategy: Inverse PCR. Primer pairs with 5' phosphorylated ends (Ph5'-) were designed to amplify the HCDR3 and framework 4 (FR4) regions of each scFv. **(A)** Typical reactions involved the reverse primer spanning the entirety of the HCDR3 region and its forward primer pair targeting the framework (F) 4 region. **(B)** Depending on the length of the HCDR3 region the reverse primer was on occasions extended into the framework 4 region and the forward primer redesigned to abut the 5' end of the new reverse primer. Adapted from Spiliotopoulos et al. (2015).

2.19.2 scFv rescue: Amplicon visualisation

The entire reaction was loaded onto a 0.9% (w/v) agarose/TAE gel, stained with Nancy-520, and run at 100 volts for one hour with a 1 KB+ ladder (Thermo Scientific Product No. SM1333). The gel was then visualised to check for PCR products with an expected size of ~5kb. Bands of the correct size were cut out of the gel and the DNA was purified using a Machery-Nagel Nucleospin Gel and PCR purification kit following the manufacturer's protocol.

2.19.3 scFv rescue: Ligation reaction

Both the HCDR3 and framework 4 primers had been phosphorylated at the 5' end to allow a subsequent ligation reaction using T4 Ligase (BioLabs Product No. M0202S) to produce circular DNA by forming a phosphor-diester bond.

In a final volume of 20 µl:

- 2 µl DNA ligase buffer (10X)
- 400U T4 DNA ligase (NEB) (1 µl of 400,000U/ml)
- 6 µl of gel-purified PCR product
- Nuclease free water up to 20 µl

The ligation reaction was conducted at room temperature and left overnight. The following day, the T4 ligase was deactivated by heating the samples to 65°C for 15 minutes.

The ligation reactions were then buffer exchanged by placing on 0.025 µm filters (Merck Product No. VSWP02500) and incubated on the surface of molecular grade purified water (HyClone Product no. 10011342) for 30 minutes before pipetting into a sterile eppendorf tube.

2.19.4 scFv rescue: TG1 Transformation

The plasmid DNA for each scFv rescue was then used to transform 50 µl of electro-competent TG1 cells using 2 µl of the ligation reaction product and cuvettes with a 0.1 cm gap. Electroporation was conducted using a BioRad GeneP Pulsar Xcell Electroporator using a pre-set exponential decay protocol for bacteria.

BioRad pulser- preset program:
- Bacterial preset protocol 1; *E. coli*; Exponential Decay
- 25 uF, 200ohm, 1,800 volts, 0.1cm gap

After electroporation the cuvettes were immediately placed on ice and 950 µl of pre-warmed SOC recovery medium (BioLabs Product No. B9020S) was added to each vial. The samples were allowed to recover for 1 hour, shaking at 37°C. Following incubation, 100 µl was spread on 2YT plates with 1% (w/v) glucose and 150 µg/ml ampicillin and were allowed to grow overnight at 37°C. A no-ligase negative control and a positive control using 2 µl of the plasmid pUC19 (400 Units) (BioLabs Product No. N3041S) were also transformed.

The following day, 4—5 colonies from each plate were grown in 10ml 2YT, 150 µg/ml ampicillin and 1% (w/v) glucose to $O.D_{600nm}$ 0.5—1 before being centrifuged at 3,000 xg and the pellets re-suspended in 2YT containing 15% (w/v) glycerol stocks at -80°C.

2.20 VHH synthesis strategy

2.20.1 VHH synthesis: primer design and inverse PCR

The HCDR3 amino acid sequences for VHH candidates selected for synthesis were identified either by Sanger sequencing or NGS analysis. The amino acid sequences were converted to optimal nucleotide codons for expression in *E. coli* using an online conversion tool (www.jcat.de). The *E. coli* strain 'K12' was selected as the model organism. This system also converted amber stop codons to glutamines (Q), thereby resolving any potential expression issues caused by premature stops within the HCDR3.

Forward and reverse primers were then designed to cover half of the HCDR3 sequence each and to also cover 18 nucleotides of the adjacent framework regions (frameworks 3 and 4) (figure 2.7). This design resulted in long 5' phosphorylated primers designed to synthesise the HCDR3 sequences straight into any vector that already contained a template VHH clone that had originated from the same library. Each primer was 5' phosphorylated to facilitate efficient ligation of the linear PCR product.

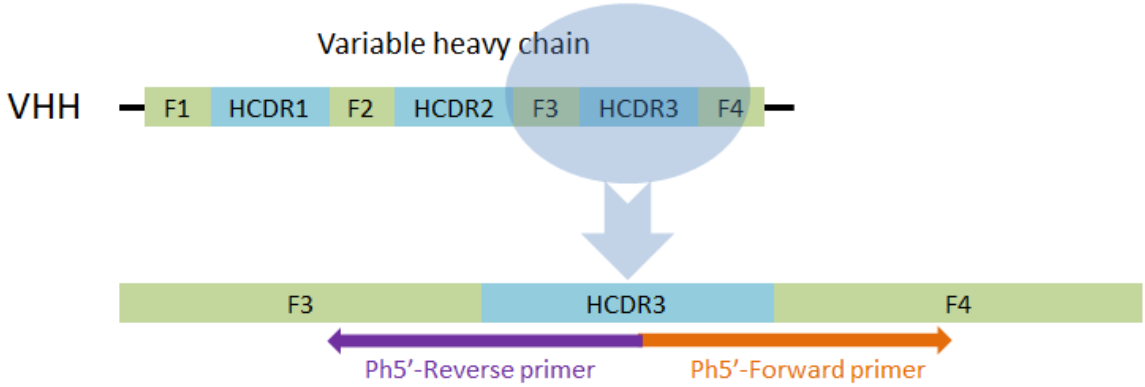


Figure 2.7. VHH synthesis strategy: Inverse PCR. Primer pairs with 5' phosphorylated ends (Ph5'-) were designed to amplify the HCDR3 and flanking framework 3 & 4 regions. The forward and reverse primers each covered ½ of the HCDR3 sequence plus 18 nucleotides of their adjacent framework regions. The inverse PCR reaction amplified the entirety of the vector, synthesising the HCDR3 sequences in the first instance. Adapted from Spiliotopoulos et al. (2015).

Mini-prepped DNA from an existing VHH clone (anti-bLAC) in the target vector was prepared at 10 ng/µl. The target vector was either the pMAL P5X or pET22B vector; refer to table 2.6. The 5' phosphorylated forward and reverse primers were prepared at a working concentration of 10 µM.

Table 2.6: The vectors in which the anti-β-lactamase VHH clone was sub-cloned into to generate a template.

Vector [1]	**Tags** [2]	***E. coli* strain** [3]
PET22B	6X HIS	BL21DE3
pMAL P5X	MBP, 6X HIS	ER2523

[1] The new expression vectors being compared using the anti-bLAC VHH clone.
[2] The detection/purification tags that will be fused with the soluble VHH antibody.
[3] The *E. coli* strain used to express each vector as recommended by the manufacturer.

The following inverse PCR reaction aimed to synthesise the HCDR3 sequence into the new P5X vector and amplify the entire construct. Similarly, to the scFv inverse PCR, the VHH synthesis PCR used a high fidelity Q5 DNA Polymerase. The inverse PCR reaction involved a final sample volume of 50 µl with the same components as described for section 2.16.1 but consistently with 25 ng of DNA Template. The inverse PCR was conducted with an initial denaturing reaction at 95°C for 5 minutes followed by 20 cycles of 95°C for 30 seconds, 52°C for 60 seconds and an extension time of 5 ½ minutes at 72°C. The following 10 cycles involved a denaturing time of 30 seconds at 95°C, 60 seconds annealing time at 55°C and an extension time of 5 ½ minutes at 72°C. The final set of 10 cycles involved a denaturing time of 30 seconds at 95°C, 60 seconds annealing time at 58°C and an extension time of 5 ½ minutes at 72°C. The reaction was then held at 72°C for 5 minutes and then chilled at 4°C until removed from the thermal cycler.

2.20.2 VHH synthesis: Amplicon visualisation

The entire 50 µl reaction was loaded onto a 1.5% (w/v) agarose/TAE gel. The agarose gel was stained with Nancy-520, and run at 120 volts for one hour with a 1 KB+ ladder. The gel was then visualised to check for PCR products with an expected size. Bands of the correct size were cut out of the gel and the DNA was purified using a Machery-Nagel Nucleospin Gel and PCR purification kit following the manufacturer's protocol. When the samples were eluted from the purification columns, two aliquots of 20 µl of pre-warmed (70°C) NE buffer were allowed to sit on the column for 5 minutes each before eluting. Gel purified products were quantified by a BR Qubit kit using 2 µl of sample.

2.20.3 VHH synthesis: Ligation reaction

The gel purified products were then used for ligations in preparation for transformation. The purified products (17 µl) were mixed with 2 µl of ligase reaction buffer (10X) and 1 µl of T4 ligase and allowed to ligate for 48 hours at 16°C, maintained by a thermal cycler. After the ligation was complete, the ligase was deactivated by heating the sample to 65°C for 10 minutes.

The ligations were buffer exchanged by placing the ligations on 0.025 µm filters. The samples were incubated on the surface of molecular grade purified water for 30 minutes before pipetting into a sterile eppendorf tube.

2.20.4 VHH synthesis: ER2523 transformation

The synthesised VHH P5X plasmids were used to transform 50 µl of electro-competent ER2523 cells using 2 µl of the ligation reaction product and cuvettes with a 0.1 cm gap. Electroporation was conducted as described under section 2.16.4.

After electroporation the cuvettes were immediately placed on ice and 250 µl of pre-warmed SOC recovery medium was added to each vial. The samples were then allowed to recover for 1 hour shaking at 37°C. Following incubation, 50 µl was spread on 2YT plates with 1% (w/v) glucose and 150 µg/ml Carbenicillin and were allowed to grow overnight at 37°C. A no-ligase negative control and a positive control using 2 µl of the plasmid pUC19 (BioLabs Product No. N3041S) were also transformed.

The following day, 4—5 colonies from each plate were grown in 10ml 2YT, 150 µg/ml Carbenicillin and 1% (w/v) glucose to $O.D_{600nm}$ 0.5—1 before being stored in 15% (w/v) glycerol stocks at -80°C. The stocks were miniprepped and the extracts were sent for Sanger sequencing to confirm the synthesised VHH HCDR3 was as expected.

2.21 Making electro-competent cells

A single TG1 or HB2151 colony was picked from a minimal agar plate and used to inoculate 10ml of 2YT. The culture was allowed to grow overnight in a shaking incubator at 30°C. The following day, the 10ml culture was added

to 1 L of 2YT and was allowed to reach an $O.D_{600nm}$ from 0.35—0.4. The culture was immediately placed on ice for 30 minutes before being split between four pre-chilled 250 ml centrifuge bottles. The cultures were then centrifuged at 3,000 xg for 20 minutes at 4°C. The supernatant was removed and the pellets re-suspended in 200ml of chilled ddH_2O per bottle. The cultures were centrifuged again at 3,000 xg for 20 minutes at 4°C. The supernatant was removed and the pellets were re-suspended in 100ml of chilled ddH_2O. The cultures were combined so there was a total of 200ml per 250ml centrifuge bottle. The cultures were then pelleted by centrifuging at 3,000 xg for 20 minutes at 4°C.

The supernatant was removed and the two pellets were re-suspended in 40ml of ice cold 10% glycerol H_2O for a total of 80ml. The solutions were allowed to chill on ice for 20 minutes before being centrifuged at 3,000 xg for 20 minutes at 4°C. The supernatant was carefully removed and the pellets were each re-suspended in 1 ml of ice cold 10% glycerol H_2O producing a total of 2 ml. The cells were divided out into 1.5ml micro-centrifuge tubes in 100 µl aliquots and stored at -80°C.

2.22 Colony PCR

A colony PCR was frequently used to confirm the presence of the scFv/VHH gene in multiple colonies, in parallel. The primers used were the same as those used for Sanger sequencing, depending on the vector.

A reaction master mix was prepared as shown below and 45 µl was added to each well of a 96-well PCR plate as required. A pipette tip was used to pick a single bacterial colony and was then dipped into an aliquot of pre-prepared PCR master mix. The pipette tip was twisted within the master mix once and was then removed. The colony was then transferred to a deep-well 96-well plate and grown in culture to maintain a colony stock.

Per sample

- 5ul 10X Standard Taq Reaction Buffer (1X)
- 1ul dNTP 10mM (200 uM)
- 4ul MgCl2 25mM (2 uM)
- 0.25ul Taq DNA Polymerase
- 32.75ul dH_2O
- 1ul Forward primer 10uM (0.2 uM)
- 1ul Reverse primer 10uM (0.2 uM)

The PCR thermal cycler program began with an initial denaturing step at 95°C for 3 minutes. After denaturation, 28 rounds were then conducted involving a denaturing step at 95°C for 30 seconds, an annealing temperature of 60°C held for 30 seconds and an extension time of 30 seconds at 68°C. The reaction was then held at 4°C.

2.23 Sanger sequencing to confirm the expected HCDR3 region

Plasmid DNA was isolated for each sample using either a Qiagen Spin Mini Prep Kit or a Sigma GenElute Plasmid Mini Prep Kit following the manufacturer's instructions. The DNA extracts were quantified using the Qubit fluorometry method and 5 µl (100 ng/µl), were sent for Sanger sequencing to confirm the HCDR3 sequence was as expected.

The Sanger sequences were annotated where applicable and the HCDR3 were identified based on VBASE alignments (http://www.vbase2.org/) and the KABAT numbering system (http://www.bioinf.org.uk/abs/#cdrid) using known flanking sequences.

The primers (3.2pmol/µl) used for Sanger sequencing were dependant on the vector in which the scFv/VHH was cloned into.

For Griffin scFv vector clones the forward 'LMB3' primer (5' CAGGAAACAGCTATGAC '3) or reverse 'FD' primer (5' GAATTTTCTGTATGAGG '3) or reverse "PHEN" primer (5' CTATGCGGCCCC ATTCA '3) was used.

For the camelid VHH pSD3 vector and PET22B vector clones, the M13R forward primer was used (5' AGCGGATAACAATTTCACACAGGA 3').

For the pMAL P5X vector clones, the forward primer (5' GGTCGTCAGACTGTCGATGAAGCC 3') or reverse primer (5' TGTCCTACTCAGGAGAGCGTTCAC 3') was used. The primer pair were used to amplify the inserts using a PCR reaction.

2.24 Phage ELISA

The specificity of isolated scFv/VHH were usually tested by a phage ELISA by incubating them with an immobilised antigen and detecting bound scFv/VHH with antibodies against phage (anti-pVIII) or a specific tag (poly HIS or MYC) downstream of the antibody fragment. The target antigens were usually immobilised on Nunc MaxiSorp® plates, or other supports if applicable

(nitrocellulose, magnetic beads, plastics etc), typically from 5 µg/ml and 100 µg/ml. Antigens were allowed to bind to the support overnight at room temperature.

Phage were grown, and their titres were determined, either as monoclonal or polyclonal preparations as previously described.

The phage supernatant and antigen-coated Nunc MaxiSorp® plates were blocked with 3% (w/v) Marvel in 1X PBS for 1 hour. The blocked phage was incubated on the Nunc MaxiSorp® plates for one hour in 100 µl aliquots. The antibodies used to detect bound phage via p8 (Anti-M13-HRP; GE Healthcare 27-9421-01) or scFv/VHH via the HIS tag (Anti-TetraHis (Mouse), Qiagen 34670; and Anti-Mouse-HRP, Invitrogen A16072) were diluted to a working concentration using the blocking solution 3% (w/v) Marvel in e.g. 1X PBS for 1 hour.

After incubating the phage, the Nunc MaxiSorp® plates were washed 3X with 1X PBS and 3X with 1X PBST. The plates were patted dry and 100 µl of blocked antibody was added to each well (in one or two steps depending on the detection system) and allowed to incubate for one hour. The Nunc MaxiSorp® plates were washed 3X with 1X PBS and 100 µl of 3,3',5,5'-Tetramethylbenzidine (TMB) substrate (Thermo Scientific 34022) was added to each well and allowed to develop for no more than 30 minutes.

Once the colour had developed (typically 10-15 minutes), the reaction was stopped with 100 µl of 2M sulphuric acid. The absorbance was measured at the 450nm wavelength and compared against positive and negative controls. The positive control was most commonly an anti-BSA scFv that was supplied alongside the library. The BSA antigen was usually immobilised at 20 µg/ml.

2.25 pMAL expression system

The pMAL™ Protein Fusion and Purification System (NEB #E8200) enables the fusion of a target protein with maltose binding protein (MBP) which can then be purified in a single step on amylose. The MBP protein can also increase the solubility and expression of many proteins when expressed in Ecoli. The pMAL vectors included a sequence coding for the factor Xa recognition site, enabling cleaving of the antibody fragment from the chaperone protein post purification. The kit also provided a murine anti-MBP

antibody. The ampicillin resistant vectors provided by the pMAL kit permit fusion proteins to be secreted either into the cytoplasm (C5X vector), or be exported to the periplasm (P5X). The C5X vector had a deletion of the malE signalling sequence, preventing exportation of the fusion proteins to the periplasm, whereas the P5X vector possessed a normal copy of the malE gene.

The pMAL expression kit also provided an *E. coli* strain ER2523 which is naturally deficient in the Lon protease (degrades mutant/abnormal proteins), a mutant in the ompT protease (membrane bound protease) and does not have the T7 RNA polymerase gene.

The P5X pMAL vector was chosen for all sub-cloning techniques used within this study to ensure the antibody-MBP fusions were exported to the periplasm. The P5X vector is a low copy number plasmid (~20 copies per cell) and contains multiple cloning sites, of which NCOI and NOTI were utilised for this study.

An scFv or VHH gene was cloned into the pMAL vector downstream of the malE gene encoding MBP so that the MBP would be a N-terminal fusion of the cloned antibody.

2.25.1 pMAL method

Antibody genes were subcloned into the pMAL P5X vector by digesting both with high fidelity restriction enzymes, ligating the products using T4 ligase and transforming them into electrocompetent ER2523 cells. To begin, 500 ng of the samples (P5X vector, Griffin pHEN2 vector, VHH pSD3 vector and/or amplicons) were double-digested using 10 units of high fidelity NCOI (NEB, product no. R3193S) and NOTI (NEB, product no. R3189S) restriction enzymes, in 1X Cutsmart buffer (NEB, product no. B7204S) up to a total of 20 µl.

The samples were vortexed and incubated at 37°C for 1 hour. The restriction enzymes were deactivated by incubating the samples at 80°C for 20 minutes. The digestion process was confirmed by analysing the digested samples on a 1% agarose gel (90 volts for 1 hour, 30 minutes) and checking for the expected band sizes.

Antibody fragment bands of the correct size were excised from the agarose gel, including the cut P5X vector, and gel purified using a Machery-Nagel Nucleospin Gel and PCR purification kit following the manufacturer's protocol. Eluted products were quantified by Qubit and 40ng of the gel purified, cut P5X vector was mixed with 20ng of the gel purified, cut antibody fragment. The samples were made up to 10 µl with dH_2O and mixed with 10 µl of 2X quick ligation buffer (NEB; product no. M2200S) and 1 µl of quick T4 DNA ligase (NEB; product no M2200S). The samples were mixed by pipetting and left to incubate at room temperature for 10 minutes. After incubation, the samples were buffer exchanged by pipetting onto 0.025 µm filters and incubated on the surface of molecular grade purified water for 30 minutes before pipetting back into sterile eppendorf tubes.

For transformation, 2 µl of the ligation product was mixed with 50 µl of electro-competent ER2523 cells, transferred to cuvettes with a 0.1 cm gap and subjected to electroporation. Electroporation was conducted using a BioRad GeneP Pulsar Xcell Electroporator using a pre-set exponential decay protocol for bacteria as previously described under section 2.16.4.

After electroporation cuvettes were immediately placed on ice and 250 µl of pre-warmed SOC recovery medium was added to each vial. The samples were then allowed to recover for 1 hour shaking at 37°C. Following incubation, 100 µl was spread on 2YT plates with 1% (w/v) glucose and 150 µg/ml Carbenicillin and were allowed to grow overnight at 37°C.

Colonies were picked the following day, grown to produce glycerol stocks and then checked for the expected insert by Sanger sequencing.

2.25.2 pMAL subcloning method

Antibody genes that proved difficult to subclone into other vectors were first cloned into an optimised sequencing vector (pCR™4-TOPO®) using a TOPO® TA Cloning® Kit (Product no. K457501). The purified amplicons were placed on ice and 5 µl of each amplicon was then incubated with 0.1 µl of a Taq Polymerase (NEB; Product no. M0273S), 2.5 µl of 10X Taq buffer ($MgCl_2$ free), 2.5 µl of 1mm dATP and 14.9 µl of dH_2O. The samples were mixed well by gentle pipetting followed by an incubation step of 15 minutes at 72°C.

After incubation, the single-A tailed amplicons were transferred to ice and used immediately for the subsequent cloning reaction into the pCR™4-TOPO® vector.

The single-A tailed amplicons were cloned into the new pCR™4-TOPO® vector by mixing 4 µl of each amplicon with 1 µl of the TOPO® TA Cloning® Kit's salt solution and 1 µl of the pCR™4-TOPO® vector. The solutions were periodically mixed by gentle pipetting for 5 minutes at room temperature. After incubation, 2 µl of the mixture was added to 50 µl of Ecoli Top10 cells and was mixed by gentle stirring. The cells were placed on ice for 5 minutes and then heat shocked at 42°C for 30 seconds. The cell were immediately placed on ice and each sample was incubated with 250 µl of SOC medium provided by the kit. The cells were allowed to grow for 1 hour at 37°C in a shaking incubator. A 50 µl aliquot of each sample was spread onto LB agar plates containing 50 µg/ml kanamycin and colonies were left to grow overnight at 37°C in a static incubator.

2.26 Soluble antibodies

Within this study, soluble antibody expression was induced using IPTG regardless of the vector or antibody fusion. The presence of IPTG, a molecule similar to allolactose, drove expression of the lac promotor gene upstream from each desired gene in this study's cases, ultimately increasing expression of the proteins. The proteins were extracted either from whole cell or from periplasmic extracts and analysed via functional ELISA and western blotting.

2.26.1 Soluble antibody expression and extraction

The bacterial glycerol stocks used to inoculate 50 ml 2YT cultures containing 0.2% glucose and 150 µg/ml Ampicillin or Carbinicillin. The cultures were grown to an ~$O.D_{600nm}$ of 0.5. The cultures were centrifuged at 3,000 xg for 10 minutes and the cell pellets were re-suspended in 50ml 2YT containing 150 µg/ml Ampicillin or Carbenicillin and 1mM IPTG. The cultures were allowed to grow overnight at 30°C in a shaking incubator. The next day, 1ml of the culture was retained in an eppendorf for whole cell extraction. The remaining 49ml of the culture was used for periplasmic extraction.

For whole cell extraction, the 1ml sample was pelleted at 3,000 xg and the supernatant was removed. The cell pellet was resuspended in 300 µl of Bugbuster protein extraction reagent containing 1 protease inhibitor tablet per 50ml and 25 U/ml Benzonase. The suspension was incubated at 37°C for 30 minutes in a shaking incubator.

For periplasmic extraction, the 49ml pellets were resuspended in 5ml of 1X PBS and re-centrifuged at 3,000 xg for 10 minutes at 4°C. The cell pellet was resuspended in 1ml of periplasmic buffer (20mM Tris-HCl, 20% sucrose and 1 protease inhibitor tablet per 50ml (Thermo Scientific Product No. A32965). The suspension was transferred to 1.5 ml eppendorfs and kept on ice for 20 minutes. The cells were centrifuged at 6,400 xg for 5 minutes at 4°C and the pellets were re-suspended in 1.5 ml of osmotic shock solution (5mM $MgSO_4$). The suspensions were placed on ice for 20 minutes before pelleting the solution at 8,000 xg and transferring the supernatant to a fresh eppendorf.

2.26.2 Soluble antibody western blots

For western blotting, the periplasmic and whole cell extracts were loaded onto pre-cast, SDS-PAGE gels after the reduction step below. The wells of the gel were rinsed with running buffer (MOPS) before loading samples. A Magic Mark protein standard ladder (7 µl) was loaded alongside the samples. For the whole cell extract, 5 µl of 2X LDS containing 5% beta-mercaptoethanol was mixed into 15 µl of extract. For the periplasmic extract, 15 µl was mixed with 5 µl of 2X LDS containing 5% beta-mercaptoethanol. The samples were then boiled for 3 minutes before loading 5 µl or 20 µl of the whole cell or periplasmic extracts respectively.

The protein gels were run at 150 volts for 90 minutes and transferred to PVDF membranes at 30 volts for 75 minutes as described in methods section 2.3. The PVDF membranes were blocked with 3% (w/v) Marvel in 1X PBST or 1M Tris buffer (pH 7.5) containing 0.1% (v/v) Tween 20 overnight. The following day, the PVDF membranes were incubated with the primary antibody for 2 hours, on a shaking platform.

The membranes were washed three times with 1XPBST or 1M Tris buffer (pH 7.5) containing 0.1% (v/v) Tween 20 for 5 minute intervals, changing the buffer at each interval. The membranes were then incubated with the secondary antibody diluted with 3% (w/v) Marvel in 1X PBST or 1M Tris buffer (pH 7.5) containing 0.1% (v/v) Tween 20 for 2 hours, on a shaking platform. The membrane was washed three times with 1X PBST or 1M Tris buffer (pH 7.5) containing 0.1% (v/v) Tween 20 for 5 minute intervals, changing the buffer at each interval. Each membrane was transferred to a new container and rinsed repeatedly with 1.5 ml of the HRP substrate ECL. The membranes were sealed in plastic and visualised using BioRad Chemi Doc.

2.26.2 Soluble antibody ELISA

For the soluble antibody ELISAs, the target antigens were immobilised overnight. The next day, the plates and soluble antibody extracts were blocked for 1 hour with 3% (w/v) Marvel in 1XPBS or 1M Tris buffer (pH 7.5). After blocking, 100 µl of the soluble antibody extracts were incubated on each plate against their intended antigens and a negative control (in triplicate). The plates were then washed as described for phage ELISAs using 1XPBS or 1M Tris buffer (pH 7.5), with or without 0.1% (v/v) Tween.

The plates were patted dry before incubating the primary antibodies for 1 hour at 18°C. The plates were washed as previously described and the secondary antibody was then incubated in each well for 1 hour at 18°C. The plates were washed again and 100 µl of the TMB substrate was incubated in each well for 15 minutes at room temperature. The reaction was then stopped with 100 µl of 2M sulphuric acid per well and the absorbance readings were taken at 450nm. The stopped solutions on the TiO_2 plates were transferred to Nunc MaxiSorp® plates before taking the absorbance values.

An anti-beta-lactamase VHH clone in the phagemid pSD3 vector was often used as a positive control. Beta-lactamase (Sigma product no. P0389-1KU) was immobilised on a Nunc MaxiSorp® plate at 20 µg/ml overnight at 18°C.

Chapter 3: Identifying antibodies (scFv) against the active phytase conformation using next generation sequencing

The present chapter aimed to devise an antibody phage display strategy that would facilitate the identification of novel antibodies, specific to the active conformation of the enzyme phytase. Furthermore, the antibodies would be isolated based on next generation sequencing data provided after only two panning rounds. Novel immunoreagents isolated would be used to distinguish between the enzyme's active and denatured conformation.

3.1 Determining the size of the Quantum Blue and Megazyme phytase

For comparison purposes, three different volumes (10, 7.5 and 5 µl) of the Quantum Blue phytase and Megazyme Total Phosphorous kit phytase were run on the same denaturing protein gel (refer to methods section 2.2). The Quantum Blue phytase consistently produced two major bands at ~45kDa and ~20kDa (figure 3.1). The Megazyme kit phytase consistently displayed a molecular mass of ~55kDa.

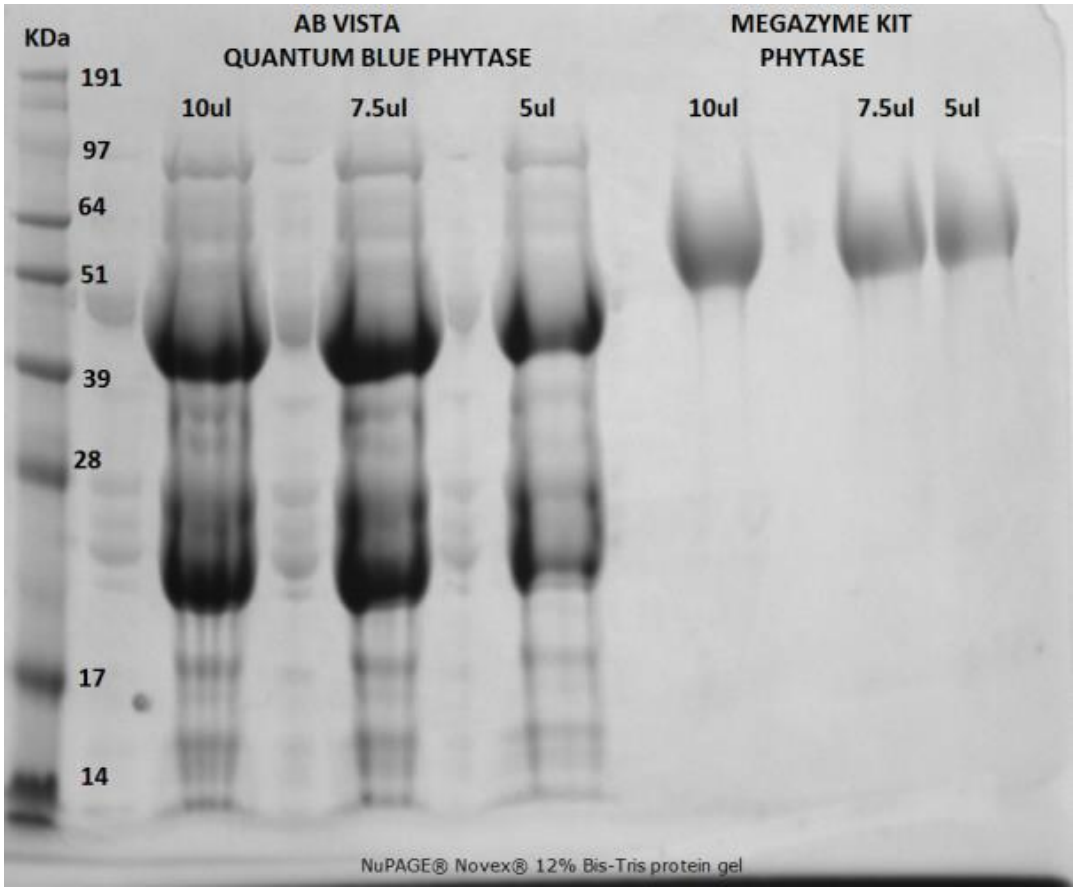


Figure 3.1. Quantum Blue phytase and Megazyme phytase molecular weight analysis. Quantum Blue phytase (supplied by AB-Vista) and Megazyme phytase (from the Megazyme Total Phosphorous kit) were analysed (5—10 µl each as indicated) on a denaturing 12 % Bis-Tris protein gel to estimate molecular weight of the enzymes. Proteins were stained with Coomassie Brilliant Blue. Molecular mass markers are shown (SeeBlue Plus 2 protein marker).

3.2 Determining the concentration of Quantum Blue & Megazyme phytase

The concentrations of protein in the Quantum Blue phytase and the Megazyme kit phytase samples were determined using a Bradford assay (refer to methods section 2.4). Based on the line of best fit generated using the BSA standard curve (figure 3.2) the Quantum Blue phytase, undiluted and with an absorbance value of 0.9, had a predicted concentration of approximately 2.1mg/ml. With an absorbance value of 0.13, the Quantum Blue phytase sample diluted 1/10 supports the predicted original concentration of approximately 2.1mg/ml. The undiluted Megazyme phytase sample had an absorbance reading of 0.8, indicating a concentration of approximately 1.9 mg/ml.

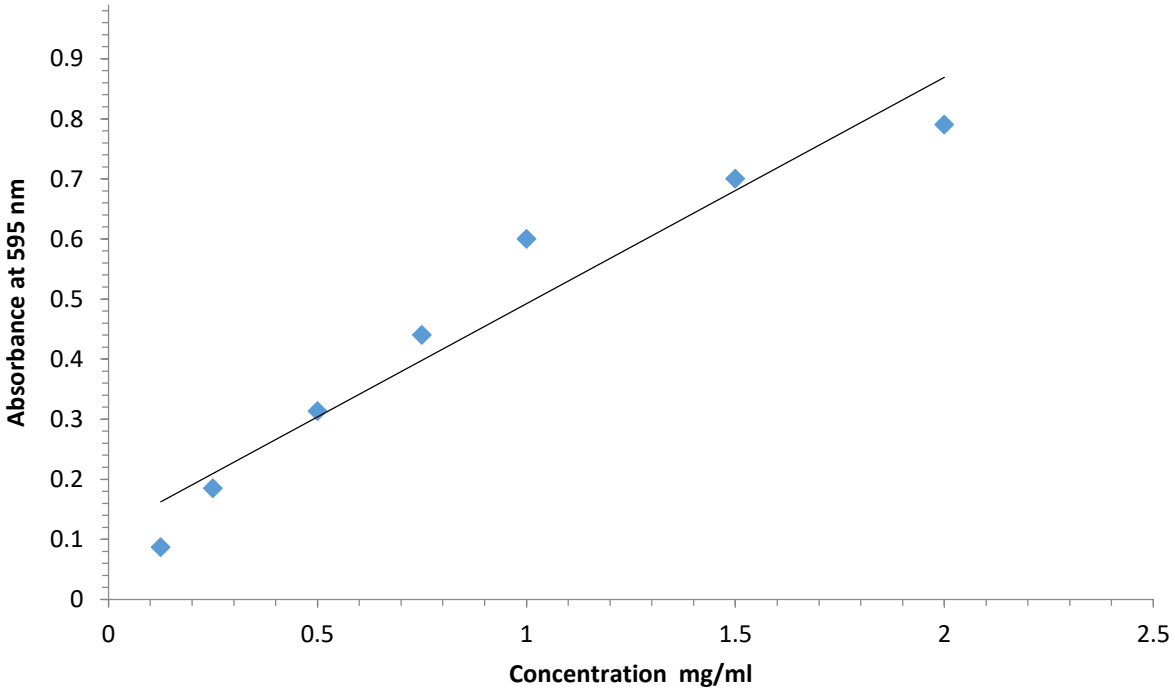


Figure 3.2. Quantum Blue phytase and Megazyme phytase concentration assay. The concentrations of the Quantum Blue phytase (supplied by AB-Vista) and Megazyme phytase (from the Megazyme Total Phosphorous kit) were predicted using a Bradford assay. The absorbance readings (595 nm) of seven BSA standards were taken to generate a standard curve which was then used to predict the concentrations of the two phytases.

3.3 Establishing conditions to produce heat-denatured or native phytase prior to panning

Using the Megazyme phytic acid total phosphorous kit (refer to methods section 2.5), treatment conditions were established for denaturing the phytase. The active and denatured phytase would serve as the immobilised targets for NGAPD. In addition to completely denaturing phytase for the negative control, for this chapter the active phytase was also treated in order to denature any potential contaminants within the Quantum Blue extract whilst retaining phytase activity. The theory being that the differences between the samples would then be the native or denatured forms of the enzyme.

Using a heat block, 100 µl aliquots of Quantum Blue phytase were treated at a variety of temperatures and incubation times in order to establish conditions that would partially, or completely, denature the enzyme or its contaminants. The Quantum Blue phytase had been engineered to be highly thermostable and would withstand the high temperatures used to denature the contaminants. The activity of the treated enzyme was then assessed using the Megazyme phytic acid total phosphorous kit.

3.3.1 Quantum Blue phytase activity assessments

To begin, the Megazyme phytic acid total phosphorous kit was used to indirectly measure phytase activity in order to establish an activity curve. The last set of activity assays aimed to determine partial and full denaturing conditions for Quantum Blue phytase in preparation for NGAPD. Samples intended to measure background levels of phosphorous within each substrate were also included (no phytase samples), but their absorbance values were not subtracted from the other readings within the same experiment unless stated otherwise.

The first activity assay used two sources of phytic acid (1g of either Megazyme kit oat powder or Sigma's sodium phytate) and two phytases (Megazyme kit phytase and Quantum Blue phytase) in an effort to test run the kit protocol. Samples were run in triplicate to produce an average absorbance reading and the enzymes were used at stock concentration (table 3.1).

Table 3.1: Phytase activity assay 1 - Megazyme total phosphorous kit

Phytase and sodium phytatesource[1]	Average absorbance[2]	
	Background phosphorous[3]	Digested[1]
Megazyme kit phytase & kit oat powder	0.23	0.82
Megazyme kit phytase & sodium phytate	1.53	>3.0
Quantum Blue phytase & kit oat powder	0.23	0.79
Quantum Blue phytase & sodium phytate	1.53	>3.0

[1] The activity of both the Quantum Blue phytase and Megazyme kit phytase were compared by digesting both the sodium phytate (Sigma) and the Megazyme kit's oat powder substrates.
[2] Absorbance at 655 nm was measured and an average absorbance value was recorded
[3] The substrate's background level of phosphorous was measured using a no-phytase control

The megazyme kit and Quantum Blue phytases showed similar levels of activity when incubated with each source of phytic acid. After using 1g of sodium phytate the absorbance readings exceeded the scale of the spectrophotometer (>3.0) when digested with both phytase sources, but the Sigma phytic acid source had a considerably higher level of background phosphorous to begin with. This was likely due to the Sigma source being a highly concentrated source of phytic acid in comparison to the concentration extracted from the oat powder.

The purpose of the second activity assay (table 3.2) was to determine a weight of Sigma's sodium phytate that could be used without exceeding the spectrophotometer absorbance scale post-digestion. The enzyme alkaline phosphatase was also removed from the assay and all future experiments in order to solely investigate Quantum Blue phytase activity. Seven different weights of sodium phytate were incubated with Quantum Blue phytase at stock concentration.

Table 3.2: Phytase activity assay 2 - Megazyme total phosphorous kit

Sodium phytate weight[1]	Average absorbance[2]	
	Background phosphorous[3]	Digested
1g	1.375	>3.0
0.75g	1.235	>3.0
0.5g	0.905	2.845
0.25g	0.405	2.37
0.1g	0.2	1.595
0.01g	0.125	0.35
0g	0.115	0.225

[1] The Quantum Blue phytase activity was analysed using different weights of the sodium phytate substrate (0—1g as indicated).
[2] Absorbance at 655 nm was measured and an average absorbance value was recorded
[3] The substrate's background level of phosphorous was measured using a no-phytase control

The background level of free phosphorous reduced by 0.2 after the removal of alkaline phosphatase in comparison to the first activity assay (1g of substrate). The 0.1g weight of Sigma's sodium phytate had a digested absorbance value of 1.3 after the background phosphorous values were subtracted. This value was considered the most acceptable given the restriction of the spectrophotometer scale (maximum absorbance value of 3) and the difference between the digested and background phosphorus values.

For the third assay, only 0.1g of Sigma's sodium phytate was used but in combination with a 5-fold dilution series of the Quantum Blue phytase to generate a standard curve of activity (figure 3.3). The starting phytase concentration was 65 µg/ml and reduced as far as 3.328×10^{-5} µg/ml. A no phytase control was also included, the absorbance value of which was subtracted from all plotted values.

The 5-fold dilution series of Quantum Blue phytase resulted in a detectable reduction in absorbance values, and thus enzyme activity, until the limit of detection (LOD) was reached at the phytase concentration 0.02 µg/ml with an absorbance value of 0.2. The peak of the activity curve was reached at the phytase concentration of 13 µg/ml.

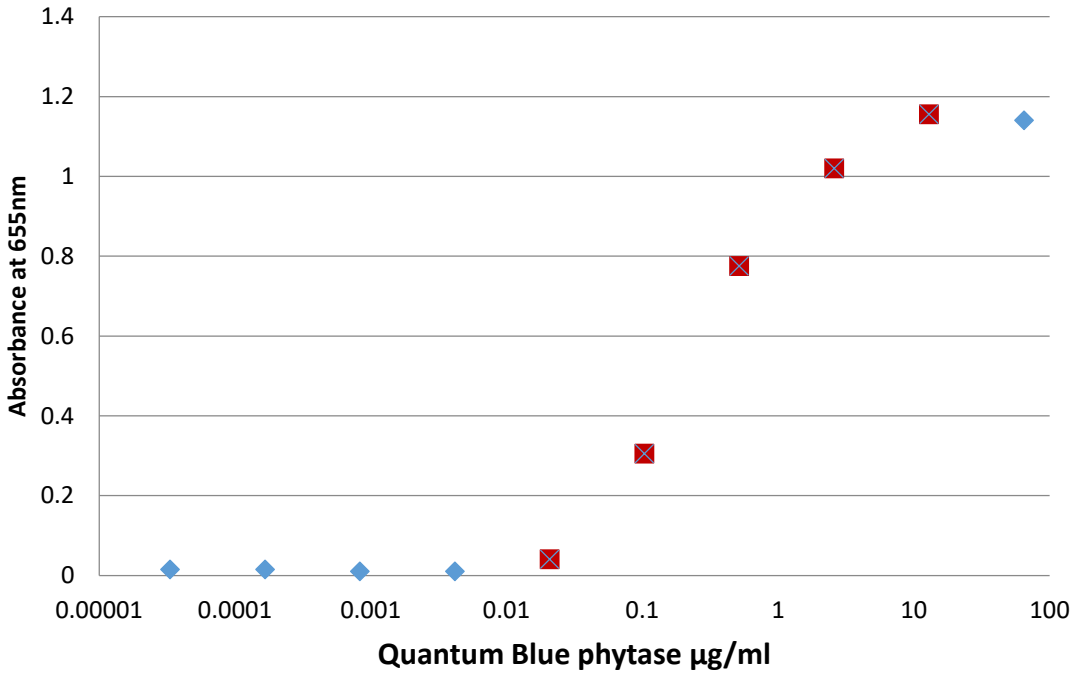


Figure 3.3: A standard curve of Quantum Blue phytase activity. The activity of a 5-fold dilution series of Quantum Blue phytase was measured with the Megazyme total phosphorous kit assay using 0.1g of sodium phytate substrate. Absorbance readings at 655 nm were taken and plotted against enzyme concentration using a logarithmic scale. The background free phosphorous levels were subtracted from all plotted values. The enzyme concentrations in the linear range of this assay are shown (red).

3.3.2 Determining denaturing conditions prior to panning

Following the establishment of an activity curve, the purpose of the final activity assays was to determine conditions that would completely or partially denature the Quantum Blue phytase. The treatments were necessary in order to maintain the positive and negative NGAPD panning targets i.e a sample that was fully denatured (negative control sample) and a sample which was treated to denaturing conditions but retained enzyme activity (active target sample).

For all of the activity assays described in this section, 0.1g of Sigma's sodium phytate was used as the source of phytic acid and the phytase used was the Quantum Blue phytase developed by AB Vista.

The first heat-denaturing assay used two concentrations of Quantum Blue phytase (100 µl aliquots at 2.1 mg/ml and 13 µg/ml) in an attempt to denature the enzyme by incubating it at 60°C, 80°C or 100°C for 15 minutes

in a heat block (table 3.3). A no phytase control was also included, the absorbance value of which was subtracted from all plotted values.

Table 3.3: Denaturing phytase assay 1 - Megazyme total phosphorous kit

Sample conditions		Average [1] absorbance
Quantum Blue phytase concentration (µg/ml) [3]	**Incubation temperature**	**Digested sample** [2]
2.1 mg/ml Positive Control	No treatment	1.4
2.1 mg/ml Incubated for 15 minutes	60°C	1.4
2.1 mg/ml Incubated for 15 minutes	80°C	1.4
2.1 mg/ml Incubated for 15 minutes	100°C	1.4
13 µg/ml Positive Control	No treatment	1.3
13 µg/ml Incubated for 15 minutes	60°C	1.3
13 µg/ml Incubated for 15 minutes	80°C	1.3
13 µg/ml Incubated for 15 minutes	100°C	0.0

[1] Absorbance at 655 nm was measured and an average absorbance value was recorded.
[2] The substrate's background level of free phosphorous was measured using a no-phytase control. The absorbance value was then subtracted from the digested sample's absorbance reading.
[3] The Quantum Blue phytase was used at two different concentrations and analysed with 0.1g of sodium phytate substrate under a range of incubation temperatures and times.

At an enzyme concentration of 13 µg/ml, an incubation step at 100°C for 15 minutes reduced the absorbance value to background levels suggesting that the enzyme had been denatured. The higher, stock enzyme concentration (2.1 mg/ml) appeared to be comparatively resistant to heat denaturation.

The second denaturing activity assay involved six different concentrations of Quantum Blue phytase incubated at 100°C for 15, 30 and 60 minutes. A fully active control was also included for comparison. The purpose of the assay was to identify concentrations of phytase that could be fully denatured or partially treated to retain phytase activity but denature any other possible contaminants (figure 3.4). A no phytase control was also included, the absorbance value of which was subtracted from all plotted values.

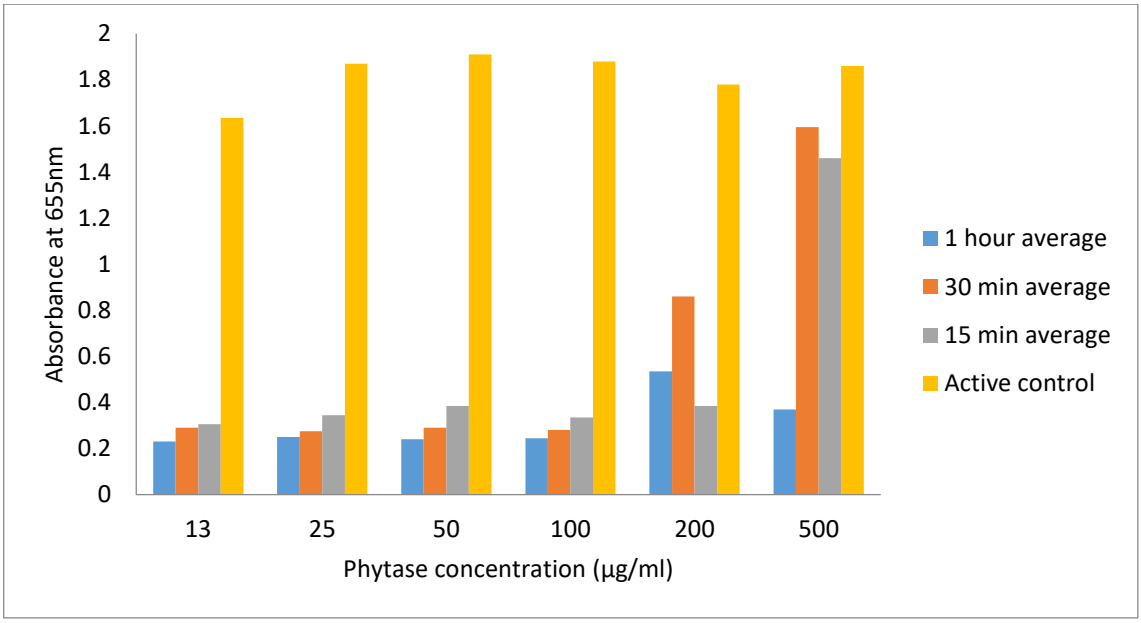


Figure 3.4: Quantum Blue phytase denaturation assay. Phytase activity following 15, 30 or 60 minutes of incubation time at 100°C. The average absorbance readings at 655 nm were plotted against phytase concentration. Enzyme activity was lost after 30 minutes incubation time up to a phytase concentration of 100 µg/ml.

After being treated for 15 minutes to one hour at 100°C, the phytase concentrations 13—100 µg/ml had final absorbance values from 0.3—0.38 suggesting that the enzymes had been fully denatured and could no longer liberate phosphates from the phytic acid molecule. At 1hr treatment with 200—500 µg/ml the phytase activity was relatively low. With 15 and 30 minutes treatments, phytase activity was retained with 200 or 500 µg/ml.

An additional activity assay was conducted to ensure that the denatured phytase at 100 µg/ml could not refold and regain activity. The phytase was treated in 100 µl aliquots for 60 minutes at 100°C the day before and a separate aliquot on the day of the activity assay. A free phosphorous background sample and an active control at 100 µg/ml were also included. The denatured phytase treated the day before was kept at 4°C overnight (the same storage conditions for the active enzyme). The absorbance values generated by the denatured phytases (denatured one day apart) were the same, suggesting that the enzyme was not able to refold under these conditions (data not shown).

The final denaturing activity assay aimed to establish conditions at which a Quantum Blue phytase concentration of 200 µg/ml could be heat treated before enzyme activity was affected. The final assay involved treating 100 µl

aliquots of Quantum Blue phytase at 100°C for 1, 2, 3, 5, 7.5, 10 and 60 minutes before conducting the activity assay (figure 3.5). A no phytase control was also included, the absorbance value of which was subtracted from all plotted values.

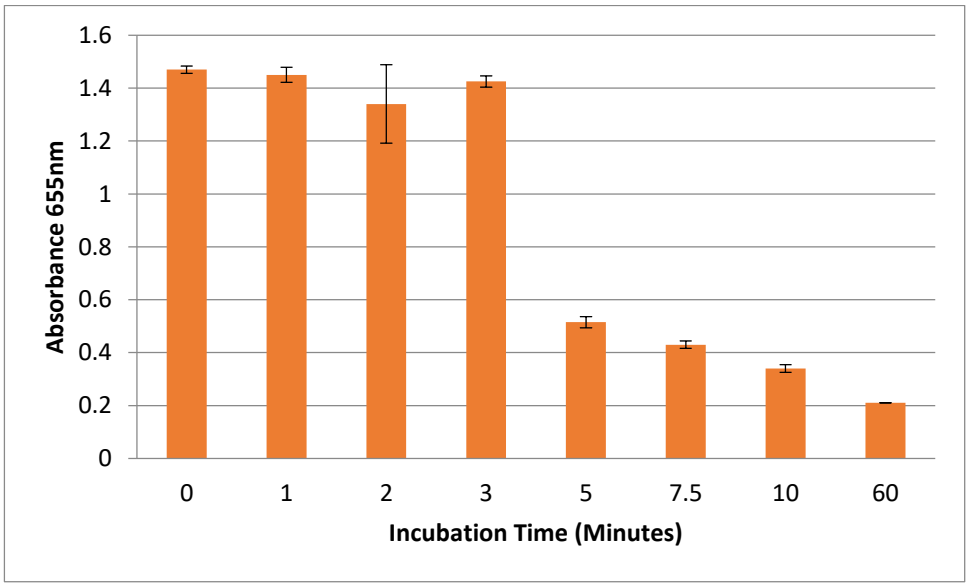


Figure 3.5: Partial denaturation assay for Quantum Blue phytase. Phytase activity (200 µg/ml) was measured following 1—60 minutes of incubation time at 100°C. The mean absorbance readings at 655 nm were plotted against incubation time.

After five minutes at 100°C, the drop in absorbance values suggested a considerable loss of phytase activity.

Based on the results from the enzyme denaturation assays, the heat treatment and enzyme concentrations for the positive and negative panning targets were chosen. To produce the negative control panning target, where the phytase would be fully denatured, the enzyme concentration of 100 µg/ml was selected and treated for one hour at 100°C in 100 µl aliquots. For the panning target, where the phytase enzyme would be heat-treated but retain enzyme activity (but the treatment would be likely to denature contaminant proteins), the treatment concentration of 200 µg/ml was treated for 3 minutes at 100°C. Treated samples were then diluted as required for immobilisation and competitive elution steps.

3.3 Activity assessment for phytase preparations used for panning

The expected activity of the active (positive) and denatured (negative) Quantum Blue phytase samples was confirmed using the Megazyme total phosphorous kit. Following treatment, aliquots of the active phytase (200 µg/ml) were diluted 20X (10 µg/ml), 100X (2 µg/ml), 500X (0.4 µg/ml) and 2500X (0.08 µg/ml) to bring enzyme activity and their subsequent absorbance readings within the scale of the standard activity curve. The previously established activity curve (figure 3.3) was also repeated within the same assay, and treated samples were plotted alongside the new curve (figure 3.6).

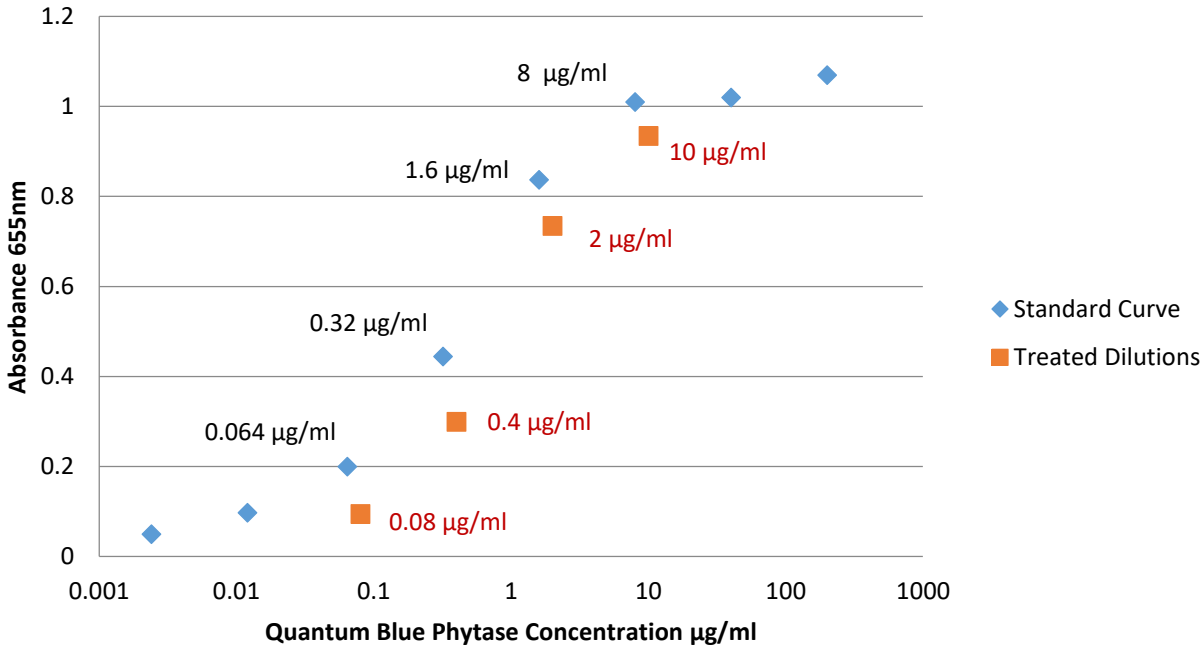


Figure 3.6: Quantum Blue standard activity curve. Untreated and partially treated samples absorbance readings were plotted on a logarithmic scale. A free phosphorous absorbance value was subtracted from the final values. An additional denatured phytase sample, at a concentration of 100 µg/ml, was also assessed for activity, and gave an absorbance reading of 0.1.

The treated 200 µg/ml (20X, 100X, 500X and 2500X) dilutions had slightly lower absorbance values than their untreated, fully active counterparts indicating a slight loss in activity due to the 3 minute treatment at 100°C. The partially treated phytase still retained activity as intended. The denatured sample (100 µg/ml) had a final absorbance value of 0.11 after the

background phosphorous absorbance value had been subtracted. Although the denatured sample had a slightly higher absorbance value than the background control it is still considered to be denatured as its fully active equivalent would have been at the peak of the activity curve with an absorbance value of ~1, indicating over 99.9% loss of activity. An absorbance value of 0.11 resembled a phytase concentration of ~0.012 µg/ml, thus 0.012 µg/ml divided by the actual concentration 100 µg/ml, multiplied by 100 for a percentage, represented 0.012% of the original concentrations activity level.

This assay confirmed that the treated-active (positive) and denatured (negative) samples had been treated successfully and were used in the following NGAPD experiment.

The active and denatured preparations of phytase were used in conjunction with the Griffin.1 library (Griffiths et al., 1994), a collection of filamentous bacteriophage displaying scFv, to select for scFv that could bind to the active conformation of Quantum Blue phytase. The typical methods for growing the phage display libraries, phage precipitation and panning rounds are described in methods; refer to sections 2.8—2.14).

3.4 Panning rounds 1 and 2

The Griffin scFv library ($5x10^{13}$ CFU/ml) was panned against 20 wells of treated-active phytase and bound phage were eluted competitively (10 wells) using an excess of 10X the bound antigen or by direct addition of the *E. coli* strain TG1 (10 wells). The phage were pooled for each elution method and produced output titres of $4x10^{7}$ CFU/ml for competitive elution and $7x10^{8}$ CFU/ml for the TG1 rescue. For round 2 panning, $5x10^{14}$ CFU/ml (competitive eluted) and $1x10^{14}$ CFU/ml (TG1 eluted) were used and panned against two sets of 5 repetitions (8 wells each) of treated-active Quantum Blue phytase and treated-denatured Quantum Blue phytase. The first set was eluted competitively and the second set was eluted using direct addition of TG1.

Every 8 wells of eluted phage were pooled to produce 5 samples per plate. Reinfected sub-libraries were grown overnight on bioassay plates, scraped the following day to produce new glycerol stocks and dsDNA was extracted in preparation for sequencing.

3.5 Preparation of DNA for Ion Torrent sequencing

The 20 dsDNA extracts were relabelled 1—20 to reflect their barcode primer. The samples were prepared for sequencing using two PCR steps (rounds 1 & 2) which were then checked using agarose gels as described in methods section 2.16. Round 1 PCR products were expected to be around 300 bp and round 2 PCR products were expected to be around ~350 bp. Round 2 PCR products were quantified by fluorometry using Qubit and concentrations ranged from 6—22ng/µl. An equal contribution by weight (ng) from each sample was used to produce a total pool of 1000 ng DNA which was purified and quantified using fluorometry, reading a final concentration of 23.8 ng/µl. The sample (200 ng) was sent for Ion Torrent sequencing at the University of Pennsylvania.

3.6 Data analysis and scFv rescue

The NGS data was analysed as described in methods section 2.18, to identify and quantify enriched scFv within each experiment. The data analysis process was summarised in figure 2.5 "stage 1" in methods, describing how each screening round was ranked and selected by number of replicates seen in and 'Z score' for each HCDR3.

The data analysis produced two primary data sets based on the elution method used: antigen eluted or TG1. A third primary data set was also produced which ignored the elution method and combined all negative samples to produce a pool.

The first data set considered scFv that were eluted competitively using active phytase. Using a Z score threshold of 10, 5 or 2 there were 32, 67 or 205 scFvs respectively, that were seen in at least 2 of 5 positive screenings. One scFv was seen in 3 of 5 positive sample replicates with a Z score of ≥10 and was selected for rescue. Six scFv were selected for rescue from the sub data set with a Z score threshold of 5. All six scFv had been seen in 2 of 5 positive experiments and contained no stop codons. Finally, two scFv with Z scores of ≥2 were seen in 3 of 5 positive sample replicates and were also selected for rescue. Of the nine scFv selected for rescue from the competitively eluted data set only two had also been seen in the negative pool.

The second data set considered scFv that had been eluted using direct infection of the bacterial host TG1. With a Z score threshold of 10, 5 or 2 there were 1, 11 or 297 scFvs respectively, that were seen in at least 2 of 5 positive sample replicates, respectively. Similar to the first data set, only one scFv had a Z score of ≥10 that was seen in more than one positive experiment; that scFv was selected for rescue. The Z5 threshold identified three scFv that had been seen in more than one positive experiment and all of them were chosen for rescue. The final sub data set had a Z score threshold of ≥2 and identified six scFvs that had been seen in four of the five positive experiments. All six were selected for rescue. For the 10 scFvs that were chosen for rescue for the TG1 eluted group, all of them had been seen within the negative pool.

For the third data set, for which the elution method was disregarded, five scFvs with a Z score of ≥2 were found in 4 or more of 10 positive experiments. The experiments that met the Z score threshold were evenly split between the two elution methods in each case, however the generic TG1 elution method often achieved the higher Z scores. Although all five scFvs had also been seen within the negative pool they were all selected for rescue.

If clones were also seen in their respective negative pools, the total read counts were at least 10 fold lower than that of the positive replicates that achieved a Z score of ≥2.

A total of 24 scFv were chosen for rescue (table 3.4). The scFv rescue strategy was described in methods section 2.19. For each rescue, 5' phosphorylated inverse PCR primers were designed based on each scFv's specific HCDR3 region, and paired with phosphorylated degenerate primers that annealed to the framework 4 (FRW4) region (table 3.5).

An alternative FRW4 primer (FRW4-Shift) was paired with the unique HCDR3 primer if the HCDR3 region consisted of ≤5 amino acids in order to extend the HCDR3 primer into the FRW4 region and improve primer annealing.

Table 3.4: The 24 scFvs selected for phagemid rescues after panning against active phytase

ScFv AA sequence[1]	Elution method | Z Score threshold[2]	Selection reasoning[3]
CGMSAVEF	Combined | Z2	Enriched in both elution methods
LKVqNS	Combined | Z2	Enriched in both elution methods
PAGWLRNP	Combined | Z2	Enriched in both elution methods
RLDNS	Combined | Z2	Enriched in both elution methods
RSTNWW	Combined | Z2	Enriched in both elution methods
SRVYqFDY	TG1 Eluted | Z2	Seen in most positive reps
QPASQ	TG1 Eluted | Z2	Seen in most positive reps
MRPQRS	TG1 Eluted | Z2	Seen in most positive reps
LLAT	TG1 Eluted | Z2	Seen in most positive reps
EVRADSTE	TG1 Eluted | Z2	Seen in most positive reps
CPTGRAT	TG1 Eluted | Z2	Seen in most positive reps
SLTW	Competitive Eluted | Z2	Seen in most positive reps
RESV	Competitive Eluted | Z2	Seen in most positive reps
MPYYRLPAMSA	Competitive Eluted | Z10	Only scFv seen in 3 positive reps
QLLAE	TG1 Eluted | Z10	Only scFv seen in 2 positive reps
ESPFWSGYV	Competitive Eluted | Z5	High Z score threshold, no Stops
HAMDGRGIAI	Competitive Eluted | Z5	High Z score threshold, no Stops
HGVGQAQRP	Competitive Eluted | Z5	High Z score threshold, no Stops
IGLDMRNALA	Competitive Eluted | Z5	High Z score threshold, no Stops
NKGCVPEVNKM	Competitive Eluted | Z5	High Z score threshold, no Stops
TNYVPPTGG	Competitive Eluted | Z5	High Z score threshold, no Stops
QSMHYTH	TG1 Eluted | Z5	High Z score threshold, no Stops
PGETCMQ	TG1 Eluted | Z5	High Z score threshold, no Stops
MFLVSPQ	TG1 Eluted | Z5	High Z score threshold, no Stops

[1] Amino acid sequence of the scFv HCDR3 selected for rescue. [2] Elution method and Z score sub-data set for each scFv chosen for rescue. [3]ScFv were selected based on their final Z score, frequency in the positive reps and respective negative pools, elution method and highest Z score threshold with no stop codons.

Table 3.5: Inverse PCR primer pairs for phagemid rescues for scFvs selected against active phytase.

Primer name [1]	FRW4[2]	Sequence 5' 3' [3]
CGMSAVEF	Universal	5'PHO-AAACTCAACAGCACTCATCCCACA-3'
LKVqNS	Universal	5'PHO-CGAATTCTAAACCTTCAG-3'
PAGWLRNP	Universal	5'PHO-CGGATTACGCAGCCAACCCGCCGG-3'
RLDNS	Universal - SHIFT	5'PHO-GCCCCAACTATTATCAAGCCT-3'
RSTNWW	Universal	5'PHO-CCACCAATTAGTACTCCT-3'
SRVYqFDY	Universal	5'PHO-ATAGTCAAACTAATAAACCCTAGA-3'
QPASQ	Universal - SHIFT	5'PHO-GCCCCACTGCGAAGCCGGCTG-3'
MRPQRS	Universal	5'PHO-CGACCTCTGCGGCCGCAT-3'
LLAT	Universal - SHIFT	5'PHO-GCCCCACGTCGCCAGCAA-3'
EVRADSTE	Universal	5'PHO-CTCAGTAGAATCCGCCCTAACCTC-3'
CPTGRAT	Universal	5'PHO-CGTAGCCCTACCCGTAGGACA-3'
SLTW	Universal - SHIFT	5'PHO-GCCCCACCAAGTAAGCGA-3'
RESV	Universal - SHIFT	5'PHO-GCCCCACACAGACTCACGA-3'
MPYYRLPAMSA	Universal	5'PHO-CGCCGACATAGCAGGAAGACGATAATACGGCAT-3'
QLLAE	Universal - SHIFT	5'PHO-GCCCCACTCCGCCAGAAGCTG-3'
ESPFWSGYV	**Unique**	5'PHO-AAACGTAGCCAGACCAAAAAGGCGACTC-3'
HAMDGRGIAI	Universal	5'PHO-AATAGCAATACCCCGACCATCCATCGCATG-3'
HGVGQAQRP	Universal	5'PHO-AGGACGCTGCGCCTGCCCCACCCCATG-3'
IGLDMRNALA	Universal	5'PHO-CGCCAACGCATTACGCATATCAAGACCAAT-3'
NKGCVPEVNKM	Universal	5'PHO-CATCTTATTAACCTCCGGCACACACCCCTTATT-3'
TNYVPPTGG	Universal	5'PHO-ACCACCCGTCGGAGGCACATAATTCGT-3'
QSMHYTH	Universal	5'PHO-ATGCGTATAATGCATACTCTG-3'
PGETCMQ	Universal	5'PHO-CTGCATACAAGTCTCCCCAGG-3'
MFLVSPQ	Universal	5'PHO-CTGAGGCGACACCAAAAACAT-3'

FRW4 [2]	Sequence 5' 3' [3]
Universal	5'PHO-TGGGGCCAAGGTACMMYGGTCACYGTCTC
Universal – SHIFT	5'PHO-CAAGGTACMMYGGTCACYGTCTC
For ESPFWSGYV	5'PHO-TGGGGCAAGGTACMMYGGTCACYGTCTC

[1] List of scFv (amino acid sequence) being rescued and their primer labels
[2] Each scFv forward primer except for ESPFWSGYV was paired with the universal degenerate primers FRW4 or FRW4-Shift
[3] 5' phosphorylated (5' PHO) primer sequences targeting the HCDR3 and FRW4 regions. Shaded primers were extended into the FRW4 region to elongate the primers, and thus paired with the Universal – SHIFT FRW4 primer.

Each scFv was rescued from the DNA extracts of the polyclonal phage sub library that had the greatest Z score for that particular scFv. The frequencies of each scFv within their most enriched experiment ranged from 0.0014—5.6% of their phagemid populations. Using the DNA extracted from round 2 panning as the template, an inverse PCR reaction, extending to include the entire pHEN 2 vector, was carried out using a high-fidelity DNA Polymerase. PCR products were checked on an agarose gel (data not shown).

Amplicons with the correct band size of approximately 5 kb were extracted and purified from the agarose gel and then quantified by fluorometry. Their quantities ranged from 2.5—19 ng/µl with two candidates (SLTW and RESV) being too low to produce a reading.

Out of 24 clones, 22 successfully produced colonies. The clones TNYVPPTGG and RSTNWW repeatedly failed to transform into electro competent TG1 cells and were excluded from further analysis.

3.7 Testing the specificity of the scFv rescues via phage ELISA

Monoclonal phage were prepared from each rescued transformant's glycerol stock and in total, 22 clones were tested for specificity in a phage ELISA (figure 3.7) as described in methods section 2.24. An anti-BSA (scFv) positive control phage, which was provided with the library, was also produced at the same time.

Of 22 candidates that successfully produced transformant colonies and phage, zero positive binders against the treated-active phytase conformation were detected by phage ELISA. All absorbance signals, including no antigen controls, remained low at background levels of ~0.3—0.4, whereas the positive control developed an absorbance reading of 2.3 (7.5X background).

The glycerol stocks were miniprepped (concentrations ranging from 12—168 ng/µl and sent for sanger sequencing, das escribed in methods section 2.23, to check whether the HCDR3 sequence was as expected. The samples were sent for sequencing using the reverse 'FD' primer compatible with the Griffin pHEN2 phagemid vector and the results were summarised in table 3.6.

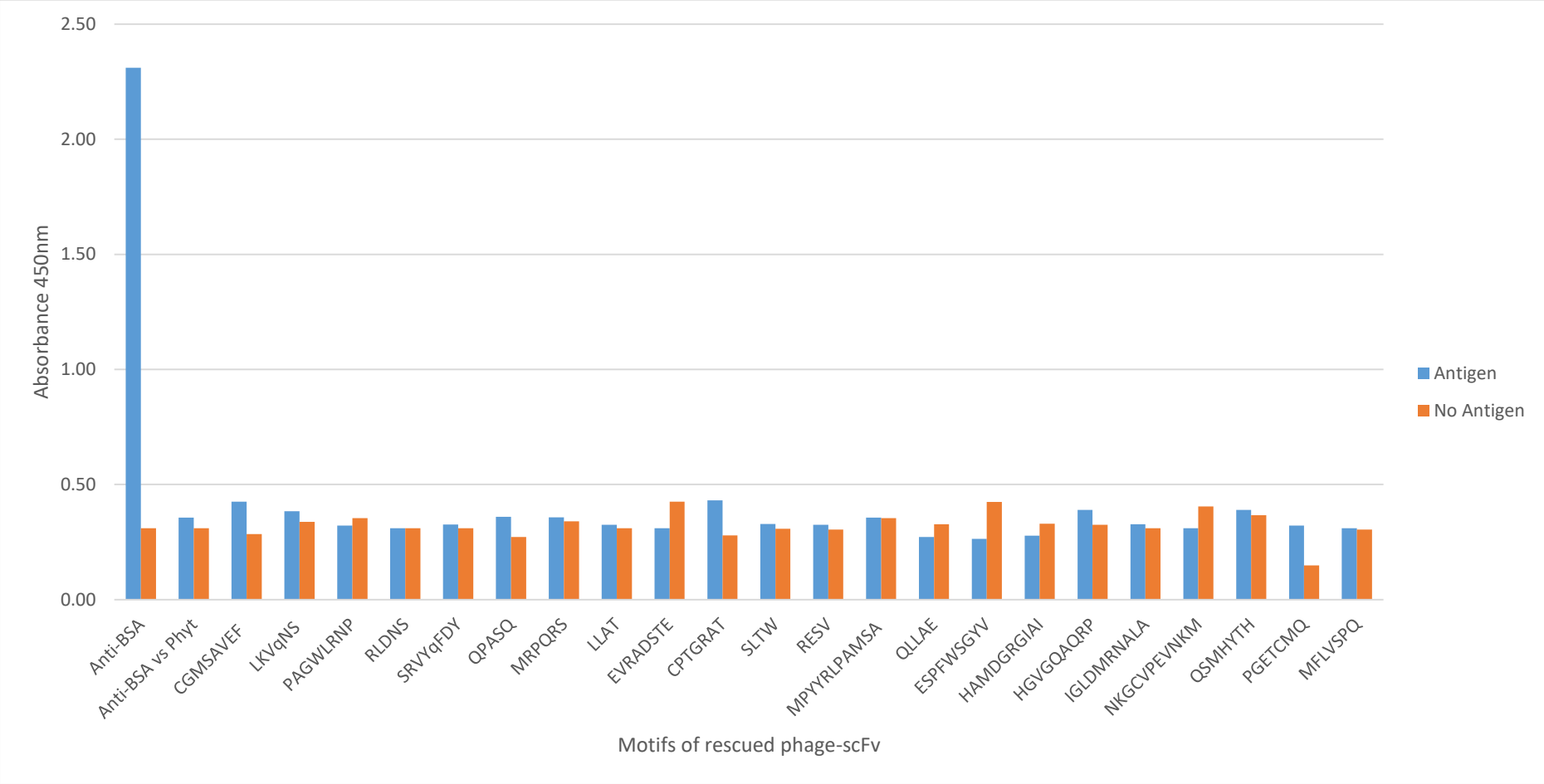


Figure 3.7: Phage-scFv ELISA against partially treated, active phytase for 22 NGS rescue candidates. Partially treated active phytase or BSA was immobilised at 20 µg/ml on a Nunc MaxiSorp® plate overnight. Monoclonal phage-scFv for each rescue were used to detect immobilised phytase or BSA (100 µl of ~10^{13} CFU/ml). Bound phage were detected by an Anti-M13-HRP antibody and a TMB substrate. After 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm.

Table 3.6: Rescued scFv transformants: Sanger sequencing results

Expected HCDR3 [1]	Sanger Sequencing result [2]
CGMSAVEF	**VYYCAK** GYYTFDY **WGQGT**
LKVqNS	No sequence
PAGWLRNP	No sequence
RLDNS	**VYYCAR** RLDNS **WGQGT**
SRVYqFDY	No heavy chain
QPASQ	No sequence
MRPQRS	No sequence
LLAT	No sequence
EVRADSTE	**VYYCAR** EVRADSTE **IDGFS**
CPTGRAT	**VYYCAK** GYYTFDY **WGQGT**
SLTW	No sequence
RESV	No sequence
MPYYRLPAMSA	No sequence
QLLAE	**VYYCAR** QLLAE **WGQGT**
ESPFWSGYV	No heavy chain
HAMDGRGIAI	No sequence
HGVGQAQRP	No sequence
IGLDMRNALA	No heavy chain
NKGCVPEVNKM	No sequence
QSMHYTH	No sequence
PGETCMQ	No sequence
MFLVSPQ	**VYYCAR** MFLVSPQ **DRLRE**

[1] List of scFvs expected HCDR3 sequences that had successfully transformed TG1 cells and produced colonies
[2] Sanger sequencing result after translation of the HCDR3 to amino acid sequences. Flanking sequences are in bold and any differences to the typical left "VYYCAR" and right "WGQGT" flanking sequences are coloured in blue.

After Sanger sequencing, the data revealed that some of the antibodies appeared to be truncated, where only the variable light chain of the scFv gene could be detected (data not shown). Only 4 of the 22 transformants had the expected HCDR3 sequence, and only 6 had evidence of a heavy chain. The majority of the samples failed to produce a sequence despite adequate template. Overall, the vast majority of the scFv candidates were not correctly retrieved from their sub-libraries as intended.

3.8 Discussion

Antibody phage display is a potentially powerful tool that can facilitate the identification of target specific antibodies, without first having to trigger an immune response. After multiple rounds of panning, individual clones can be isolated from the enriched sub-libraries by random colony picking. Colony picking is an effective way of isolating the most enriched clones within the sub-library; however, they may not necessarily be the best target binders, but may simply be the least toxic to their phage and bacterial hosts. Random colony picking is also limiting in the sense that realistically, only a few hundred colonies are picked of which the majority will be duplicates of the same clones. In this study, next generation sequencing was coupled with antibody phage display so that a much greater range of antibody candidates could be considered, rather than those that were simply enriched enough to be picked by chance. More importantly, the bioinformatics approach provides a method which can potentially be used to interrogate antibody enrichment from earlier rounds of panning, and still identify target specific binders as has been demonstrated previously for both antibodies (Zhang et al., 2011) and peptides (AC't Hoen et al., 2012, Naqid et al., 2016). This approach can potentially reduce the number of panning rounds necessary before target binders can be identified, and thereby reduce the bias towards parasitic clones with growth advantages unrelated to the selection process.

The present chapter set out to identify scFv specific for the active form of the commercially available phytase known as Quantum Blue, produced by AB Vista. Quantum Blue phytase is produced for use as a feed additive and is a modified version of the wild-type *E. coli* phytase with six amino acid substitutions designed to increase the thermostability of the protein.

The molecular weight of the phytase was assessed using a denaturing SDS-PAGE gel, and two major bands at ~45kDa and the second at ~20 kDa, were identified. Several weaker bands in between and below the major bands were also detected. It is likely that considering the enzyme was analysed using a denaturing gel that the smaller bands were simply degraded enzyme, but there remained the possibility of contaminants within the stock solution. Complete and partial denaturing conditions were determined for phytase to

compensate for possible contaminants but retain a functional phytase conformation. The conditions were determined indirectly, using a total phosphate kit by Megazyme. The denaturing conditions determined by the total phosphate assay reflected the thermostable properties of the phytase.

The panning strategy was designed to select for phage displaying scFv with affinity for the partially treated, active phytase conformation. The fully denatured phytase antigen acted as a negative control. Phage display was conducted with the coupling of the Griffin phagemid (pHEN2) library (semi-synthetic human VH + VL scFv library) (Griffiths et al., 1994) and the helper phage known as Exphage (Baek et al., 2002). Two rounds of panning were conducted using two elution methods; competitive and direct bacterial infection. The first panning round involved the incubation of the Griffin scFv library with only the heat-treated, active phytase antigen. Bound phage were eluted either competitively using the active phytase, providing another opportunity for the phage to specifically bind to active phytase, or eluted by direct bacterial infection, a more generic elution method. The second round of panning used the enriched sub-library glycerol stocks from round 1 panning, and incubated the two libraries (competitive or direct elution) with both the partially treated, active phytase and the fully denatured phytase antigens in 5 repeats each and the output scFv-phage was sequenced at the HCDR3 region.

Colony picking and polyclonal ELISAs were not conducted for this chapter because positive clones were not expected to be enriched enough after two rounds of panning. The most abundant HCDR3 sequences within the 10 positive replicates accounted for only 1.09—8.07% of the total reads.

Even after considering the HCDR3 sequences with the highest read counts as a percentage of the replicate's total read count, the highest frequencies were only 6%.

When the work in this chapter was conducted, only a handful of antibody phage display strategies had been successfully paired with next generation sequencing methods to facilitate the identification of novel target binders (Ravn et al., 2010b, Zhang et al., 2011, D'Angelo et al., 2014). Furthermore, Zhang et al. (2011) and D'Angelo et al. (2014) successfully used NGS data to identify target specific scFv after only two rounds of panning.

The bioinformatics strategy used within this chapter firstly applied Z scores (Zhang et al., 2011) for each HCDR3 sequence present within each experiment/replicate. This method was used to define scFv enrichment within the original library and also the enrichment against a specific target (active phytase) compared to a control (denatured phytase). Two rounds of panning against the target antigen, active phytase, were enough to provide a selection of HCDR3 sequences that had become enriched after NGS analysis against a specific targtet.

It was to be expected that the generic elution method would capture a greater number of clones that were prevalent in both the positive and negative experiments. This is reflected by all of the best candidates from the TG1 eluted dataset being present within the negative pool, albeit with less than 100 reads each in comparison to thousands of reads within the positive replicates. All of the TG1 eluted selected scFv had still been seen in multiple positive replicates against the intended target. In contrast, the competitive elution method provided an additional step to screen the phage for scFv with specificity for the active phytase conformation.

The competitively eluted candidates that had not been seen in the negative pool (7 of 9) were expected to have greater selectivity for the active phytase having gone through an additional affinity screening step. For the two candidates that were also seen in the negative pool, their read counts were 10—100 fold higher in the positive replicates that met the Z score threshold than the negative pool read count.

Generally, the competitively eluted scFv candidates were seen considerably fewer times in the negative pool than the TG1 eluted candidates. However, the TG1 eluted candidates often had much higher frequencies within their replicates with the highest Z score including the datasets most enriched candidate MFLVSPQ (5.6%).

Although the NGS approach provides substantial information regarding the level of enrichment of each HCDR3 sequence after panning, it does not physically isolate clones of interest from the diverse sub-libraries. Unlike peptides, chemical synthesis of the scFv (typically ~350–400 amino acids) candidates is not economically viable. Gene synthesis of such a coding sequence is possible, but it would be costly to screen a high number of candidates. This would also rely on the full sequence of the scFv being known,

whereas only short read sequencing data (~300bp) around the HCDR3 region was recovered from the bioinformatics, as is the case from most of the current DNA sequencing platforms. The rescue method relied on 5' phosphorylated primers that annealed to a specific heavy chain CDR3 and framework 4 region for each scFv that were chosen for rescue. The HCDR3 primers were unique for each candidate however the primer annealing to the FRW4 region was usually a degenerate, 'universal' primer. The FRW4 primer accounted for the variation within the framework sequence between clones, enabling multiple intended targets to be rescued using the same FRW4 primer but paired with a unique HCDR3 primer. The primers were used as part of an inverse PCR reaction designed to amplify the target clone and the entirety of the pHEN 2 vector (Spiliotopoulos et al., 2015).

Of the 24 initial scFv candidates that were chosen, 22 produced transformed colonies after ligating and transforming *E. coli* with the inverse PCR products. The remaining 2 clones failed to transform after repeated attempts. Phage were produced from the 22 candidates that successfully produced colonies, however no binding activity against the active conformation of phytase was detected by phage ELISA. The candidates were Sanger sequenced to confirm the HCDR3 sequences were correct but only 4 of the 22 transformants had the expected HCDR3 sequence. The incorrect HCDR3 sequences likely explains the negative phage ELISA results as the promising candidates had not been rescued as intended. Many of the picked colonies failed to sequence, returning empty files or if a sequence was retrieved, it did not contain a heavy chain. The clones EVRADSTE and MFLVSPQ were rescued correctly in terms of their HCDR3 sequences however their 3' flanking sequences were completely different to the typical 'WGQGT' sequence. The bioinformatics analysis only permitted one change in the typical flanking sequences so the mutations likely occurred during the inverse PCR rescue process.

The efficacy of the bioinformatics approach could not be verified as the ELISA data reflected the binding activity of clones that were not intended to be rescued as potential binders.

The inverse PCR method had been previously shown (Spiliotopoulos et al., 2015) to surpass other methods in the literature in terms of sensitivity, rescuing phage-scFv targets with a frequency as low as 0.026%. The other

methods such as the two step PCR method described by Ravn et al. (2010b) rescued scFvs with frequencies of 0.5%, whereas the method involving biotinylated oligonucleotides described by Zhang et al. (2011) rescued scFv with frequencies as low as 1% and even other inverse PCR adaptations, such as D'Angelo et al. (2014), rescued scFv candidates with frequencies from 0.5-1.5%. Here we were looking to rescue scFvs with frequencies down to 0.0014% and so the method by Spiliotopoulos et al. (2015) and coworkers was applied. In contrast, the methods by Ravn et al. (2010b) and D'Angelo et al. (2014) would have only been sensitive enough to rescue 14 of the 24 rescue candidates.

This method still proved to be ineffective for the present panning experiments. The combination of low stringency inverse PCR conditions due to the amplification of a large product, low frequency of the intended candidates after only two panning rounds and the degenerate nature of the target vector and FRW4 primer may have accounted for the low recovery rate.

It is a possibility that further panning rounds and the addition of subtractive steps intending to remove non-specific clones from the sub-libraries may enable more efficient scFv rescue.

Chapter 4: Identifying antibodies (scFv) against active and denatured phytase conformations and phytic acid using traditional and next generation phage display

The following study intended to develop a new panning strategy to isolate antibodies against three target antigens in parallel: active phytase, denatured phytase and phytic acid. One aim of the project was to isolate conformation specific antibodies against the phytase to help determine whether phytase remains active or becomes denatured when used as a feed additive.

In chapter 3, the active phytase was partially treated to denature potential contaminating proteins. Before conducting the second panning experiment, the Quantum Blue phytase was analysed for potential contaminants using a native and denaturing protein gel coupled with short-gradient liquid chromatography mass spectrometry (LC-MS).

4.1 Determination of purity of Quantum Blue phytase

In preparation for the LC-MS analysis, the Quantum Blue phytase was treated with PNGase F to remove potentially glycosylated residues as described in methods section 2.6. Untreated Quantum Blue phytase samples were also included for comparison.

Upon PAGE analysis (figure 4.1) one consistent band was identified using a native gel. Upon analysis under denaturing conditions, the same band splits into two major bands. The analysis of the denatured sample suggested that the Quantum Blue phytase extract contained two major bands at ~45 kDa and several other bands that spanned several molecular markers. The bands on both the denaturing and native gels look similar, suggesting that there were no major shifts in molecular weight, thus it is unlikely that the protein was glycosylated to begin with.

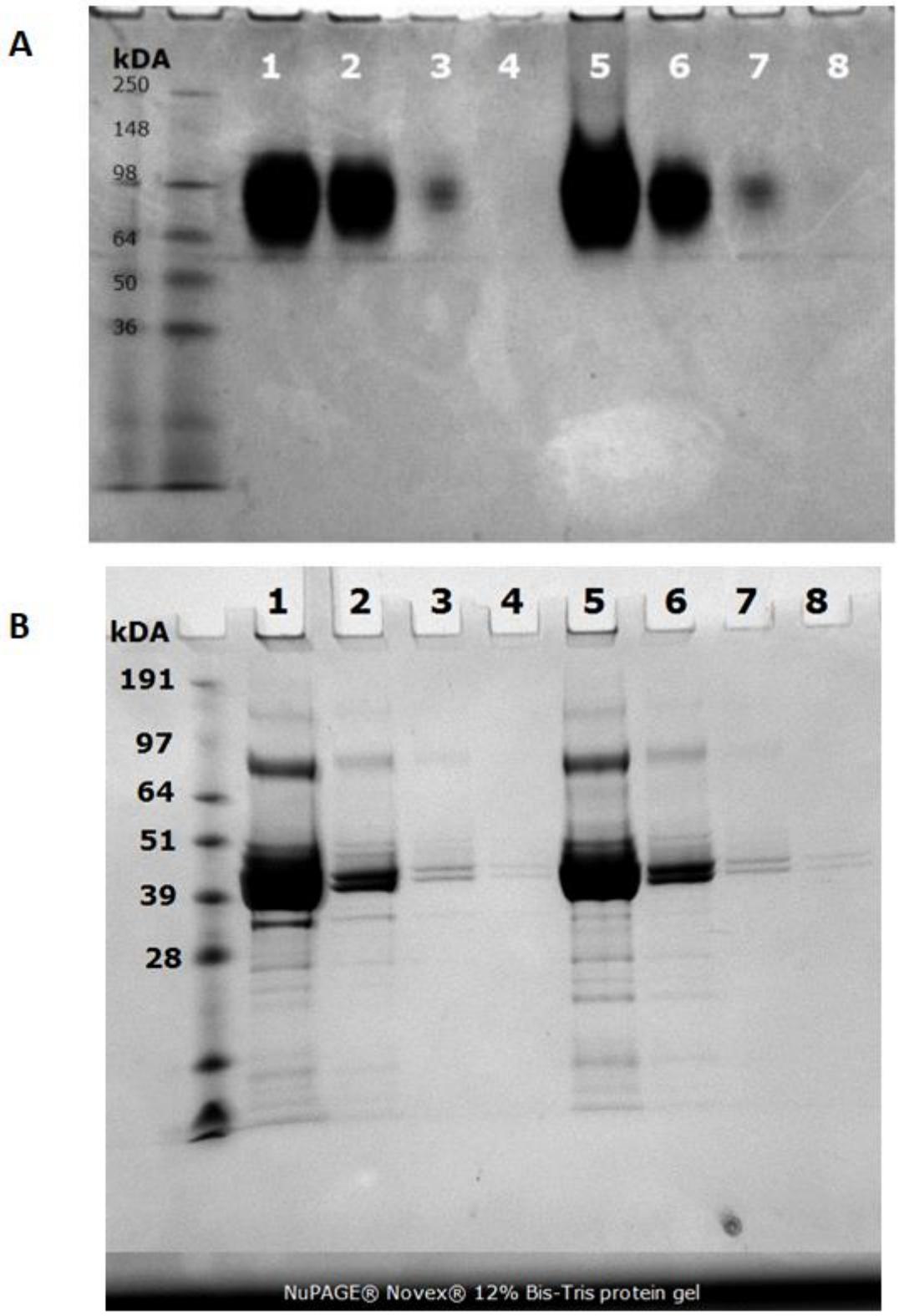


Figure 4.1. PNGase F treated and untreated Quantum Blue phytase molecular weight analysis on a native and denaturing protein gel. PNGaseF treated and untreated Quantum Blue phytase (supplied by Ab-Vista) samples were analysed (4—0.032 µg) on a 12% Tris-Glycine protein gel (A), using native running and sample buffers or a 12% Bis-Tris protein gel, (B) to estimate molecular weight of the treated and untreated phytases. Proteins

were stained with coomassie blue. Molecular mass markers are shown (SeeBlue Plus 2 protein marker). Lanes 1-4 represent 4—0.032 µg (5-fold dilution) of PNGaseF treated Quantum Blue phytase and lanes 5-8 represent 4—0.032 µg of untreated Quantum Blue phytase.

To further determine the purity of the Quantum Blue phytase, four of the PNGase F treated bands (band samples 1—4) on the denaturing gel were sent to the University of Leicester for analysis by short-gradient LC-MS (figure 4.2).

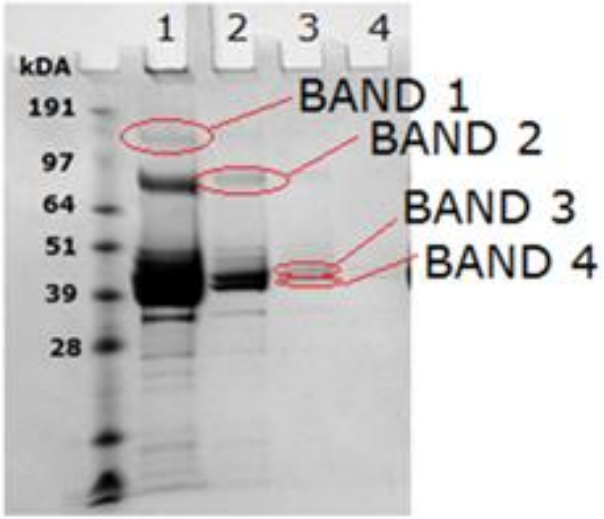


Figure 4.2. PNGase F treated Quantum Blue phytase samples selected for LC-MS. PNGaseF treated Quantum Blue phytase (supplied by Ab-Vista) samples were extracted from the 12% Bis-Tris protein gel and sent to the University of Leicester for analysis by short-gradient LC-MS.

The samples and expected amino acid sequence were provided in an unidentifiable manner with no indication of the product identity.

The % coverage of the expected sequence, within each gel band, was determined and summarised in table 4.1.

Table 4.1: Expected amino acid sequence coverage for gel extracts

Bio sample identifier [1]	Gel Band [2]	% Coverage [3]
UL02021-RR-01	1	12%
UL02021-RR-02	2	30%
UL02021-RR-03	3	36%
UL02021-RR-04	4	36%

[1] The new sample identifier given to each denatured gel band submitted for analysis.
[2] The original band number as extracted from the 12% Bis-Tris protein gel
[3] Percentage coverage of the expected amino acid sequence for Quantum Blue phytase by each band extracted from the denaturing protein gel.

The extracted gel samples covered from 12—36% of the expected amino acid sequence. The results confirmed that the bands that were extracted all had sequences matching Quantum Blue phytase. As all of the major bands on the denaturing gel were phytase, the sample was considered highly likely to be free of contaminating proteins.

4.2 Antigen preparation: Phytases

The active phytase antigen was used as provided by AB Vista, with no additional treatment. The denatured phytase antigen was treated in a benchtop autoclave (121°C for ~30 min) at 200 µg/ml in 30 ml aliquots. The autoclave process was used to obtain larger volumes of denatured phytase.

A complete loss of activity was confirmed using the Megazyme total phosphorous kit as demonstrated by the denatured control in figure 4.4. An alternative denaturant method using a pH shift (pH 12, 10, 4 or 2) was also investigated, but it failed to denature the Quantum Blue phytase (25 µg/ml). Benchtop-autoclaved (chapter 4) and heatblock (chapter 3) denatured Quantum Blue phytase were compared as inactive controls. They were prepared as previously described and then diluted to 25 µg/ml. The Megazyme total phosphorous kit assay used 0.1g of Sigma's sodium phytate as the source of phytic acid. The activity of the pH-shift phytases were also compared to a fully active phytase control at pH 7 (figure 4.3).

All of the pH-shifted phytases retained some degree of activity throughout the pH scale. The phytases pre-incubated at an acidic pH overnight demonstrated a considerable increase in activity in comparison to those incubated at alkaline pHs, and were equal to the absorbance value attained

by the untreated, fully active control. Denaturation of the Quantum Blue phytase using a pH-shift was unsuccessful and the high volume bench-top autoclave method was used to produce denatured phytase as a panning target.

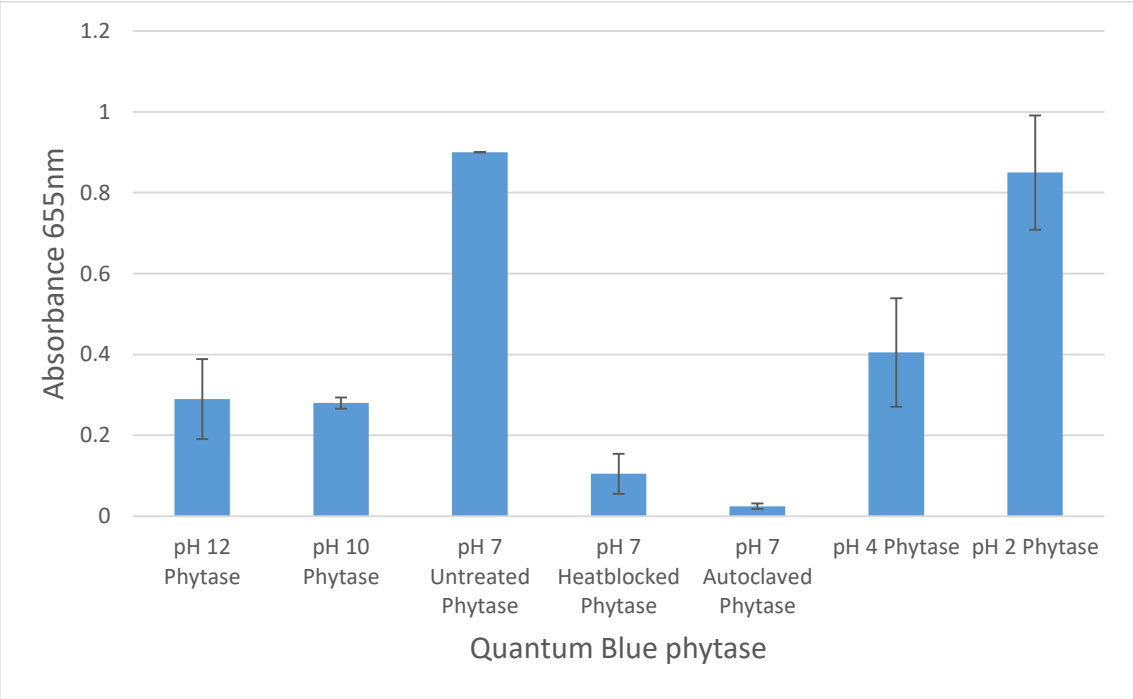


Figure 4.3: pH-shift Quantum Blue phytase activity assay. The activity of phytases at different pHs (25 µg/ml) was measured with the Megazyme total phosphorous kit assay and compared against active (untreated), heatblock denatured and autoclave denatured phytases (25 µg/ml). Absorbance readings at 655 nm were taken and the background free phosphorous levels were subtracted.

4.3 Optimising the immobilising phytic acid

The immobilisation strategy for phytic acid was adapted from Wilson et al. (2015), who demonstrated the use of titanium dioxide (TiO_2) 5µm beads.

This study also used TiO_2 coated, polystyrene 96-well plates or alternatively, with a zirconium dioxide (ZiO_2) coating. Initial immobilisation tests were conducted to ensure that phytic acid was binding, and then to ensure that the capacity of the beads and plates was exceeded to maximise the presentation of the target antigen.

4.3.1 TiO_2 Beads (Titansphere)

A starting concentration of 0.001M phytic acid was selected based on previous publications (Ulusoy and Simsek, 2005).

Aliquots of the TiO_2 beads (10mg) were incubated with 1 ml of different concentrations of phytic acid solution. Phytic acid was digested directly from the beads in order to confirm that the phytic acid had bound as expected. An aliquot of each phytic acid solution was used as the control 'start' solution for each concentration. The 'after binding' solutions were also digested to determine whether there was less phytic acid remaining in the solution after incubation with the TiO_2 beads. Aliquots of the stock 'start' and 'after binding solutions' were used as the phytic acid source in the standard Megazyme protocol. To digest phytic acid from the beads, the full 10 mg of beads were used as the phytic acid source and were vortexed periodically during incubation steps to assist with enzyme access. Free phosphorous samples were also included for each phytic acid concentration, using aliquots of the start solutions but without incubating the kit's alkaline phosphatase or the Quantum Blue phytase. The values generated by these controls were subtracted from the final absorbance values for their respective digests (figure 4.4).

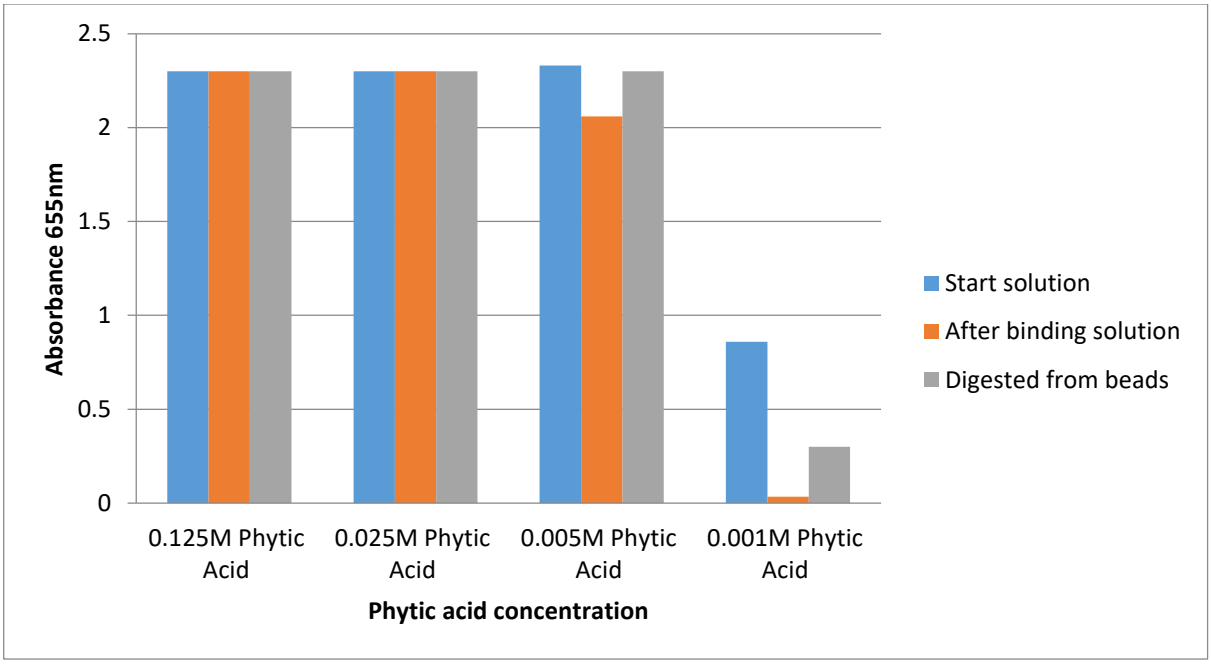


Figure 4.4: Digestion of phytic acid from 10mg of Titansphere TiO_2 beads. A range of different phytic acid concentrations (0.001M, 0.005, 0.025 & 0.125M) were incubated with 10mg of Titansphere TiO_2 beads. The bound

phytic acid was digested from the beads using the Megazyme total phosphorous kit alkaline phosphatase and Quantum Blue phytase. Absorbance readings at 655 nm were taken for the control samples and digested start, after binding and bead-digests. The background free phosphorous values were subtracted from all plotted values.

The 'starting' stock solution contained freely accessible phytic acid (varying concentration) that the Quantum Blue phytase could readily dephosphorylate.

The 'after binding' supernatants reflected how much of the phytic acid present within the start solution had been lost through TiO_2 adsorption. The absorbance value from the direct bead-digest samples suggested that the enzymes were able to access some, if not all, of the immobilised phytic acid and release the phosphates into solution and thereby indicating that the phytic acid had bound to the beads. The maximum signal of the total phosphorous assay was reached (2.25) for both the 0.125M and 0.025M phytic acid concentrations digested from the beads and considering there were still a maximum signal for free phosphates/phytic acid available within the 'after binding' solution, this also suggested that the binding capacity of the beads was considerably exceeded. For the 0.005M phytic acid concentration the level of free phosphates was below the assay's maximum signal for the 'after binding' solution. The digested solution remains at the assay's maximum signal, suggesting that the beads binding capacity was still exceeded. The 0.001M phytic acid concentration demonstrated near-complete adsorption of the phytic acid onto the TiO_2 beads. The 'after binding' solution had an absorbance value of 0.035 suggesting that there were no phosphates available within the solution after the beads had been incubated. The start solution however had an absorbance value of 0.86 and the bead-digest sample had a value of 0.3. This data indicates that although the phytic acid had been adsorbed onto the beads, not all of what was bound was accessible to the enzyme as the digested absorbance value was still lower than the starting solution. Nevertheless, the binding capacity of 10mg of Titansphere TiO_2 beads was likely exceeded somewhere from 0.005—0.001M phytic acid.

To ensure complete saturation of the beads, the concentration of 0.125M phytic acid was selected for future immobilisation assays.

4.3.2 TiO_2 and ZiO_2 coated 96-well plates (Titansphere)

Following confirmation that the phytic acid was binding to the TiO_2 beads, it was determined whether the same strategy could be applied successfully to the TiO_2 and ZiO_2 coated 96-well plates (Fisher).

To bind the phytic acid, 100 µl of the phytic acid solution (0.125M, pH 1) were incubated in each pre-washed well overnight at 4°C. The immobilisation was then confirmed by direct digestion of the phytic acid from the wells. After binding phytic acid to the metal coated wells overnight, the supernatant was removed from the well and used as the 'after binding' control. The Megazyme total phosphorous protocol volumes were scaled down to suit digestion within the wells (400 µl capacity) and digestion stages were conducted at 40°C and with agitation (Varishaker). Absorbance values from the free phosphorous control (background) were subtracted from the readings of the other values before they were plotted (figures 4.5).

The TiO_2 plate had the strongest digested immobilised-phytic acid signal (A_{630} of 0.89) compared to ZiO_2 (A_{630} of 0.16). The latter had little accessible immobilised-phytic acid present. Both the TiO_2 and ZiO_2 well capacities were likely exceeded because the 'after bound' solution signals were approximately the maximum signal for the assay.

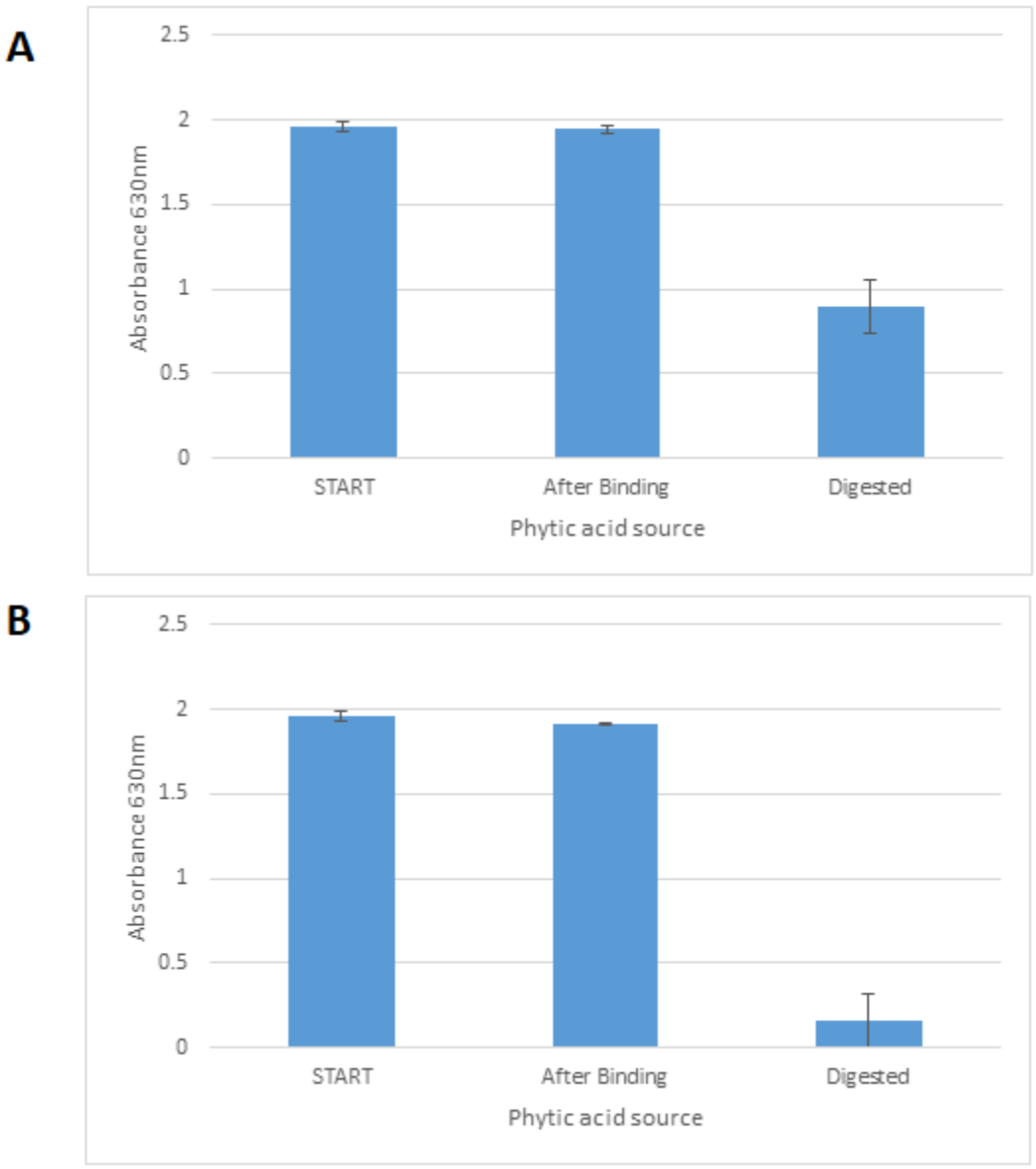


Figure 4.5: Digestion of immobilised phytic acid from TiO_2 or ZiO_2 coated plates. Phytic acid (100 µl of 0.125M, pH 1) was bound to the pre-washed TiO_2 (A) or ZiO_2 (B) plates overnight and washed the following day with 1M Tris buffer (pH 7.5) and keeping aside the overnight supernatant. Aliquots of the start and after binding solution were used as the source of phytic acid for their respective Megazyme digestion samples. Immobilised phytic acid was digested directly from the well. All volumes were scaled down to accommodate the plates 400 µl capacity. The digests were incubated with colour solution for one hour and then stopped with 50% trichloroacetic acid. Stopped solutions were then moved to a non-metal coated, 96-well plate to be read at 630 nm. Background values were subtracted.

4.4 Staining

To confirm that the phytic acid molecules had been immobilised on the metal supports, the wells (TiO_2 & ZiO_2) and beads (TiO_2) were stained with Toluidine Blue as described in methods section 2.13. Wells and beads with no phytic acid bound were used as negative controls (figure 4.6).

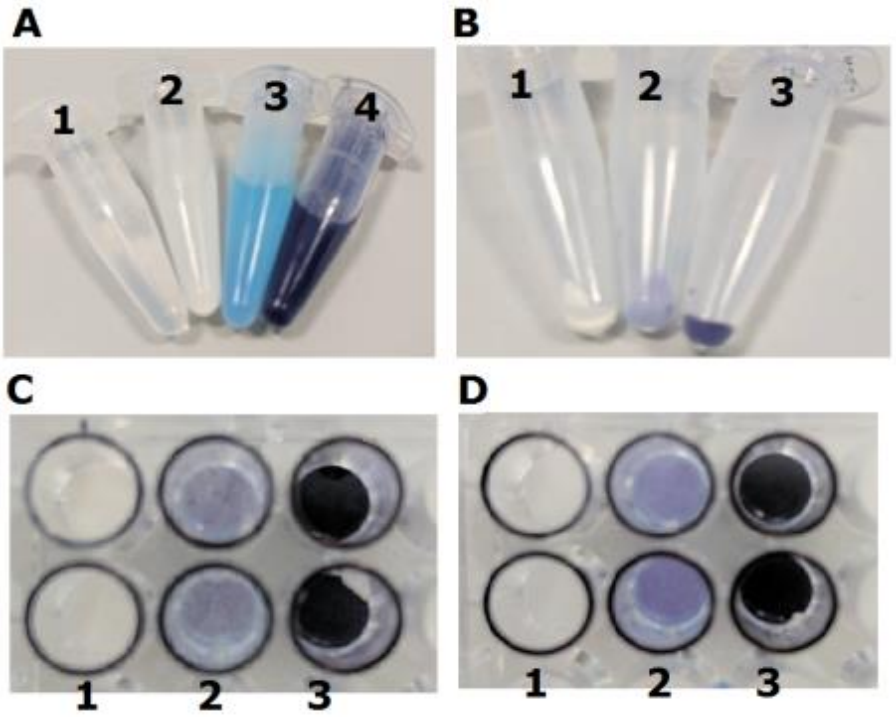


Figure 4.6: Stained phytic acid bound to TiO2 and ZiO2 supports. Phytic acid (100 µl of phytic acid solution 0.125M, pH 1) was immobilised on 10 mg of TiO_2 beads (**A** & **B**) or TiO_2 (**C**) and ZiO_2 (**D**) coated wells. A1 = de-stain solution, A2 = TiO_2 beads in de-stain solution, A3 = TiO_2 beads in de-stain solution after staining with Toluidine blue, A4 = TiO_2 beads with immobilised phytic acid in de-stain solution after staining with Toluidine blue. B1 = TiO_2 beads, B2 = De-stained TiO_2 beads, B3 = De-stained TiO_2 beads with immobilised phytic acid. C & D 1 = no staining and no phytic acid bound. C & D 2 = De-stained metal-coated wells. C & D 3 = De-stained metal-coated wells with phytic acid immobilised.

The beads and wells with phytic acid bound stained a dark navy blue whereas wells without phytic acid returned to near white/pale blue after de-staining. In combination with the digestion experiments, the staining technique confirmed that phytic acid was binding to the metal supports, either in the bead format or the metal coated 96-well plates.

The panning experiments that followed were conducted using 96-well plates with the TiO_2 coating. The TiO_2 TitanSphere beads were also used during panning but for subtraction steps only. The ZiO_2 plates were reserved

for later experiments to confirm whether any isolated antibodies were able to bind to phytic acid specifically and without the TiO_2 support.

4.5 Panning Rounds 1—4 and screening

The panning strategy for each antigen involved 4 rounds of panning with subtraction steps as summarised by figure 4.7.

The Griffin scFv library was panned against 20 Nunc MaxiSorp® wells of active phytase and 20 Nunc MaxiSorp® wells of denatured phytase for each round of panning. The Griffin scFv library was also panned against 20 wells of phytic acid immobilised on TiO_2 for the first round followed by 12 wells for panning rounds 2—4. The number of wells were reduced to compensate for the volume of TiO_2 beads available for subtraction. The TiO_2 beads were used as the subtractive antigen, with nothing else bound. Phage were then eluted from the plates competitively and pooled for each target to create a 4ml solution. Subtraction steps were conducted before each panning round and the remaining phage supernatants were used for panning.

Screening rounds were conducted for panning rounds 1—4 output phage sub-libraries for denatured phytase and rounds 3—4 for active phytase and phytic acid. Screening rounds were conducted in a similar way to panning with the purpose of generating multiple positive and negative replicate samples from the panning sub-libraries. The positive and negative replicate samples were to be compared during the NGS analysis.

After each panning round glycerol stock was grown to produce phage with input titres ranging from 10^{12}—10^{14}. For each phytase panning round, phage were screened against 40 wells coated with the target antigen. An additional 40 wells were also coated with the negative control antigen.

For each phytic acid panning round, 24 phytic acid bound TiO_2 coated wells were used to screen the phage sublibrary. An additional 24 wells were incubated with Tris buffer and served as the negative control replicates.

All phage were eluted competitively with the target antigen. For the phytases, every 4 wells (800 µl) were pooled to generate 10 replicates for the target wells and 10 replicates for the negative control wells. For phytic acid, every 2 wells (400 µl) were pooled to generate 12 replicates for the

target wells and 12 replicates for the negative control wells. Output phage titres ranged from 10^6—10^7.

For labelling purposes, screening replicates generated using round 3 panning glycerols, became known as round 4 replicates, and those generated using round 4 glycerols became known as round 5 replicates.

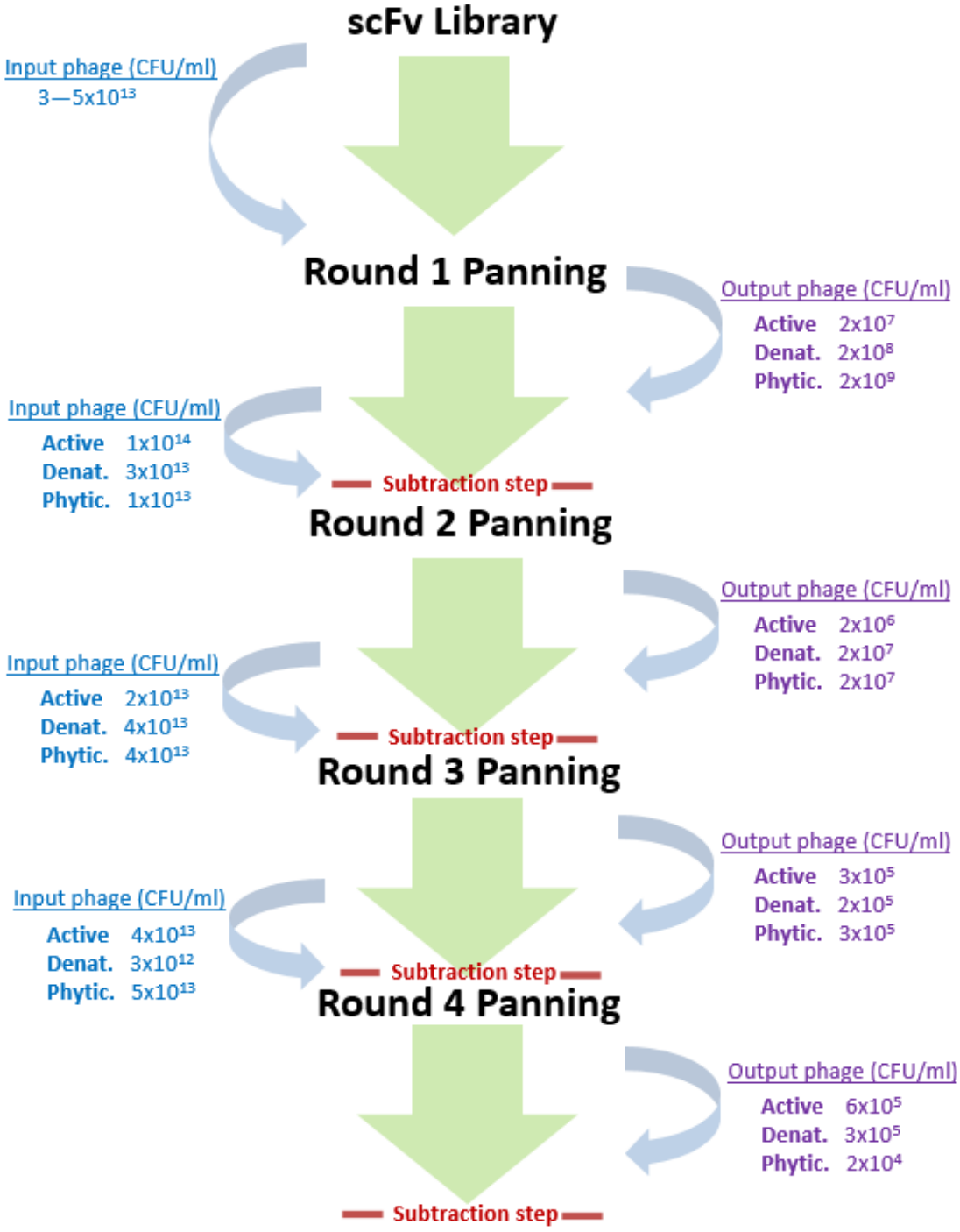


Figure 4.7: Diagrammatic representation of the panning strategy used with the Griffin scFv library for all three target antigens. Four rounds of panning were conducted for each target antigen (active phytase "**Active**", denatured phytase (**Denat.**) and phytic acid (**Phytic.**) with

subtraction steps. Input and output phage titres are shown at each step (CFU/ml).

4.6 Preparation of DNA for Ion Torrent sequencing

The ssDNA extracts from eluted phage were all allocated sample numbers and barcodes and then sent for next generation sequencing at the same time. The DNA (ss or ds) extracts were then used for round 1 PCR to generate ~300 bp amplicons. The round 2 PCR produced expected amplicon sizes of ~350 bp.

An equal contribution from each round 2 PCR product was used to generate a 1000 ng DNA pool. Due to the limited number of barcodes (96), the samples were distributed between two separate 1000 ng DNA pools to avoid duplicate barcodes within each pool.

The sample types, numbers and barcodes were summarised in table 4.2. Samples that had failed round 2 PCR upon multiple repeat attempts were excluded from the NGS preparation (sample numbers 69, 134, 136); their barcodes were reassigned as required.

4.7 Colony picks

In addition to the NGS method, the colony picking approach was also used to help identify the most highly enriched antibodies against active and denatured phytase and phytic acid from the round 4 panning sub libraries. The random colony picking approach was described in methods section 2.15.

Sixty colonies were randomly picked and grown to produce phage for each target antigen. The phage titres ranged from 10^{10}—10^{13} and were tested by phage ELISA (figures 4.8). The absorbance signals were compared against a monoclonal Anti-BSA phage-scfv serving as a positive control. Out of 60 randomly picked colonies, none of the phage-scFv that were produced were able to bind to active phytase. The highest absorbance value was only 0.28 in comparison to the positive control value of 1.3 (6.5X background).

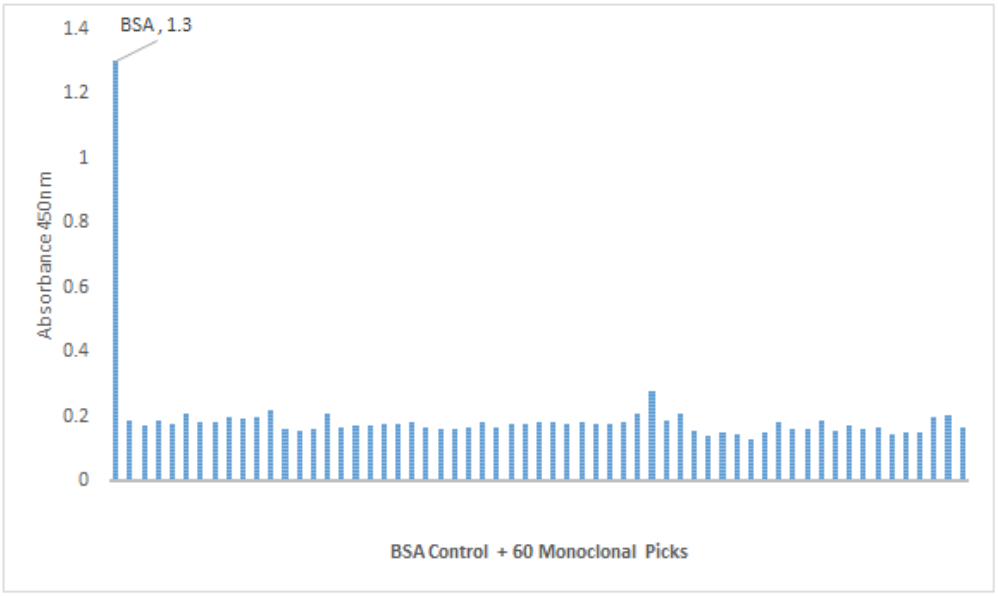
Table 4.2: Summary of samples prepared for next generation sequencing for active and denatured phytase and phytic acid

Samples summary [1]	Sample no. range [2]	Barcode no. range [3]
Active Phytase Bacterial Glycerols Panning Rounds 1-4	1 to 4	1to3+69 *(Sample 2 reassigned)*
Active Phytase Panning Output Phage Glycerols Rounds 1-4	5 to 8	4to7
Denatured Phytase Bacterial Glycerols Panning Rounds 1-4	9 to 12	8to11
Denatured Phytase Panning Output Phage Glycerols Rounds 1-4	13 to 16	12to15
Phytic Acid Bacterial Glycerols Panning Rounds 1-4	17 to 20	16to19
Phytic Acid Panning Output Phage Glycerols Rounds 1-4	21 to 24	20to23
Phytic Acid R4 positive replicates	25 to 36	24to35
Phytic Acid R4 negative replicates	37 to 48	36to48
Phytic Acid R5 positive replicates	49 to 60	49to60
Phytic Acid R5 negative replicates	61 to 72	61to72 (-69)
Denatured Phytase R2 positive replicates	73 to 82	73to82
Denatured Phytase R2 negative replicates	83 to 92	83 to 92
Denatured Phytase R3 positive replicates	93 to 102	93to6
Denatured Phytase R3 negative replicates	103 to 112	7to16
Denatured Phytase R4 positive replicates	113 to 122	17to26
Denatured Phytase R4 negative replicates	123 to 132	27to35
Denatured Phytase R5 positive replicates	133 to 142	36to46 (-37 & 39)
Denatured Phytase R5 negative replicates	143 to 152	47to56
Active Phytase R4 positive replicates	153 to 162	57to66
Active Phytase R4 negative replicates	163 to 172	67to76
Active Phytase R5 positive replicates	173 to 182	77to86
Active Phytase R5 negative replicates	183 to 192	87to96

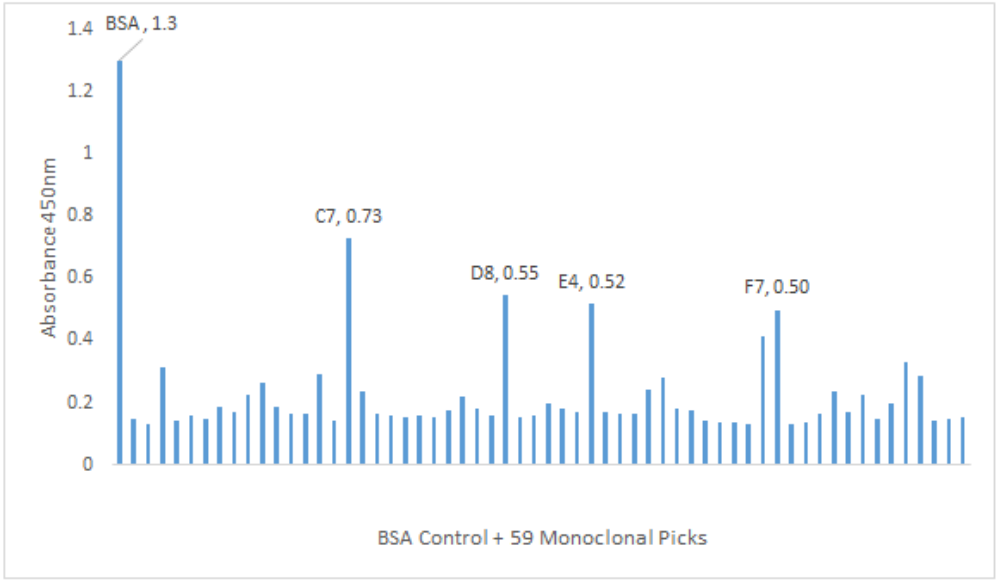
[1] Description and type of sample (grouped).
[2] Sample numbers assigned to each sample type (group ranges).
[3] Barcode number assigned to each sample. Codes/samples highlighted in yellow were used to make the first 1000 ng pool. The range of codes/samples highlighted in orange overlap both pools. Codes/samples highlighted in green were used to make the second 1000 ng pool.

The pooled DNA samples had a final concentration of 22.6 ng/µl and 10.4 ng/µl. Aliquots of the pooled DNA samples (200ng) were sent for Ion Torrent sequencing at the University of Pennsylvania.

A


BSA , 1.3
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Absorbance 450nm
BSA Control + 60 Monoclonal Picks

B


BSA , 1.3
C7, 0.73
D8, 0.55
E4, 0.52
F7, 0.50
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Absorbance 450nm
BSA Control + 59 Monoclonal Picks

C

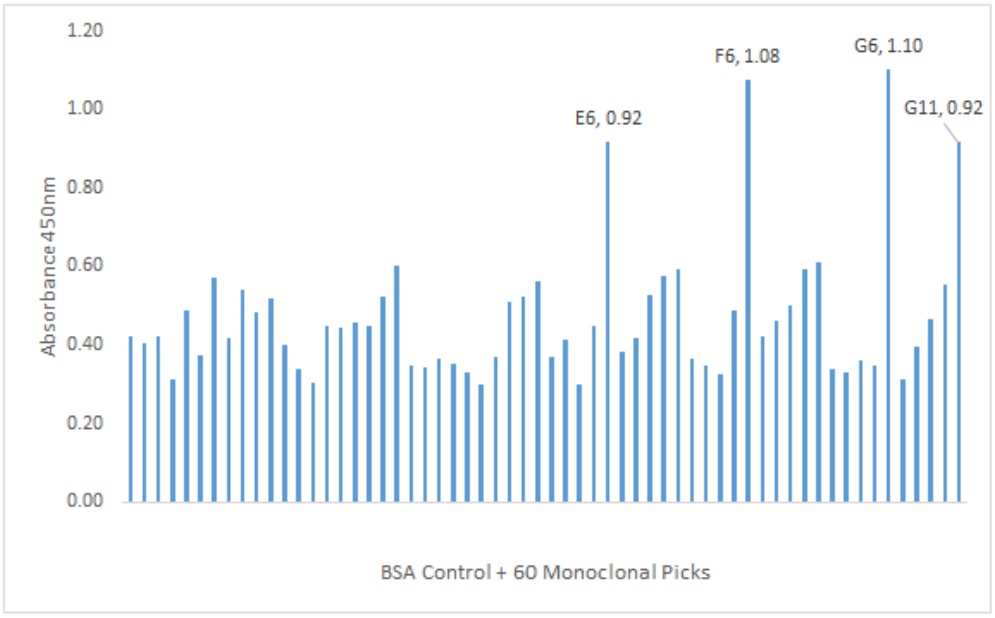

F6, 1.08
G6, 1.10
E6, 0.92
G11, 0.92
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Absorbance 450nm
BSA Control + 60 Monoclonal Picks

Figure 4.8: Colony picking phage-ELISA. Phage-scFv were produced from 60 randomly picked colonies from round 4 panning against active and denatured phytase and phytic acid. Active or denatured phytase or BSA was immobilised overnight in 100 µl aliquots at 20 µg/ml on a Nunc MaxiSorp® plate. Phytic acid was immobilised (100 µl of 0.125M Phytic acid (pH 1) on a TiO2 coated 96-well plate (Fisher). Monoclonal phage-scFv for each colony pick were used to detect immobilised active or denatured phytase or phytic acid (100 µl of ~10^9—10^{13} CFU/ml). Bound phage-scfv were detected with an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. 'No antigen' values (0.2—0.45) were subtracted.

For denatured phytase, four clones were identified (C7, D8, E4, F7) with absorbance values ranging from 0.5—0.7 (2—3X background).

For phytic acid, four clones were identified (E6, F6, G6, G11) with absorbance values ranging from 0.9—1.1 (2X background).

All four denatured phytase clones (C7, D8, E4, F7) were then tested using new phage supernatants in a phage ELISA against both active phytase and denatured phytase using both heat-denaturing methods (benchtop autoclave and the original heat-block method used in chapter 3 (figure 4.9). Phage were also produced from randomly (RDM) picked colonies from the base Griffin library and were used as negative controls.

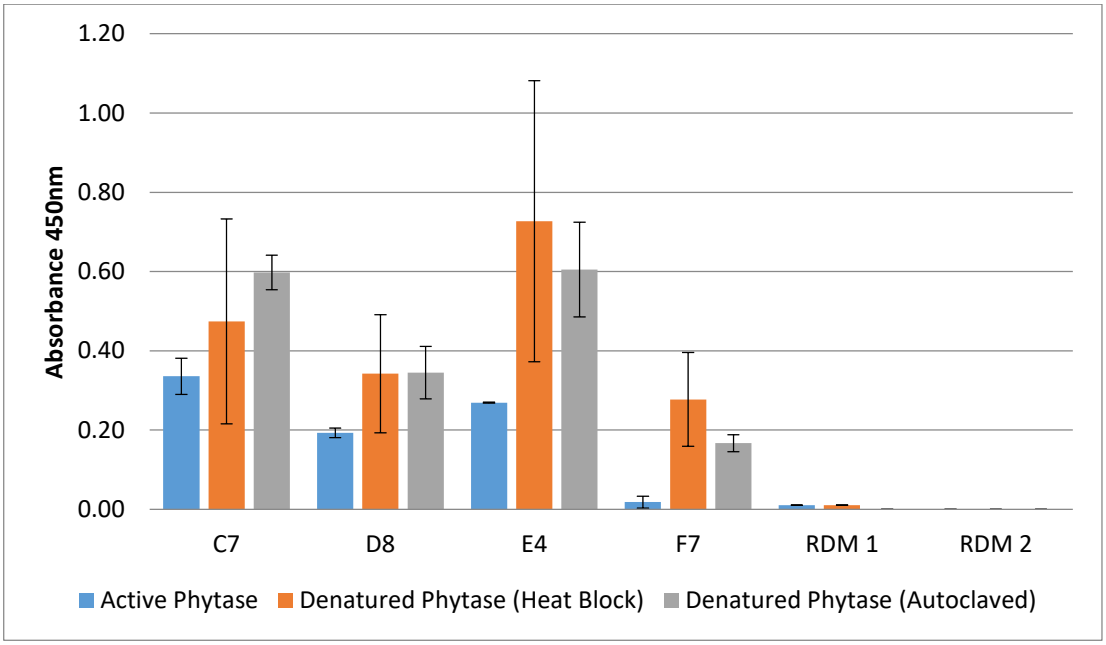


Figure 4.9: Repeat phage-scFv ELISA using the four monoclonals picked from round 4 panning against denatured phytase. RDM (randomly picked clone from the Griffin library used as a negative control). Denatured phytase (heat block method & autoclave method) and active phytase was immobilised overnight at 20 µg/ml on a Nunc MaxiSorp® plate. Monoclonal phage-scFv for each clone were used to detect immobilised phytase (100 µl of ~10^{11}—10^{13} CFU/ml). Bound phage were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. No antigen values (0.2) were subtracted.

The four clones (C7, D8, E4, F7) were able to bind to denatured phytases (>2X background) produced by both denaturation methods whereas random clones picked from the Griffin library could not bind to either of the conformations. The four clones all consistently produced lower absorbance values against the active conformation of phytase, with the F7 clone producing no signal at all (although the signals for F7 were relatively low).

To test the specificity of the phytic acid clones, other molecules containing multiple phosphates were immobilised via their phosphate groups on either TiO_2 or ZiO_2 coated 96-well plates. The plates were incubated with 0.1M Phytic acid (pH 1), 0.1M Adenine Tri-phosphate (ATP), 0.1M PolyP 25 and 0.1M PBS (free phosphate). Each antigen was incubated overnight at 4°C in two wells each. The following day the wells were rinsed with 1M Tris buffer (pH 7.5) and stained with toluidine blue (figure 4.10).

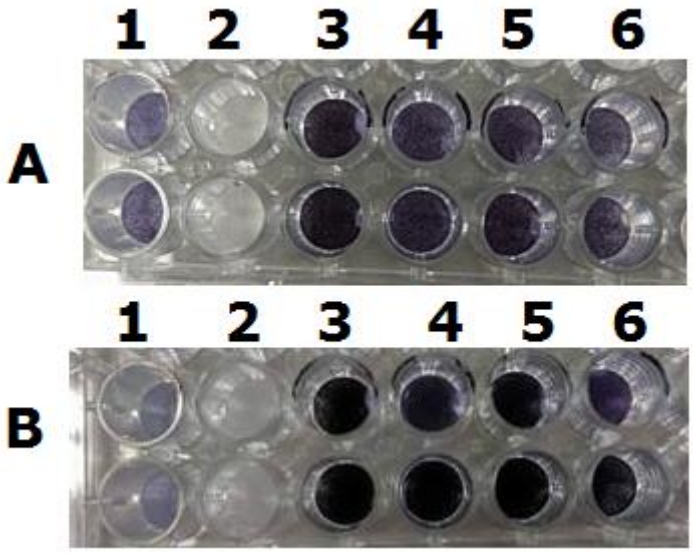


Figure 4.10: Immobilised phytic Acid, ATP, PolyP 25 and free phosphate stained with toluidine blue. The various phosphate containing molecules were immobilised (100 µl of 0.125M Phytic acid (pH 1), 0.1M ATP, 0.1M PolyP 25, 10X (0.1M) PBS) on a TiO_2 **(A)** or ZiO_2 **(B)** coated 96-well plate (Fisher) and stained with toluidine blue. Column 1 wells were destained with no phosphates bound. Column 2 had unstained wells with no antigen. Column 3 was incubated with phytic acid 0.1M pH1, column 4 with ATP 0.1M, column 5 with PolyP 25 0.1M and column 6 with free phosphates (PBS).

All of the wells on both the TiO_2 and ZiO_2 coated plates that had been incubated with a phosphate-containing molecule stained dark blue whereas the no antigen control stained a pale blue. The results suggested that all of the phosphate molecules were able to bind to the metal supports and their phosphates were able to accept and retain the stain.

Following confirmation that the alternative phosphate molecules could bind to the metal supports, phage were produced for the four phytic acid monoclonal picks, plus random picks from the Griffin library (RDM) and also exphage (pIII only) to be used as controls. The phage supernatants were then used in a phage ELISA against all four phosphate-containing molecules to check for specificity to phytic acid when bound to TiO_2 or ZiO_2 (figure 4.11).

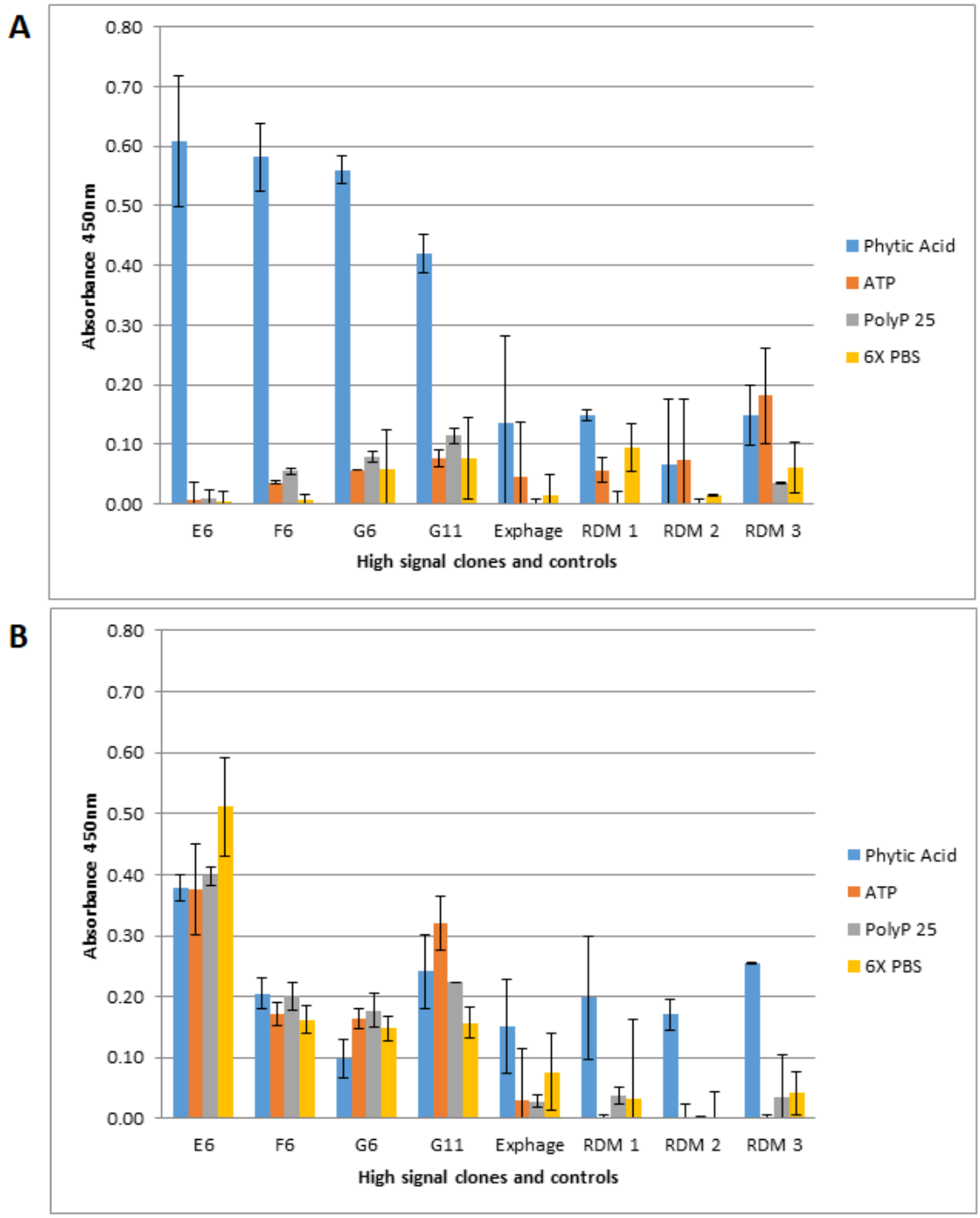


Figure 4.11: Phage-scFv ELISA using four monoclonal picks from round 4 panning against phytic acid and other phosphate sources immobilised on TiO_2 and ZiO_2. The various phosphate containing molecules were immobilised (100 µl of 0.125M Phytic acid (pH 1), 0.1M ATP, 0.1M PolyP 25, 0.1M PBS/free phosphate) on a TiO_2 (a) or ZiO_2 (B) coated 96-well plate (Fisher). Monoclonal phage-scFv for the high signal clones, Griffin controls and exphage were used to detect immobilised phosphate-molecules (100 µl of ~10^{10}—10^{12} CFU/ml). RDM (randomly picked clone from the Griffin library used as a negative control). Bound phage were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Stopped solutions were transferred to a Nunc

MaxiSorp® plate and absorbance values were taken at 450 nm. No antigen values of 0.4—0.45 were subtracted.

The four clones (E6, F6, G6, G11) that were originally binding to phytic acid were only able to bind to phytic acid (absorbance signals of 2X background) and not other phosphate containing molecules when immobilised on TiO_2.

When the same antigens were immobilised on ZiO_2, the clones bound to all of the substances to the same extent, producing similar absorbance readings against each target.

4.8 Sub cloning scFv candidates

Four phage-scFv clones (C7, D8, E4, F7) bound specifically to denatured phytase when immobilised on a Nunc MaxiSorp® plate. Another four phage-scFv clones (E6, F6, G6, G11) bound specifically to phytic acid when it was immobilised on TiO_2. The next objective was to sub-clone the positive scFv candidates into a new expression vector to produce soluble scFv.

The pMal expression system was chosen due to its previous application for the production of maltose binding protein (MBP)-scFv fusion. The pMAL expression vector, 'P5X', would result in the expression of scFv-MBP fusion proteins that would be exported to the periplasm of the host bacterium.

The Griffin pHEN2 vector had an NCOI restriction site at the 5' end of the scFv gene and NOTI restriction site at the 3' end. The restriction sites were compatible with the new P5X vector.

Multiple P5X vector aliquots were double digested using NCOI and NOTI and cleavage confirmed on an agarose gel. As a control, an aliquot of uncut P5X vector (500 ng) was analysed for comparison (figure 4.12).

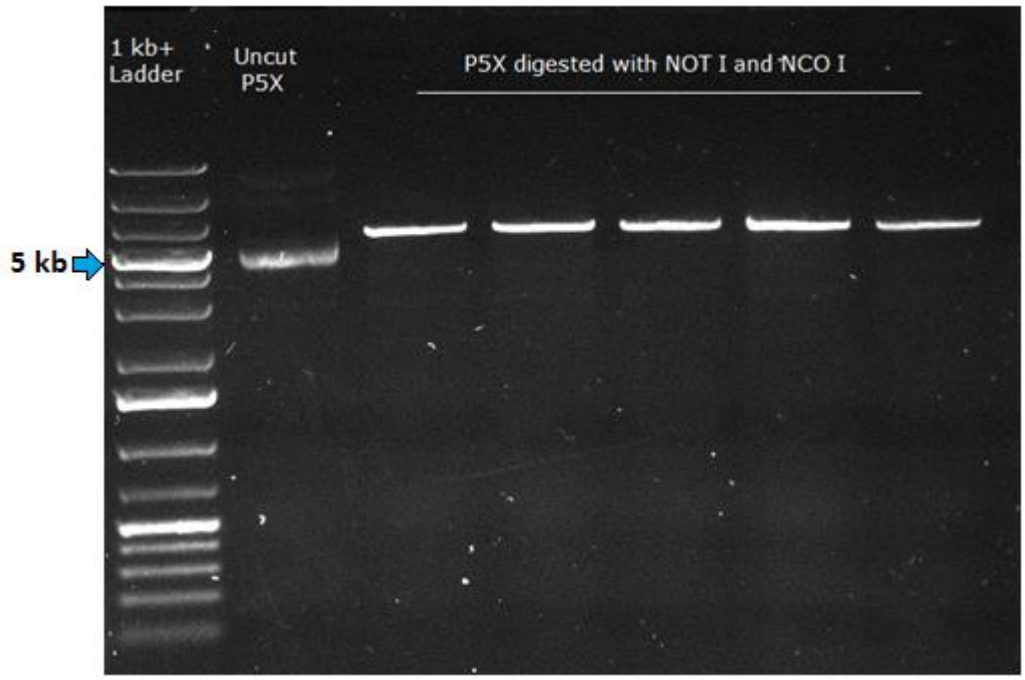


Figure 4.12: Double digested pMal P5X vector. The uncut and double digested P5X vector was analysed on a 0.9% (w/v) agarose gel using 1 KB+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6X loading buffer.

The coiled, uncut P5X vector was 5.7kb in size whereas the five, double-digested samples appeared to be slightly larger due to the decoiling effect

caused by restriction enzyme activity. Single cut experiments with each restriction enzymes showed the same decoil effect on an individual basis (data not shown) suggesting that each enzyme was cutting as expected. The cut P5X bands were gel purified and concentrated.

The four Griffin scFv clones for denatured phytase (C7, D8, E4, F7) were also double digested and visualised on an agarose gel to check whether the scFv genes had been released from the pHEN2 vector (figure 4.13).

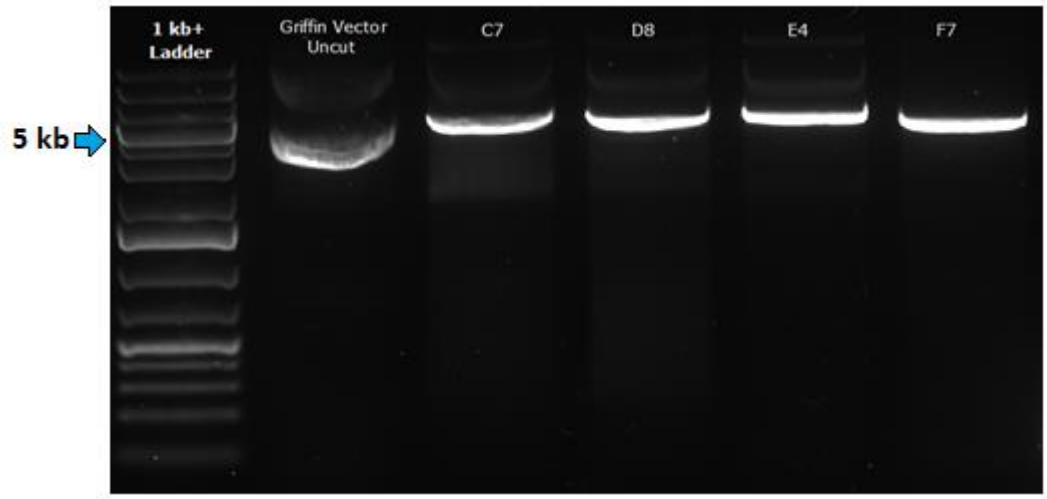


Figure 4.13: Double digested high signal clones to denatured phytase (C7, D8, E4, F7). The uncut and double digested Griffin clones were analysed on a 0.9% (w/v) agarose gel using LR Gene and 1 KB+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6X loading buffer.

None of the four denatured phytase clone's scFv's showed evidence of being released from the pHEN2 vector based on having only one visible band at the 5kb marker.

The Griffin scFv clones for phytic acid (E6, F6, G6, G11) were also double digested and visualised on an agarose gel. The digests were compared against uncut vector (figure 4.14).

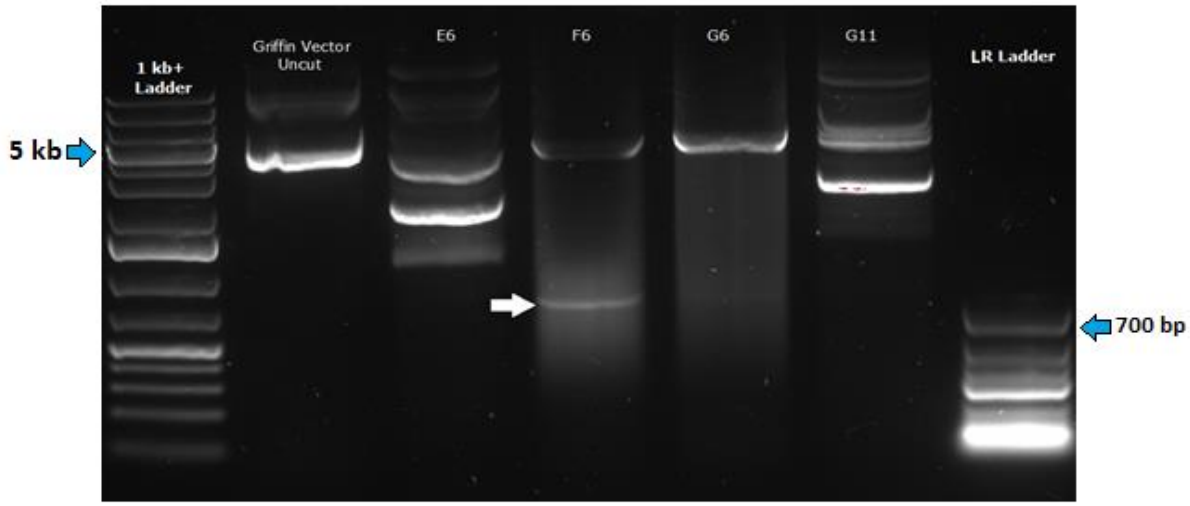


Figure 4.14: Double digested high signal clones to phytic acid (E6, F6, G6, G11). The uncut and double digested Griffin clones were analysed on a 0.9% (w/v) agarose gel using LR Gene and 1 KB+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6X loading buffer. The expected fragment sizes were indicated by the white arrow.

Only the F6 clone showed evidence of being released from the pHEN2 vector with two visible bands on the gel with the pHEN2 vector at 5kb and the expected scFv fragment at ~800bp. The digests were repeated several times giving equivalent results.

4.9 Sanger sequencing high signal clones from monoclonal picks

To troubleshoot the lack of restriction enzyme digestion of the pHEN clone, all eight Griffin scFv clones were sent for Sanger sequencing to check for the presence of a full length scFv and identify their HCDR3 sequences. The samples were sequenced using a reverse primer 'FD' compatible with the Griffin pHEN2 vector (table 4.3).

The G6 clone was sequenced multiple times but would not return a sequence that could be recognised as an antibody, or any of the other vector derived expected sequences such as the pelB leader or linker sequence.

The F7 clone had no evidence of an scFv gene at all and only the pelB leader sequence could be identified.

The clones E6, G11, C7, D8 and E4 were all truncated and had lost their heavy chains entirely. There was no evidence of a pelB leader sequence for the E6 or G11 clones but the light chain ORFs covered most of the domain, including the LCDR3. The ORF for all three of the denatured clones, covering

the LCDR3, were also out of frame with the pelB leader sequence. Only the F6 clone proved to be a full length scFv with all the expected antibody fragments, linker and leader sequences and restriction sites. The HCDR3 amino acid sequence for F6 was identified as DECLPFDY but the ORFs were also out of frame from the pelB leader sequence.

Based on the light chain, the denatured phytase clones C7, D8 and E4 were displaying the same antibody whereas the phytic acid clones E6, F6 and G11 were displaying different antibodies (table 4.4).

Table 4.3: Sanger sequencing HCDR3 amino acid sequence results for the four high signal clones for denatured phytase and phytic acid.

High Signal Clone [1]	Sanger Sequence (Amino acid translation) [2]
C7	No heavy chain. Truncated scFv. Light chain only.
D8	No heavy chain. Truncated scFv. Light chain only.
E4	No heavy chain. Truncated scFv. Light chain only.
F7	No scFv gene at all.
E6	No heavy chain. Truncated scFv. Light chain only.
F6	Full length scFv. VYYCAR DECLPFDY WGQGT
G6	Sequence unrecognisable.
G11	No heavy chain. Truncated scFv. Light chain only.

[1] High signal clone isolated from round 4 panning for denatured phytase (C7, D8, E4 & F7) and phytic acid (E6, F6, G6 and G11).

[2] Translated Sanger sequencing results for each clone's HCDR3 region. Expected heavy chain CDR3 flanking regions were highlighted in yellow.

Table 4.4: Sanger sequencing LCDR3 amino acid sequence results for the four monoclonal picks for denatured phytase and phytic acid.

High Signal Clone [1]	Sanger Sequence (Amino acid translation) [2]
C7	VYYC QQDYNLLFPT FGQGT
D8	VYYC QQDYNLLFPT FGQGT
E4	VYYC QQDYNLLFPT FGQGT
F7	No scFv gene at all.
E6	VYYC MQEAQDTPT FGQGT
F6	VYYC QQYGSSQWLT FGQGT
G6	Sequence unrecognisable.
G11	TYYC QQANSFHST FGQGT

[1] High signal clone isolated from round 4 panning, monoclonal picks for denatured phytase (C7, D8, E4 and F7) and phytic acid (E6, F6, G6 and G11).
[2] Translated Sanger sequencing results for each clone's LCDR3 region. Flanking regions were highlighted in yellow.

4.10 ScFv gene amplification

To further confirm the sanger sequencing findings, the scFv genes for all 8 clones were amplified in a standard PCR reaction using the primers 'PHEN' and 'LMB3' (figure 4.15).

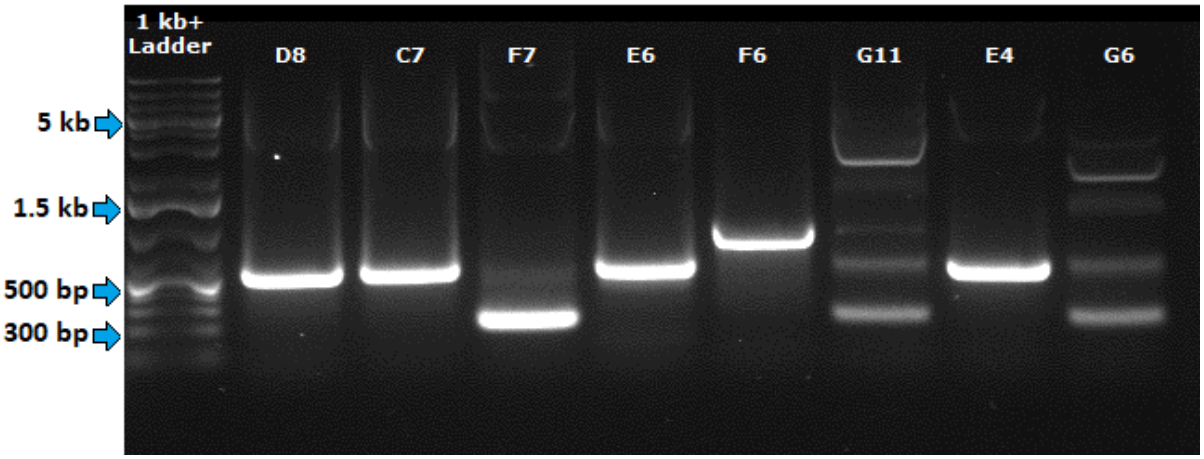


Figure 4.15: Denatured phytase and phytic acid monoclonal picks – amplified scFv genes (C7, D8, E4, F7, E6, F6, G6, G11). The scFv genes for the 8 monoclonal picks with binding activity against denatured phytase and phytic acid were amplified and analysed on a 0.9% (w/v) agarose gel using 1 KB+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6X loading buffer.

The product size for a full length scFv was expected to be ~1,000 bp whereas with no scFv gene at all, the product size was expected to be ~300bp. The clones isolated to denatured phytase D8, C7 and E4 all produced bands at ~700bp suggesting that part of the scFv was missing – the sanger sequencing correlated with this showing the lack of heavy chain in these cases. The F7 clone produced a product size at ~300 bp suggesting that no scFv gene was present. The phytic acid clone F6 produced a 1,000 bp product indicating a full size scFv gene, again consistent with the Sanger sequencing result. The clone E6 produced a band at ~700bp suggesting that part of the scFv was missing – the Sanger sequencing confirmed the lack of heavy chain in this case. The clones G6 and G11 produced faint bands at multiple sizes that may indicate mis-priming and that part of the primer binding site was missing. The presence of an altered vector may explain the difficulties experienced when sequencing the G6 clone. With the truncated, or absence of, antibodies confirmed by PCR and Sanger sequencing, the lack of cutting

in the restriction digests was attributed to the lack of an NCOI restriction site usually located upstream from the variable heavy chain.

What was also concerning was the fact that all of the antibodies were seemingly out of frame from the pelB leader, suggesting that they were never exported to the periplasm for folding and display on recombinant phage. Without the antibody/scFv the phage should not have been able to bind the antigen so specifically and consistently. There would be no significant feature that would have separated them from the other phage in the monoclonal phage ELISAs. The western blot assays that follow aimed to determine whether the phage were actually displaying the truncated antibody fragments.

4.11 Confirmation of antibody display

In a series of western blot analyses, the display of scFv antibodies were confirmed by either an anti-PIII antibody or by an anti-6X-HIS tag antibody N.B the DNA sequencing confirmed the presence of the expected HIS tag for the denatured clones C7, D8 and E4 and phytic acid clones F6, G6, E6 and G11. The clone F7 could not be confirmed due to a complete loss of the antibody gene and the clone G6 could not be confirmed due to the inability to sequence the required region. The standard western blot method is described in methods section 2.3.

The first western blot involved detecting the pIII protein and determining whether there was a size shift between pIII, with and without the scFv displayed. To begin, phage was produced for the eight clones (C7, D8, F7, E4, F6, G6, E6 and G11) and were acid precipitated. The maximum volume (25µl) for each phage sample was loaded onto a SDS-PAGE gel (figure 4.16).

Wild type PIII was clearly shown by the helper phage 'Exphage' with a band at ~50 kDa. The same band was present in all 8 of the monoclonals for denatured phytase and phytic acid as well as the polyclonal Griffin library. The expected pIII-scFv fusion band at ~75 kDa was only visible for the polyclonal Griffin library, suggesting that the monoclonals were either not expressing the antibodies at all, or that their expression rates were considerably lower than the library. The Griffin library sample was a

polyclonal mixture of clones, some of which would undoubtedly have higher expression rates than others, making them easier to detect by western blot.

The experiment was repeated in full with the intention of maximising PIII-scFv expression by including both Exphage and IPTG in the media. Phage were acid precipitated and run on two western blots, one using an anti-PIII antibody as before (figure 4.16) and the second using an anti-X6His-HRP antibody (1:500) (data not shown).

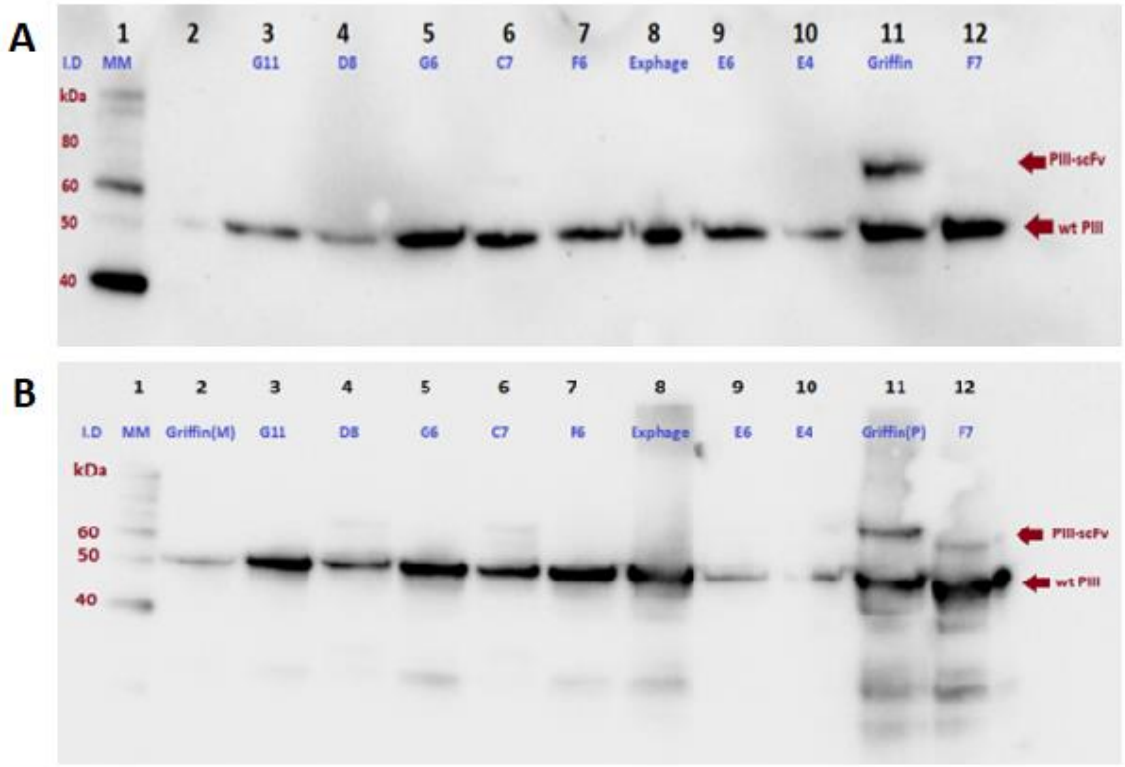


Figure 4.16: Western blot analysis for denatured phytase and phytic acid clones with anti-pIII antibody. The western blot compares the 4 denatured phytase clones (C7, D8, F7 and E4) and the 4 phytic acid clones (F6, G6, E6 and G11) display of pIII-scFv fusions, compared to the wild type pIII control (Exphage) and polyclonal, pIII-scFv Griffin library. Antibody-Piii expression was facilitated by the use of Exphage (A) and Exphage and IPTG (B). All of the phage samples, including the Griffin library (Griffin, Griffin P), had wild type pIII present. MM = magic marker.

There was a noticeable improvement in the detection of pIII-scFv fusions after induction of expression by IPTG for the denatured phytase clones but not for the phytic acid clones. The pIII fusions for all four of the denatured phytase clones were visible, if only faintly, suggesting that the antibodies were being displayed in some instances. In some cases the molecular weights were slightly smaller than the expected, full-length pIII fusion (refer to

polyclonal Griffin banding pattern) which correlates with the suspected truncation of the scFvs.

The HIS tags could not be detected on any of the monoclonal or polyclonal Griffin samples however the positive control (an unrelated poly-HIS tagged protein) confirmed the assay was working as intended (data not shown). It is possible that the phage entity would reduce access to the HIS tag, thereby preventing the antibody from being able to bind to it.

Based on the anti-pIII westerns using recombinant phage, it is possible that the scFv were being displayed despite the frameshifts suggested by the Sanger sequencing. It is possible that the shifts were caused by an error in sequencing, although strange that it should affect all the picked clones for both phytic acid and denatured phytase. The investigative process for all eight clones is summarised in tables 4.5 and 4.6.

Table 4.5: Troubleshooting summary for the subcloning of monoclonal picks after panning against denatured phytase.

Clone[1]	Release by digest?[2]	Sanger Sequencing[3]								Expression detected in anti-Piii western blots [4]	
		PelB Leader	PelB in frame?	NCOI	VH	Linker	VL	NOTI	His Tag	Complete scFv?	PIII expression (Exphage)	PIII expression (Exphage & IPTG)
C7	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes
D8	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes
F7	No	Yes	?	?	No	No	No	No	Yes	Absent	No	Yes
E4	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes

[1] High signal clone isolated from round 4 panning, monoclonal picks for denatured phytase.

[2] After the clone was double digested with NOTI and NCOI restriction digest enzymes, the scFv should have been released from the pHEN2 vector.

[3] Translated Sanger sequencing results for each clone and whether the scFv was full length or truncated as confirmed by amplification of the scFv gene, and whether there was evidence of the expected sequences (pelB leader, linker sequence and HIS tag) and sites (restriction sites NCOI and NOTI).

[4] Summary of the western blot results with anti-pIII.

Table 4.6: Troubleshooting summary for the subcloning of monoclonal picks after panning against phytic acid

Clone[1]	Release by digest?[2]	Sanger Sequencing[3]									Expression detected in anti-Piii western blots [4]	
		PelB Leader	PelB in frame?	NCOI	VH	Linker	VL	NOTI	His Tag	Complete scFv?	PIII expression (Exphage)	PIII expression (Exphage & IPTG)
F6	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
G6	No	?	?	?	No	No	No	No	No	?	No	No
E6	No	No	N/A	No	No	Yes	Yes	Yes	Yes	?	No	No
G11	No	No	N/A	No	No	Yes	Yes	Yes	Yes	?	No	No

[1] High signal clone isolated from round 4 panning, monoclonal picks for phytic acid.

[2] After the clone was double digested with NOTI and NCOI restriction digest enzymes, the scFv should have been released from the pHEN2 vector.

[3] Translated Sanger sequencing results for each clone and whether the scFv was full length or truncated as confirmed by amplification of the scFv gene, and whether there was evidence of the expected sequences (pelB leader, linker sequence and HIS tag) and sites (restriction sites NCOI and NOTI).

[4] Summary of the western blot results with anti-pIII.

4.12 Sub cloning scFv: Continued

Two of the positive clones, 'C7' (denatured phytase binder) and 'E6' (phytic acid binder), were pursued further for sub-cloning into P5X.

The C7 and E6 clones both had truncated scFv genes where only the light chain was present. This loss of genetic material also resulted in the lack of the NCOI restriction site typically used to release the antibody gene from the Griffin pHEN2 vector.

The following experiment intended to add the NCOI restriction site through a PCR reaction that would amplify the scFv genes for C7 and E6.

The NCOI restriction enzyme recognises the sequence 5'..CCATGG..3' (highlighted in yellow) and this was included in the new forward primer intended to partially bind and amplify the scFv (highlighted in blue).

C7 Forward Primer: 5' T GCG TCC ATG GGA ATT GTA ATG ACA 3'
E6 Forward Primer: 5' T GCG TCC ATG GGA ATT GTG ATG ACC 3'

The forward primer was designed to ensure that the inserted scFv gene, after double digestion, would be in-frame with the P5X vector's MBP gene. The forward primers were coupled with the Griffin reverse primer PHEN.

Griffin Reverse Primer PHEN: 5' CTA TGC GGC CCC ATT CA 3'

Using mini-prepped dsDNA, the C7 and E6 scFv genes were amplified in a PCR reaction in order to add the NCOI restriction site, and were visualised on an agarose gel (figure 4.17).

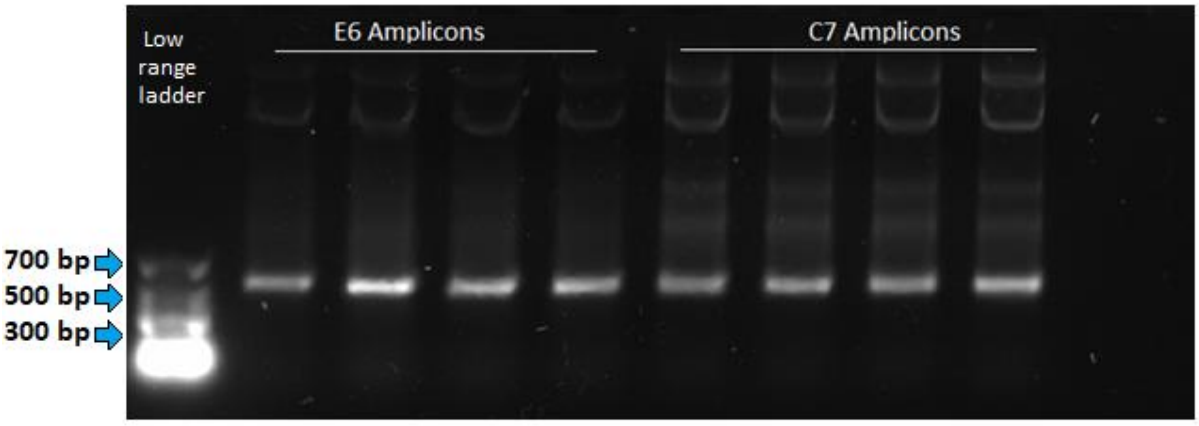


Figure 4.17: Amplification of 'C7's and 'E6's scFv gene to add an NCOI restriction site. The C7 and E6 scFv gene amplicons with the newly

added NCOI sites were analysed on a 2% (w/v) agarose gel and checked against a LR Gene Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6x loading buffer. The gene was amplified in multiples so they could be concentrated into one sample at a later step.

The cleaned amplicons were then A tailed and cloned into an optimised sequencing vector (pCR™4-TOPO®) in preparation for sub-cloning into the pMAL system. Eight pCR™4-TOPO® colonies were picked and grown from the C7 and E6 transformant plates. To check for the scFv gene insert, the eight C7 and E6 TOPO-vector colonies were double digested using NCOI and NOTI restriction enzymes. The products were visualised on an agarose gel to check for evidence of cutting. The uncut pCR™4-TOPO® vector was loaded as a no-insert control (figure 4.18).

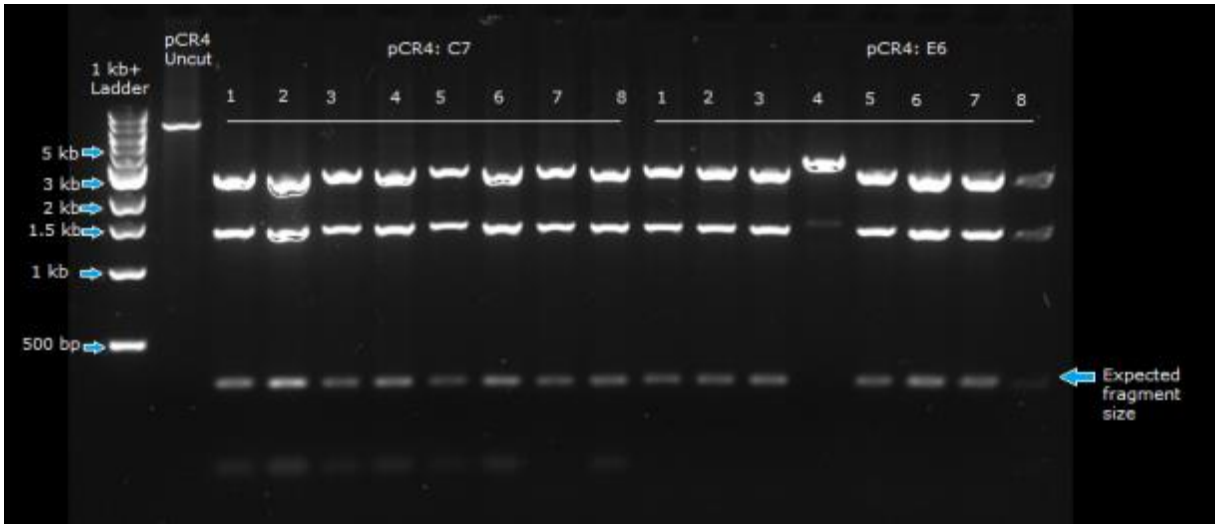


Figure 4.18: Double digested pCR™4-TOPO vector containing C7 and E6 scFv genes. The eight double digested pCR™4-TOPO vector C7 and E6 colonies were analysed on a 1.5% (w/v) agarose gel using a 1 KB+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6X loading buffer.

The 300 bp scFv fragment genes were released from the TOPO vector and their presence was confirmed in all but one of the sixteen colonies.

The released double digested scFv bands were gel purified for colonies 3, 4 and 5 (C7) and 1, 3 and 5 (E6) and then inserted into the gel purified, cleaved pMAL vector P5X.

Three single colonies were picked for C7 and E6 and their dsDNA was extracted and digested using NCOI and NOTI. The digests were visualised on

an agarose gel to check for evidence of cutting. The empty P5X vector was loaded as a no-insert control (figure 4.19).

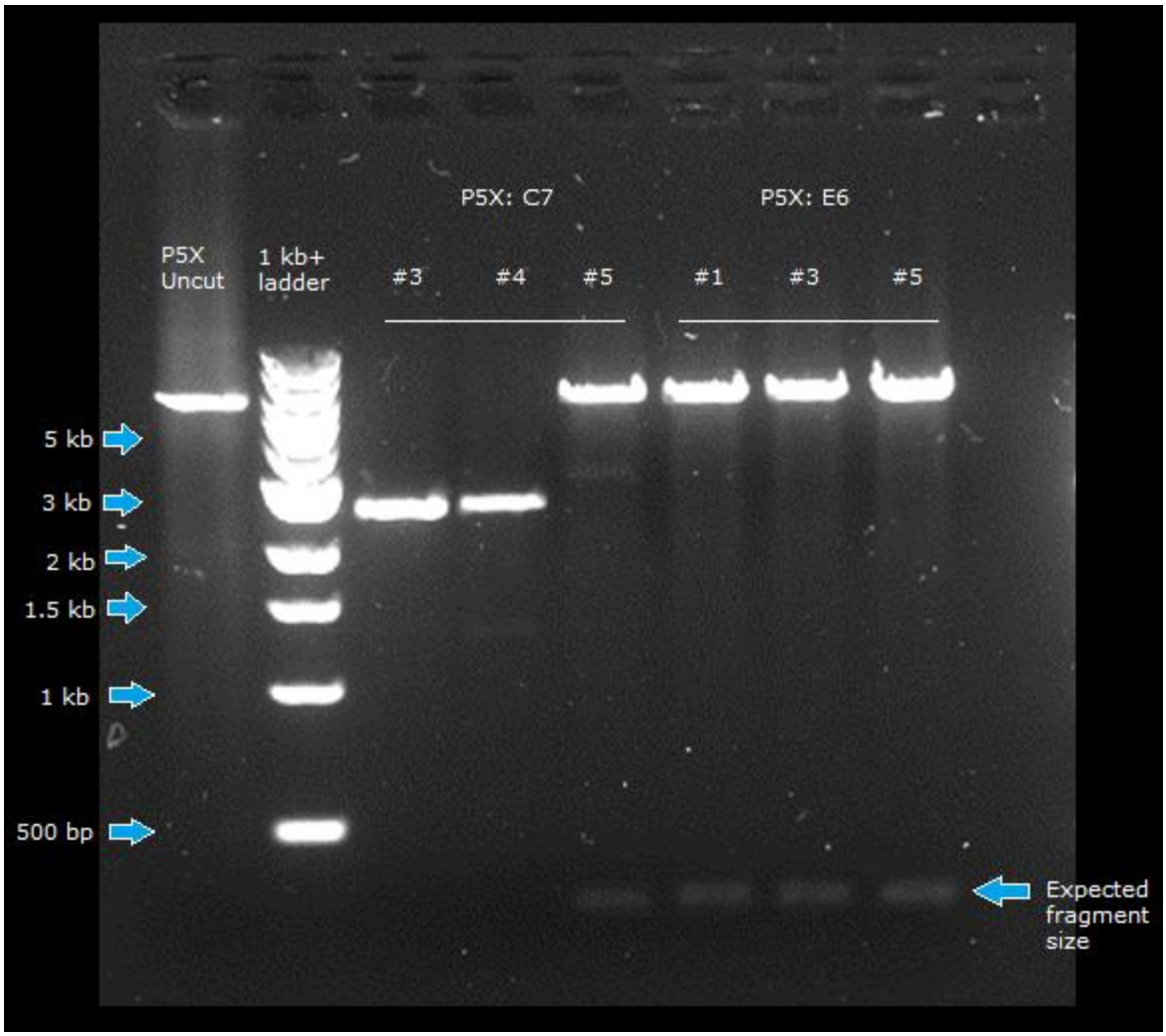


Figure 4.19: Double digested P5X vector containing C7 and E6 scFv genes. The six double digested P5X vectors containing either C7 or E6 scFv genes were analysed on a 1.5% (w/v) agarose gel using a 1 KB+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6X loading buffer.

The digests confirmed the presence of the scFv gene in 4 of the 6 colony picks with the release of the expected 300bp fragments.

All 6 of the clones were then tested to see whether the scFv fragment antibodies were being expressed as MBP fusion proteins, as described in methods section 2.26. An anti-beta-lactamase VHH (described in chapter 5), in the P5X vector, was also grown as a positive control. The antibody-MBP fusions were extracted from both the whole cell lysate and periplasm (figures 4.20).

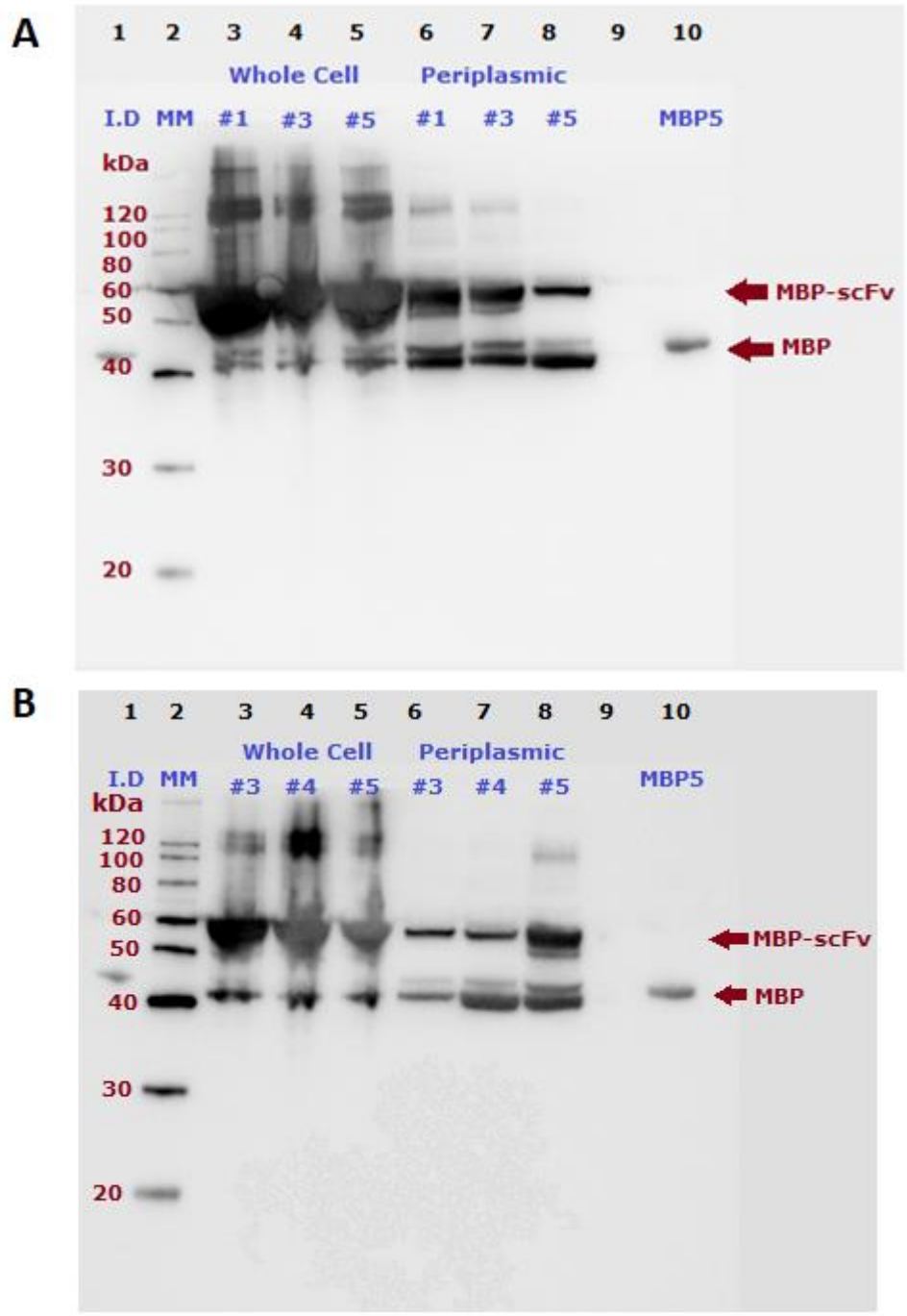


Figure 4.20: Western blot analysis for the expression of three anti-phytic acid and denatured phytase P5X:E6 clones with anti-MBP and anti-mouse-HRP. The western blot compares the ER2523 P5X:E6 colony picks for the transformed clones #1, #3 and #5 against phytic acid (a) and #3, #4 and #5 against denatured phytase (B). Expression of the expected scFv-MBP fusions was detected with a mouse anti-MBP primary and anti-mouse-HRP secondary antibody from both whole cell and periplasmic extracts. Fusions were compared against the control MBP (MBP5) protein with no scFV fusion. MM = magic marker.

Based on the western blot, all of the C7 and E6 clones had evidence of the MBP-scFv fusions (~67 kDa). Sanger sequencing also confirmed that the clones were correct (data not shown).

The same periplasmic extracts were then tested against their targets in an ELISA to check whether the VHH or scFv fragment fusions bound to their intended targets. Unblocked periplasmic extract was also incubated on the Nunc MaxiSorp® plate to immobilise the VHH/scFv-MBP fusions for 1 hour (figures 4.21 & 4.22).

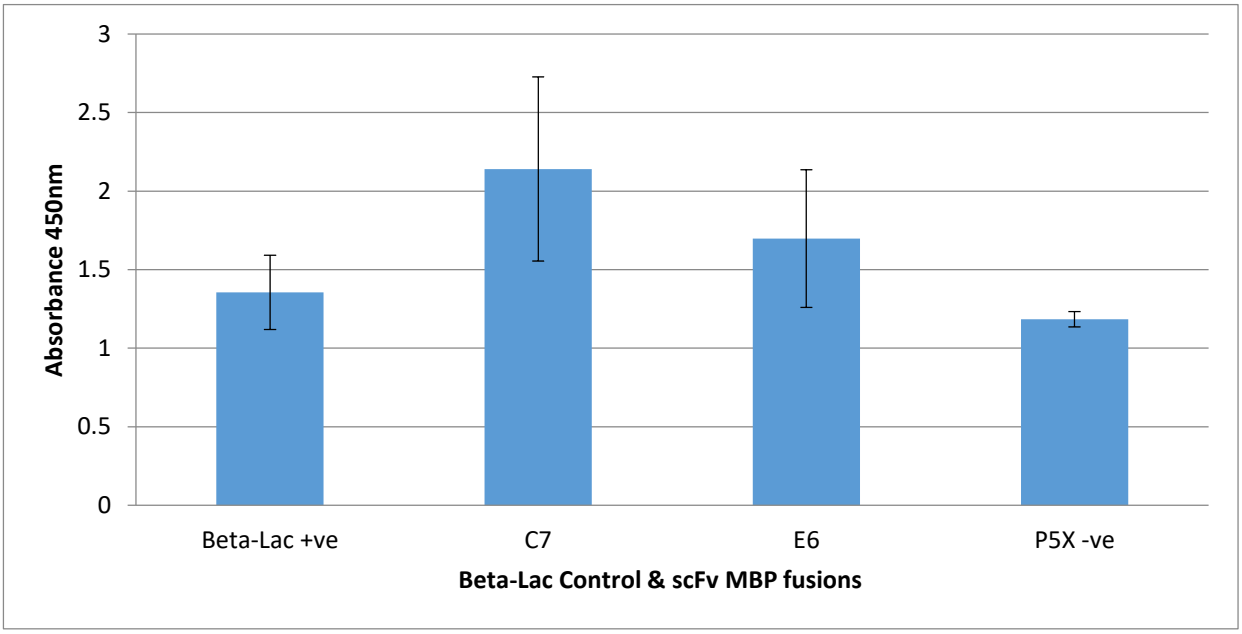


Figure 4.21: Immobilised soluble MBP-scFv ELISA using P5X:E6 (anti-phytic acid) and P5X:C7 (anti-denatured phytase). The MBP/antibody fusions for bLAC (VHH), P5X:C7 (scFv), P5X:E6 (scFv) and empty P5X (MBP only) periplasmic extracts were immobilised on a Nunc MaxiSorp® plate overnight. Bound MBP/fusions were detected by an anti-MBP primary antibody (1:10,000) and an anti-mouse-HRP secondary antibody (1:2,000). The HRP substrate TMB was incubated in each well and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. No antigen values (0.2) were subtracted.

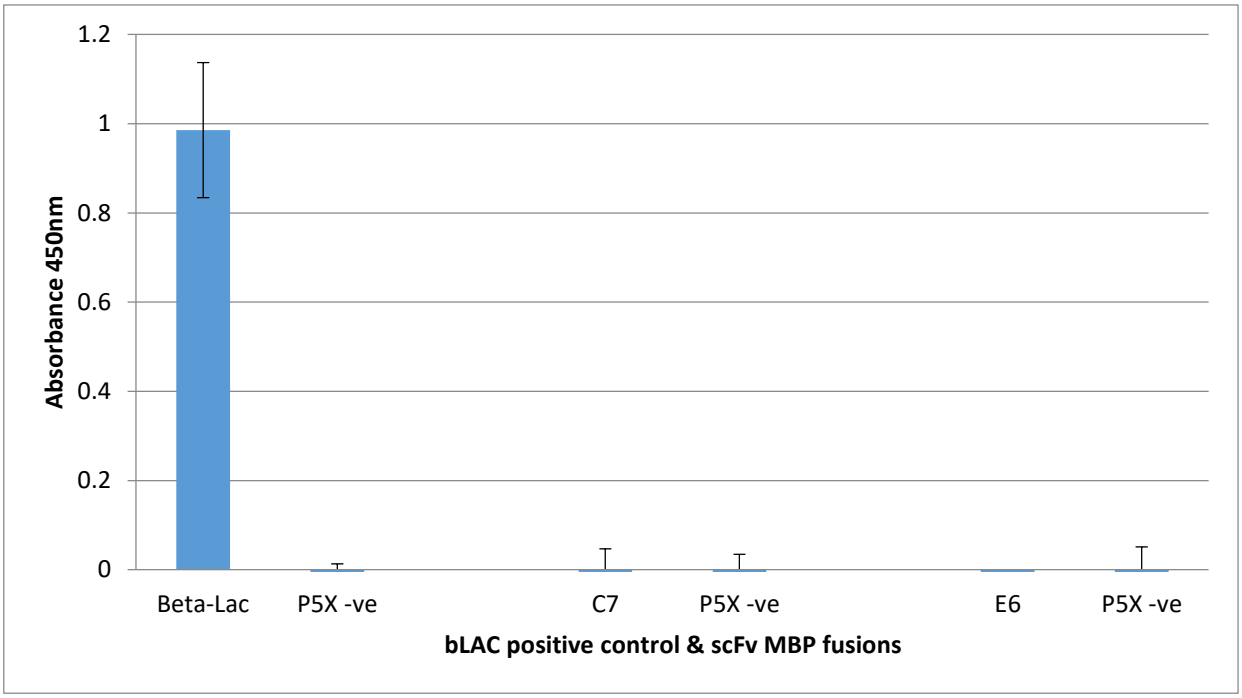


Figure 4.22: Soluble MBP-scFv ELISA using P5X:E6 (anti-phytic acid) and P5X:C7 (anti-denatured phytase). Denatured phytase and beta-lactamase were immobilised at 20 µg/ml (100 µl per well) on a Nunc MaxiSorp® plate and phytic acid was immobilised on a TiO_2 plate (100 µl of 0.125M Phytic acid (pH 1) overnight as previously described. Soluble MBP-scFv/VHH fusions or MBP were extracted from the periplasm for the empty P5X vector (expressing MBP), anti-beta-lactamase (bLAC) P5X VHH clone and the P5X scFv clones C7 and E6. Each MBP fusion was incubated against their respective target antigen. The MBP extract (P5X –ve) was also incubated against each target antigen as a negative control. Bound MBP/fusions were detected by an anti-MBP primary antibody (1:10,000) and an anti-mouse-HRP secondary antibody (1:2,000). The HRP substrate TMB was incubated in each well and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Stopped solutions were transferred to a Nunc MaxiSorp® plate and absorbance values were taken at 450 nm. No antigen values (0.2) were subtracted.

The immobilised VHH/scFv-MBP ELISA confirmed the presence of VHH/scFv-MBP as seen in the western blots.

The soluble VHH/scFv-MBP fusions were then tested for specificity against their target antigens as described in methods section 2.26.2. The anti-beta-lactamase-MBP VHH served as a positive control and produced a positive signal (5X background) against its antigen beta-lactamase. The empty P5X control (MBP only) served as a negative control and failed to produce a signal against all three target antigens (beta-lactamase, denatured phytase and phytic acid). The C7 scFv-MBP clone did not produce a positive signal against

denatured phytase and the E6 scFv-MBP also failed to produce a positive signal against phytic acid despite their presence being confirmed by both western blotting and detection of the immobilised MBP-fusions.

4.13 NGPD data analysis and scFv rescue

The NGS data was received from the University of Pennsylvania in a compressed, FASTQ file format. The data was de-compressed and de-multiplexed. The data analysis process was summarised in figure 2.5 in methods, describing how each screening round were ranked and selected by 'Z score' for each HCDR3.

The high signal clones identified through randomly picking colonies from round 4 panning for denatured phytase could not be found in the NGS data possibly as they would not have been amplified due to the loss of their heavy chains. The phytic acid clones E6 and G11 were also not found due to their lack of heavy chains. The phytic acid G6 clone was unrecognisable so had no reference to search for within the data.

Depending on where the phytic acid F6 clone fell within the data (prevalence in replicates and final Z scores), it may have been useful to set the potential search criteria that would help identify other potential binders. All *.LR.*fa files were screened for the presence of the HCDR3 sequence DECLPFDY using the following command in Linux terminal.

*grep -c -w HCDR3sequence *.LR.fa > HCDR3sequence.Zscores.table*

The search identified the F6 clone in a positive replicate (barcode 28) file from round 3 screening. The barcode 28 file had 246,890 reads in total, with 2 of those reads representing the F6 clone. When barcode 28 was compared against the pool of negative replicates, the DECLPFDY sequence produced a Z score of 1.41. The search also identified the F6 clone in round 4 screening barcodes 66, 71 and 72 but they represented negative controls. Considering the same clone had been picked randomly from round 4 panning, within 60 picks, the data analysis results for the high signal clones were not used to narrow down the search criteria for alternative clones within the data

Briefly, the candidates that were seen in at least 4 positive replicates with a Z score of ≥2 were retained for dataset comparisons. Duplicates were removed between Z score thresholds and screening round shortlists. The final lists of candidates for each target antigen were compared against each other and any duplicates (suggesting non-specific binding) were removed. The remaining candidates were ranked according to the number of positive reps that they were seen in followed by the highest Z score threshold.

Following the data analysis, the top 20 HCDR3 sequences that best fit the selection criteria for each target were shortlisted for further analysis.

The 20 shortlisted HCDR3 sequences, for each target, were then subjected to further 'reversed' analysis. Out of the top 20 candidates, 12 of the anti-active phytase, 17 of anti-denatured phytase and 12 of anti-phytic acid HCDR3 sequences met the final selection criteria (table 4.7).

Table 4.7: The final list of intended scFv rescues for active phytase and denatured phytase post-NGS analysis.

Data Set & Round [1]	AA Motif [2]	BC Number (Z score) [3]
Active Phytase (R4)	TSLLGVNR	65 (16.26)
Active Phytase (R4)	YLAEFG	61 (9.5)
Active Phytase (R4)	GWFFV	59 (9.74)
Active Phytase (R4)	IKPACNTMA	59 (12.7)
Active Phytase (R4)	VGESQ	59 (28.52)
Active Phytase (R5)	PDDLIG	82 (27.05)
Active Phytase (R5)	PPSVHFDY	82 (69.08)
Active Phytase (R5)	STQLP	82 (29.5)
Active Phytase (R5)	ASCS	84 (15.99)
Active Phytase (R5)	DSEFDTK	84 (25.66)
Active Phytase (R5)	HMLFTS	82 (66.96)
Active Phytase (R5)	LMSQQ	84 (25.09)
Denatured Phytase (R4)	PEPQPFDY	19 (21.57)
Denatured Phytase (R4)	GGASPLHN	19 (6.45)
Denatured Phytase (R4)	LLDGQG	26 (16.46)
Denatured Phytase (R4)	RVFLLFDY	19 (11.71)
Denatured Phytase (R4)	VLGDRH	21 (12.69)
Denatured Phytase (R4)	WGPGSE	20 (8.27)
Denatured Phytase (R5)	AEAAD	38 (4.58)
Denatured Phytase (R5)	AETAD	36 (66.56)
Denatured Phytase (R5)	ALNLR	36 (13.12)
Denatured Phytase (R5)	ASMSPVH	38 (27.85)
Denatured Phytase (R5)	CQIEW	38 (28.29)
Denatured Phytase (R5)	GAMEYG	38 (24.03)

Denatured Phytase (R5)	GAYEDYVQ	38 (16.13)
Denatured Phytase (R5)	GDPS	38 (35.01)
Denatured Phytase (R5)	GTYLY	38 (23.26)
Denatured Phytase (R5)	IGATNPREISSL	38 (20.04)
Denatured Phytase (R5)	KPGPLFASS	38 (19.01)
Phytic Acid (R4)	KIYTH	28 (5.49)
Phytic Acid (R4)	QGGGW	26 (8.66)
Phytic Acid (R4)	AVDL	33 (6.11)
Phytic Acid (R4)	CGSGL	33 (6.99)
Phytic Acid (R4)	GLKHFYRPVWID	28 (6.19)
Phytic Acid (R4)	LNPA	24 (3.51)
Phytic Acid (R4)	LTPSQG	32 (5.76)
Phytic Acid (R4)	TRSPK	33 (8.83)
Phytic Acid (R5)	AFFIQTPD	60 (8.74)
Phytic Acid (R5)	QWSYSKDFSPN	51 (12.1)
Phytic Acid (R5)	IHRMVYL	51 (8.25)
Phytic Acid (R5)	REICR	49 (13.14)

[1] The NGS data set separated by screening round number.
[2] The amino acid sequence of the HCDR3 being rescued.
[3] The positive replicate, barcode number with the (highest Z score) for the intended rescue.

For each rescue, 5’ phosphorylated inverse PCR primers were designed based on each clones specific HCDR3 sequence (table 4.8).

Table 4.8: Inverse PCR primer pairs for phagemid rescues for scFvs selected against active and denatured phytase.

Primer name [1]	FRW4[2]	Sequence 5' 3' [3]
TSLLGVNR	Universal	5'PHO-ACGATTAACACCCAACAAACTAGT-3'
YLAEFG	Universal	5'PHO-CCCAAACTCGGCCAAATA-3'
GWFFV	Universal - SHIFT	5'PHO-GCCCCAAACAAAAAACCAACC-3'
IKPACNTMA	Universal	5'PHO-CGCCATAGTATTACACGCAGGCTTAAT-3'
VGESQ	Universal - SHIFT	5'PHO-GCCCCAACGATTAACACCCAACAAACTAGT-3'
PDDLIG	Universal	5'PHO-CCCAATCAAATCATCAGG-3'
PPSVHFDY	Universal	5'PHO-ATAGTCAAAATGCACAGAAGGAGG-3'
STQLP	Universal - SHIFT	5'PHO-GCCCCAAGGCAGCTGCGTAGA-3'
ASCS	Universal - SHIFT	5'PHO-GCCCCAAGAACACGAAGC-3'
DSEFDTK	Universal	5'PHO-CTTAGTATCAAACTCCGAATC-3'
HMLFTS	Universal	5'PHO-AGACGTAAACAACATATG-3'
LMSQQ	Universal - SHIFT	5'PHO-GCCCCACTGCTGCGACATAAG-3'
PEPQPFDY	Universal	5'PHO-ATAGTCAAAAGGCTGCGGCTCAGG-3'
GGASPLHN	Universal	5'PHO-ATTATGAAGAGGCGAAGCCCCCCC-3'
LLDGQG	Universal	5'PHO-CCCCTGCCCATCAAGCAA-3'
RVFLLFDY	Universal	5'PHO-ATAGTCAAAAAGCAGGAAAACGCG-3'
VLGDRH	Universal	5'PHO-ATGACGATCCCCAAGCAC-3'
WGPGSE	Universal	5'PHO-CTCAGACCCAGGCCCCCA-3'
AEAAD	Universal - SHIFT	5'PHO-GCCCCAATCCGCAGCCTCCGC-3'
AETAD	Universal - SHIFT	5'PHO-GCCCCAATCCGCAGTCTCCGC-3'
ALNLR	Universal - SHIFT	5'PHO-GCCCCAACGCAAATTCAACGC-3'
ASMSPVH	Universal	5'PHO-ATGCACCGGACTCATACTCGC-3'
CQIEW	Universal - SHIFT	5'PHO-GCCCCACCACTCAATCTGACA-3'
GAMEYG	Universal	5'PHO-ACCATACTCCATCGCACC-3'
GAYEDYVQ	Universal	5'PHO-CTGCACATAATCCTCATAAGCACC-3'
GDPS	Universal - SHIFT	5'PHO-GCCCCAACTCGGATCCCC-3'
GTYLY	Universal - SHIFT	5'PHO-GCCCCAATACAAATAAGTCCC-3'
IGATNPREISSL	Universal	5'PHO-AAGACTCGAAATCTCCCGCGGATTCGTAGCCCCAAT-3'
KPGPLFASS	Universal	5'PHO-CGAACTCGCAAACAGCGGCCCAGGCTT-3'
KIYTH	Universal - SHIFT	5'PHO-GCCCCAATGCGTATAAATCTT-3'
QGGGW	Universal - SHIFT	5'PHO-GCCCCACCAACCACCACCCTG-3'
AVDL	Universal - SHIFT	5'PHO-GCCCCACAGATCCACCGC-3'
CGSGL	Universal - SHIFT	5'PHO-GCCCCACAACCCACTACCACA-3'
GLKHFYRPVWID	Universal	5'PHO- ATCAATCCAAACAGGCCTATAAAAATGCTTCAGCCC-3'
LNPA	Universal - SHIFT	5'PHO-GCCCCAAGCAGGATTCAA-3'
LTPSQG	Universal	5'PHO-CCCCTGACTAGGCGTCAG-3'
TRSPK	Universal - SHIFT	5'PHO-GCCCCACTTCGGACTCCTCGT-3'
AFFIQTPD	Universal	5'PHO-ATCAGGCGTCTGAATAAAAAACGC-3'
QWSYSKDFSPN	Universal	5'PHO-ATTCGGAGAAAAATCCTTCGAATAACTCCACTG-3'
IHRMVYL	Universal	5'PHO-CAAATAAACCATCCGATGAAT-3'
REICR	Universal - SHIFT	5'PHO-GCCCCACCTACAAATCTCCCG-3'
FRW4 [2]		**Sequence 5' 3'** [3]
Universal		5'PHO-TGGGGCCAAGGTACMMYGGTCACYGTCTC
Universal – SHIFT		5'PHO-CAAGGTACMMYGGTCACYGTCTC

[1] List of scFv (amino acid sequence) being rescued and their primer labels

[2] Each scFv forward primer except for ESPFWSGYV was paired with the universal degenerate primers FRW4 or FRW4-Shift

[3] 5’ phosphorylated primer sequences targeting the HCDR3 and FRW4 regions. Shaded primers were extended into the FRW4 region to elongate the primers, and thus paired with the Universal – SHIFT FRW4 primer.

Each scFv was rescued from the DNA extracts of the polyclonal phage sub library that had the greatest Z score for that particular scFv. The frequencies of each scFv within their most enriched experiment ranged from 0.008—1.956% of their phagemid populations.

Using the DNA extracted from round 2 panning as the template, an Inverse PCR reaction, extending to include the entire pHEN 2 vector, was carried out using a high-fidelity DNA Polymerase. PCR products were checked on an agarose gel (data not shown). Amplicons with the correct band size of 5kb were extracted and purified from the agarose gel and then quantified by fluorometry (concentrations ranged from 2—6 ng/µl however many were too low to read). After repeated attempts, PCR products at around ~5kb could not be generated for the rescues AEAAD, AETAD, GAMEYG, GDPS, KPGPLFASS, GLKHFYRPVWID, IHRMVYL, LTPSQG, TRSPK or RVFLLFDY and they were not taken further.

4.14 Testing the specificity of the scFv rescues via colony PCR and phage ELISA

The denatured phytase rescues GAYEDYVQ, GTYLY, IGATNPREISSL, ALNLR, ASMSPVH, WGPGSE and active phytase rescues IKPACNTMA, ASCS, DSEFDTK, GWFFV, STQLP, VGESQ and YLAEFG produced 1, 4, 1, 1, 1, 1, 2, 9, 4, 1, 1, 18 and 1 transformants respectively. The phytic acid rescues AVDL, LNPA and QGGGW produced 9, 5 and 3 colonies respectively. The remaining rescues failed to produce any colonies despite repeated attempts. All of the inverse-PCR transformants were picked from each plate the following day and were used to inoculate a colony PCR master mix, before being grown to $O.D_{600nm}$ 0.5-1 in deep 96-well plates. The colony PCRs were used to check for the presence of the scFv gene in each bacterial colony as described in methods section 2.22.

Monoclonal phage were also produced overnight for each colony by sub-culturing the master plate, and their phage-scFv fusions were tested in a functional phage ELISA.

Colony PCR

An scFv amplicon of ~1,000 bp was expected for a full length antibody. The PCR products were analysed on an agarose gel to check for the presence of the expected scFv genes (figures 4.23). Evidence of an scFv gene was detected in 35 out of 62 picks. However, all of the picks appeared to have truncated antibodies with amplicon sizes around ~700 bp.

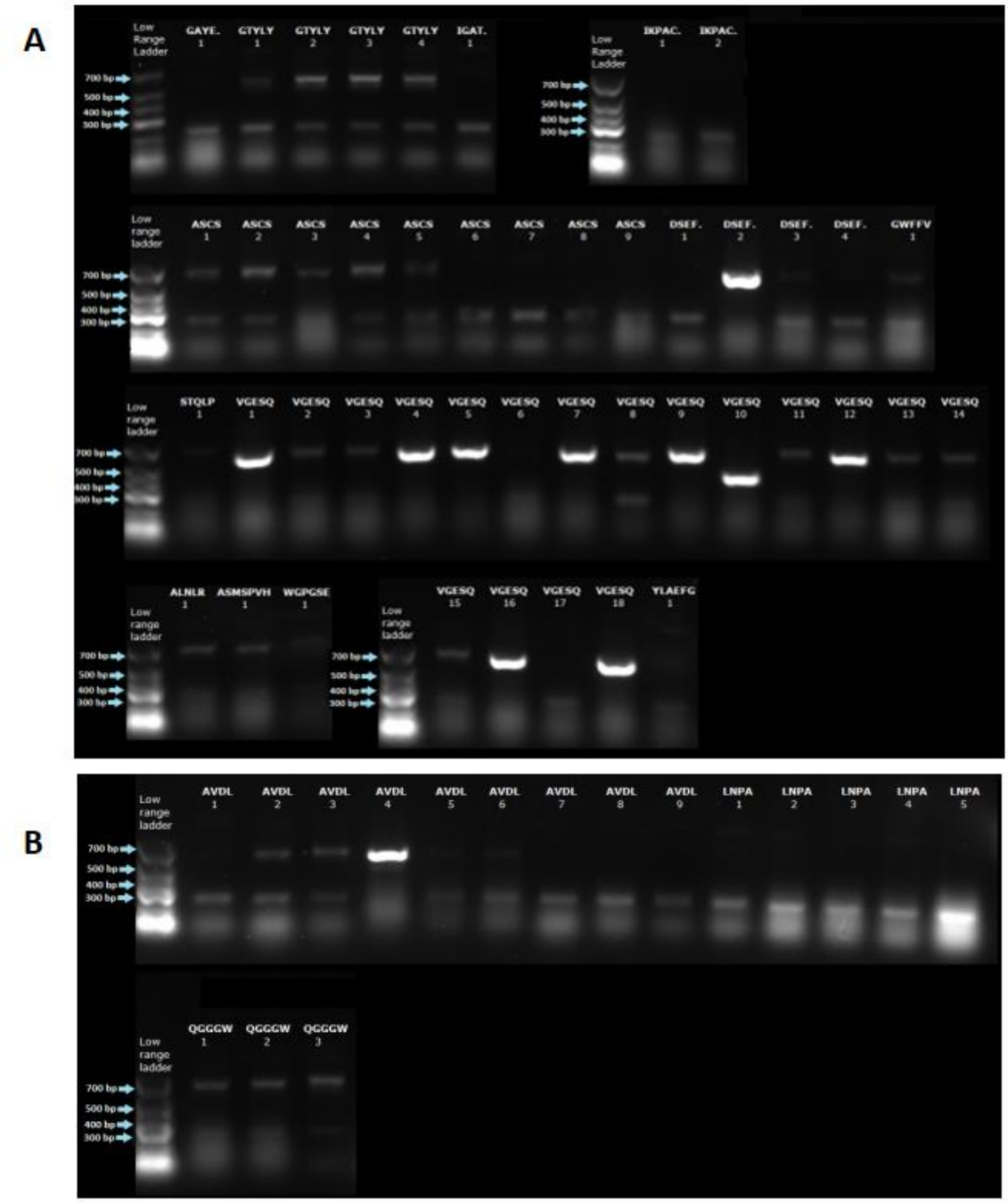


Figure 4.23: Amplification of the scFv genes present in the Inverse PCR transformants for clones isolated against Active and Denatured Phytase and Phytic Acid. The Inverse PCR transformants for active and denatured phytase (A) and phytic acid (B) were used as the template for a

colony PCR. The scFv genes were amplified and the resulting products were analysed on a 2% (w/v) agarose gel and checked against a LR Gene Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6x loading buffer.

Phage ELISA

The inverse-PCR transformant colonies were picked for each rescue and used as template for a colony PCR to confirm the presence of the scFv gene. The same colonies were then grown in a deep 96-well plate to produce phage overnight. The phage supernatant was used in an Anti-M13-HRP phage ELISA, however all of the phage cultures failed to produce a positive absorbance reading in an ELISA (data for active phytase and phytic acid candidates not shown). The absorbance signals of the denatured phytase rescues were compared against the 'C7' clone from the monoclonal picking section (figure 4.24), which produced a positive signal of 0.61 (3X background). The graph is shown only for comparative purposes.

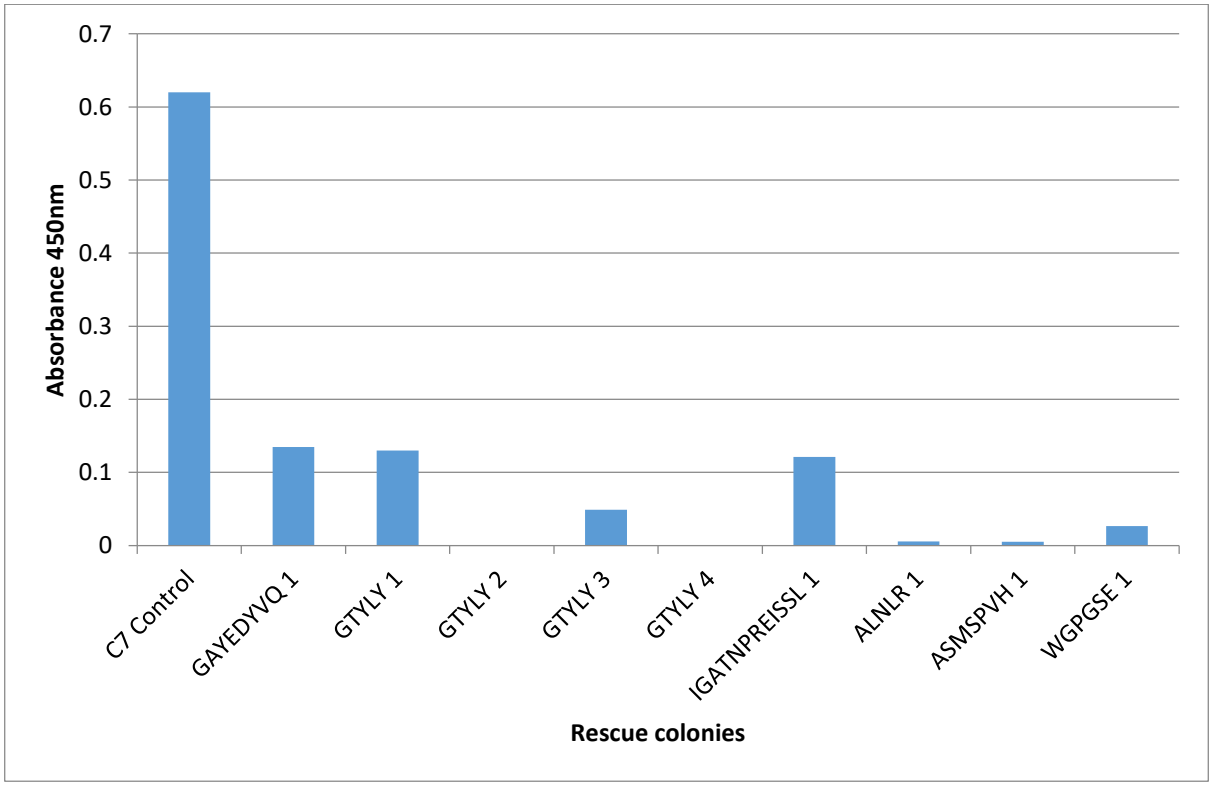


Figure 4.24: Phage-scFv ELISA for the denatured phytase inverse PCR rescues. Denatured phytase (autoclave method) was immobilised overnight at 20 µg/ml on a Nunc MaxiSorp® plate. Monoclonal phage-scFv for each clone were used to detect immobilised, denatured phytase (100 µl of ~10^{11}—10^{13} CFU/ml). Bound phage were detected by an Anti-M13-HRP antibody and TMB substrate. After 10 minutes the reaction was stopped with

100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. No antigen values (0.2) were subtracted.

Sanger sequencing

The colony picks were mini-prepped and sent for Sanger sequencing to check for the scFv sequence as described in methods section 2.23. Each sequence was aligned against VBASE to check for the expected antibody frameworks and confirm the presence of the variable chains (table 4.9).

Table 4.9: Sanger Sequencing for Inverse PCR transformant colonies

scFv Rescue[1]	No. Inv. PCR Colonies[2]	Evidence of scFv Amplicon? [3]	Sanger Sequencing[4]
GAYEDYVQ	1	No	?
GTYLY	4	Yes	?
IGATNPREISSL	1	No	?
ALNR	1	Yes	?
ASMSPVH	1	Yes	?
WGPGSE	1	Yes	VYYCMQ ALQTLLT FGQGT Light chain only. LCDR3
IKPACNTMA	2	No	?
ASCS	9	Yes	?
DSEFDTK	4	Yes	?
GWFFV	1	Yes	?
STQLP	1	Yes	?
VGESQ	18	Yes	?
YLAEFG	1	No	VYYCMQ ALQTRT FGQGT Light chain only. LCDR3
AVDL	9	Yes	?
LNPA	5	No	?
QGGGW	3	Yes	TYY*QQ YDNLPSDT FGQGT Light chain only. LCDR3

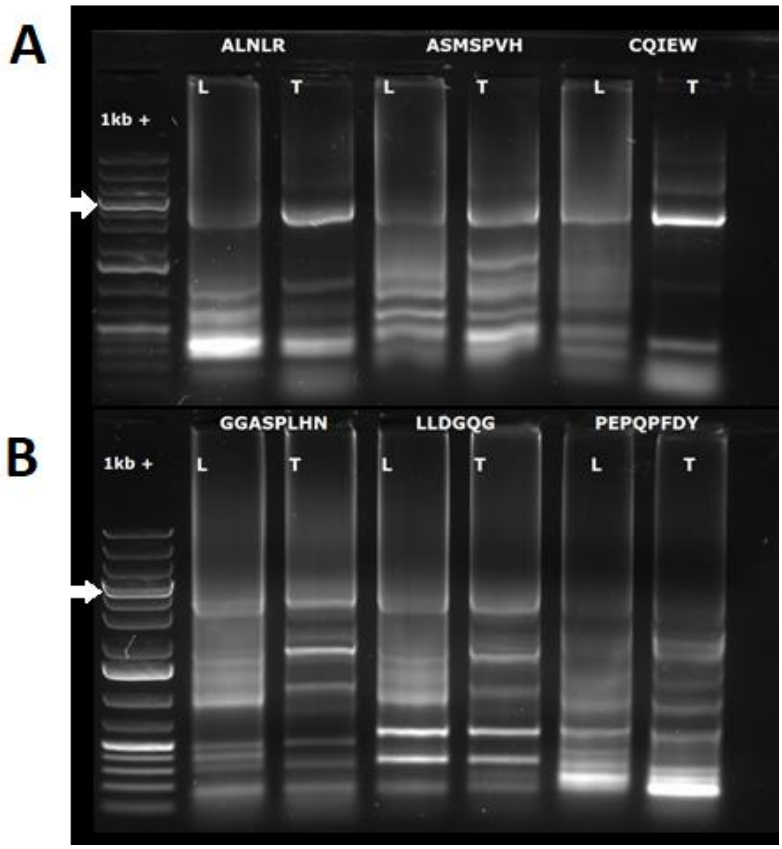
[1] The amino acid sequence of the HCDR3 being rescued.
[2] The total number of inverse PCR transformant colonies.
[3] Evidence of an scFv gene based on the colony PCRs agarose gels.
[4] (H/L)CDR3 Sanger sequencing result covering the scFv gene

Many of the DNA extracts came back from Sanger Sequencing with no data. Other colonies were mini-prepped and sent for sequencing as repeats but the majority encountered the same problem. Alternative pHEN2 primers were also tested (forward and reverse LMB3 and PHEN) primers but they also failed to generate sequence data.

4.15 Inverse PCR primer specificity

The inverse PCR step was unsuccessful at retrieving the correct clones from the polyclonal sub libraries. Similar difficulties were encountered in chapter 3 where the incorrect HCDR3 was amplified, if at all.

The following experiment aimed to address whether the primers were annealing with low specificity. A new Inverse PCR reaction was conducted using the 11 rescue primers designed for the denatured phytase clones that had previously produced ~5kb amplicons (ALNLR, ASMSPVH, CQIEW, GGASPLHN, LLDGQG, PEPQPFDY, GAYEDYVQ, GTYLY, IGATNPREISSL, VLGDRH and WGPGSE). They were used in combination with the degenerate FRW4 primers but were then combined with ssDNA from the experiment with the clone's best Z score (target template) as before, or with dsDNA extracted from the Griffin library as the template (that should have negligible concentrations of the target HCDR3). The resulting Inverse PCR products were visualised on an agarose gel with the Griffin base library template (L) and target template (T) samples side by side for each rescue (figure 4.25).



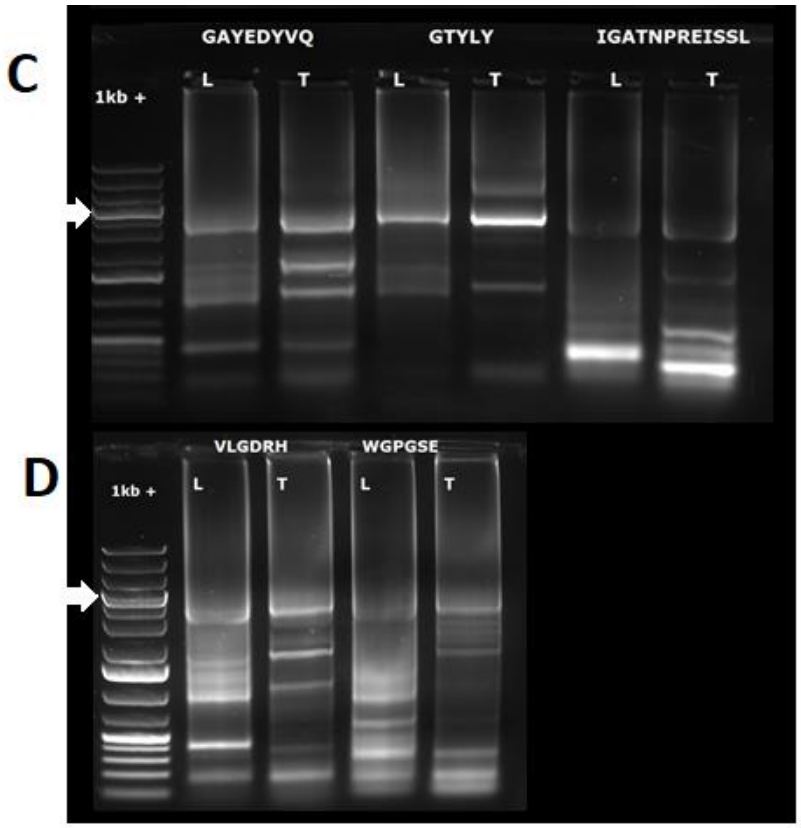


Figure 4.25: Amplification of the pHEN2 phagemid by Inverse PCR using library and target template. The pHEN2 phagemid was amplified from the base Griffin library (L) or target sample template (T) using HCDR3 specific and degenerate FRW4 primers. The resulting products were analysed on a 2% (w/v) agarose gel and checked against a 1kb+ Ladder. Nancy-520 was used to stain the gel and samples were loaded using 6x loading buffer. The products were intended to be ~5kb; the 5kb marker is indicated by the white arrow.

The intended rescues were thought to be too dilute within the Griffin library, without pre-selection and enrichment, to produce a PCR product; unless the primers were able to anneal non-specifically. The clones ALNLR, ASMSPVH, CQIEW, GGASPLHN, LLDGQG, GAYEDYVQ, GTYLY and WGPGSE (8/11) all produced noticeably more intense PCR products around ~5kb using the target DNA (T) as template rather than the Griffin library (L). Despite the presumably small quantity of target DNA present in the library, an amplicon was produced for most of the rescues.

4.16 Discussion

The present chapter built upon the NGAPD strategy conducted in chapter 3. The present chapter increased the number of panning rounds and introduced subtractive steps with the intention of reducing the presence of nonspecific phage within the panning round sub-libraries. After trying to rescue candidates from their bacterial glycerol stocks in chapter 3, the present chapter increased the number of panning rounds to increase the frequency of the target clones before trying to rescue them. The present chapter also used the colony picking method for comparison against the bioinformatics approach.

Prior to panning, the Quantum Blue phytase was found to be free from contaminating proteins as demonstrated by the LC-MS analysis conducted on multiple SDS-PAGE gel extracts. The extracts all resembled the expected amino acid sequence for Quantum Blue phytase. The gel extract with the highest molecular weight resembled Quantum Blue phytase the least, this is possibly due to aggregation caused by the denaturing process, and the exposure of reactive groups causing the typical/expected sequence to become distorted. In combination with the single band shown by the native protein gel, the data suggested that the Quantum Blue phytase solution was free from contaminants and when the active enzyme was the target in panning experiments it could be used in its most active, and untreated state.

In addition to the phytase conformations, a NGAPD strategy was developed to identify antibodies that bound specifically to phytic acid. Phytic acid is known as a hapten, a small (660 daltons) non-immunogenic molecule that does not trigger an immune response by itself. Haptens such as phytic acid can also be modified (Xu et al., 2005, Sathe et al., 2011) to facilitate antibody recognition but it remains complicated and expensive to engineer them so that they can become compatible with Hybridoma technology. In order to induce an immune response haptens are often conjugated to a large carrier protein. The antibodies that are identified often bind to the carrier protein or the conjugated form of the hapten rather than to the hapten itself. By removing the need to trigger an immune response entirely, antibody phage display technology can potentially overcome the limitations of

hybridoma technology and serve as a more suitable alternative for antibody discovery against more difficult targets such as haptens. Nevertheless, phage display experiments aiming to isolate binders against hapten targets continue to conjugate them to carrier proteins such as keyhole limpet hemocyanin (KLH) or BSA (Charlton et al., 2001, Charlton and Porter, 2002, Al Qaraghuli et al., 2015, Patlewicz et al., 2016). In contrast to hybridoma technology, phage display allows a greater number of antibodies to be investigated at any one time and therefore increases the likelihood of isolating antibodies that are specific to the lone hapten. The carrier conjugates still present a larger, non-specific selection target and present haptens in a potentially altered state and should ideally be removed from the experiments entirely. In this chapter, phytic acid was not conjugated to a carrier protein in order to facilitate immobilisation and hapten specific recognition. Instead, the panning strategy was reliant on the immobilisation of phytic acid by its phosphate groups. Phytic acid will not bind to plastic plates due to a lack of hydrophobic/hydrophillic domains so instead, five possible supports were identified from the literature - poly 4-vinyl pyridine (PVP) (Tsao et al., 1997), bentonite (Ulusoy et al., Ulusoy and Simsek, 2005), zeolite (Ulusoy and Simsek, 2005, Faghihian and Farsani, 2013) and metals such as titanium dioxide (Wilson et al., 2015) and zirconium dioxide (Kweon and Håkansson, 2006). With 6 phosphate groups, the phytic acid can interact with positively charged metal ions such as titanium or zirconium (Yang et al., 2007). Each of the supports had been shown to adsorb phytic acid within their respective publications however, titanium dioxide (TiO_2) and zirconium dioxide (ZiO_2) were successfully replicated during the present study. The immobilisation of phytic acid on TiO_2 and ZiO_2 was confirmed by both phytase digestion and staining with Toluidine blue. For TiO_2, the digested signal for immobilised-phytic acid was not as high as the start solution due to hindered ability of the phytase to access bound phytic acid and digest it. This may reflect the lower quantity of phytic acid present on the plates possibly due to the binding capacity of the wells. Only single wells were assessed for digestibility thus it is unsurprising that the absorbance signals achieved for the plates was considerably lower than that of the TiO_2 beads (10 mg).

After panning, colonies were randomly picked from round 4 bacterial glycerol stock. Four clones against denatured phytase and four clones against

phytic acid demonstrated binding activity when screened by phage ELISA. The clones were tested in multiple ELISAs against similar antigens and remained specific to their targets. Sanger sequencing revealed that 7 of the 8 picks were truncated antibodies that had lost at least their heavy chains. Based on the light chains, the Sanger sequencing data identified 3 of the 4 denatured phytase as the same clone, which had simply been picked on three separate occasions. In contrast, 3 of the 4 phytic acid clones each had different LCDR3 sequences. The high prevalence of truncated antibodies suggested either a large number of mutations occurring within the library itself, or that the panning strategies were selecting for smaller antibody fragments. The smaller antibody fragments may have been more successful during the growth stages due to their lower toxicity to the host. Where visible, the pelB leader sequence was out of frame for all of the truncated antibodies this would suggest that the expressed proteins were never exported to the periplasm where they could be packaged into new virions. It is peculiar that the clones should demonstrate such specificity during phage ELISAs if they were unlikely to have been exported to the periplasm. After inducing expression with both IPTG and exphage, the anti-denatured phytase clones were faintly visible by western blot, suggesting that they were infact being expressed. It is possible that the region was a difficult sequencing target and the frameshifts were sequencing errors.

Only the 'C7' clone (anti-denatured phytase) and 'E6' clone (anti-phytic acid) were pursued further as soluble antibodies. Antibodies were expressed as MBP fusions after having successfully subcloned the genes into the pMAL P5X vector. Expression was confirmed by western blot and each entity was detected when immobilised on a Nunc MaxiSorp® plates. As soluble antibodies extracted from the periplasm, both the 'C7' and 'E6' clones failed to produce a detectable absorbance signal against their targets, whereas the 'bLAC' positive control produced a strong signal. It is unclear as to why the novel clones lost their binding ability after being expressed as MBP fusions, however similar problems have been seen by other studies. The study by Ravn et al. (2010a) tested six scFv candidates derived from both random colony picking and the present studys bioinformatics approach. The scFv clones were expressed as both phage fusions and soluble antibodies extracted from the periplasm. The study found that as phage fusions, all six of the

clones bound to their target whereas only the clones derived from colony picking continued to bind to their target as soluble antibodies. The authors suggest that in some cases the phage may have become an essential part of the binding process considering that is the format in which the antibodies are panned against their intended target. In other cases, the author suggested that any lack of binding was likely due to suboptimal growth and expression because the clones were grown in a low volume culture and less aeration in a 96-well format. In the present study, the antibodies were expressed as MBP fusions in 100 ml cultures so volume and aeration were unlikely causative factors.

The major pVIII protein on the phage surface is composed of three domains. The hydrophilic N terminal domain is exposed to the external environment of the virus, giving the outer coat a negative charge (Chung et al., 2014). It is possible that the phage had become part of the binding complex, if the antibodies only had a low affinity for the targets then in combination with the negative charge of the phage complex, the avidity effect may be enough to retain the clones during the vigorous wash steps during the panning process. Alternatively, there is some distance between the pVIII proteins and the antibodies and perhaps it is more likely that the pIII N-terminal region is contributing towards the binding interaction. This has been seen multiple times for peptide display experiments (personal communication, Gough). Removing the phage entity may reduce the affinity of the antibody-antigen interactions enough that they can easily be washed away during a soluble antibody ELISA.

Previous NGAPD studies had shown that known positive clones could be used to probe NGS data as a means of narrowing down the criteria for identifying new positive binders from the data (Turner et al., 2016). Only the anti-phytic acid clone 'F6' could be found within the data as it was the only clone that had retained its heavy chain. When the positive replicates were compared against the negative pool, the 'F6' clone only had a Z score of 1.41 in a single replicate. The Z score was lower than the intended thresholds of the study ($Z \geq 2$, $Z \geq 5$, $Z \geq 10$) and likely just a reflection of the sequence being present in a positive replicate. The low Z score would not necessarily discourage the likelihood of the candidate being a true binder despite the lack of activity as a soluble antibody. A study by Lovgren et al. (2016) also

demonstrated that some of the positive binders identified during the random colony picking method also had a low frequency within the NGS data or were completely absent from it, despite being picked by chance and producing positive ELISA signals. The authors remarked that the frequency of some of the picked clones were abnormally low in the NGS data, despite being picked once or twice, but were also unable to determine why. As in the current chapter, it is possible that these picked clones were simply done so by chance despite their potentially low frequency within the sub-library. Alternatively, it is possible that those particular clones had some unforeseen difficulties during the amplification steps prior to sequencing thus reducing their frequency considerably within the NGS data. For the present study, the positive phage ELISA signals for the colony picked clones suggested that any candidates with Z scores greater than the lowest threshold ($Z \geq 2$) would be strong candidates.

The present chapter built upon the bioinformatics strategy used in chapter 3 by analysing screening round data originating from later panning rounds. Positive binders were assumed to be more frequent within the later rounds, facilitating their discovery by data analysis and Z score rankings. To combat HCDR3 sequences that had become prevalent in both the positive and negative replicates, suggesting that they were binding to both the intended antigen and the negative control antigen, the data was also 'reverse' analysed. Any HCDR3 sequences with Z scores of 2 or more, in 2 or more negative replicates were omitted from further analysis, resulting in 12 of the anti-active phytase, 17 of the anti-denatured phytase and 12 of the anti-phytic acid top 20 candidates being retained after the analysis was complete. In addition to the reversed analysis, the present chapter also increased the stringency of the search by comparing the three unrelated datasets (active phytase, denatured phytase and phytic acid). In theory, any HCDR3 sequences that were enriched in 2 or more datasets suggests that they were not specific for their target.

Out of the 41 candidates identified by the bioinformatics approach, only 16 produced transformant colonies that could be cultivated and tested in phage ELISA. None of the candidates demonstrated binding activity against their intended antigens. After Sanger sequencing, only 3 candidates returned data however all 3 candidates appeared to be truncated, having lost their heavy chains. The rescue strategy was the same as conducted for chapter 3

however the DNA template was single stranded as it had been extracted directly from the panning/screening phage eluates. This strategy was more rapid and amenable to cloning large number of target clones, and also removed a propagation step that would bias the sub-library diversity towards clones with growth advantages. The recovery rate had worsened considerably despite the use of subtractive steps during panning to remove nonspecific phage, and the higher number of panning rounds. It is possible that the ssDNA did not provide enough of a starting template for the inverse PCR reaction. The combination of amplifying a large amplicon and the degenerate nature of the Griffin library scFv FWR and the FRW4 primer may have been the reason why the clones could not be recovered successfully. The specificity of the HCDR3 (denatured phytase candidates only) and FRW4 primers were tested on both the most enriched sub-library template for the individual HCDR3 candidate and then the Griffin library template. The enriched sub-library templates often produced stronger amplicons than the Griffin library but surprisingly, the Griffin library template was able to produce substantial amplicons in many cases. The low copy number of the intended candidates in the Griffin library should have severely restricted the success of the PCR and the band intensity indicates non-specific binding of the primers to other clones. The non-specific binding of the primers probably resulted in the incorrect clones being amplified.

Similarly, to chapter 3, the bioinformatics approach could not be disproven as the candidates were not rescued as intended. The rescue strategy remains a major challenge for the large, scFv candidates but synthesising a large number of them, such as the 41 candidates in this study, is not a financially viable option. Other studies have resorted to full gene synthesis such as Saggy et al. (2012) who synthesised 3 candidates and Lovgren et al. (2016) who synthesised 10 candidates. A number of different recovery strategies for scFv exist in the literature. The study by Zhang et al. (2011) used biotinylated primers to target the HCDR3 region. The regions were amplified and then captured by streptavidin-coated magnetic beads. The process allowed the recovery of individual clones by physically removing them from the enriched sub-library, enabling the authors to transform *E. coli*. Alternatively, some studies (Ravn et al., 2010b, Ravn et al., 2013, Yang et al., 2017) used a two-step PCR approach to rescue scFv candidates from their sub-libraries. The VH

and VL chains were amplified separately and then bridged together by the second PCR step using primers targeting the specific HCDR3 sequence. The joined amplicons were then digested and cloned back into the expression vector. The study by D'Angelo et al. (2014) used an inverse PCR strategy similar to that used by the present study. The present study used a primer covering the entirety of the HCDR3 sequence in combination with a degenerate FRW4 primer that could be applied to a range of rescue candidates as described by Spiliotopoulos et al. (2015). The study by D'Angelo et al. (2014) split the HCDR3 sequence between both primers which extended into the adjacent framework regions. The degeneracy of the framework regions would likely affect the specificity of the primers unless the HCDR3 sequence was particularly long and encompassed the majority of the primer sequence. By covering the entire HCDR3 sequence with one primer, the specificity of the primer used in the present study was expected to be much greater. Similarily, the study by Sasso et al. (2015) rescued scFv clones from their enriched sub-libraries using a QuickChange II XL site-directed mutagenesis kit. The target HCDR3 sequences were amplified by overlapping primers which extended to cover the entire vector. The sample was then digested by Dpn1 to remove the methylated DNA from the sample but retain the synthetic, un-methylated amplicon which could be used directly to transform new cell lines.

After testing the inverse PCR primer pairs with target and library DNA templates, there were a high number of non-specific bands present in all of the reactions which suggested that the cycler conditions and/or PCR master mix components were perhaps not stringent enough.

The rescue strategy proved even more difficult than the first attempt in chapter 3. The strategy showed little improvement between panning experiments, with clones randomly being rescued from the enriched sub-libraries despite specific HCDR3 primers and the targets themselves being present at unexpectedly low frequencies after multiple panning rounds (0.008—1.956%). This highlights the difficulties of rescuing scFv from the sub-libraries after being identified by NGS. Although NGS gives us the additional read depth and ability to compare multiple replicates and maximise the chances of identifying novel binders, the technique still requires clones to be physically rescued from the polyclonal sub-library.

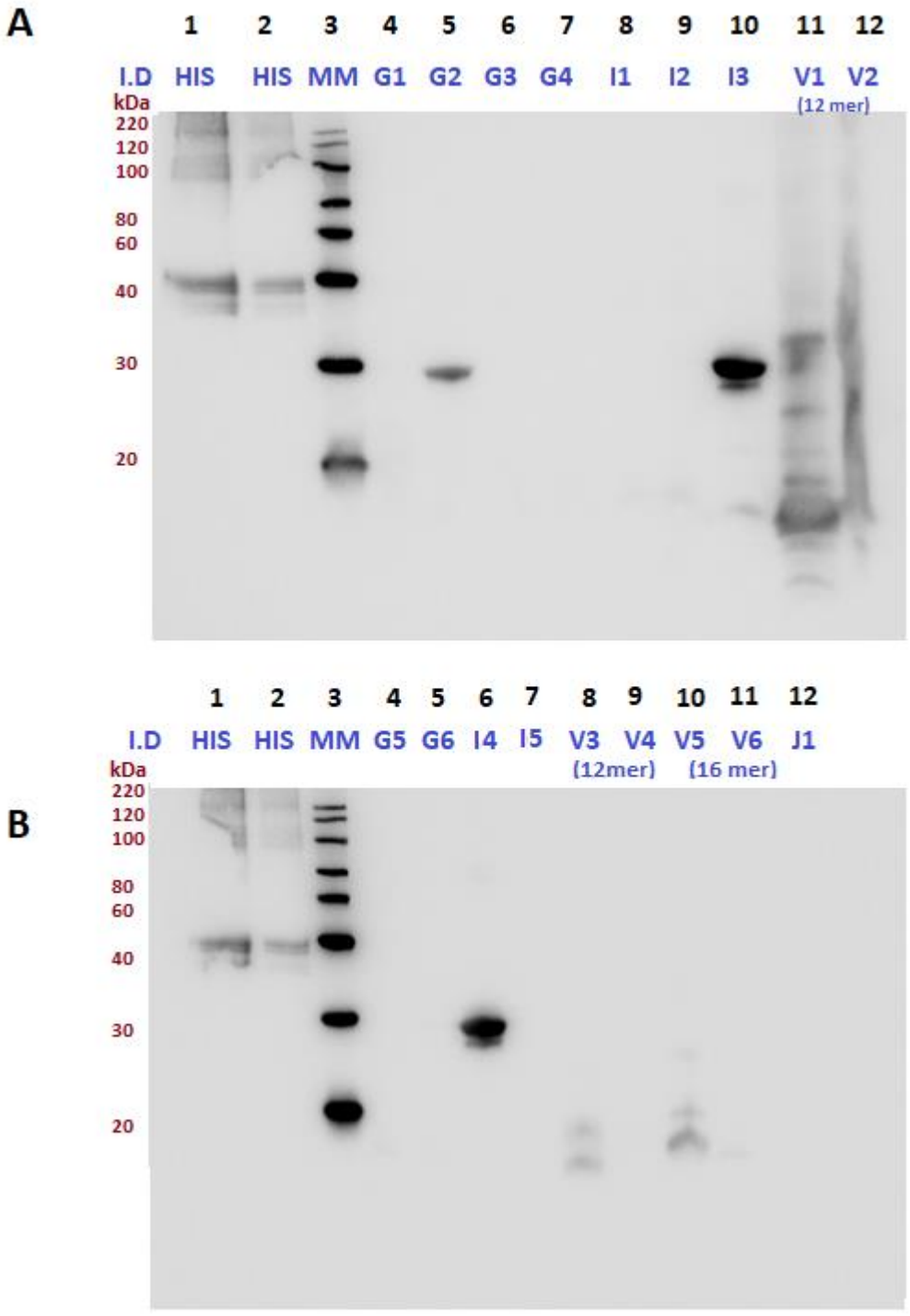
The colony picking technique used in this chapter successfully identified phage-scFv binders against denatured phytase and phytic acid however, they failed to bind as soluble MBP fusions. The inability to accurately retrieve scFv clones was a major setback. The Griffin library itself appeared to contain multiple truncated antibodies and it is unclear whether the panning strategy specifically selected for them. The truncated clones may have had a lower toxicity than full length scFvs, or they may have been more suited to the target antigens. Alternatively, the library itself may have mutated during storage and repeated propagations. Furthermore, the lack of sequences returned by Sanger sequencing suggested that many of the clones were likely containing mutations in the vector sequence and the primers were not annealing. These problems severely hindered the usefulness of the library and the ability to test the bioinformatics strategies used within the study.

Chapter 5: Identifying single domain antibodies from a naïve VHH library using traditional and next generation phage display

The next chapter of the study compared the expression levels of the Griffin library to other antibody phage display libraries available to the research group and devised a new NGAPD panning strategy. The library with the highest expression rate would be used for further panning experiments. Two commercially available human scFv libraries Griffin (Griffiths et al., 1994) and the Tomlinson library made up of the I & J libraries (Nissim et al., 1994) and an in-house camelid VHH antibody were compared. The VHH library was based on a pSD3 (Li et al., 1999) which resulted in the expression of pIII-VHH fusions. With diversity introduced into the VHH library's HCDR3, it comprised of three sub-libraries containing different lengths of HCDR3 sequences – 12mers, 16mers and 21mers.

5.1 Comparison of antibody expression levels between three phage display libraries

A total of ten colonies from each library were transformed into electro-competent HB2151 cells and were used to express soluble antibodies which were extracted from both the whole cell and the periplasm; refer to methods section 2.26. Antibody expression was induced for 10 colonies from each library. Ten samples were analysed for the Griffin scFv library, five samples for the Thomlinson scFv I library and five samples for the Thomlinson scFv J library, 4 samples for the VHH 12mer library, 3 samples for the VHH 16mer library and 3 samples for the VHH 21mer library (figures 5.1—5.2). The scFv antibodies from the Griffin and Thomlinson libraries were expected to be ~30 kDa and the VHH antibodies from the camelid libraries were expected to be ~15 kDa.


A
1 2 3 4 5 6 7 8 9 10 11 12
I.D HIS HIS MM G1 G2 G3 G4 I1 I2 I3 V1 V2
(12 mer)
kDa
220
120
100
80
60
40
30
20
B
1 2 3 4 5 6 7 8 9 10 11 12
I.D HIS HIS MM G5 G6 I4 I5 V3 V4 V5 V6 J1
(12mer) (16 mer)
kDa
220
120
100
80
60
40
30
20

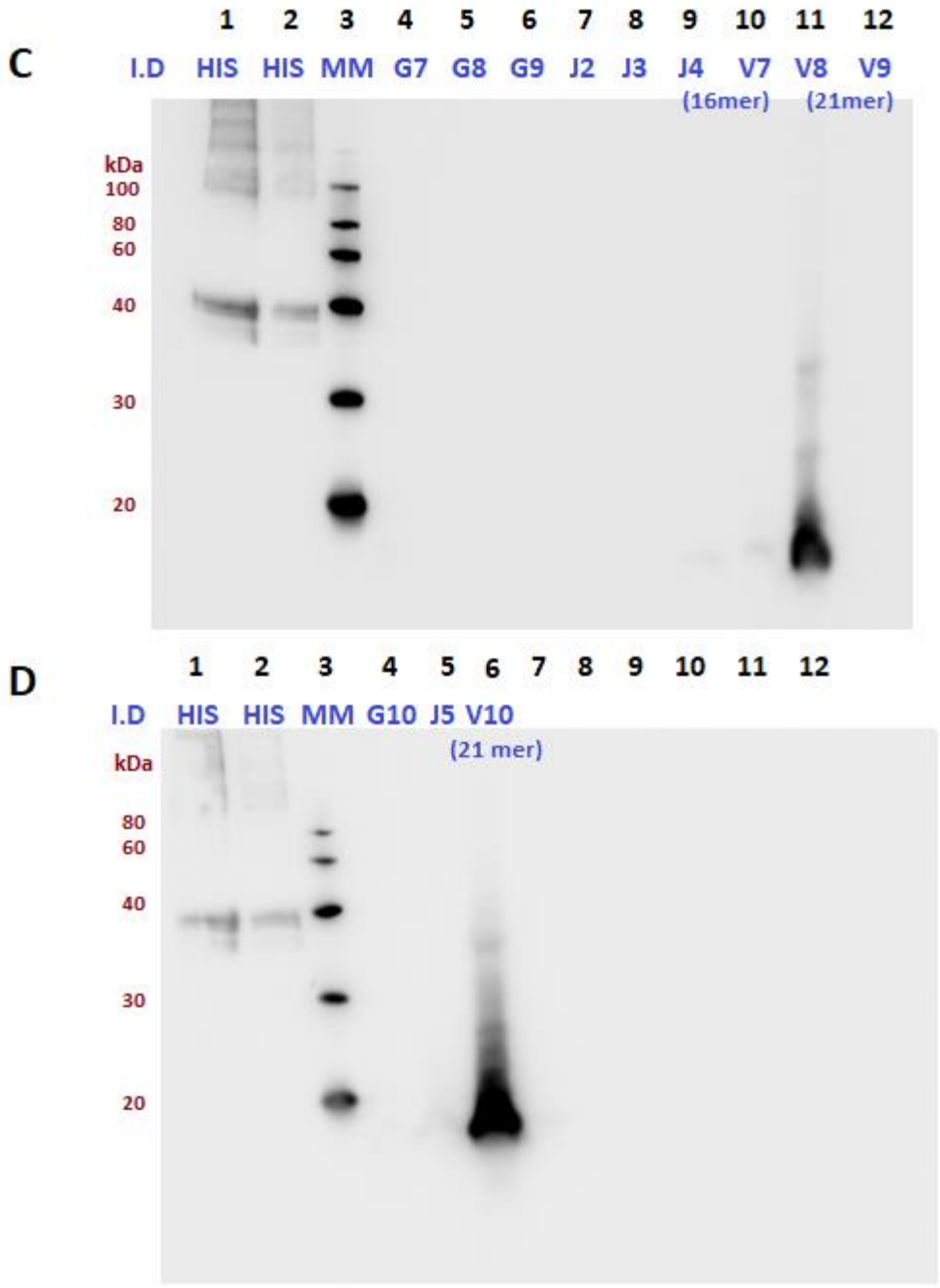
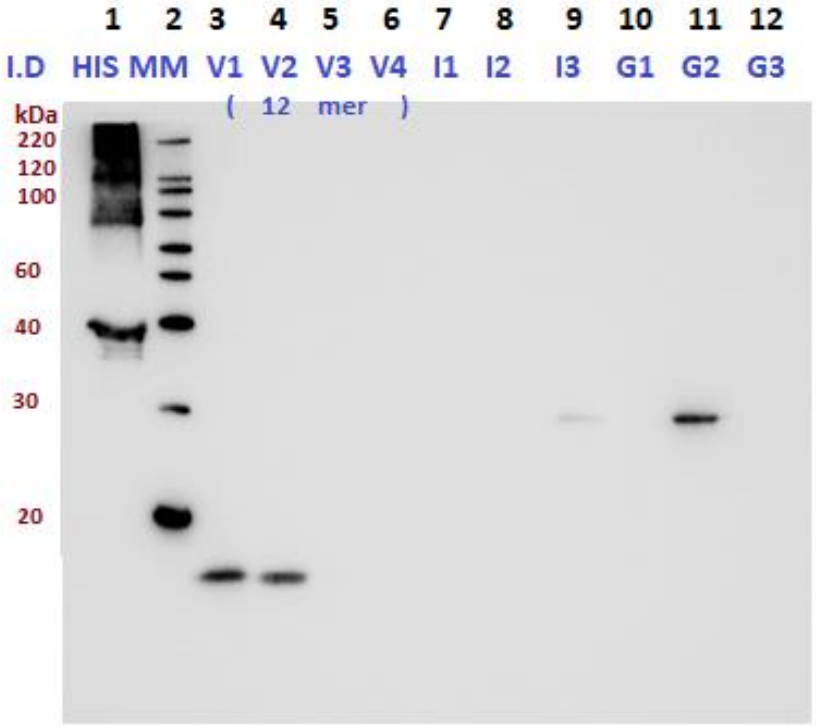
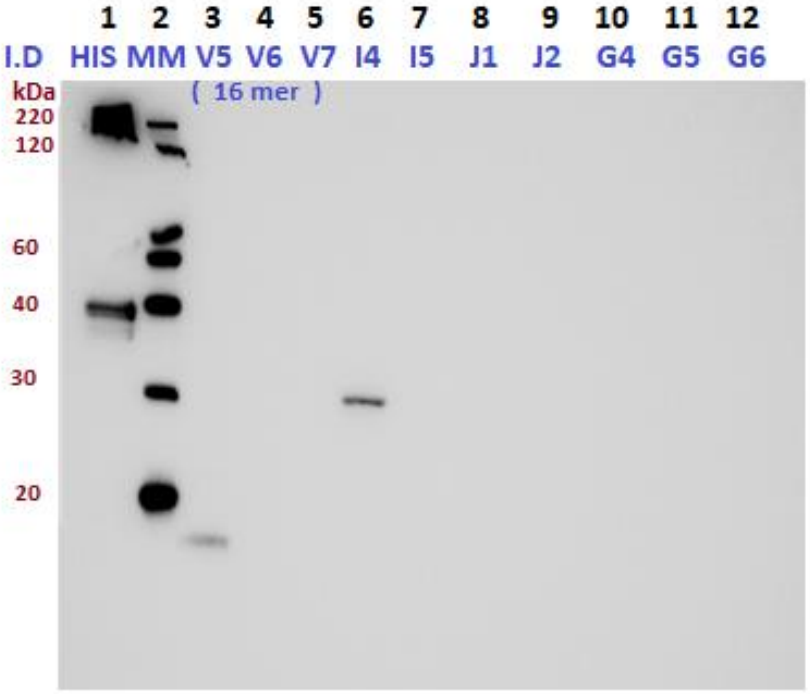


Figure 5.1: Western blot analysis of soluble extracts from whole cell extracts. The western blots A—D compares the expressed soluble antibodies derived from whole cell extracts from the Griffin (**G**), Thomlinson (**I** or **J**) and VHH (**V**; **12**mer, **16**mer, **21**mer) libraries (lanes 4—12). Lane 1 and 2 showed a control HIS-tagged protein (containing 20 µl and 10 µl respectively). Lane 3 contained the molecular weight ladder, magic marker (MM). I.D refers to identifier. Soluble antibodies were detected via their HIS tags using an anti-tetra-his primary antibody (1:2000) and an anti-mouse-HRP secondary antibody (1:2000).

A


1 2 3 4 5 6 7 8 9 10 11 12
I.D HIS MM V1 V2 V3 V4 I1 I2 I3 G1 G2 G3
(12 mer)
kDa
220
120
100
60
40
30
20

B


1 2 3 4 5 6 7 8 9 10 11 12
I.D HIS MM V5 V6 V7 I4 I5 J1 J2 G4 G5 G6
(16 mer)
kDa
220
120
60
40
30
20

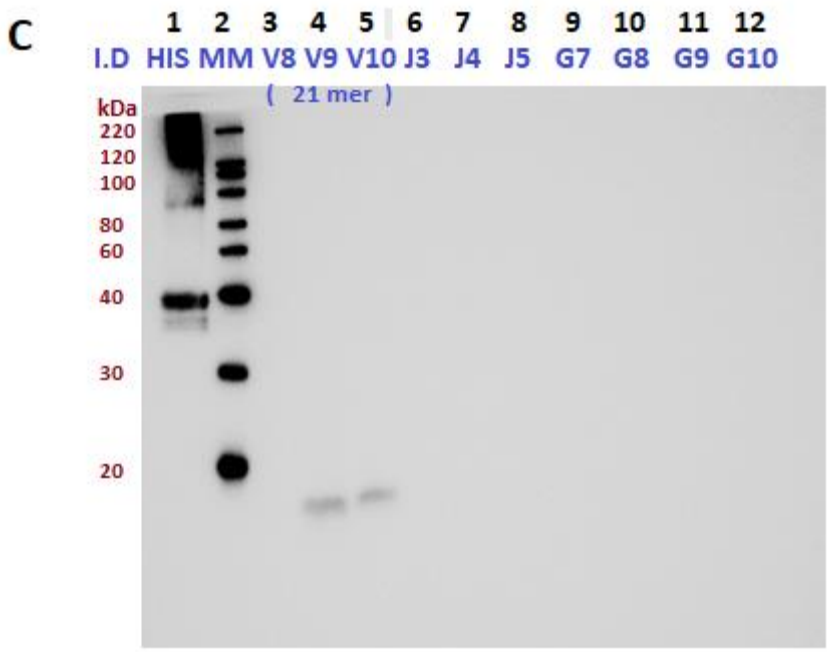


Figure 5.2: Western blot analysis of soluble extracts from the periplasm. The western blots A—D compares the expressed soluble antibodies derived from periplasmic extracts from the Griffin (**G**), Thomlinson (**I** or **J**) and VHH (**V**; **12**mer, **16**mer, **21**mer) libraries (lanes 4—12). Lane 1 showed a control HIS-tagged protein (10 µl). Lane 3 contained the molecular weight ladder, magic marker (MM). I.D refers to identifier. Soluble antibodies were detected via their HIS tags using an anti-tetra-his primary antibody (1:2000) and an anti-mouse-HRP secondary antibody (1:2000).

Out of the 10 samples loaded for the Griffin library, only 1 whole cell extracts and 1 periplasmic extract displayed evidence of soluble scFv. The Thomlinson I+J library had 2 samples with detectable soluble scFv in the whole cell and periplasmic extracts. The VHH camelid library had the highest number of clones being expressed, with 6 samples spanning all three sub-libraries displaying a detectable level of soluble VHH in whole cell extracts and 5 samples detectable as periplasmic extracts.

The scFv libraries only gave 10% of clones expressing detectable levels of antibody, whereas this was ≥50% of clones for the VHH library. The in-house camelid VHH library was therefore selected for use due to the higher expression rates. In addition, the design of the VHH library meant that clones did not need to be rescued. Instead, the VHH antibodies only varied within the HCDR3 region and could be synthesised into a new scaffold using specific primers.

5.2 NGAPD using the camelid VHH library

5.2.1 Optimising phytase denaturation: thermal cycler method

Phytic acid and active phytase were prepared in the same way as described in chapter 4 with the exception that phytase antigen was denatured by incubating 100 µl aliquots of active phytase (100 µg/ml) per well of a 96-well PCR plate. The 96-well PCR plate was incubated in a thermal cycler at 99.9°C for 60 minutes. The new thermal cycler treatment successfully denatured phytase to an extent that the activity could not be detected by the Megazyme activity assay using 0.1g of Sigma's phytic acid as the source (figure 5.3).

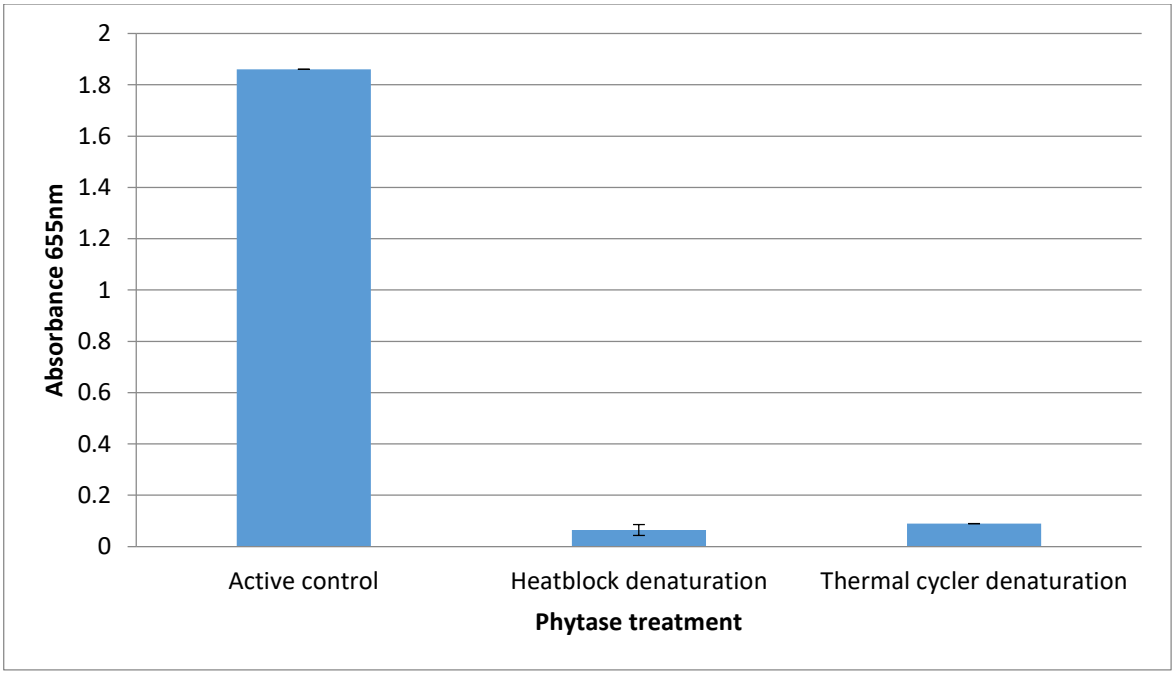


Figure 5.3: Phytase activity assay after thermal cycler denaturation. The activity of the thermal cycler denatured phytase was measured with the Megazyme total phosphorous kit assay. Sigma's phytic acid (0.1g) was digested by phytases that had been denatured by a thermal cycler (99.9°C for 60 minutes; 100 µl aliquots of phytase (100 µg/ml) or a heat block (as described in chapter 3, 100°C for 60 minutes; 100 µl aliquots of phytase (100 µg/ml), or by an untreated active phytase, active control. Absorbance readings at 655 nm were taken and the background free phosphorous levels (0.2) were subtracted from the final plotted values.

5.2.2 Panning rounds 1—4 and screening

The panning strategies using the camelid VHH library for all three target antigens were the same as that described for chapter 4 with the following exceptions. An additional subtraction step was conducted using phage eluted from round 4. Furthermore, phage isolated against phytic acid were subtracted against TiO_2 beads with bound ATP. Phage were eluted in the first and final round of panning using a pH shift induced by 0.2M glycine-HCl. For the second and third panning rounds, the phage were eluted competitively. Lastly, the phytase conformations were immobilised using maleic anhydride plates for all panning and screening rounds. The panning strategy is summarised in figure 5.4.

Screening rounds were conducted for panning rounds 1—4 for active phytase and panning rounds 3—4 for denatured phytase and phytic acid.

Each panning round glycerol stock was grown to produce phage with input titres ranging from 10^{10}—10^{11}. For each phytase screening round, the phage were screened exactly as described for chapter 4 screening rounds, only using maleic anhydride wells instead. For each phytic acid screening round, the phage were screened as described for chapter 4 screening, however the negative control replicates used TiO_2 wells coated with ATP. Output phage titres ranged from 10^5—10^9.

For labelling purposes, screening replicates generated using round 3 panning glycerols, became known as round 4 replicates, and those generated using round 4 glycerols became known as round 5 replicates.

For each experiment (active and denatured phytase and phytic acid), 50 µl of eluted phage from each round 4 screen replicate, both positive and negative, were pooled and used to create a new round 5 glycerol stock. This stock was then streaked out for the purpose of colony screening described later in the chapter.

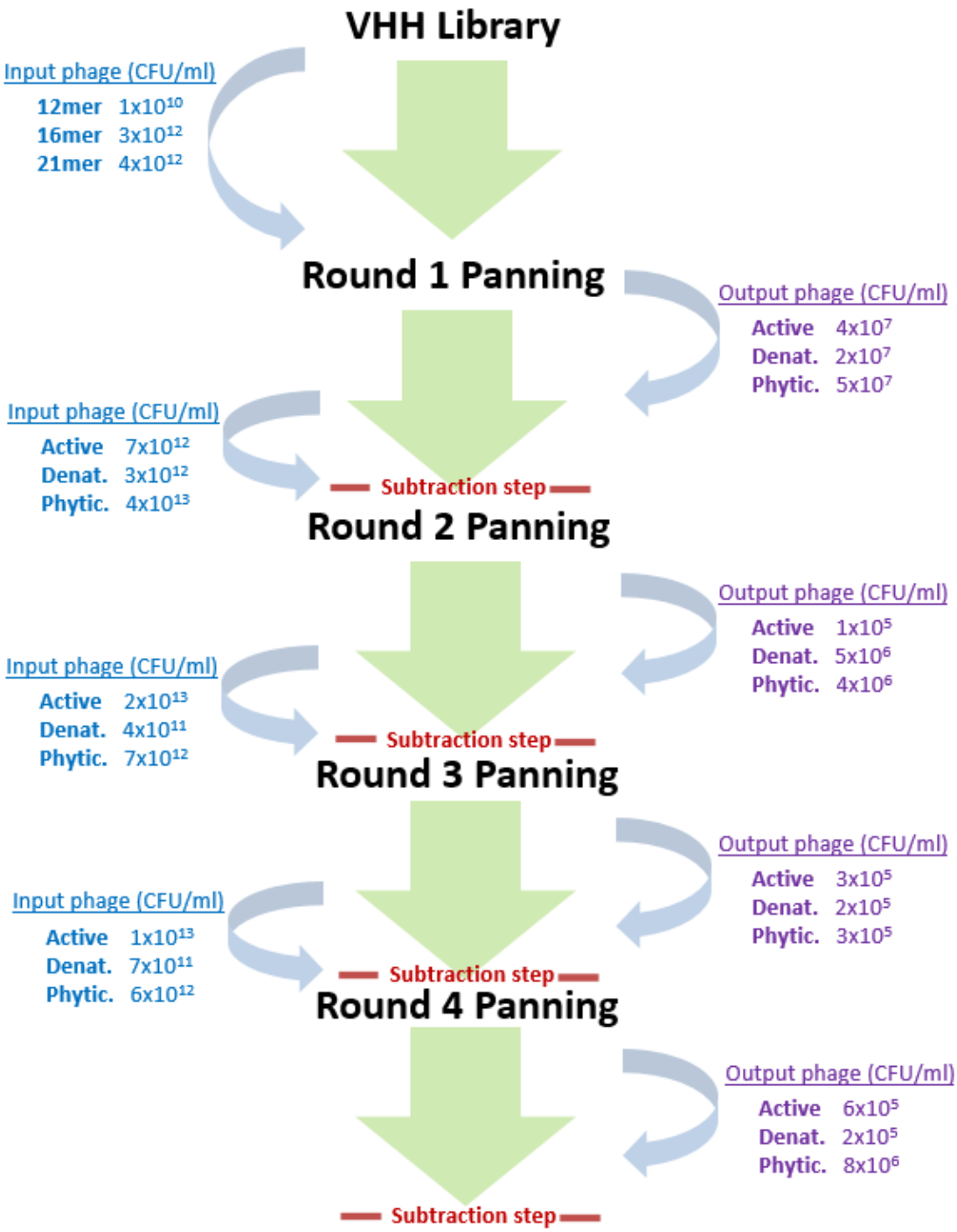


Figure 5.4: Diagrammatic representation of the panning strategy used with the VHH library for all three target antigens. Four rounds of panning were conducted for each target antigen, active phytase ("**Active**"), denatured phytase ("**Denat.**") and phytic acid ("**Phytic.**") with subtraction steps. Input and output phage titres were shown at each step (CFU/ml).

5.2.3 Additional panning rounds

In addition to the panning rounds described, a further panning experiment was conducted for comparison purposes. The new panning experiment involved repeating round 2 panning exactly as described earlier, however the phage were eluted using glycine HCl instead of by antigen competition. The new experiment was conducted for all three target antigens using their existing round 1 panning glycerol stock. The input titres were $7x10^{11}$ CFU/ml, $5x10^{11}$ CFU/ml and $7x10^{12}$ CFU/ml for active phytase, denatured phytase and phytic acid respectively. The output titre for active phytase was $7x10^{5}$ CFU/ml, denatured phytase was $2x10^{5}$ CFU/ml and phytic acid was $4x10^{6}$ CFU/ml.

In addition to the new panning round, the round 2 acid eluted glycerol stock was also used for a screening step. The screening step was conducted exactly as described earlier, however the phage were also eluted using glycine HCl. This step generated another 20 samples for active and denatured phytase and 24 samples for phytic acid. The input and output phage titres were $2x10^{10}$—$7x10^{12}$ CFU/ml and $4x10^{6}$—$1x10^{8}$ CFU/ml respectively. The additional panning glycerols and screening samples were referred to as "acid eluted" samples from hereafter.

5.3 Preparation for NGS: Phage reinfection and DNA extractions

The dsDNA extracts from overnight, re-infected liquid cultures (10ml) were assigned sample numbers and barcodes as summarised in table 5.1. The round 1 and 2 PCR reactions produced ~300 bp and ~350bp amplicons respectively. The products were then checked on agarose gels and quantified by fluorometry (data not shown).

The samples were pooled between within separate NGS runs with a total of 96 different barcodes/samples per run. Each pool had a total of 1,000 ng, with an equal contribution from each sample. The three pools, NGS1, NGS2, NGS3 had final concentrations of 49 ng/µl, 24 ng/µl and 29 ng/µl respectively. Aliquots of the pooled DNA samples (200ng) were sent for Ion Torrent sequencing at the University of Pennsylvania.

Table 5.1: Summary of samples prepared for next generation sequencing for active and denatured phytase and phytic acid

Samples Summary[1]	Sample No. Range[2]	Barcode No. Range[3]
Active Phytase Bacterial Glycerols Panning Rounds 1-4	1 to 4	1 to 4
Active Phytase Panning Output Phage Glycerols Rounds 1-4	5 to 8	5 to 8
Denatured Phytase Bacterial Glycerols Panning Rounds 1-4	9 to 12	9 to 12
Denatured Phytase Panning Output Phage Glycerols Rounds 1-4	13 to 16	13 to 16
Phytic Acid Bacterial Glycerols Panning Rounds 1-4	17 to 20	17 to 20
Phytic Acid Panning Output Phage Glycerols Rounds 1-4	21 to 24	21 to 24
Phytic Acid R4 positive replicates	25 to 36	25 to 36
Phytic Acid R4 negative replicates	37 to 48	37 to 49 (- #42)
Phytic Acid R5 positive replicates	49 to 60	50 to 61
Phytic Acid R5 negative replicates	61 to 72	62 to 73
Active Phytase R2 positive replicates	73 to 82	74 to 83
Active Phytase R2 negative replicates	83 to 92	84 to 93
Active Phytase R3 positive replicates	93 to 102 **(-1)**	**94 to 96 : 1 to 6**
Active Phytase R3 negative replicates	103 to 112	7 to 16
Active Phytase R4 positive replicates	113 to 122	17 to 26
Active Phytase R4 negative replicates	123 to 132	27 to 36
Active Phytase R5 positive replicates	133 to 142	37 to 47 (-#42)
Active Phytase R5 negative replicates	143 to 152	48 to 57
Denatured Phytase R4 positive replicates	153 to 162	58 to 67
Denatured Phytase R4 negative replicates	163 to 172	68 to 77
Denatured Phytase R5 positive replicates	173 to 182	78 to 87
Denatured Phytase R5 negative replicates	183 to 192 **(-1)**	88 to 96
Active Phytase Bacterial Glycerols Panning Round 2 (Acid eluted)	193	1

Active Phytase Panning Output Phage Glycerols Round 2 (Acid eluted)	194	2
Denatured Phytase Bacterial Glycerols Panning Round 2 (Acid eluted)	195	3
Denatured Phytase Panning Output Phage Glycerol Round 2 (Acid eluted)	196	4
Phytic Acid Bacterial Glycerols Panning Round 2 (Acid eluted)	197	5
Phytic Acid Panning Output Phage Glycerols Round 2 (Acid eluted)	198	6
Active Phytase R3 positive replicates (Acid Eluted)	199 to 208	7 to 16
Active Phytase R3 negative replicates (Acid Eluted)	209 to 218	17 to 26
Denatured Phytase R3 positive replicates (Acid Eluted)	219 to 228	27 to 36
Denatured Phytase R3 negative replicates (Acid Eluted)	229 to 238	37 to 47 (- #42)
Phytic Acid R3 positive replicates (Acid Eluted)	239 to 250	**48 to 50 & 70 to 78**
Phytic Acid R3 negative replicates (Acid Eluted)	251 to 262	79 to 90

[1] Sample descriptions including panning experiment and screening round (R2-5)
[2] Sample numbers assigned to each sample type.
[3] Barcode number assigned to each sample. Codes/samples highlighted in yellow were used to make the first 1000 ng pool. The range of codes/samples highlighted in orange overlap both pools. Codes/samples highlighted in green were used to make the second 1000 ng pool. Codes/samples highlighted in blue were used to make the third 1000 ng pool.

5.4 Colony picking

As described in chapter 4, the NGS method was also paired with polyclonal ELISAs and colony picking to help identify positive VHH binders against each of the three targets. As described in 5.2.2, 50 µl of eluted phage from each round 4 screen replicate was used to create a round 4 polyclonal phage glycerol stock.

5.4.1 Polyclonal ELISA

Phage were produced (10^{11}—10^{13} CFU/ml) for the monoclonal scFv, anti-BSA positive control, the initial VHH library (12mer, 16mer, 21mer combined) and all of the three target antigen's subsequent panning rounds, including the new round 2 library produced by acid eluting phage at each round. The phage were used in a polyclonal phage ELISA. The absorbance readings were compared for each solid support (Nunc MaxiSorp® or maleic anhydride) and also against the anti-BSA positive control (figures 5.5—5.6).

The first polyclonal ELISA for active phytase, on maleic anhydride plates, suggested an enrichment of binders due to the rise in signals (2X background) between panning rounds 2—4, although the signal from round 2 came from the acid eluted panning round only. Denatured phytase and phytic acid polyclonal ELISA signals did not reach twice the background signal and were not indicative of polyclonal antibody binding.

Although the signals were considered positive, the signals produced for panning rounds against active phytase and the anti-BSA positive control didn't produce particularly high signals as seen in previous assays. Due to the low signals, the experiment was repeated using fresh phage preparations. The anti-BSA positive control was re-transformed from an older stock and new phage were produced. Only maleic anhydride and TiO_2 plates were used in the second ELISA. Each phage preparation was also incubated against the panning experiments negative control to test for specificity against the intended target (figures 5.7—5.8).

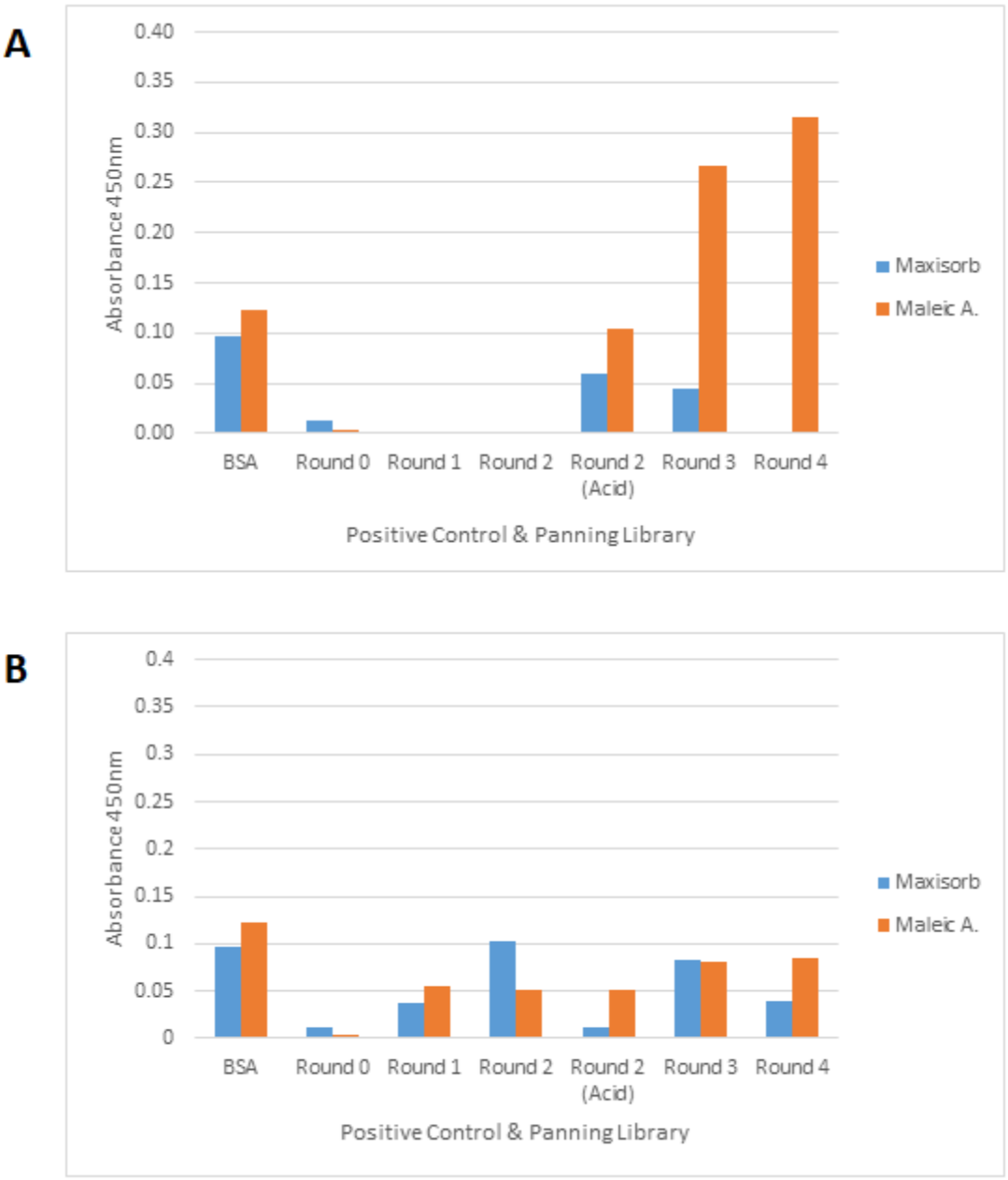


Figure 5.5: Polyclonal phage ELISA for the active and denaturd phytases. The phage-VHH were produced for each panning round for active phytase (A) and denatured phytase (B) plus the anti-BSA positive control and camelid VHH library. Active or denatured phytase or BSA was immobilised at 10 µg/ml on a Nunc MaxiSorp® or maleic anhydride plate. Polyclonal phage-VHH were used to detect the immobilised antigens and those remaining after the wash steps were detected by an Anti-M13-HRP antibody. The TMB substrate was allowed 10 minutes incubation with the well before the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm and no antigen values (0.25—0.27) were subtracted.

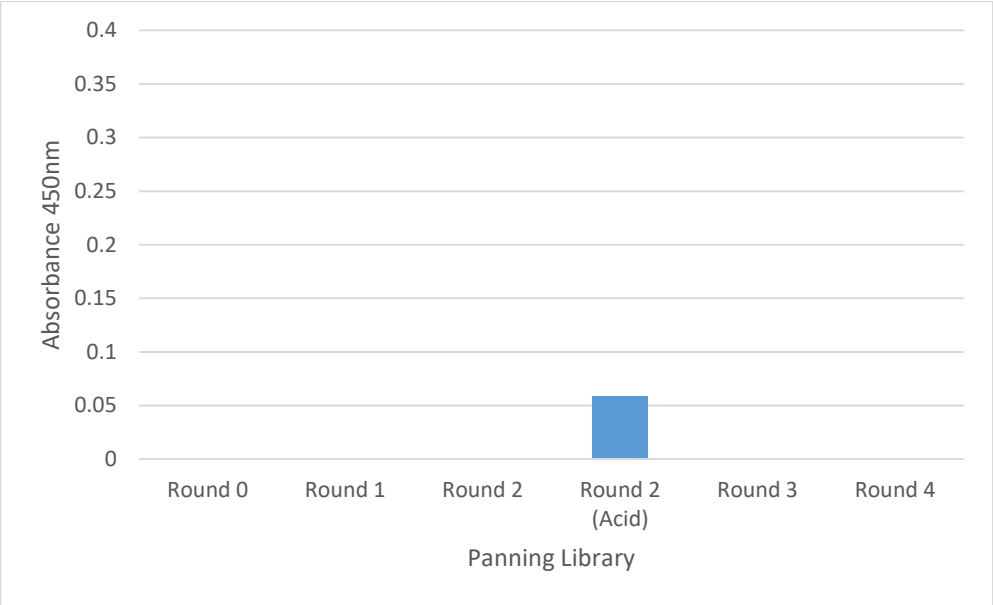


Figure 5.6: Polyclonal phage ELISA for phytic acid. The phage-VHH were produced for each panning round for phytic acid plus the anti-BSA positive control and camelid VHH library. Phytic acid was immobilised on a TiO_2 plate and BSA was immobilised at 10 µg/ml on a Nunc MaxiSorp® plate. Polyclonal phage-VHH were used to detect the immobilised antigens and those remaining after the wash steps were detected by an Anti-M13-HRP antibody. The TMB substrate was allowed 10 minutes incubation with the well before the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm and no antigen values (0.5) were subtracted.

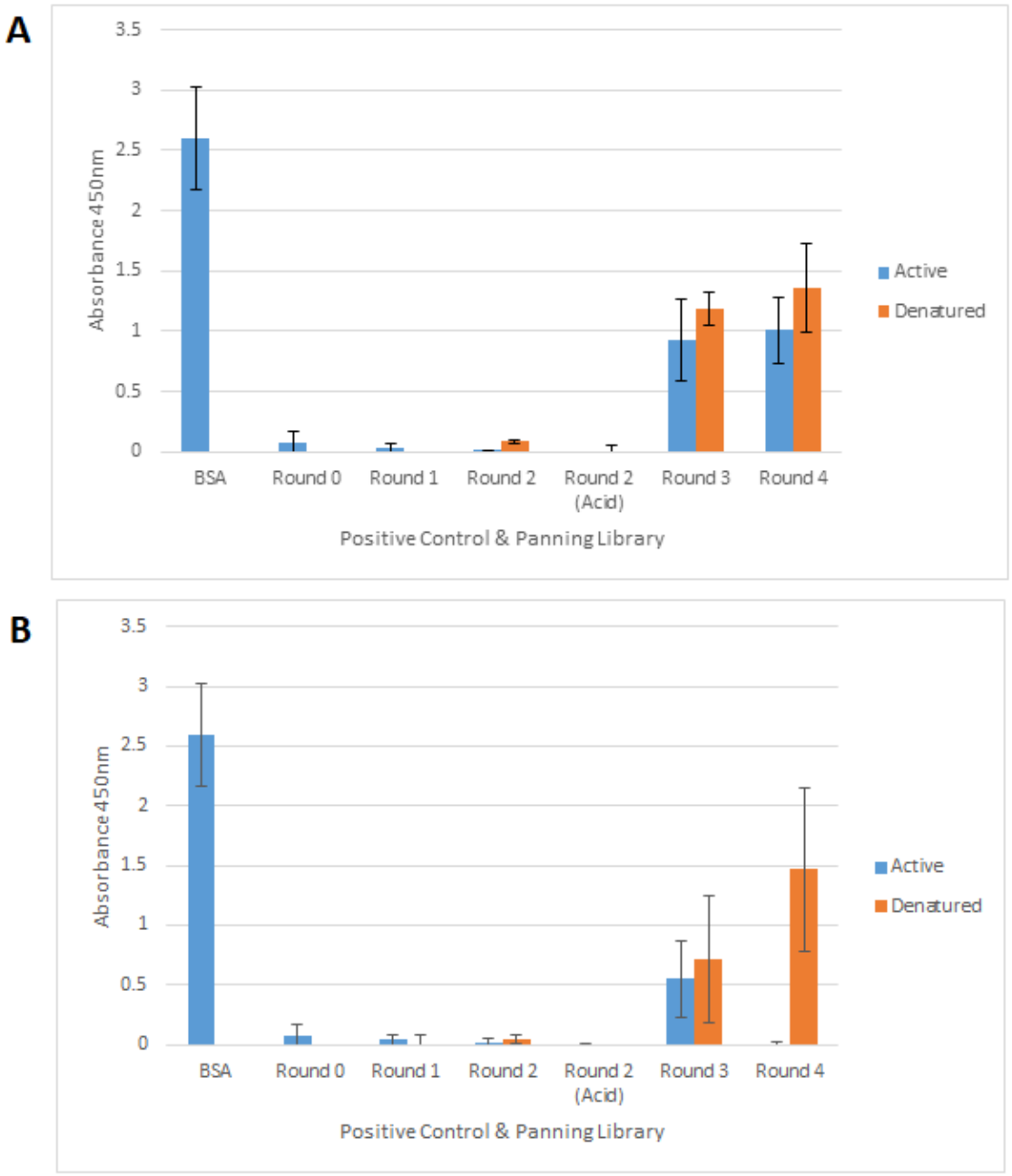


Figure 5.7: Polyclonal phage ELISA for active and denaturd phytase. The phage-VHH were produced for each panning round for active phytase (A) and denatured phytase (B) plus the anti-BSA positive control and camelid VHH library. Active and denatured phytase, or BSA, was immobilised at 10 µg/ml on a maleic anhydride plate. Polyclonal phage-VHH were used to detect the immobilised antigens and those remaining after the wash steps were detected by an Anti-M13-HRP antibody. The TMB substrate was allowed 10 minutes incubation with the well before the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm and no antigen values (0.25—0.4) were subtracted.

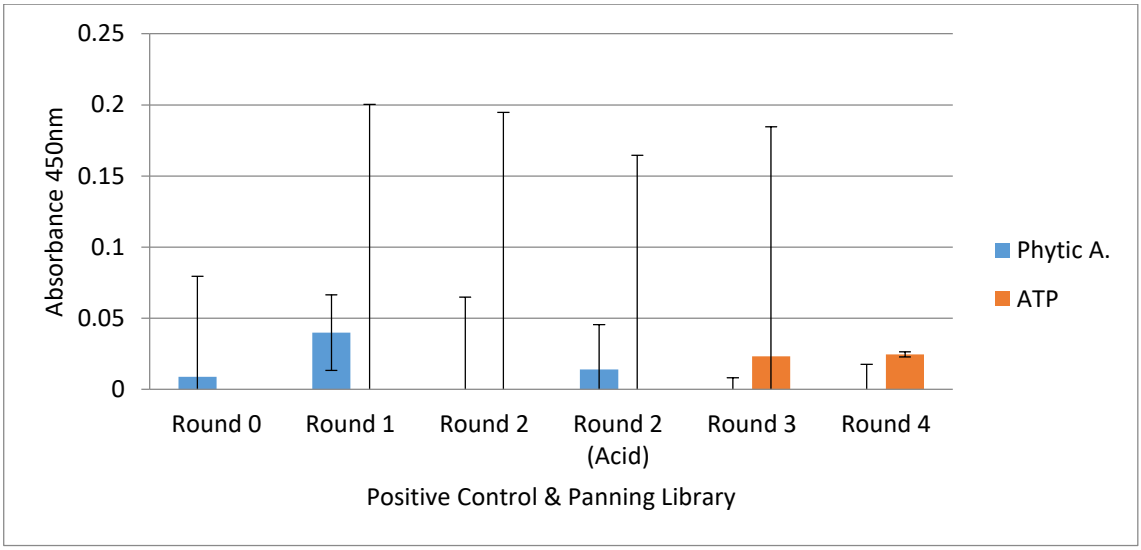


Figure 5.8: Polyclonal phage ELISA for phytic acid. The phage-VHH were produced for each panning round for phytic acid plus the anti-BSA positive control and camelid VHH library. Phytic acid and ATP were immobilised on a TiO_2 plate. BSA was immobilised at 10 µg/ml on a maleic anhydride plate. Polyclonal phage-VHH were used to detect the immobilised antigens and those remaining after the wash steps were detected by an Anti-M13-HRP antibody. The TMB substrate was allowed 10 minutes incubation with the well before the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm and no antigen values were subtracted.

On the second attempt, the anti-BSA positive control worked as intended with a positive absorbance signal of 2.5 (6X background).

The first, third and fourth rounds of panning for active phytase had positive absorbance signals (2X background), rising from third to the fourth round. Interestingly, the positive signals were against both the active and denatured conformations of phytase. The second round of panning (both normal and acid eluted) for active phytase did not produce a signal.

For denatured phytase, the panning rounds 3 and 4 absorbance signals rose between rounds (3—6X background), however only round 3 showed some binding activity (2X background) against the negative control antigen (active phytase).

The phytic acid panning rounds did not show any activity against either the intended or control antigens.

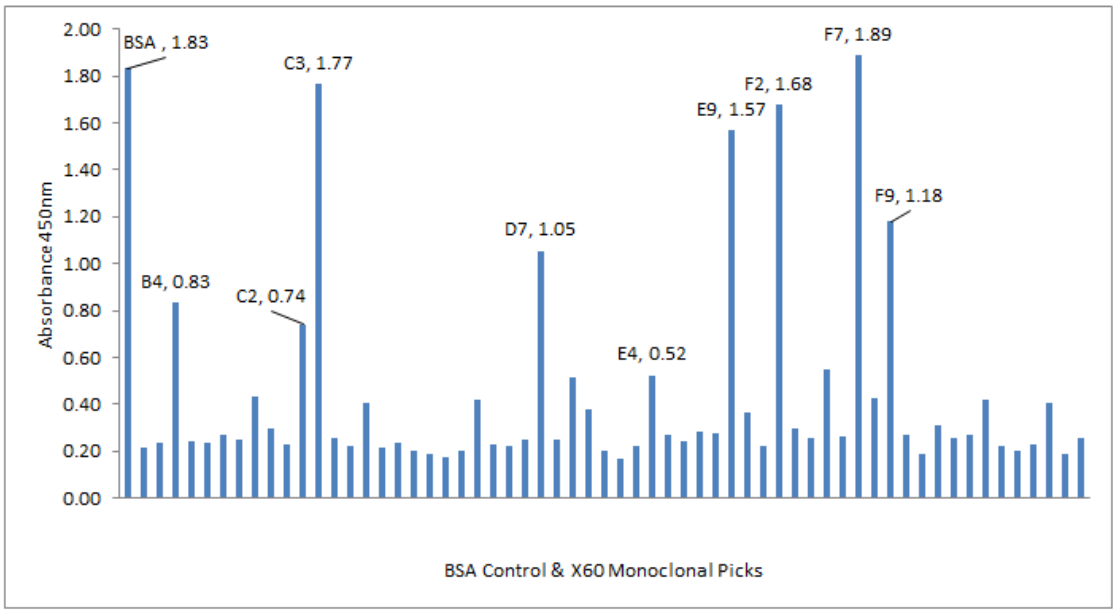
5.4.2 Monoclonal phage ELISAs

Following the polyclonal ELISAs for the two phytase conformations producing positive signals, monoclonal phage ELISAs were conducted for all three target antigens. The round 4 polyclonal glycerol stocks were streaked onto 2YT agar plates and incubated overnight at 37°C in a static incubator. The following day, X60 colonies were picked and grown to produce phage for each target antigen. The anti-BSA positive control was also used to produce phage. The specificities of the phage-VHH fusions were tested by phage ELISA (figures 5.9) as described for the polyclonal ELISAs.

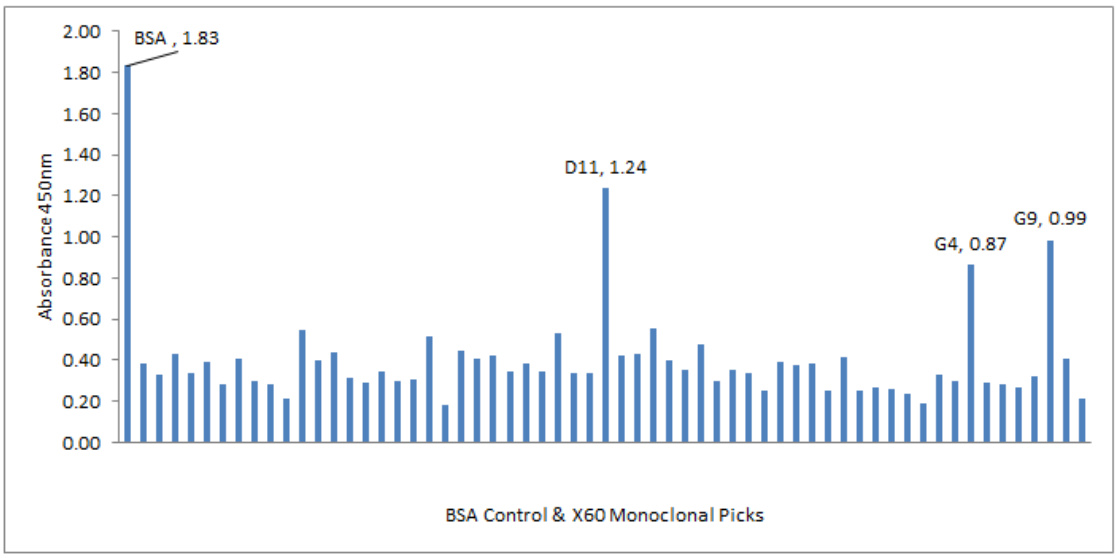
The positive control produced a strong absorbance value of 1.8. Eight picks against active phytase produced absorbance values of 0.6 or higher (1.5—4.5X background) (B4, C2, C3, D7, E9, F2, F7 and F9) as opposed to only three picks against denatured phytase (D11, G4 and G9; 2—3X background) and zero picks against phytic acid.

The picks with absorbance values of 0.6 or higher were regrown in 96-well plates from glycerol stocks and used to produce fresh phage overnight; this included the anti-BSA positive control. Each phage preparation was tested in triplicates against both the intended, and negative control antigens (figure 5.10). No picks were tested for phytic acid.

A


2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Absorbance 450nm
BSA , 1.83
B4, 0.83
C2, 0.74
C3, 1.77
D7, 1.05
E4, 0.52
E9, 1.57
F2, 1.68
F7, 1.89
F9, 1.18
BSA Control & X60 Monoclonal Picks

B


2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Absorbance 450nm
BSA , 1.83
D11, 1.24
G4, 0.87
G9, 0.99
BSA Control & X60 Monoclonal Picks

C

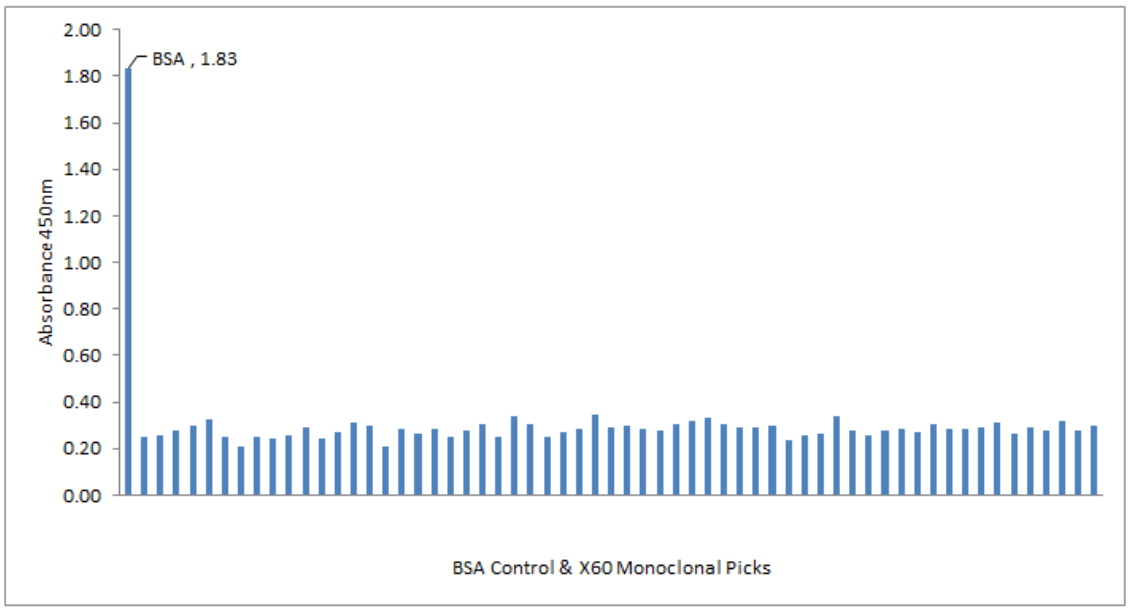

2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Absorbance 450nm
BSA , 1.83
BSA Control & X60 Monoclonal Picks

Figure 5.9: Colony picking phage-ELISA. Phage-VHH were produced from 60 randomly picked colonies after panning against active and denatured phytase and phytic acid. Active or denatured phytase or BSA was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Phytic acid was immobilised overnight in 100 µl aliquots (0.125M, pH 1) on a TiO_2 plate. Monoclonal phage-VHH were used to detect immobilised active and denatured phytase or phytic acid (100 µl of ~10^{10}-10^{13} CFU/ml). Bound phage-VHH were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. A no antigen control value (0.4—0.7) was subtracted.

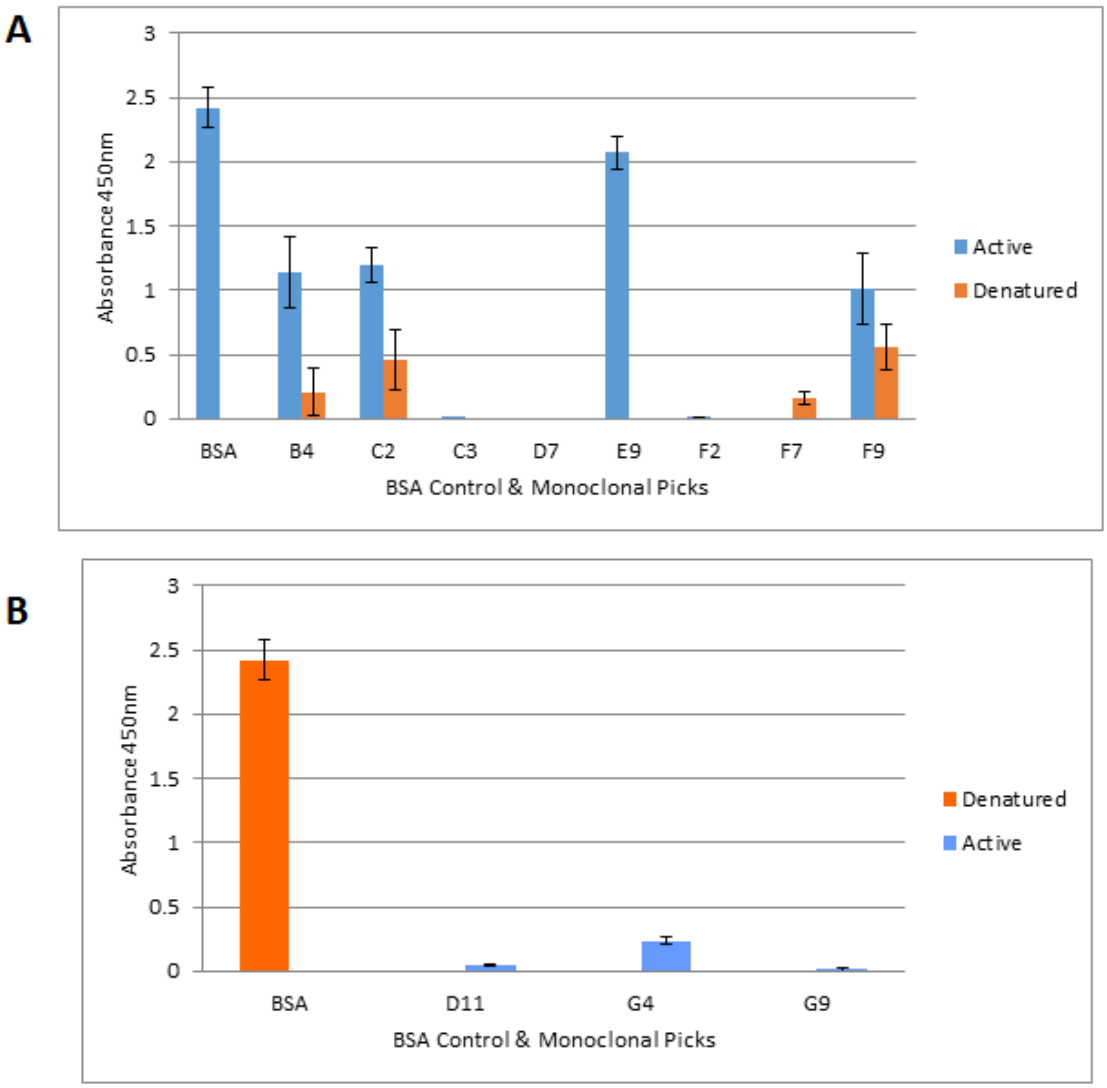


Figure 5.10: Repeat phage ELISA for the phytase VHH colony picks. Phage-VHH were produced again from the colony picks glycerol stocks, that produced absorbance signals of 0.6 or greater against active (A) or denatured (B) phytase. Active phytase, denatured phytase and BSA was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Monoclonal phage-VHH were used to detect the immobilised antigens (100 µl of ~10^{10}-10^{12} CFU/ml). Bound phage-VHH were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. A no antigen control value (0.2) was subtracted.

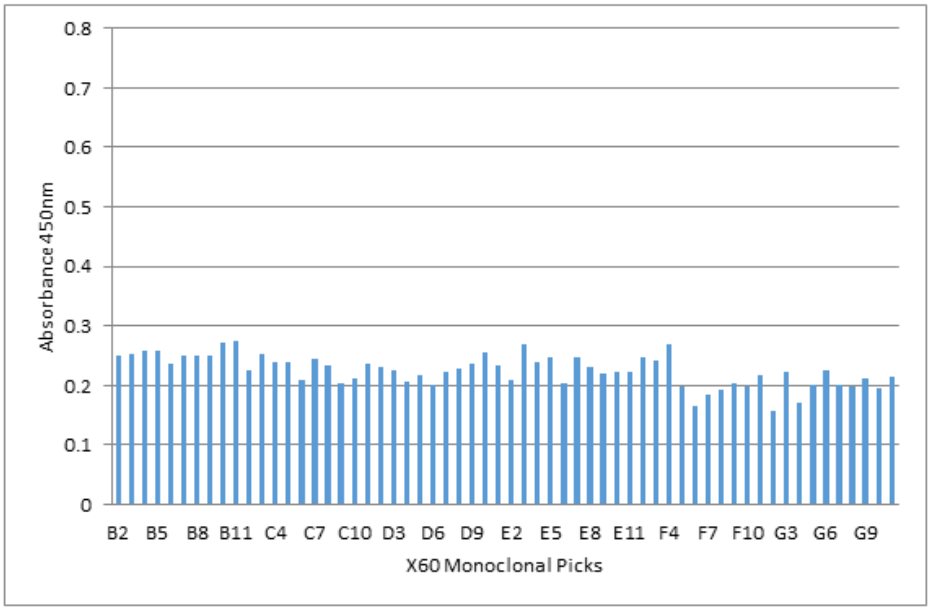
The positive control produced a strong signal of 2.4 (12X background). For active phytase, the samples B4, C2, E9 and F9 maintained a strong absorbance signal (5—10X background) against the active conformation of phytase, with some binding activity against the denatured conformation (1—3X background) for B4, C2 and F9. The C3, D7 and F2 samples failed to produce any signal. The F7 sample only produced a faint signal (<1X background) against the denatured conformation. The C3, D7, F2 and F7 clones were excluded from further analyses.

For denatured phytase, all three of the originally high signal clones failed to produced strong signals against denatured phytase despite evidence of phage. The G4 clone produced a faint signal (1X background) against the active conformation. All three of the denatured phytase candidates (D11, G4 and G9) were excluded from further analyses.

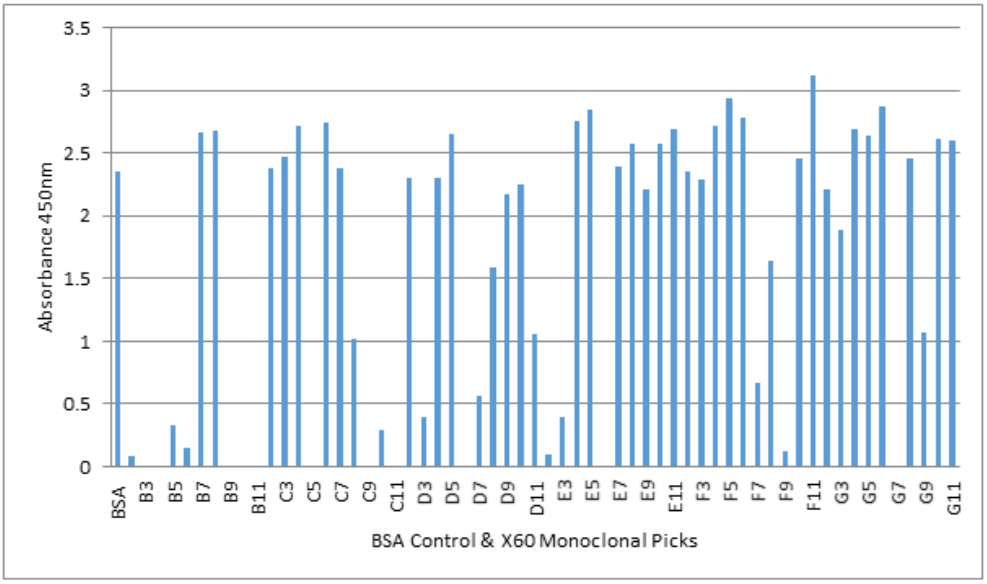
Due to the inconsistency of binding between ELISAs and apparent loss of some of the binding clones, another set of X60 colony picks were taken from the same glycerol stocks as the first set, for all three targets. The colonies were grown in the same way as the first set (96-well plates) and used to produce both glycerol stocks and phage as before. A new culture for the positive anti-BSA control was also grown alongside them. The new picks were first tested against their intended antigen in single wells (figures 5.11).

For the second set of colony picks, the phytic acid phage ELISA did not produce any positive absorbance signals (<2X background) as also seen in the first set of X60 picks. The active and denatured phytase picks produced many positive signals (2—7X background). To retest all of the clones, the stock plates were used to inoculate a fresh 96 deep-well plate and all of the clones were used to produce fresh phage exactly as described earlier. The phage were then tested in single wells against both the intended and control antigens (figures 5.12).

A


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Absorbance 450nm
B2 B5 B8 B11 C4 C7 C10 D3 D6 D9 E2 E5 E8 E11 F4 F7 F10 G3 G6 G9
X60 Monoclonal Picks

B


3.5
3
2.5
2
1.5
1
0.5
0
Absorbance 450nm
BSA B3 B5 B7 B9 B11 C3 C5 C7 C9 C11 D3 D5 D7 D9 D11 E3 E5 E7 E9 E11 F3 F5 F7 F9 F11 G3 G5 G7 G9 G11
BSA Control & X60 Monoclonal Picks

C

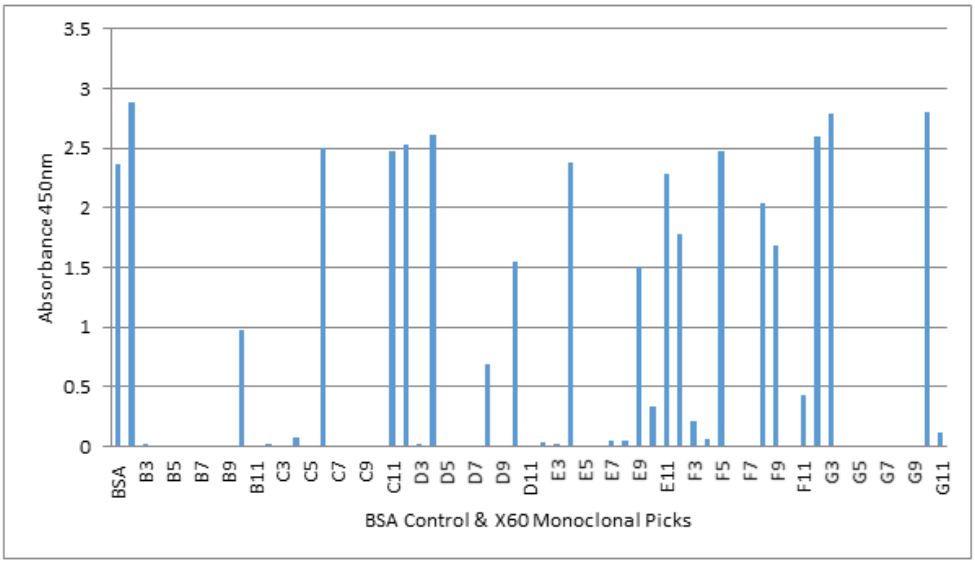

3.5
3
2.5
2
1.5
1
0.5
0
Absorbance 450nm
BSA B3 B5 B7 B9 B11 C3 C5 C7 C9 C11 D3 D5 D7 D9 D11 E3 E5 E7 E9 E11 F3 F5 F7 F9 F11 G3 G5 G7 G9 G11
BSA Control & X60 Monoclonal Picks

Figure 5.11: Colony picking phage-ELISAs: set 2. Phage-VHH were produced from another X60 randomly picked colonies after panning against phytic acid (A), active (B) and denatured phytase (C). Phytic acid was immobilised overnight in 100 µl aliquots (0.125M, pH 1) on a TiO_2 plate. Active and denatured phytase and BSA was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Monoclonal phage-VHH were used to detect immobilised phytic acid or the phytases (100 µl of ~10^{11}-10^{13} CFU/ml). Bound phage-VHH were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. A no antigen control value (0.45—0.7) was subtracted

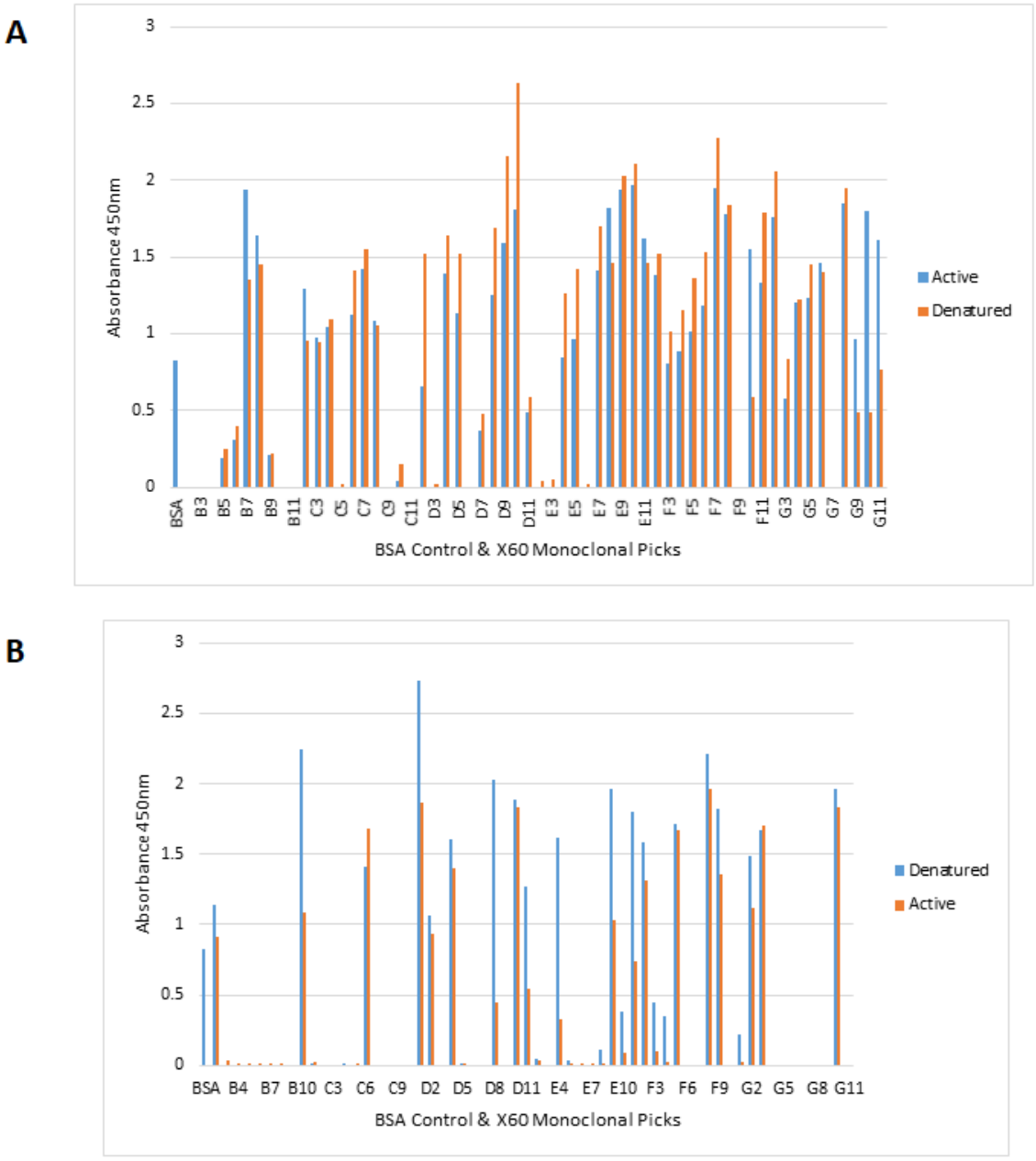


Figure 5.12: Active and Denatured phytase colony picking phage-ELISA: set 2 repeat. Phage-VHH were produced from the second set of randomly picked colonies (X60) after panning against active and denatured phytase. Active phytase, denatured phytase and BSA was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Monoclonal phage-VHH were used to detect the immobilised antigens (100 µl of ~10^{10}-10^{13} CFU/ml). Bound phage-VHH were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. A no antigen control value (0.4) was subtracted.

Many clones continued to produce high signals (>2X background) however only a few displayed the intended binding specificity against their primary antigens. The clones F10, G10 and G11 were selected for active phytase and the clones D8, D11, E4 and E11 were selected for denatured phytase. In addition to the antigen-specific clones, other clones were also selected from the data that showed binding activity against both the intended and control antigens; for active phytase, the clones E9 and E10 were selected, and for denatured phytase, the clones F5 and F8 were selected.

In summary, fifteen clones were pursued from the monoclonal picks section as possible binders to the two phytase conformations. No clones were pursued for phytic acid. The fifteen VHH clones were sent for Sanger sequencing using Source Bioscience's in-house primer '30nt_M13R'.

30nt_M13R Primer: 5' AGC GGA TAA CAA TTT CAC ACA GGA 3'

All of the clones were in frame with the pelB leader sequence and were expected to express as normal. A summary of the HCDR3 sequences for each clone was shown in table 5.2.

Table 5.2: Summary of samples prepared for next generation sequencing for active and denatured phytase and phytic acid

Active Clones [1]	ELISA Specificity [2]	FRW3 Flank [3]	HCDR3 [4]	FRW4 Flank [3]
Set 1: B4	Active	I Y Y C A A	* H R P R P H R P T K R R K S Y	W G Q G T
Set 1: C2	Active	I Y Y C A A	* H R P R P H R P T K R R K S Y	W G Q G T
Set 1: E9	Active	I Y Y C A A	* H R P R P H R P T K R R K S Y	W G Q G T
Set 1: F9	Active	I Y Y C A A	* N ? ? P R ? ? ? K ? ? ? ? ?	?
Set 2: F10	Active	I Y Y C A A	? ? ? ? P H ? ? T ? ? R K S ?	W ? Q ? T
Set 2: G10	Active	I Y Y C A A	* H R ? ? ? ? ? ? ? ? R R K ? ?	W G Q G T
Set 2: G11	Active	I Y Y C A A	* H R P R P H R P T K R R K S Y	W G Q G T
Set 2: E9	Act & Den	I Y Y C A A	* N S K K P R R N S K V Q	W G Q G T
Set 2: E10	Act & Den	I Y Y C A A	* H R P R P H R P T K R R K S Y	W G Q G T
Denatured Clones [1]	**ELISA Specificity [2]**	**FRW3 Flank [3]**	**HCDR3 [4]**	**FRW4 Flank [3]**
Set 2: D8	Denatured	I Y Y C A A	R K K F A R A P N H N E T K W R	W G Q G T
Set 2: D11	Denatured	I Y Y C A A	R K K F A R A P N H N E T K W R	W G Q G T
Set 2:E4	Denatured	I Y Y C A A	R T T K S H * H S S R K R F K H	W G Q G T
Set 2: E11	Denatured	?	?	?
Set 2: F5	Act & Den	?	?	?
Set 2: F8	Act & Den	I Y Y C A A	R K K F A R A P N H N E T K W R	W G Q G T

[1] The VHH clones derived from colony picking. Set 1 refers to the first set of X60 clones picked for the colony picking approach. Set 2 refers to the second set of X60 clones picked for the colony picking approach.
[2] The binding specificity of each clone based on ELISA data. Active phytase was abbreviated to 'Act' and denatured phytase was abbreviated to 'Den'.
[3] The left and right flanking sequences for each VHH clone. The correct left motif was IYYCAA and the right motif was WGQGT.
[4] The HCDR3 sequence for each clone. Amber stop codons were indicated by * and unknown amino acids were indicated by ?.

The anti-phytase (active conformation) clones B4, C2, E9, G10, G11, possibly F10 and the non-specific clone E10 shared the same HCDR3 sequence. The clones F9 and the non-specific clone E9 may share the same HCDR3 sequence. For the denatured picks, the clones D8, D11 and the non-specific clone F8 shared the same HCDR3 sequence. The clone E4 had a unique HCDR3 sequence. The HCDR3 sequences for the clone E11 and the non-specific clone F5 could not be identified.

Based on the Sanger sequencing data covering the HCDR3, there were 4 different clones in total; *HRPRPHRPTKRRKSY, *NSKKPRRNSKVQ, RKKFARAPNHNETKWR and RTTKSH*HSSRKRFKH.

Interestingly, the clones that were considered to be non-specific binders, showing binding activity against both phytase conformations, had the same HCDR3 sequences as the other clones considered to be specific to their respective phytase conformations.

In light of the Sanger sequencing data, the binding specificities of the clones were re-tested in another phage ELISA. The clones were grown from glycerol stocks in 100 ml 2YT cultures as previously described and phage were produced overnight. The phage supernatants were then tested in a phage ELISA (figures 5.13).

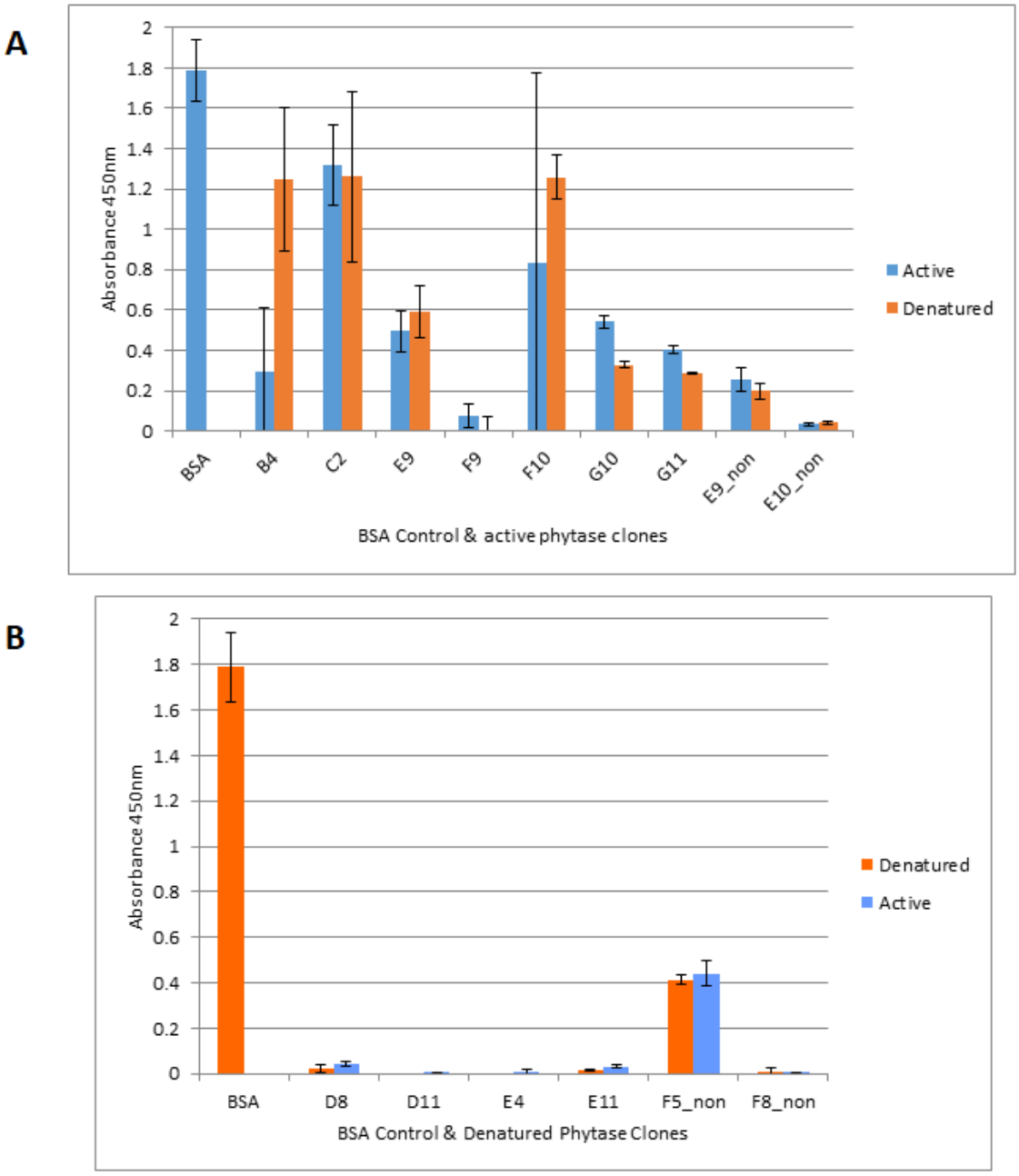


Figure 5.13: Repeat phage ELISA for the VHH clones derived from the colony picking approach for active and denatured phytase. Phage-VHH were produced from the 9 and 6 clones that had been identified through the colony picking approach against active and denatured phytase respectively. The identifier '_non' was used to identify the previously non-specific clones which demonstrated binding activity against both phytase conformations. Active phytase and BSA was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Monoclonal phage-VHH were used to detect the immobilised antigens (100 µl of ~10^{10}-10^{12} CFU/ml) in triplicates. Bound phage-VHH were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm. A no antigen control value (0.4) was subtracted from the plotted data.

The clones derived from the colony picking approach against active phytase showed varying degrees of binding activity against the active and denatured conformations of phytase despite having only two different HCDR3 sequences amongst them. The clones derived from colony picking approach against denatured phytase lost their binding signals in the repeat ELISA, apart from the non-specific control clone F5, which showed similar binding activity against both phytase conformations as expected.

Due to the inconsistencies, the four different HCDR3 sequences identified by the colony approach were synthesised into a new expression vectors, alongside the VHH NGS candidates as described later in the chapter.

No clones were identified against phytic acid using the colony picking approach. Although the potential NGS candidates for phytic acid were yet to be identified and tested, a second panning experiment was also conducted for phytic acid.

5.5 Phytic Acid panning using Covalink support

Following the unsuccessful conventional colony picking method for the phytic acid experiment, the purpose of the following experiments was to investigate the use of an alternative support called CovaLink™ for phytic acid.

5.5.1 Immobilising phytic acid and ATP on CovaLink

Phytic acid and ATP was bound to Covalink wells. The following day, the wells were stained with toluidine blue (figure 5.14).

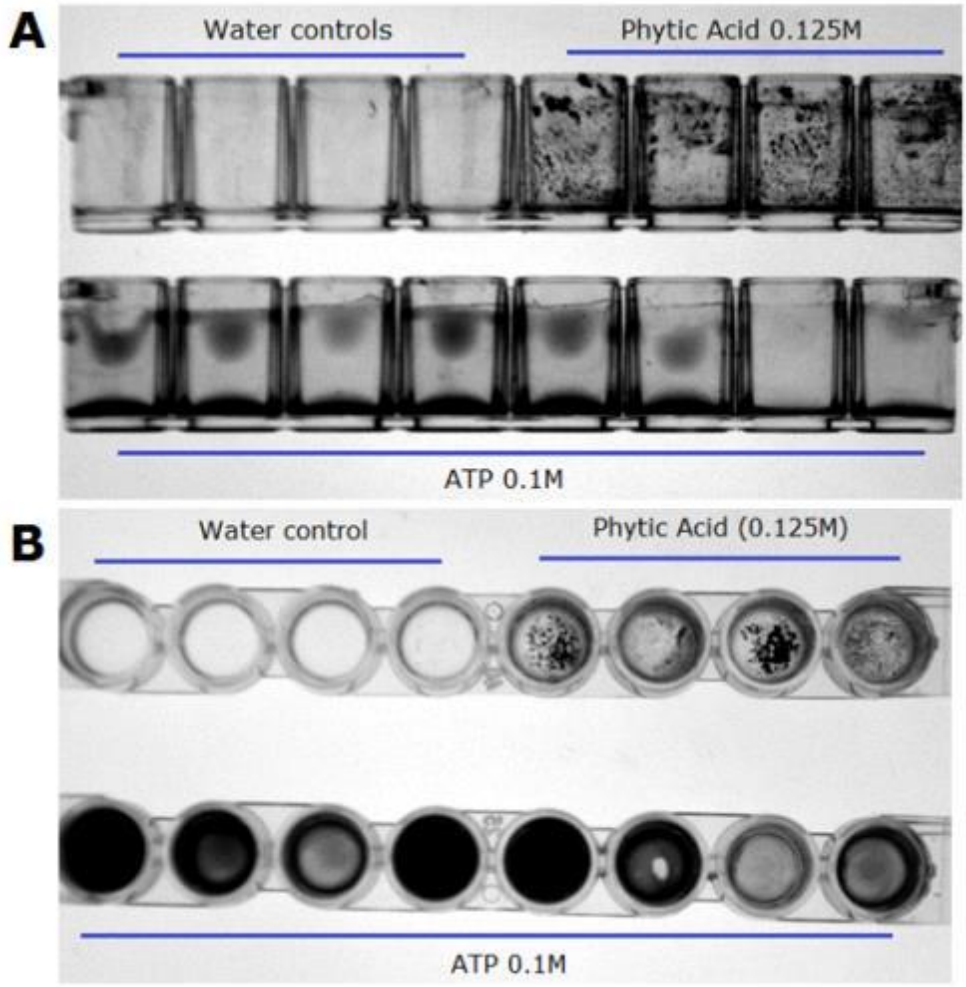


Figure 5.14: Phytic acid and ATP immobilised on Covalink wells. Phytic acid (0.125M pH 1) and ATP (0.1M) were immobilised (100 µl) on Covalink wells overnight at 4°C. Distilled water was also incubated as a negative control. The wells were stained with Toluidine Blue dye for 1 hour and then destained repeatedly. The dye bound to the phosphate groups present on both phytic acid and ATP – visible as a dark stain from the side (A) and the top (B) of the Covalink wells.

The staining confirmed that phytic acid and ATP had both become immobilised on the Covalink support as intended. The Covalink support was then used in the subsequent panning experiment to try and isolate antibodies against phytic acid.

5.5.2 Growing the VHH library and phage precipitation

The acid precipitated VHH libraries ($8x10^{12}$ CFU/ml, $3x10^{11}$ CFU/ml and $4x10^{11}$ CFU/ml) were used for 4 panning rounds with subtraction steps against the negative control antigen (ATP) in-between. The combined VHH libraries were panned against 20 Covalink wells containing phytic acid. Phage were then eluted from the plates using Glycine HCl. Output phage for panning against phytic acid was $1x10^8$ CFU/ml, $5x10^9$ CFU/ml, $2x10^7$ and $9x10^8$ CFU/ml for rounds 1 to 4 respectively. Input phage for each round was ~$1x10^{13}$ CFU/ml.

Due to time constraints, the sub-libraries were not subjected to the next generation sequencing approach and were instead only used for the colony picking method.

5.5.3 Polyclonal and Monoclonal ELISA

The phage ELISAs were conducted using the Covalink plates as a means of immobilising phytic acid.

For the polyclonal phage ELISA, phage-VHH fusions were produced for all four panning sub-libraries and also the original three VHH libraries (12, 16, 21mer). The three original VHH libraries (12, 16, 21mer), phage supernatants were pooled before use. Once the background signal had been subtracted, the data showed no evidence of any interaction between the polyclonal VHH and immobilised phytic acid (data not shown).

For the monoclonal picks, 120 random colonies were picked from the round 4 panning glycerol stock and tested by phage ELISA. The 120 clones showed no evidence of binding to phytic acid when it was immobilised on the Covalink support (data not shown).

When the phage preparations were immobilised and detected on a Nunc MaxiSorp® (figures 5.15) plate the signals were all positive, confirming the presence of phage and that the detection system was working as intended.

The VHH covalink panning experiment was unable to identify phytic acid binders through the colony picking approach similarly to the VHH TiO_2 panning experiments. For phytic acid, only the NGS approach remained as a potential means of identifying new binders.

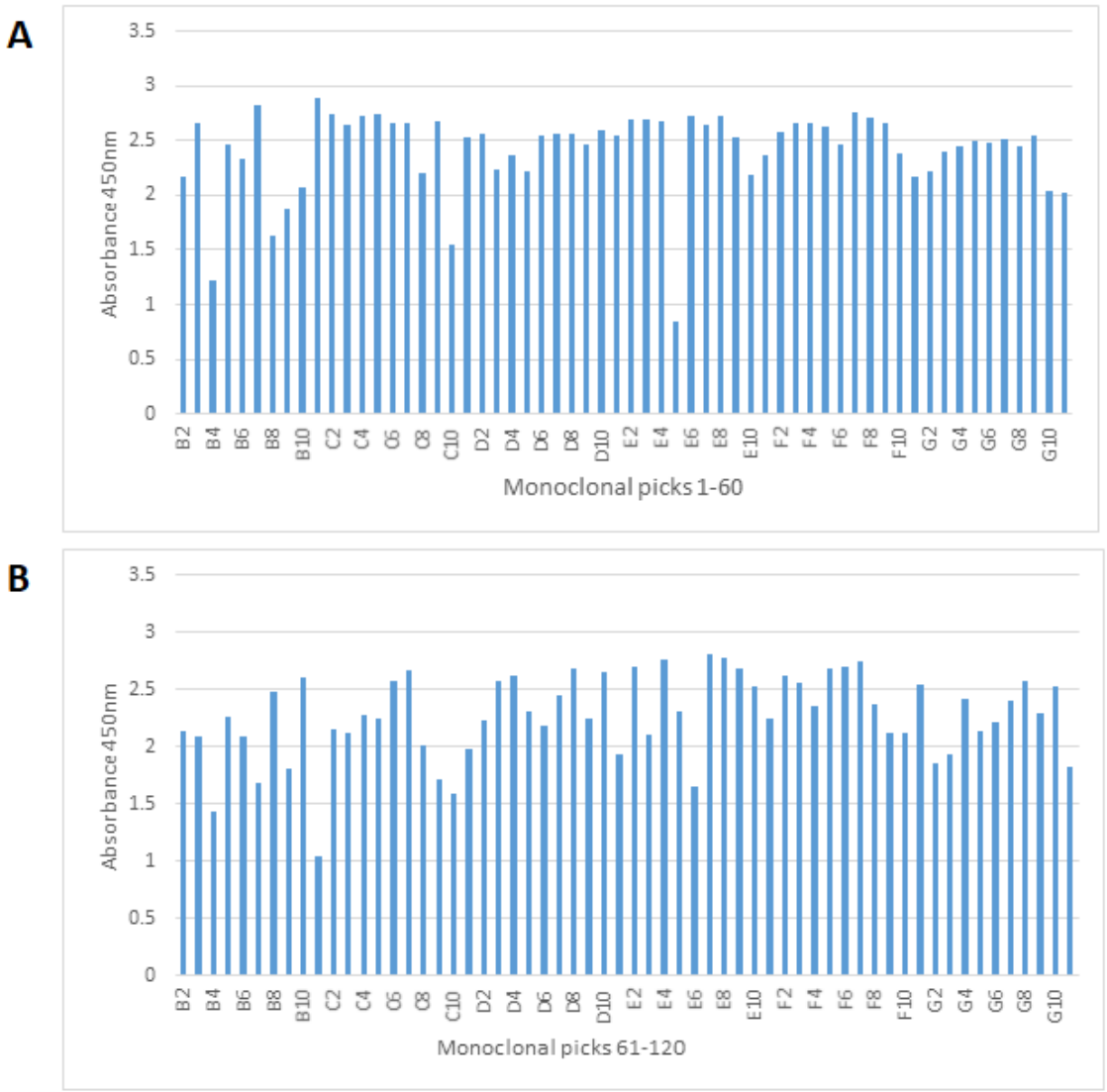


Figure 5.15: Phytic acid colony picking phage-ELISA – picks 1—120 immobilised. Phage-VHH were produced from 120 randomly picked colonies from round 4 panning against phytic acid immobilised on Covalink plates; (A) colony picks 1—60 and (B) colony picks 61—120. Monoclonal phage-VHH for each colony pick were immobilised on a Nunc MaxiSorp® plate for 1 hour. Bound phage-VHH were detected by an Anti-M13-HRP antibody, TMB substrate and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid. Absorbance values were taken at 450 nm.

5.6 NGS Analysis

The NGS data was received from the University of Pennsylvania in a compressed, FASTQ file format. The data was de-compressed and de-multiplexed, making the described alterations for the VHH library; refer to section 2.18.1 in methods.

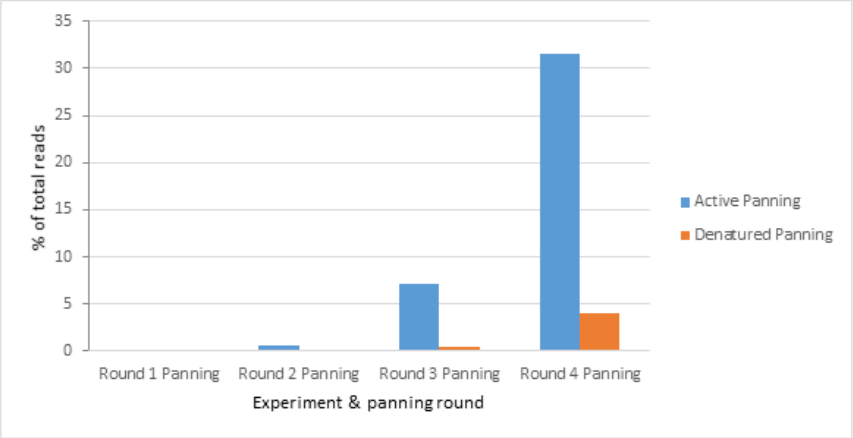
5.6.1 NGS Analysis: Monoclonal Picks

The active and denatured phytase panning round files (.LR.fa) were searched for the presence of the 4 HCDR3 sequences identified during the colony picking approach. Amber stop codons were substituted with lowercase q. The command below was used to identify the exact HCDR3 sequences in each .LR.fa file:

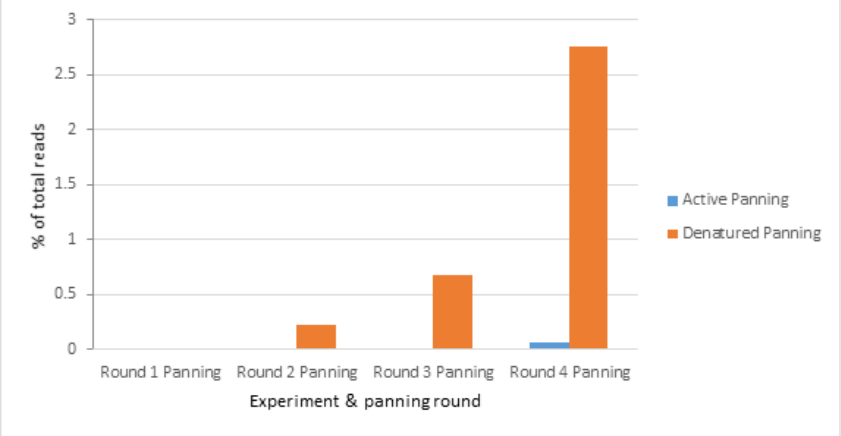
*grep -c -w HCDR3sequence *.LR.fa > HCDR3sequence.table*

The number of reads for each HCDR3 were converted to percentages of the total number of reads within each panning file searched, so that they could be compared between experiments (figures 5.16).

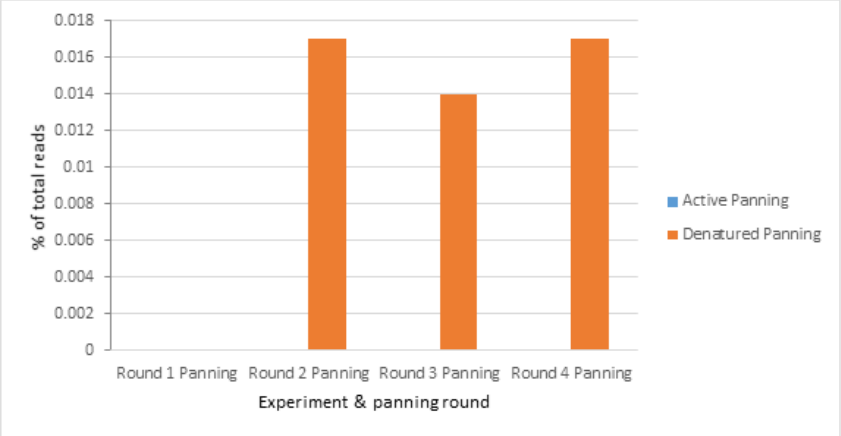
A


35
30
25
20
15
10
5
0
% of total reads
Round 1 Panning
Round 2 Panning
Round 3 Panning
Round 4 Panning
Experiment & panning round
Active Panning
Denatured Panning

B


3
2.5
2
1.5
1
0.5
0
% of total reads
Round 1 Panning
Round 2 Panning
Round 3 Panning
Round 4 Panning
Experiment & panning round
Active Panning
Denatured Panning

C


0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
% of total reads
Round 1 Panning
Round 2 Panning
Round 3 Panning
Round 4 Panning
Experiment & panning round
Active Panning
Denatured Panning

D

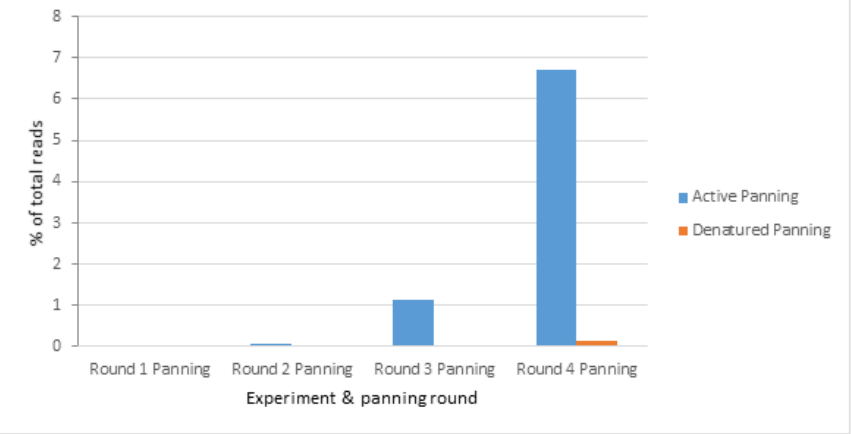

8
7
6
5
4
3
2
1
0
% of total reads
Round 1 Panning
Round 2 Panning
Round 3 Panning
Round 4 Panning
Experiment & panning round
Active Panning
Denatured Panning

Figure 5.16: The number of reads, as a percentage of the total, within each panning round for the picked clones. The panning round .LR.fa files were searched for the presence of the picked clone 'qHRPRPHRPTKRRKSY' (A), 'RKKFARAPNHNETKWR' (B), 'RTTKSHqHSSRKRFKH' (C) and 'qNSKKPRRNSKVQ' (D). The number of reads were converted to percentages of the total number of reads for each panning round .LR.fa file for both active and denatured phytase experiments.

The clone 'qHRPRPHRPTKRRKSY' demonstrated clear enrichment throughout the panning rounds, accounting for 31% of the total reads in round 4 panning for the active phytase experiment. In contrast, the clone only accounted for 4% of the total reads in round 4 panning for the denatured phytase experiment, which also represents considerable enrichment of the clone during the panning. This data concurs with this clone being picked repeatedly from round 4 of the active phytase experiment.

The clone 'RKKFARAPNHNETKWR' was clearly enriched during the denatured phytase experiment, increasing in number between each rounds however accounting for considerably less of the total number of reads (2.7% in round 4 panning) in comparison to the 'qHRPRPHRPTKRRKSY' clone for active phytase. In contrast, there was little evidence of enrichment throughout the active phytase panning experiment. This follows a similar pattern as the first clone, having been isolated by random picks from the denatured phytase panning experiment.

The clone 'RTTKSHqHSSRKRFKH' was isolated from the denatured phytase panning experiment and thus shows no enrichment during the active phytase experiment as expected. Within the denatured phytase experiment, the clone accounts for very little of the total number of reads in all of the panning rounds, showing little to no meaningful increase between rounds. This clone represented 'E4' and was only picked once in the sequenced clones. The clone was completely undetected in round 1 panning.

The clone 'qNSKKPRRNSKVQ' was isolated from the active phytase panning experiment and shows a considerable increase in read counts between panning rounds 3 and 4 where it accounts for 1% then 6.7% of the total read count. The clone failed to demonstrate meaningful enrichment within the denatured phytase panning data.

5.6.2 NGS Analysis: identifying new candidate VHH clones with target specificities

The data analysis process was summarised in figure 2.5 in methods, describing how each screening round was ranked and selected by 'Z score' for each HCDR3.

After applying all of the filters and comparing the three datasets (active phytase, denatured phytase and phytic acid), this would have resulted in the omission of the four clones derived from colony picking. For the two clones picked from panning against active phytase: the clone 'qHRPRPHRPTKRRKSY' was present in all three datasets, but was represented by sequences most enriched in the round 5 screen against active phytase, Z≥10 threshold, the acid eluted screening data for denatured phytase, Z≥10 threshold and the round 4 screen against phytic acid, Z≥10 threshold. The clone 'qNSKKPRRNSKVQ' was best represented by the active phytase round 5 screen with a Z≥5 threshold, denatured phytase round 4 screen (Z≥2 threshold) and phytic acid round 5 screen (Z≥5 threshold). For clones isolated against denatured phytase: the 'RKKFARAPNHNETKWR' clone had its most enriched representatives in the round 5 screen against denatured phytase (Z≥5 threshold) and the round 4 screen against phytic acid (Z≥2 threshold). The clone 'RTTKSHqHSSRKRFKH' was absent from all three lists because it failed to meet the basic criteria of being present in 4 or more positive replicates, with Z scores of 2 or greater. The positive Z scores within the phytase datasets reflect the likelihood of binding activity against each antigen, which correlates with the phage ELISA data that suggested non-specific binding to both phytase conformations. However, based on these filters the four non-distinguishing clones identified by random colony picking should have been omitted from further analysis. Instead they were synthesised alongside the new NGS candidates as known binders that could serve as positive controls.

Briefly, the candidates that were seen in at least 4 positive replicates with a Z score of ≥2 were retained for dataset comparisons. Duplicates were removed between Z score thresholds and screening round shortlists. The final lists of candidates for each target antigen were compared against each other and any duplicates (suggesting non-specific binding) were removed. The

remaining candidates were ranked according to the number of positive reps that they were seen in followed by the highest Z score threshold.

Once the final list of NGS candidates were generated for each target antigen, the top 20 candidates were then subjected to a reverse analysis.

The new files were searched for the top 20 candidates from the original analysis to check their new Z scores. The active and denatured phytase datasets retained all 20 candidates whereas their acid eluted counterparts lost 2 and 3 candidates respectively. The phytic acid dataset lost 8 candidates after the reversed analysis, leaving only 12 of the top candidates remaining. The phytic acid, acid eluted counterpart retained all 20 candidates.

In the list of remaining candidates, considering both combinational (acid eluted and competitive elution) and acid eluted datasets, the top 10 candidates were selected for each target (table 5.3). The candidates from the acid eluted datasets were excluded as they were outcompeted by their equivalent HCDR3's seen in rounds 4 and 5 screening (combinational approach). The top HCDR3 candidates were selected based on the number of positive replicates that had met the Z score threshold (Z≥10, Z≥5 or Z≥2), followed by the highest Z score threshold e.g a candidate seen in 10 of 10 replicates in the Z≥2 was considered a better candidate than one that was seen in 8 of 10 replicates in the Z≥5 threshold.

Table 5.3: The top 10 NGS candidates for active phytase, denatured phytase and phytic acid.

Data Set & Round [1]	AA sequence [2]	No. of positive reps [3]
Active, Round 4	qGKKSGRGPNTKLRRK	10 of 10 (Z≥5)
Active, Round 4	RSPKqTRANKSRRRMR	10 of 10 (Z≥2)
Active, Round 4	SRRSqGPRWRYRLPAK	10 of 10 (Z≥2)
Active, Round 5	qSARCSKNCRVRKMSH	10 of 10 (Z≥10)
Active, Round 4	ARPKKNqQFRTDRPLR	9 of 10 (Z≥2)
Active, Round 4	qKTKKVTSIKSSSSRR	9 of 10 (Z≥2)
Active, Round 4	AKHRKRKQRLKRRRWY	8 of 10 (Z≥2)
Active, Round 4	AKRVNYFKTPNHRYIL	8 of 10 (Z≥2)
Active, Round 4	FKRRGLqSQKRRWKSR	8 of 10 (Z≥2)
Active, Round 4	GRLRRTqLPLSFSNRK	8 of 10 (Z≥2)

Data Set & Round [1]	AA sequence[2]	No. of positive reps [3]
Denatured, Round 4	AqGRWKSRRTWNQSAQ	10 of 10 (Z≥2)
Denatured, Round 4	AYPVCTCSKKGAPRALRqRSV	10 of 10 (Z≥10)
Denatured, Round 4	AYPVCTCSKKGGSEGLLRqRSV	10 of 10 (Z≥2)

Denatured, Round 4	GRRqSASGLAKWqKKRASKDG	10 of 10 (Z≥2)
Denatured, Round 4	KIMLNqRLSKKRKSKI	10 of 10 (Z≥2)
Denatured, Round 4	QKKQKALLPRPKNVKR	10 of 10 (Z≥2)
Denatured, Round 4	qPKRVHCKNFCLKKNMRRCRA	10 of 10 (Z≥2)
Denatured, Round 4	RqRRKRKNMNPQLTNKRNQNF	10 of 10 (Z≥2)
Denatured, Round 4	RRFKRNRYPRKKRKCK	10 of 10 (Z≥2)
Denatured, Round 4	RRqRTTTKAPQLqKPRSLKLK	10 of 10 (Z≥2)

Data Set & Round [1]	**AA sequence** [2]	**No. of positive reps** [3]
Phytic Acid, Round 4	qSGFCRNSqDqK	9 of 12 (Z≥10)
Phytic Acid, Round 4	qLTRSSNIqMPTSIRIHACYA	8 of 12 (Z≥10)
Phytic Acid, Round 4	DHLQVGFGPTL	7 of 12 (Z≥2)
Phytic Acid, Round 4	DHLQVVFGPTL	7 of 12 (Z≥2)
Phytic Acid, Round 4	LSKYCLAGCPRQRLTIVKSqP	7 of 12 (Z≥10)
Phytic Acid, Round 4	LVSPPEHSTKDNPQEK	7 of 12 (Z≥2)
Phytic Acid, Round 4	qRLRHNFSTPII	7 of 12 (Z≥10)
Phytic Acid, Round 4	qRTLMWIQPTTLFFLSQLqST	7 of 12 (Z≥10)
Phytic Acid, Round 4	qSPASLKPAKKQDSCM	7 of 12 (Z≥10)
Phytic Acid, Round 4	RCNPMAqDqSYKqFSGPITWq	7 of 12 (Z≥5)

[1] The dataset and best screening round that each NGS candidate was selected from followed by the identification strategy 'Top 10' based on Z score.
[2] The amino acid translation for each NGS candidate's HCDR3 sequence.
[3] The number of positive replicates, within the screening round (maximum of 10 for phytases and 12 for phytic acid), that had a Z score meeting the threshold (Z score threshold).

In addition to the Z score analysis the VHH datasets were also used for a common motif analysis, as an alternative method for identifying candidates with the desired binding properties.

The analysis was conducted on the entire list of potential candidates that had been seen in 4 replicates or more (rounds 4 and 5), with a minimum Z score of 2. Prior to the motif analysis, the list was subjected to a reversed analysis as described for the top 20 NGS candidates above. The candidates that remained (>600 for each dataset) were included in the final common motif search. Each dataset was combined with its acid eluted counterpart (round 3) and duplicate HCDR3 were removed before the motif analysis.

The motif analysis was conducted using the online Multiple EM for Motif Elicitation (MEME) suite (http://meme-suite.org; Bailey and Elkan (1994)). Only the first ten significant motifs were considered. Any candidates that were already selected based on the previous Z score analysis were omitted from the final list of common motif candidates.

The analysis produced an image, or logo, for each motif common to the dataset provided (figures 5.17). The logos for each dataset were ranked based on the number of contributing sequences (sites) within the dataset and the E-value. Each logo's width (number of amino acids making up the motif) and statistical E-value were provided by the MEME suite software. The logo was more statistically significant if it possessed a low E-value. If motifs possessed an E-value greater than 0.05 then the image was listed partially transparent as an additional indicator. The E-value was generated based on the log likelihood ratio, width, contributing sites (peptide sequences), background letter frequencies and also the size of the dataset provided.

Essentially, the E-value is an estimate of the likelihood of finding a similar motif, within a random set of sequences of the same size, where each position is independent and each letter was chosen based on the background frequencies of each letter.

A

	Logo	E-value	Sites	Width
1.		2.4e-087	11	15
2.		9.5e-004	3	12
3.		2.8e+000	2	10
4.		9.4e+000	2	6
5.		5.0e+001	2	8
6.		8.1e+001	2	9
7.		1.4e+002	2	6
8.		2.0e+002	2	7
9.		3.0e+002	2	9
10.		3.2e+002	2	6

B

	Logo	E-value	Sites	Width
1.		4.0e-161	15	21
2.		1.2e-138	14	21
3.		9.9e-053	16	11
4.		4.3e-030	6	15
5.		1.0e-020	6	15
6.		1.1e-002	3	15
7.		3.5e+000	2	6
8.		1.4e+001	2	6
9.		1.6e+001	2	8
10.		5.2e+001	3	10

C



Figure 5.17: Common motif analysis. The active phytase (including acid eluted) (A) denatured phytase (including acid eluted) (B) and phytic acid (including acid eluted) (C) NGS candidates, seen in 4 replicates or more with a minimum Z score of 2, were subjected to reversed analysis. The final list was used as the input for a common motif analysis. The list was analysed using the online suite MEME generating common motif 'logos'. The logos were ranked based on their statistical significance (E-value) and number of contributing sites within the data.

Based on the motif analysis (figure 5.17), the most common motifs were summarised in table 5.4. Any major motifs represented by 2 or less sites (peptide sequences) were excluded.

Table 5.4: The most common motifs identified by the MEME suite for all three target antigens.

Major Motif [1]	Target [2]	Already selected by top 10? [3]
SARCSKNCRVRKMSH	Active	Yes; QSARCSKNCRVRKMSH
YKMRSTTRPFRP	Active	No
AYPVCTCSKKG	Denatured	Yes; AYPVCTCSKKGAPRALRQRS and AYPVCTCSKKGGSEGLLRQRSV
QFSKPKSVK	Denatured	No
RRGPGGKSVIR	Denatured	No
TFARRVNHGLTRPSL	Denatured	No
QGTQRKPPRTRHARL	Denatured	No
CRQMTAR	Denatured	No
HWQSRQHPQAPSTNR	Phytic Acid	No
LKKNK	Phytic Acid	No
AYTSSQSQPYAAIGLYTSLTS	Phytic Acid	No

[1] The statistically significant motifs identified by the MEME suite analysis.
[2] The target antigen whose dataset the major motif was derived from. The datasets were separated by a grey background for ease of interpretation.
[3] The major motif was queried against the top 10 NGS candidates to check whether the motif was already selected as a final candidate. If the candidates were already chosen by the top 10 NGS analysis, they were omitted from the motif analysis selection.

The selected motifs were investigated further by considering which peptide sequences (sites) were contributing to each motif, how many positive replicates they had been seen in and what Z scores they were achieving in the dataset. All of the candidates had already been subjected to a reversed analysis, thus the Z scores within the negative pools could be disregarded.

Major motif: YKMRSTTRPFRP (Active phytase)

Contributing AA Motif	No. of positive reps
WKYKMRSTTRPFRPRR	4 of 10 (Z≥2)
RKYKMRSTTRPFRPRR	4 of 10 (Z≥2)
GKYKMRSTTRPFRPRR	6 of 10 (Z≥2)

Major motif: QFSKPKSVK (Denatured phytase)

Contributing AA Motif [1]	**No. of positive reps** [2]
QFSKPKSVEPLKQNKPMKRRY	8 of 10 (Z≥2)
QFSKPKSVKPLKQNKPMKRGY	7 of 10 (Z≥2)
QFSKPKSVKPLKQDKPMKRRY	7 of 10 (Z≥2)
QFSKPKSVKPLRQNKPMKRRY	7 of 10 (Z≥2)
QFSKPKSVKPLKQNRPMKRRY	9 of 10 (Z≥2)
QFSKPKGVKPLKQNKPMKRRY	7 of 10 (Z≥2)
QFSKPKSVKRLKQNKPMKRRY	7 of 10 (Z≥2)
QFSKPKSAKPLKQNKPMKRRY	7 of 10 (Z≥2)
QFSKPKSVKPLKQNKPMRRRY	8 of 10 (Z≥2)
KFSKPKSVKPLKQNKPMKRRY	7 of 10 (Z≥2)
QFSKPNECEALKQNKPMKRRY	6 of 10 (Z≥2)
QFSKPKSVKRPSRISLLKRRY	6 of 10 (Z≥2)
QFSKPKSVKRPSRISLPSGRY	6 of 10 (Z≥2)
QFSKPKSVKRPSRISLPSGGY	8 of 10 (Z≥2)

Major motif: RRGPGGKSVIR (Denatured phytase)

Contributing AA Motif [1]	**No. of positive reps** [2]
SKRKKRRGPGGKSVIR	4 of 10 (Z≥2)
RKRKKRRGPGGKSVIR	4 of 10 (Z≥2)
qKRRKRRGPGGKSVIR	4 of 10 (Z≥2)
qERKKRRGPGGKSVIR	4 of 10 (Z≥2)
qRRKKRRGPGGKSVIR	5 of 10 (Z≥2)
qKRKKRRGPGGESVIR	4 of 10 (Z≥2)
qKRKKRRGPGGKSVVR	5 of 10 (Z≥2)
qKRKKRRGPGGRSVIR	5 of 10 (Z≥2)
qKRKKRRGPGGKSAIR	4 of 10 (Z≥2)
qKRKKRRGPGGKPVIR	5 of 10 (Z≥2)
qKRKKRRGSGGGKSVIR	4 of 10 (Z≥2)
qKRKKASGSGGKSVIR	6 of 10 (Z≥2)
qKRKKRRGSGGGRSVIR	6 of 10 (Z≥2)
qKRKKASGVRGGKSVIR	5 of 10 (Z≥2)
RNKRRPPGGKQVKWKR	6 of 10 (Z≥2)
qKRKKRRGSGGGSLPFR	4 of 10 (Z≥5)

Major motif: TFARRVNHGLTRPSL (Denatured phytase)

Contributing AA Motif [1]	**No. of positive reps** [2]
TFAGRVNHGLTRPSLR	4 of 10 (Z≥2)
AFARRVNHGLTRPSLR	5 of 10 (Z≥2)
PFARRVNHGLTRPSLR	5 of 10 (Z≥2)
TLARRVNHGLTRPSLR	4 of 10 (Z≥2)
TFARRANHGLTRPSLR	5 of 10 (Z≥2)

TLRGVVNHGLTRPSLR	6 of 10 (Z≥2)

Major motif: QGTQRKPPRTRHARL (Denatured phytase)

Contributing AA Motif [1]	**No. of positive reps** [2]
GQGTQRKPPRTRHARL	6 of 10 (Z≥2)
RQGTQRKPPRTWHARL	4 of 10 (Z≥2)
RQGAQRKPPRTRHARL	4 of 10 (Z≥2)
RWGTQRKPPRTRHARL	4 of 10 (Z≥2)
RQGTQGKPPRTRHARL	6 of 10 (Z≥2)
RQGTQRKPPRTRRARL	4 of 10 (Z≥2)

Major motif: CRQMTAR (Denatured phytase)

Contributing AA Motif [1]	**No. of positive reps** [2]
CRQMTARRAAVSGGGVK	4 of 10 (Z≥2)
CRQMTARAAAVSGGGVK	4 of 10 (Z≥2)
CRQMTARRRQCRGRVK	5 of 10 (Z≥10)

Major motif: HWQSRQHPQAPSTNR (Phytic acid)

Contributing AA Motif [1]	**No. of positive reps** [2]
PHWQSRQHPQAPSTNR	4 of 12 (Z≥5)
AHWQSRQHPQAPSTNR	4 of 12 (Z≥10)
THWQSRQHPWAPSTNR	4 of 12 (Z≥5)
THWQSRQHPQAPSMNR	4 of 12 (Z≥5)
THWQSRQHPQAPFTNR	4 of 12 (Z≥5)
THWQSRQHPQAPSANR	4 of 12 (Z≥5)
THWQSRQYPQAPSTNR	4 of 12 (Z≥5)
THWQPRQHPQAPSTNR	4 of 12 (Z≥5)
THWQSRQHPQTPSTNR	4 of 12 (Z≥2)
THWQSRQHPQAPPTNR	4 of 12 (Z≥5)
THWQSRQHPQAPSTDR	4 of 12 (Z≥5)
THRQSRQHPQAPSTNR	4 of 12 (Z≥5)
THPQSRQHPQAPSTNR	4 of 12 (Z≥5)

Major motif: LKKNK (Phytic acid)

Contributing AA Motif [1]	**No. of positive reps** [2]
LKKNKPGNSDSEIFFAWRQS	4 of 12 (Z≥2)
LKKNKPGEFRFPIFFAWRQS	6 of 12 (Z≥2)
LKKNKPGNSDSESFLLGRQS	6 of 12 (Z≥2)
LKKNKTWELQILNLFAWRQS	4 of 12 (Z≥2)

Major motif: AYTSSQSQPYAAIGLYTSLTS (Phytic acid)

Contributing AA Motif [1]	No. of positive reps [2]
AYTQSQSQPYAAIGLYTSLTS	4 of 12 (Z≥5)
ANTQSQSQPYAAIGLYTSLTS	4 of 12 (Z≥5)
RYTQSQSQPYAAIGLYTSLTS	4 of 12 (Z≥5)

The best contributing motifs were selected from each list, highlighted and summarised in table 5.5. The motifs with the highest number of positive replicates seen in and highest Z score threshold were selected as the best contributors for each major motif (table 5.7). Where motifs were equally ranked, those with amber stop codons (q) were avoided (e.g RNKRRPPGGKQVKWKR), or the best contributor was randomly chosen (e.g GQGTQRKPPRTRHARL, LKKNKPGEFRFPIFFAWRQS and AYTQSQSQPYAAIGLYTSLTS) as it was not feasible to test all of the combinations.

Table 5.5: The major motifs for each target, as identified by MEME suite, and their selected contributing sequence.

Major Motif [1]	Best Contributor [2]	Target Antigen [3]
YKMRSTTRPFRP	GKYKMRSTTRPFRPRR	Active phytase
QFSKPKSVK	QFSKPKSVKPLKQNRPMKRRY	Denatured phytase
RRGPGGKSVIR	RNKRRPPGGKQVKWKR	Denatured phytase
TFARRVNHGLTRPSL	TLRGVVNHGLTRPSLR	Denatured phytase
QGTQRKPPRTRHARL	GQGTQRKPPRTRHARL	Denatured phytase
CRQMTAR	CRQMTARRRQCRGRVK	Denatured phytase
HWQSRQHPQAPSTNR	AHWQSRQHPQAPSTNR	Phytic acid
LKKNK	LKKNKPGEFRFPIFFAWRQS	Phytic acid
AYTSSQSQPYAAIGLYTSLTS	AYTQSQSQPYAAIGLYTSLTS	Phytic acid

[1] The major amino acid motif identified by the MEME suite analysis.
[2] The best contributing sequence for each major motif, based on the number of positive replicates seen in, meeting each Z score threshold. The datasets were separated by a grey background for ease of interpretation.
[3] The target antigen whose dataset the motif was common to.

The common motifs search concluded the NGS data analysis, with a total of 39 VHH candidates selected spanning all three target antigens. In combination with the clones derived from colony picking, this provided a total of 43 VHH candidates.

5.7 VHH synthesis

Unlike the scFv candidates from chapters 3 and 4, the VHH candidates were not tested as phage fusions, except for those derived from colony picks. Instead, the VHH candidates were tested as antibody-MBP fusions, removing the phage construct entirely with the intention of testing them as soluble antibodies, the format that would be used in most applications for phage-panning derived antibodies.

The NGS approach provided the HCDR3 sequences for a range of potential binding candidates (Z score and motif analyses) however as for chapters 3 and 4, the strategy does not physically rescue those candidates from their polyclonal, bacterial glycerol stocks. The recovery strategy for the scFv proved very challenging and usually failed to rescue the correct HCDR3 sequences. The new VHH strategy involved synthesising the intended HCDR3 as opposed to recovering the clones from the polyclonal sub-libraries as described in methods section 2.20. The VHH libraries had the benefit of having the same antibody scaffold sequences surrounding the randomised VHH HCDR3 sequences. The HCDR3 sequences were short enough to be encompassed by two primers designed for an inverse PCR reaction, whilst also covering the adjacent framework sequences. This strategy enabled the HCDR3 sequences to be synthesised into any expression vector already containing the VHH scaffold from the same library.

In addition to the new synthesis strategy, the VHH HCDR3 nucleotide sequences were also converted to the optimal codons for expression in *E. coli* and any stop codons were changed to code for a Q.

5.7.1 Optimal amino acid codons for expression in e. coli

All of the VHH candidate's HCDR3 sequences were converted to optimal codon combinations for the expression of proteins in *E. coli*. Using the optimal nucleotide sequences for each VHH HCDR3 candidate, inverse PCR primers were designed for each candidate, including those derived from colony picks (as controls). Each VHH was also assigned an identifier (I.D) number.

Initially, the primers would only anneal partially to the existing framework of the VHH clone already present in the new expression vector. The DNA polymerase was expected to extend the PCR product into the remaining

portion of the primer covering the HCDR3. Thereafter the PCR reaction would repeatedly amplify the intended HCDR3 sequence within a linear vector amplicon. This system bypasses the recovery strategy used for the scFv formats.

5.7.2 Creating the template for the VHH synthesis PCRs

To synthesise the VHH candidates into a new vector, the reaction first needed a template of the new vector, which already contained a VHH clone from the same library. This allowed the new primers to anneal to the existing framework sequences in the new vector, but then synthesise the new HCDR3 sequences into the existing clone's place.

The template was created using a VHH HCDR3 (VRGYFMRLPSSHNFRY) sequence known to bind to β-lactamase (Conrath et al., 2001). The anti-β-lactamase (bLAC) VHH clone was provided within the pSD3 vector. The following experiments cloned the desired VHH clone into both the pMAL P5X and PET22B vectors (methods table 2.5) using inverse PCR. The two vectors expression levels and the resulting VHH binding properties were compared before selecting a vector for the subsequent experiments.

Based on the anti-bLAC clone's Sanger sequence, forward and reversed primers were designed to flank the VHH gene, whilst also containing restriction enzyme sites.

- Anti-bLAC forward primer containing an NCOI restriction site:
  5' CCAACCGGCCATGGCCCAAGTTC 3'
- Anti-bLAC reverse primer containing an ECORI restriction site:
  5' CTCGAATTCTCAGTGGTGGTGGTGGTGGTGTGC 3'

The two primers would create a 496 bp amplicon using the anti-bLAC VHH clone within the pSD3 vector as the template. The anti-bLAC VHH clone was miniprepped and 10ng was amplified in duplicate in the following PCR reaction (figure 5.18).

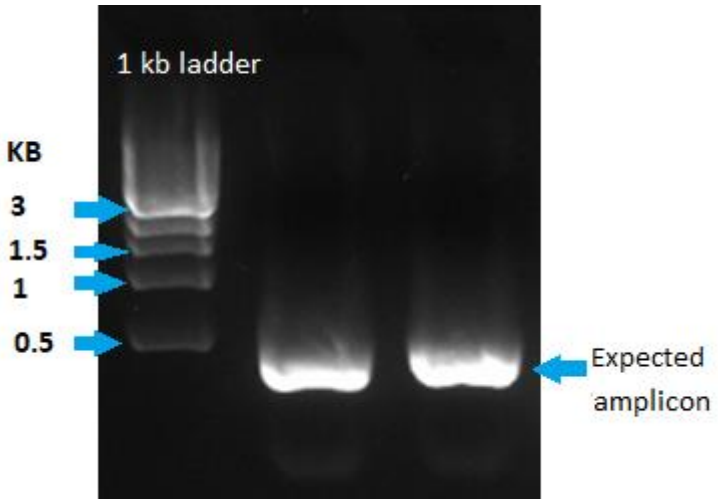


Figure 5.18: Anti-bLAC VHH gene amplicon. The duplicate anti-bLAC VHH gene amplicons were visualised on a 1.5% agarose gel and checked alongside a 1,000 bp ladder (product no. N3232S; NEB). The gel was stained with Nancy-520 and the samples were loaded using a 6x loading buffer.

The PCR products were confirmed and the gel purified amplicons and P5X/PET22B vectors were digested with NCOI and ECORI restriction enzymes. The amplicons were gel purified and ligated into the double-cut P5X or PET22B vector. The two vectors containing the anti-bLAC VHH clone were used to transform their respective *E. coli* strain.

Four colonies from each plate (p5X and pET22B) were used to check for the insert, 1 µg of the minipreps for all 8 cultures was double digested with high fidelity NCOI and ECOR I restriction enzymes and 10 µl of each digest was analysed on a 1.5% agarose gel (figure 5.19). The uncut vectors were run alongside the samples as controls.

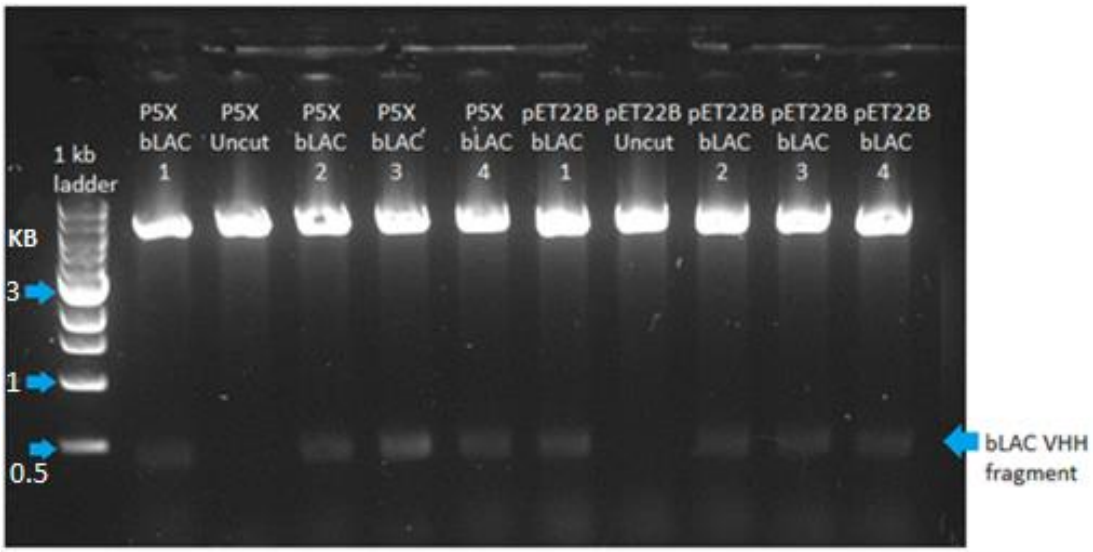


Figure 5.19: Anti-bLAC VHH digests from the P5X and PET22B vectors. Four P5X and PET22B minipreps were digested with high fidelity NCOI and ECOR I and were visualised on a 1.5% agarose gel. The gel confirmed the release of the anti-bLAC VHH gene from the two vectors. The digests were checked alongside a 1,000 bp ladder (product no. N3232S; NEB) and uncut vector controls. The gel was stained with Nancy-520 and the samples were loaded using a 6x loading buffer.

The agarose gel confirmed the presence of the bLAC VHH gene in all eight colony picks for both vectors.

5.7.3 Expression and binding comparison: bLAC in P5X and PET22B

Two anti-bLAC VHH clones were randomly picked for each vector (P5X and PET22B) for the comparative analyses and were used to produce soluble VHH. The antibodies were extracted from the whole cell and periplasm and assessed by western blot and ELISA (figure 5.20).

The anti-bLAC VHH antibodies are clearly detected at the expected molecular weights for both the PET22B (~15 kDa) and P5X (~57 kDa) vectors. The weight of the P5X VHH were considerably larger due to the MBP fusion however bands were detected at 15 kDa suggesting degradation during overnight incubation. The anti-bLAC VHH antibodies are also detected at the expected molecular weights. The samples were much cleaner with little degradation detected in comparison to the whole cell extracts. Some VHH without the MBP fusion were detected again for the P5X clones, with bands still visible at 15 kDa.

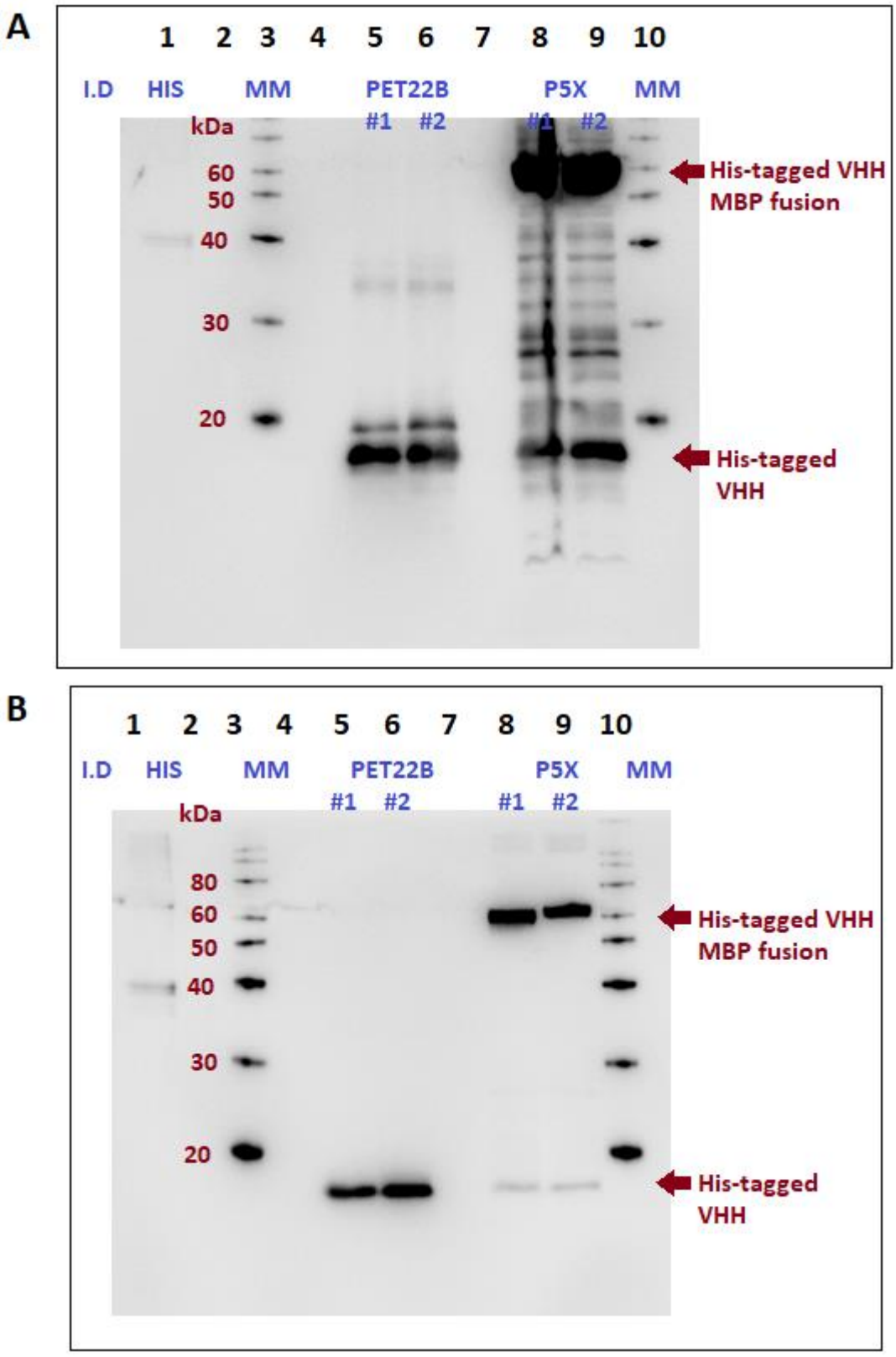


Figure 5.20: Western blot analysis of the anti-bLAC soluble VHH. The western blot compares the expression of the bLAC VHH antibody, derived from whole cell extracts (A) and the periplasm (B) from BL21DE3 (PET22B) and ER2523 (P5X) *E. coli*. Lane 1 showed a control HIS-tagged protein (10 µl). Lane 3 and 10 contained the molecular weight ladder, magic marker (MM). I.D refers to identifier. Soluble VHH antibodies (from both P5X

and PET22B) were detected via their HIS tags using an anti-tetra-his primary antibody (1:2000) and an anti-mouse-HRP secondary antibody (1:2000).

After confirming that both the P5X and PET22B vectors were expressing in all 4 clones as expected, the periplasmic extracts were tested for binding specificity by ELISA. The plate detected with anti-tetra-his failed to produce any signals for both the VHH specificity test (binding to the beta-lactamase immobilised on the ELISA plate) and for immobilised VHH (where the periplasmic extract was directly bound to the ELISA plate). Neither the P5X or PET22B produced VHH were detected by anti-tetra-his ELISA (data not shown).

The plate detected with the anti-MBP failed to detect the wells containing the PET22B extracts as expected, due to the lack of MBP. However, the anti-MBP detection system produced strong signals for both the immobilised MBP-VHH and the MBP-VHH specificity test.

An alternative VHH clone with the HCDR3 sequence 'QRLKKRKHHITSTSWK' was determined to be an anti-Elastin binder by a previous panning experiment carried out in our labs. The clone was present in the original pSD3 vector enabling the antibody to be expressed with a 6X HIS tag. To determine whether any VHH clone with a 6X HIS tag could be detected, or whether it was a problem unique to the anti-bLAC clone, the anti-Elastin clone was transformed into both TG1 (phage production) and HB2151 (soluble VHH production) *E. coli* cell lines. The anti-Elastin VHH was expressed as phage (from TG1) and soluble HIS-tagged VHH (from HB2151). Phage fusions and soluble VHH formats were both detected by western blot (data not shown), although the soluble VHH format signal was notably weak despite a high exposure. For the functional ELISAs, the antigen Tropoelastin (product no. T0706; Sigma) was immobilised on maleic anhydride and Nunc MaxiSorp® plates in duplicates. Aliquots of each VHH format (phage and soluble) were also immobilised on the Nunc MaxiSorp® plate (figure 5.21).

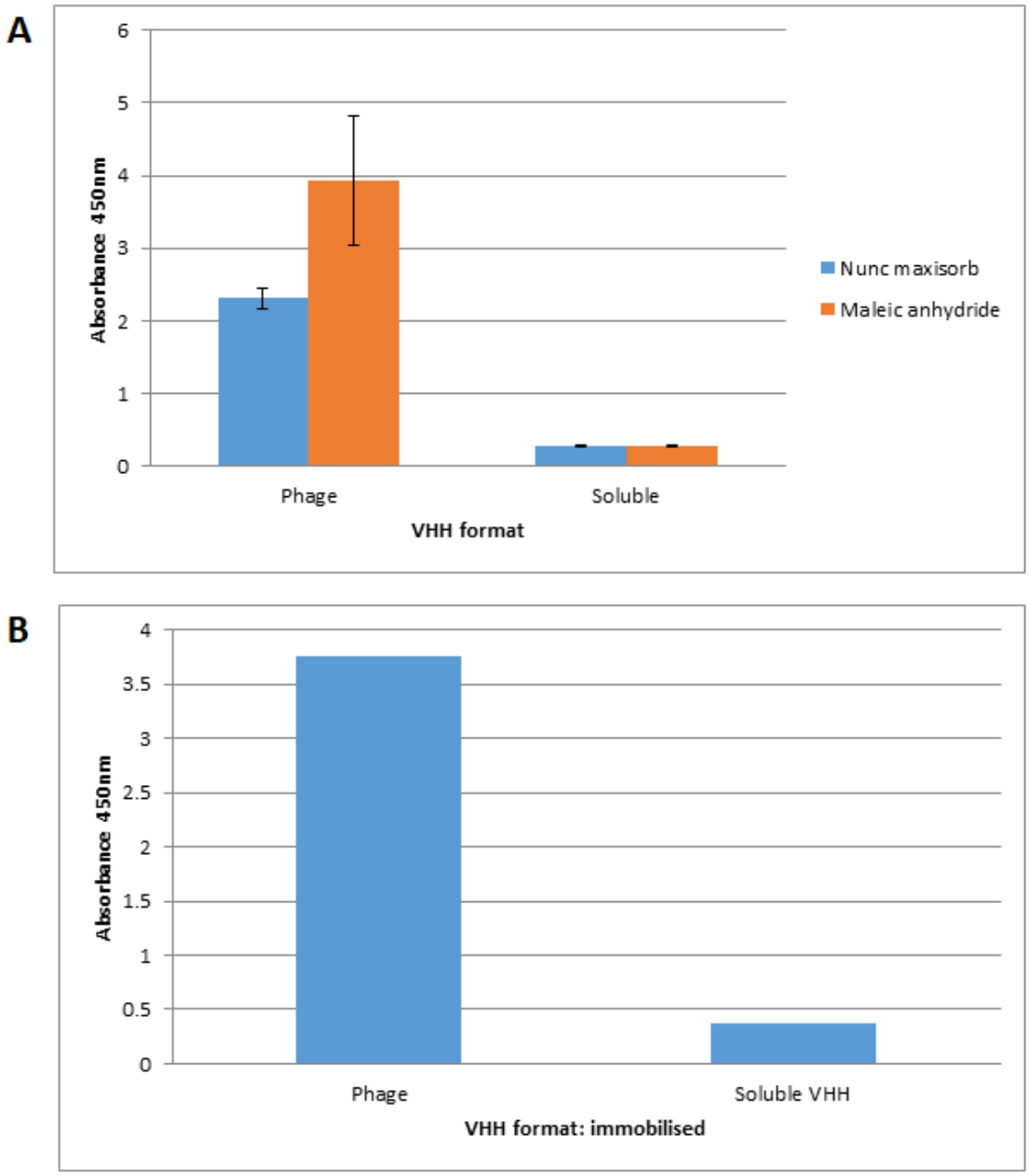


Figure 5.21: Anti-Elastin VHH ELISA: phage and soluble VHH formats. The anti-Elastin phage fusion and soluble VHH were immobilised (A) or tested for specificity (B) by ELISA. Tropoelastin (10 µg/ml) and the antibody formats were immobilised on a Nunc MaxiSorp® plate and maleic anhydride plates in 100 µl aliquots. Phage and soluble VHH fusions were used to detect the antigen. Immobilised or bound phage were detected by an anti-pIII primary antibody (1:4000) and immobilised or bound soluble VHH were detected by an anti-tetra-his primary antibody. An anti-mouse-HRP secondary antibody (1:2000) was used to detect both primary antibodies. The TMB substrate was used to develop positive signals, and after 15 minutes the reaction was stopped with 100 µl of 2M sulphuric acid per well. Absorbance values were taken at 450 nm. The average no antigen control value (0.2) was subtracted from the plotted data.

As a phage fusion, the anti-Elastin clone bound to immobilised elastin on both supports (absorbance signal on Nunc maxisorp was 10X background and on maleic anhydride was 19X background). As a soluble-pIII fusion, the binding activity was not apparent. Directly immobilised phage fusions were detected as expected, however the soluble format could not be detected. This was a similar setback to results with the PET22B vector, where perhaps the 6X HIS tag was inaccessible unless in a denatured state, hence its detection on a denaturing protein gel and subsequent western blot, but not as a native structure as detected during an ELISA.

Due to the lack of positive signals in the anti-tetra-his ELISA, the pMAL, P5X vector containing the anti-bLAC VHH clone was selected as the final template for the cloning of the VHH clone identified in the NGS data.

5.7.4 VHH synthesis

The 43 VHH candidates were re-synthesised into the pMAL P5X vector. The candidates were resynthesized by conducting an inverse PCR reaction using 5' phosphorylated primers; the products were checked on an agarose gel (data not shown). The anti-bLAC VHH clone in the P5X vector was used as the DNA template.

The inverse PCR products were expected to be around ~6,225 bp (p5X = 5,752 bp; VHH gene = ~473bp) depending on the size of the VHH HCDR3 (12mer, 16mer, 21mer). Samples #5 and 42 failed to amplify and were not pursued any further. The concentrations of the gel-purified extracts ranged from 1—17 ng/ul.

The samples with identifiers 16 and 19 failed to produce any transformants. Their transformations were repeated but still didn't produce any colonies so were abandoned. Four colonies from each VHH clone were sent for Sanger sequencing. All four picks for sample #3 failed to sequence after multiple attempts and was omitted from further analyses. In total, 5 of the original 43 VHH candidates had been lost due to lack of PCR amplicons, transformant or inability to sequence. The remaining VHH candidates were confirmed as the intended HCDR3 sequences by Sanger sequencing (data not

shown) and one of the 4 picked colonies were chosen for the functional assays that followed.

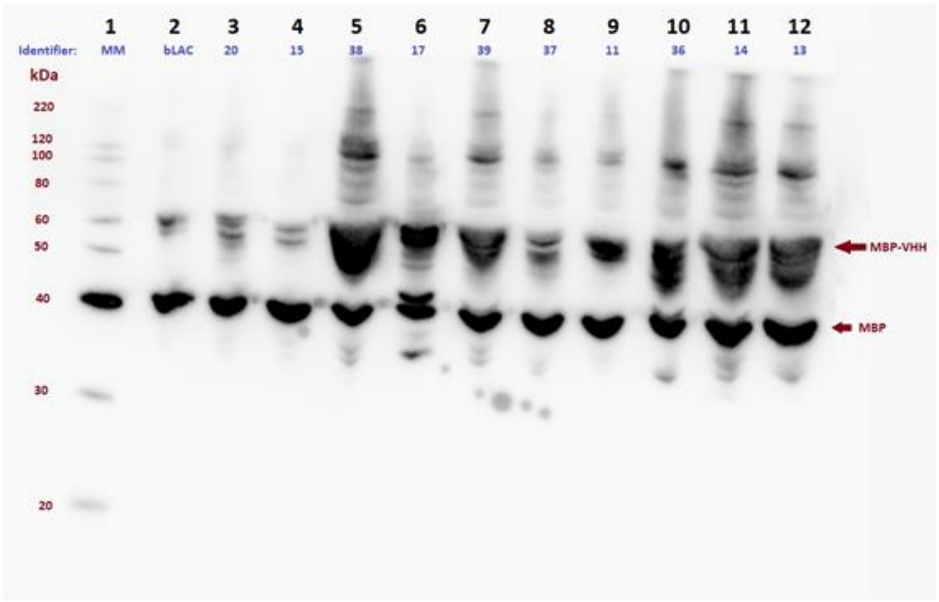
5.7.5 VHH production

Expression of the VHH candidates was analysed within the whole cell extracts and periplasmic extracts alongside the anti-bLAC VHH clone in P5X, serving as a positive control. The methods are described in section 2.26.

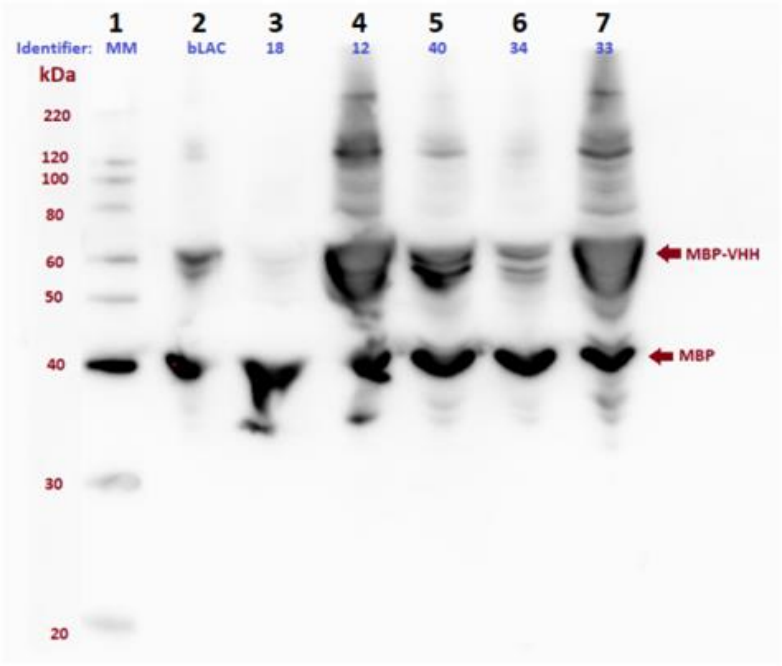
5.7.6 VHH expression and binding specificities

The first set of western blots and ELISAs were conducted using the whole cell extracts. The first group of VHH candidates to be tested were raised against denatured phytase (15 candidates; figure 5.22), the second group against active phytase (11 candidates; figure 5.23) and the third group against phytic acid (12 candidates; figure 5.24). Each VHH candidate was referenced by their identifier number.

A


1 2 3 4 5 6 7 8 9 10 11 12
Identifier: MM bLAC 20 15 38 17 39 37 11 36 14 13
kDa
220
120
100
80
60
50
40
30
20
MBP-VHH
MBP

B


1 2 3 4 5 6 7
Identifier: MM bLAC 18 12 40 34 33
kDa
220
120
100
80
60
50
40
30
20
MBP-VHH
MBP

C

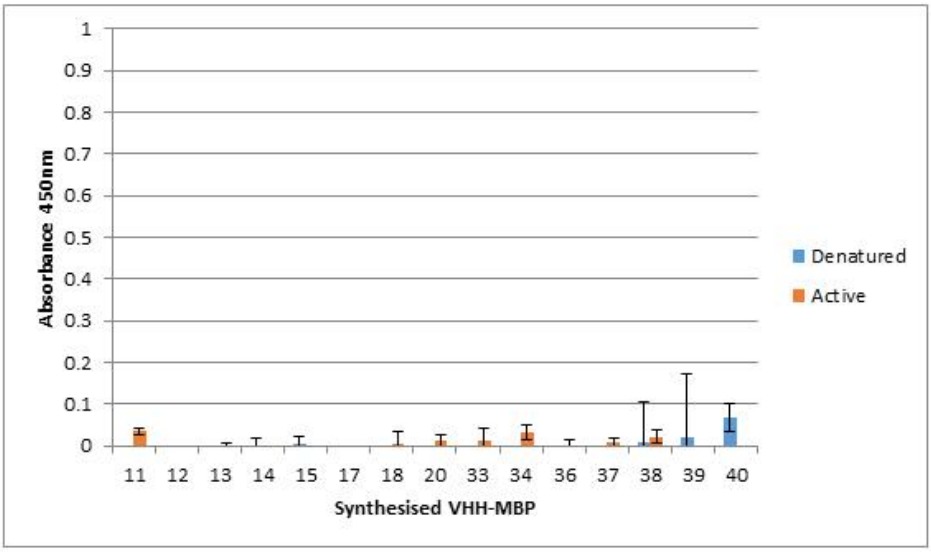

Absorbance 450nm
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
11 12 13 14 15 17 18 20 33 34 36 37 38 39 40
Synthesised VHH-MBP
Denatured
Active

Figure 5.22: Western blot analysis and Soluble VHH-MBP ELISA for the 15 clones raised against denatured phytase: whole cell extracts. The western blot compares the expression of the VHH candidates raised against denatured phytase and the positive control, an anti-bLAC VHH antibody. The soluble VHH-MBP fusions were derived from whole cell extracts from ER2523 (P5X) *E. coli.* Lane 1 contained the molecular weight ladder, magic marker (MM). Soluble VHH antibodies were detected via their MBP fusions using an anti-MBP primary antibody (1:10,000) and an anti-mouse-HRP secondary antibody (1:2000). The ELISA assessed the functionality of the clones (B). Denatured and active phytase and beta-lactamase were immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Soluble VHH-MBP fusions were used to detect the immobilised antigens in triplicates. Bound VHH-MBP were detected by an anti-MBP (1:10,000) and anti-mouse-HRP antibody (1:2,000). The TMB substrate was used to develop positive signals, and after 15 minutes the reaction was stopped with 100 µl of 2M sulphuric acid per well. Absorbance values were taken at 450 nm. The average no antigen control value (0.4) was subtracted from the plotted data.

Expression of the soluble VHH-MBP fusions was confirmed by western blot for all denatured phytase candidates, including the positive anti-bLAC control. All of the VHH candidates failed to produce strong, positive signals in the subsequent ELISA, including the positive control (not shown) and the positive clones identified by colony picking (I.D #33 & 34). For the active phytase candidates, extracted from the whole cell, the ELISA was conducted the same way as for denatured phytases, except for a slight alteration to the anti-MBP concentration. The first 30 minutes of the primary antibody's incubation step used the anti-MBP at a concentration of 1:10,000, and the second set of 30 minutes used the anti-MBP antibody at a concentration of 1:2,000. This was to try and offset the potential interference from other cell debris. The secondary antibody was used as normal (1 hour, 1:2000).

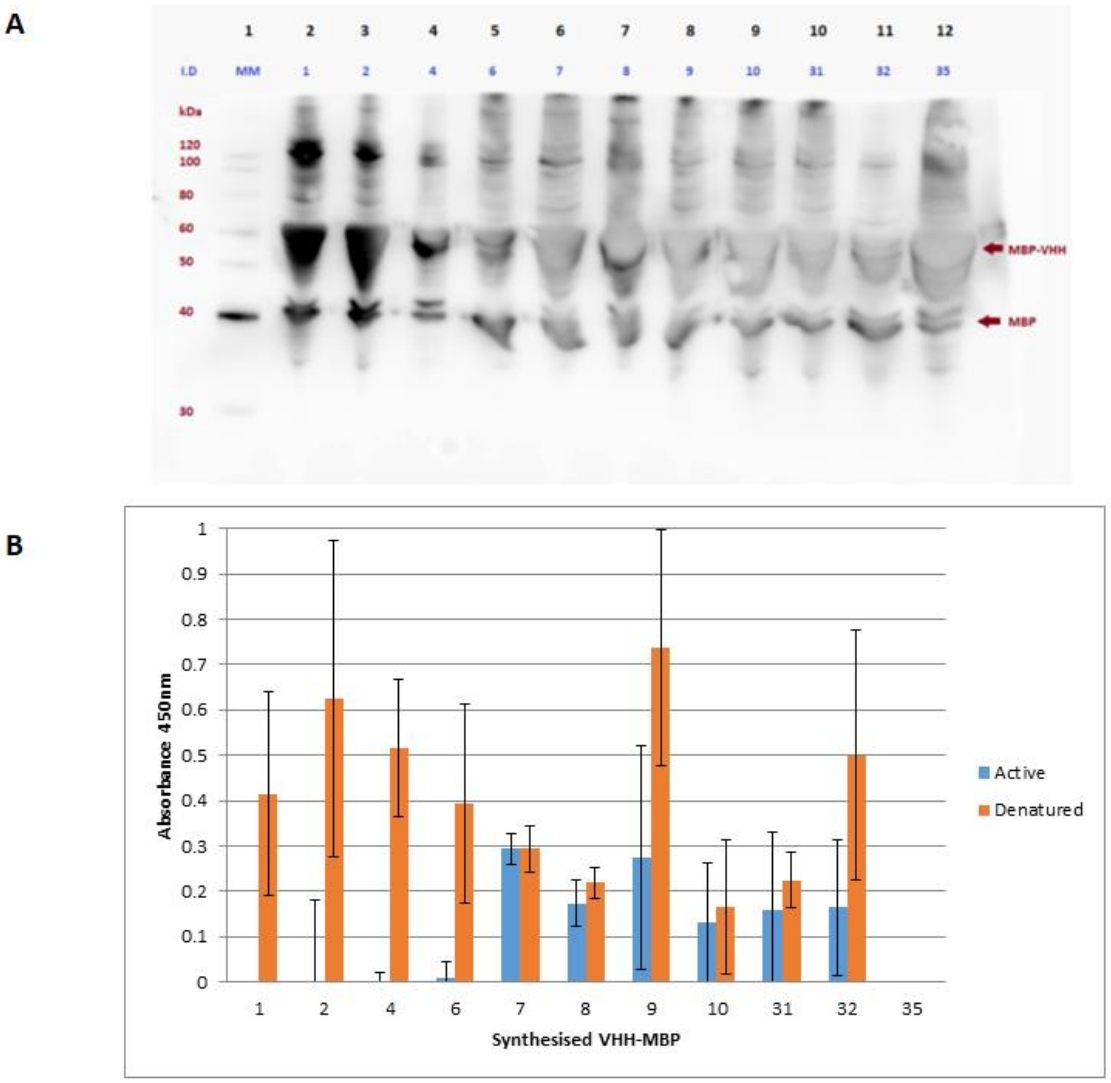


Figure 5.23: Western blot analysis and Soluble VHH-MBP ELISA for the 11 clones raised against active phytase: whole cell extracts. The western blot (A) compares the expression of the VHH candidates raised against active phytase. The soluble VHH-MBP fusions were derived from whole cell extracts from ER2523 (P5X) E. coli. Lane 1 contained the molecular weight ladder, magic marker (MM). Soluble VHH antibodies were detected via their MBP fusions using an anti-MBP primary antibody (1:10,000) and an anti-mouse-HRP secondary antibody (1:2000). The ELISA assessed the functionality of the clones (B). Active and denatured phytase was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Soluble VHH-MBP fusions were used to detect the immobilised antigens in triplicates. Bound VHH-MBP were detected by an anti-MBP (1:10,000 & 4,000) and anti-mouse-HRP antibody (1:2,000). The TMB substrate was used to develop positive signals, and after 5 minutes the reaction was stopped with 100 µl of 2M sulphuric acid per well. Absorbance values were taken at 450 nm. The no antigen control value (1.91) was subtracted from the plotted data.

Expression was confirmed for all 11 of the active phytase candidates. The increase in anti-MBP concentration led to a very high background value (~1.91) which was subtracted from all of the plotted values for the active phytase candidates. The reaction also had to be stopped after 5 minutes but it would usually be left for from 10—15 minutes. Most of the candidates still demonstrated some degree of binding activity, however this favoured the denatured conformation of phytase rather than the active. The clones identified from colony picking (I.D #31 & 32) demonstrated weak binding activity against both conformations. Replicates had high standard deviations that was potentially caused by the higher concentration of the anti-MBP antibody, and the absorbance values achieved by the active phytase VHH candidates were very low and not even twice that of the background signal. In addition, none of the candidates favoured the active conformation of phytase.

The use of the anti-MBP antibody at the recommended concentration (1:10,000) led to very low absorbance values however this was sufficient to detect the positive control as a periplasmic extract as shown previously in chapter 4. The higher concentration (1:2,000) of the anti-MBP antibody led to especially high background signals and the final absorbance values were still lower than twice the background signal.

For the phytic acid ELISAs, the anti-MBP antibody was tested at 1:4,000. However, the target and immobilisation system differed considerably to that of the phytases and was less comparable.

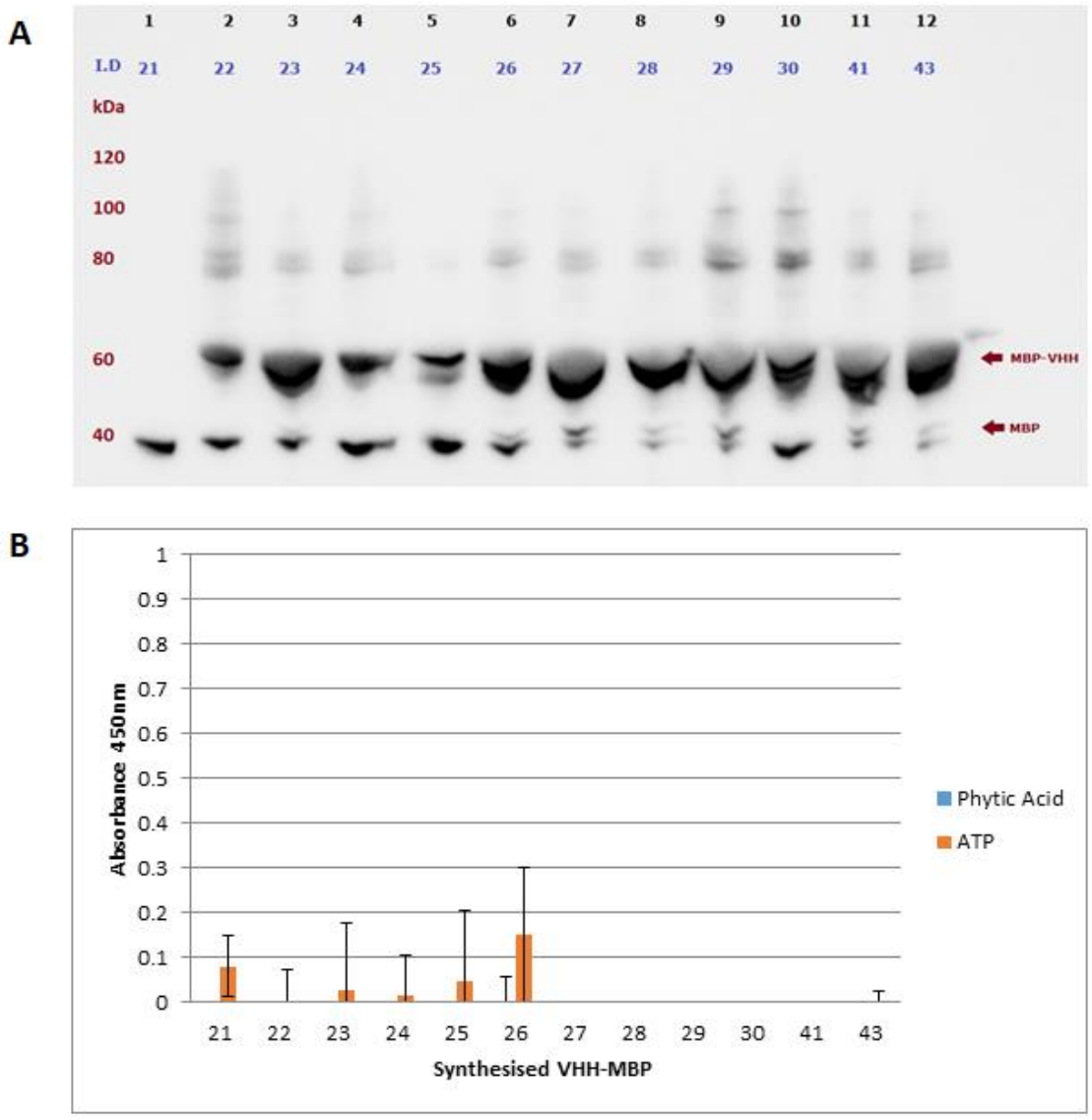


Figure 5.24: Western blot analysis and soluble VHH-MBP ELISA for the 12 clones raised against phytic acid: whole cell extracts. The western blot (A) compares the expression of the VHH candidates raised against phytic acid. The soluble VHH-MBP fusions were derived from whole cell extracts from ER2523 (P5X) E. coli. Lane 1 contained the molecular weight ladder, magic marker (MM). Soluble VHH antibodies were detected via their MBP fusions using an anti-MBP primary antibody (1:10,000) and an anti-mouse-HRP secondary antibody (1:2000). The ELISA assessed the functionality of the clones (B). Phytic acid and ATP was immobilised overnight in 100 µl aliquots (0.1M) on a TiO_2 plate. Soluble VHH-MBP fusions were used to detect the immobilised antigens in triplicates. Bound VHH-MBP were detected by an anti-MBP (1:4,000) and anti-mouse-HRP antibody (1:2,000). The TMB substrate was used to develop positive signals, and after 15 minutes the reaction was stopped with 100 µl of 2M sulphuric acid per well. Absorbance values were taken at 450 nm. The no antigen control value (0.55) was subtracted from the plotted data.

The expression of all but one (I.D #21) of the VHH candidates raised against phytic acid was confirmed by western blot. The ELISA data failed to demonstrate any convincing binding activity for the VHH clones. The background absorbance values were within the usual range for the TiO_2 system as achieved throughout the study. This value was subtracted from all plotted data.

None of the VHH candidates bound to their antigens convincingly as whole cell extracts, but neither did the positive anti-bLAC control. It was possible that whole cell extracts had too much cell debris interference. Before conducting the western blots and ELISAs for the VHH candidate's periplasmic extracts, a control experiment was run using the anti-bLAC positive control.

Fresh anti-bLAC VHH was extracted from both the whole cell and periplasm. A no insert P5X control was grown alongside it, expressing lone MBP. An aliquot of each extract was also nickel purified using a Qiagen Ni-NTA Spin Kit (product no. 31314) following the manufacturer's protocol under native conditions. The extracts were then tested by western blot and ELISA (functional and immobilised VHH) using both anti-MBP and anti-tetra-his detection systems (figures 5.25). Immobilised VHH/MBP were done so in single replicates only whereas the functionality ELISA was conducted in triplicates. For the ELISAs, the dilution of the anti-MBP antibody was changed to 1:6,000 as opposed to the recommended 1:10,000.

Expression was confirmed for all extracts as intended. The bLAC clone produced VHH-MBP fusions as detected by both anti-MBP and anti-tetra-his demonstrating the full length VHH was present as an MBP fusion. The confirmed presence of the HIS-tag enabled the subsequent experiment during which the soluble VHH were purified using nickel columns (figure 5.26).

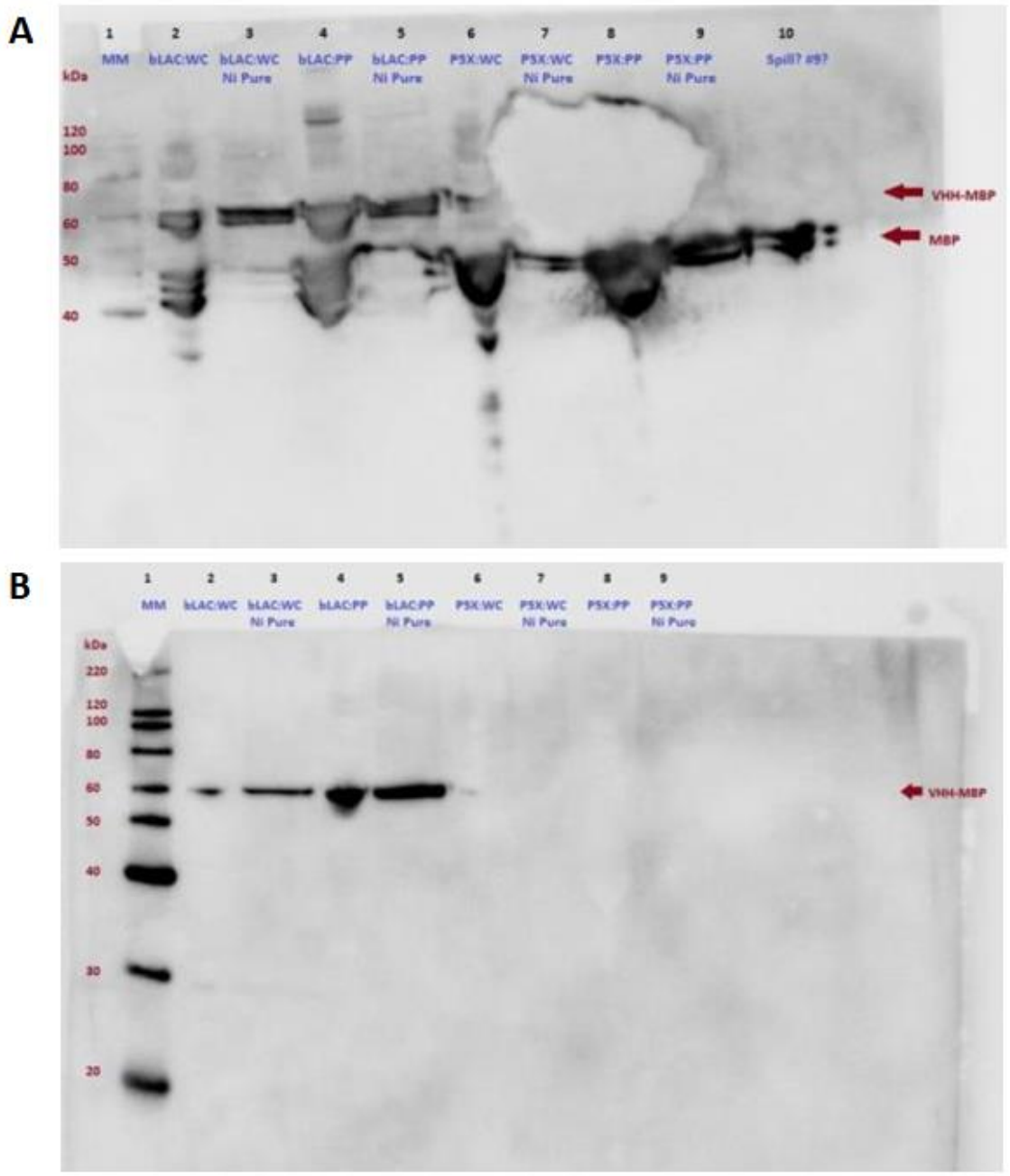


Figure 5.25: Western blot analysis of the anti-bLAC VHH clone. The western blot compares the expression of the anti-bLAC VHH clone (bLAC) and a no insert P5X control (P5X). The soluble VHH-MBP fusions were derived from whole cell (WC) and periplasmic (PP) extracts from ER2523 (P5X) E. coli. Aliquots of each extract were also nickel purified (Ni Pure). Lane 1 contained the molecular weight ladder, magic marker (MM). Soluble VHH antibodies were detected via their MBP fusions using an anti-MBP primary antibody (1:10,000) or their 6X HIS tags by an anti-tetra-his antibody (1:2,000) followed by an anti-mouse-HRP secondary antibody (1:2000).

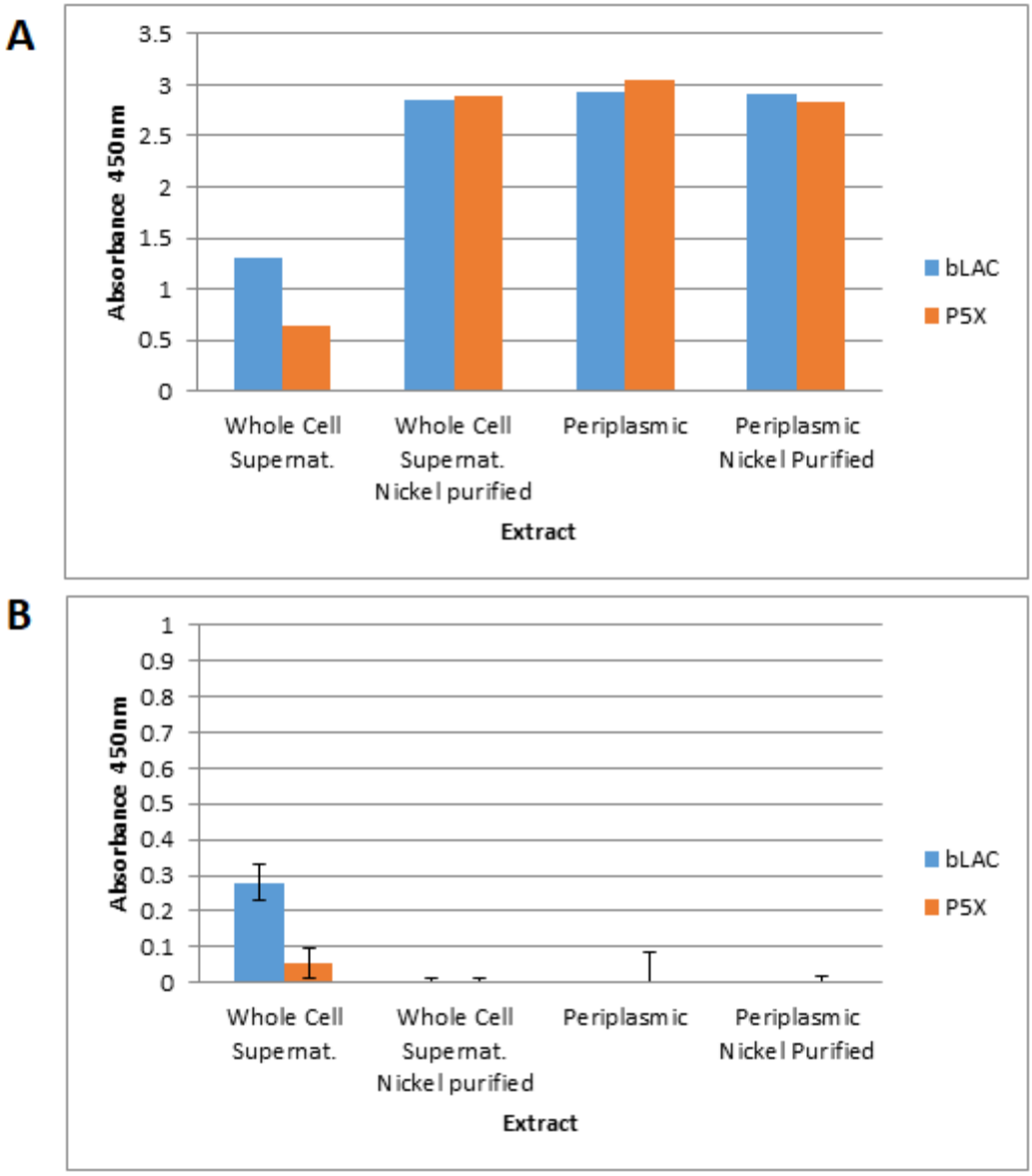


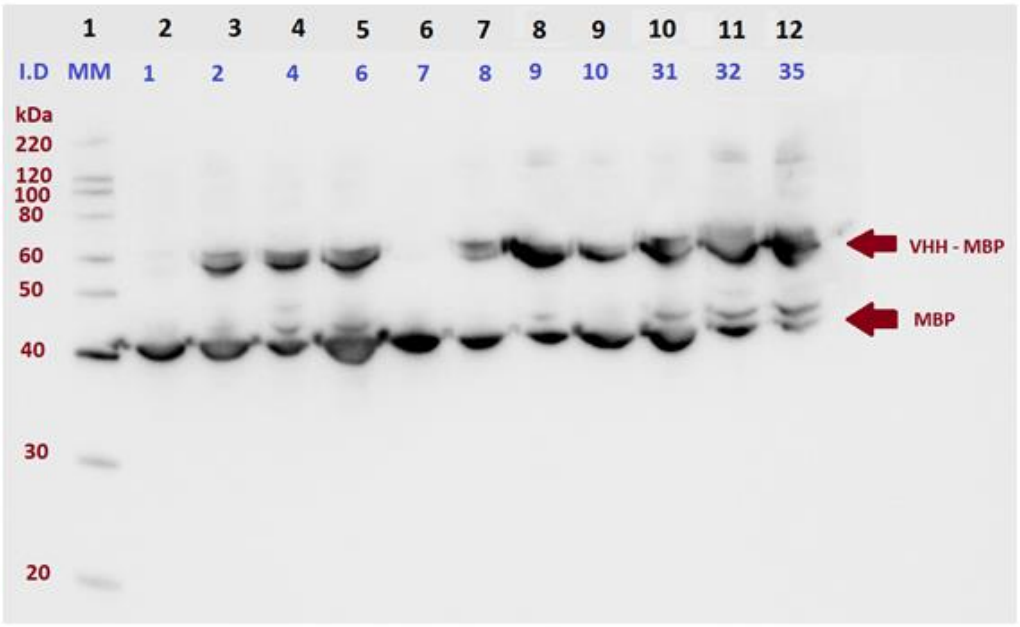
Figure 5.26: Soluble VHH-MBP ELISAs for anti-bLAC clone and no insert P5X control. The expressed VHH/MBP proteins from the anti-bLAC clone and an empty P5X control were used in an ELISAs. The VHH/MBP were immobilised (A) on a Nunc MaxiSorp® plate. Beta-Lactamase was immobilised on a maleic anhydride plate at 10 µg/ml (B). The beta-lactamase was detected using VHH-MBP fusions or lone MBP. Immobilised proteins were detected by anti-MBP (1:6,000) and anti-mouse-HRP antibody (1:2,000). The TMB substrate was used to develop positive signals, and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid per well. Absorbance values were taken at 450 nm. The average no antigen control value (0.2—0.4) was subtracted from the plotted data.

The anti-MBP ELISA system detected immobilised VHH-MBP/MBP proteins from all sources however produced a much lower signal for the whole cell supernatant. This suggested some interference, preventing the antibody from binding as easily as when it was nickel purified or from a cleaner extract (periplasmic). It is possible that because the sample was extracted from the whole cell, the cell debris would have reduced the amount of MBP-VHH immobilised on the plate itself, causing a lower absorbance signal. In contrast, the only extract preparation to show functional binding against beta-lactamase for the positive control was the whole cell supernatant (non-purified). Previously whole cell extract failed to produce a positive ELISA signal in the functional ELISA, whereas a previous periplasmic extract had a positive ELISA signal. Here, the nickel purification step also failed to improve the functional binding signals for the anti-bLAC clone.

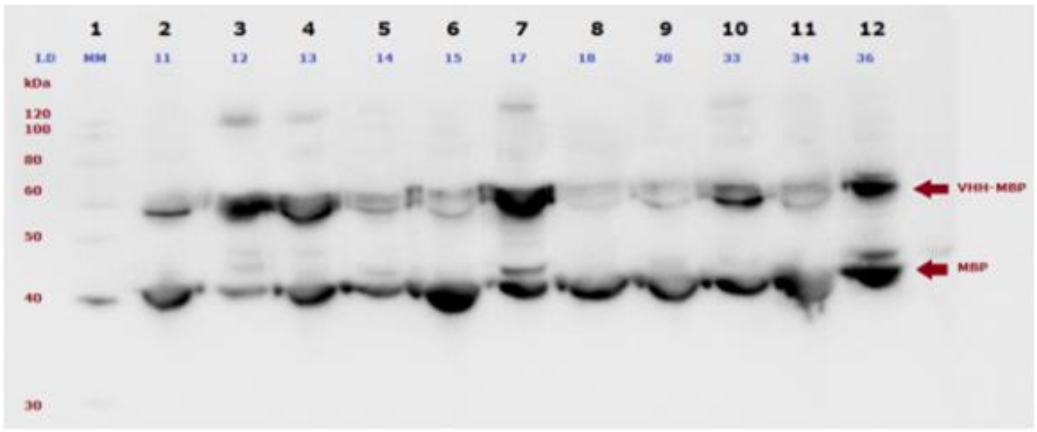
The western blots and immobilised VHH ELISAs confirmed the expression of the intended anti-bLAC VHH, however the functional ELISAs failed to produce convincing binding signals. Over several experiments it was apparent that the binding of the anti-bLAC VHH-MBP for its antigen was highly variable which may be due to stability issues of the antibody fusion or the antigen. Due to time constraints, the periplasmic extracts for each VHH candidate to phytic acid and to phytase were tested by western blot and ELISA using the anti-MBP assay as it stood (figure 5.27). The anti-bLAC control was not included. The anti-MBP antibody was used at a dilution factor of 1:6,000.

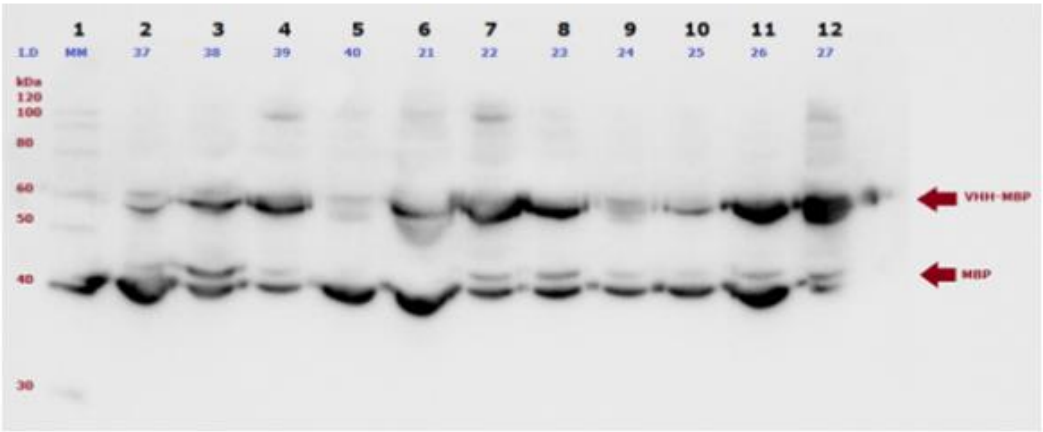
Expression was confirmed for all but two of the VHH candidates (active phytase candidates 1 and 7) using periplasmic extracts. The candidates were then tested for functional by ELISA (figure 5.28).

A


1 2 3 4 5 6 7 8 9 10 11 12
I.D MM 1 2 4 6 7 8 9 10 31 32 35
kDa
220
120
100
80
60
50
40
30
20
VHH - MBP
MBP

B


1 2 3 4 5 6 7 8 9 10 11 12
I.D MM 11 12 13 14 15 17 18 20 33 34 36
kDa
120
100
80
60
50
40
30
VHH-MBP
MBP


1 2 3 4 5 6 7 8 9 10 11 12
I.D MM 37 38 39 40 21 22 23 24 25 26 27
kDa
120
100
80
60
50
40
30
VHH-MBP
MBP

C

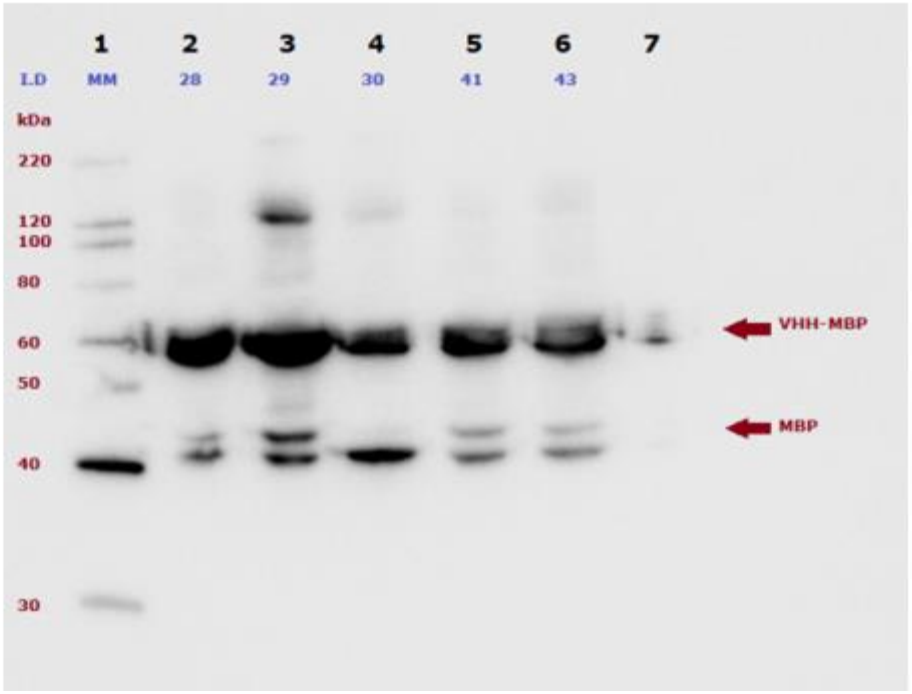


Figure 5.27: Western blot analysis of the active and denatured phytase and phytic acid VHH candidates extracted from the periplasm. The western blot compares the expression of the active (A) and denatured phytase (B) and phytic acid (C) VHH candidates. The soluble VHH-MBP fusions were derived from periplasmic extracts from ER2523 (P5X) *E. coli*. Lane 1 contained the molecular weight ladder, magic marker (MM). Soluble VHH antibodies were detected via their MBP fusions using an anti-MBP primary antibody (1:6,000) and an anti-mouse-HRP secondary antibody (1:2000).

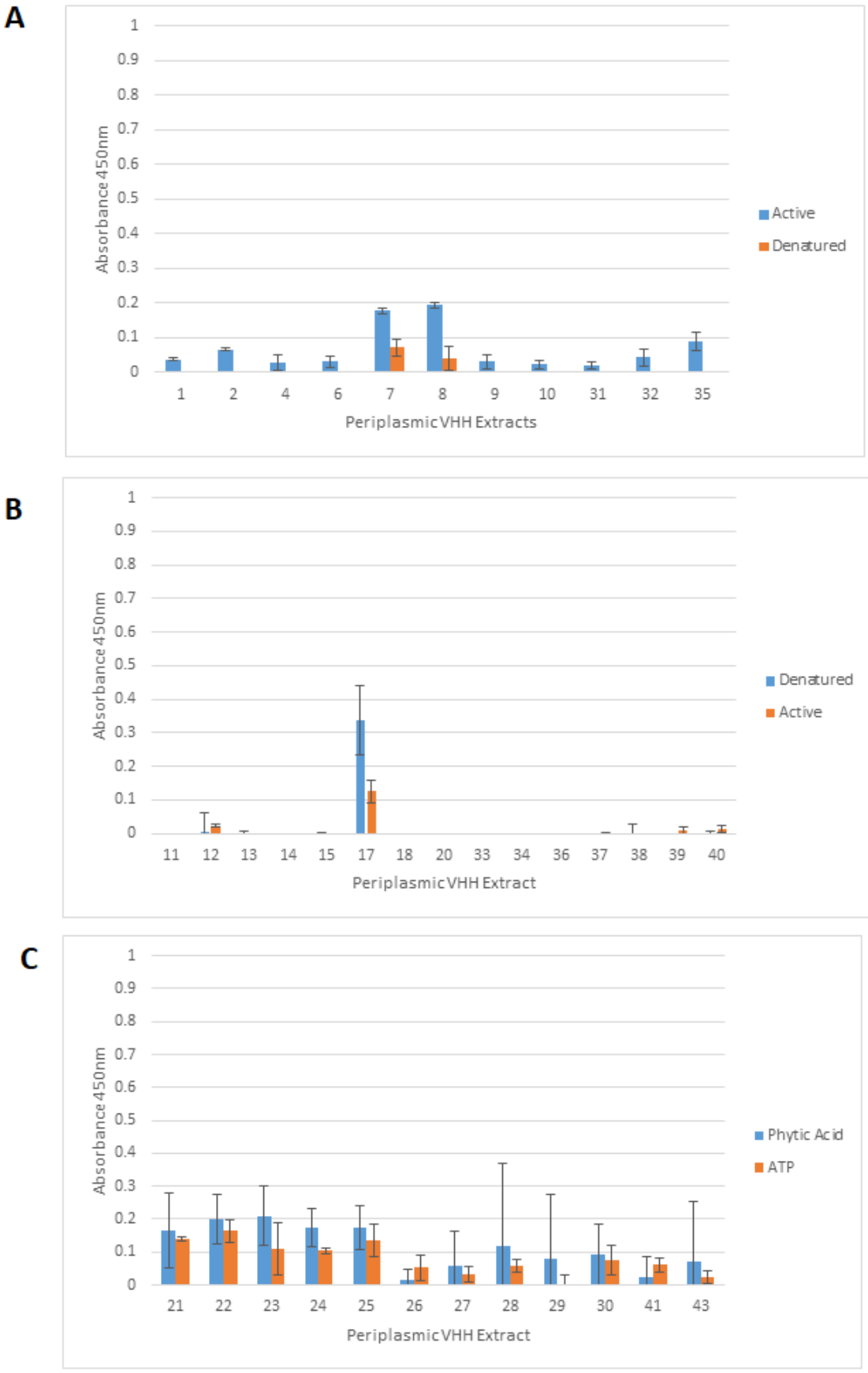


Figure 5.28: Soluble VHH-MBP ELISA for the clones raised against active and denatured phytase: periplasmic extracts. The 11, 15 and 12

VHH clones raised against active phytase (A), denatured phytase (B) and phytic acid (C) respectively, were tested for functionality by ELISA. Active phytase and denatured phytase was immobilised overnight in 100 µl aliquots at 10 µg/ml on a maleic anhydride plate. Phytic acid and ATP was immobilised overnight in 100 µl aliquots (0.1M) on a TiO_2 plate. Soluble VHH-MBP fusions were used to detect the immobilised antigens in triplicates. Bound VHH-MBP were detected by an anti-MBP (1:6,000) and anti-mouse-HRP antibody (1:2,000). The TMB substrate was used to develop positive signals, and after 10 minutes the reaction was stopped with 100 µl of 2M sulphuric acid per well. Absorbance values were taken at 450 nm. The no antigen control value (0.4) was subtracted from the plotted data.

The functional ELISA signals were not convincing (<1X background). The denatured phytase candidate I.D #17 had a slightly elevated signal against the denatured conformation of phytase as intended but the signal was still below 1X the background signal. The active phytase candidate I.D #'s 7 and 8 showed similar activity to that of the denatured phytase candidate I.D #17. There was a slight elevation in the absorbance signal against active phytase in comparison to denatured phytase but the signals were still not very strong (<1X background). The phytic acid candidates had maximum absorbance signals of 0.2 after the background was subtracted, but with minimal difference between the phytic acid and ATP absorbance signals. Due to time constraints, these data concluded the lab-based aspect of the study.

5.8 Discussion

The camelid VHH library demonstrated much greater expression levels when compared to the Griffin and Thomlinson (I & J) library. After comparing the expression of 10 clones from each library, the VHH library consistently had a higher expression rate (when analysing either whole cell or periplasmic extracts). The lack of expression in some clones may have been due to the presence of stop codons within the antibody genes. If this were the only preventative factor, then the Griffin and I & J libraries would be expected to have a much higher prevalence of stop codons within the region of interest than the VHH library, possibly because they contain larger genes with more regions of synthetic diversity. It is possible that the other clones were expressing but the level of production was too low, or the antibodies degraded more readily or were simply too toxic to the host to reach detectable levels. Due to the higher chances of antibody production, the camelid VHH library was selected for use in this part of the study. Furthermore, it would be much simpler to synthese VHH instead of rescuing them from a polyclonal sub-library.

The camelid heavy chain antibody format possesses many useful traits in comparison to scFv. They are highly soluble, thermostable and can rapidly refold after the removal of some denaturants. Their small size (~15 kDa) enables them to reach less accessible epitopes, an ideal trait considering the study's difficult conformational targets. Similarly to vNARs, the VHH antibodies naturally possess longer CDR3 sequences (15—18 residues (Kovalenko et al., 2013) that not only contribute to the level of variation between antibodies, but may also form pockets that are ideal for capturing smaller target antigens such as haptens.

In addition to the use of a VHH library, this chapter introduced two new immobilisation supports. The first being maleic anhydride for the phytases and the second being Covalink for phytic acid. The phage display panning method involves multiple wash steps intended to remove non-specific antibodies, whereas target binders are expected to bind strongly enough to the target to withstand the stringent wash steps. Although Nunc MaxiSorp® plates are commonly used in phage display experiments; this chapter

investigated the alternatives due to their stronger covalent or ionic interactions with the target antigens. Lysine amino acids are especially nucleophilic due to their amine side chain and the AB Vista Quantum Blue phytase contains 13 within its polypeptide chains; this presented the opportunity of a stronger immobilisation interaction between the phytase and maleic anhydride that may have been more suitable for phage display applications. Likewise, the Covalink plates offered covalent bonding for phytic acid immobilisation, forming a phosphoramidate bond (NR_2) between the phytic acid phosphates and the plate's nitrogen linker. The covalent bonding was more likely to withstand the vigorous washing procedures during panning but also allow better orientation of phytic acid so that it might be more accessible to the VHH antibodies.

The experiments involved four panning rounds with subtraction steps in-between. Polyclonal ELISAs were conducted for each round of panning to establish whether there was an abundance of positive binders within the pool. The phytase conformations had positive signals for rounds 3 and 4 however the signals were positive against both conformations. This was to be expected as the majority of positive binders were unlikely to bind to the rarer epitopes unique to each conformation. As for the polyclonal ELISA, the monoclonal ELISA results for phytic acid were negative, including for the Covalink experiment.

For the phytases, a total of 4 different clones were identified from monoclonal ELISA data however the data was often conflicting when establishing whether the clones were conformation specific. It was decided that the 4 clones would be tested as soluble antibodies.

The candidates identified by the bioinformatics analysis, based on Z scores and common motif analysis, and the 4 clones identified by random colony picks were synthesised into the pMAL p5X vector. The in-house camelid library used within this study randomised the HCDR regions whilst conserving the surrounding sequences thereby facilitating the rapid synthesis of new VHH candidates into other vectors already containing a scaffold clone from the same library. The synthesis strategy has significant advantages over the rescue of naïve or semi-synthetic scFv clones from panning sublibraries, as previously attempted. A study by Lovgren et al. (2016) had a similar approach for an scFv library where only the CDR regions of the VH and VL

chains were randomised. The study sequenced the scFv candidates using the V3 2x 300bp Illumina sequencing kit, thereby covering both the VH and VL chains individually. Lovgren et al. (2016) synthesised ten clones via Integrated DNA technologies (USA), with restriction sites at the 5′ and 3′ ends enabling their digestion and subsequent ligation into a phagemid vector. In the present study's VHH strategy aimed to bypass the reintroduction of clones into a phagemid vector, and instead progress to the immediate testing of promising candidates as a soluble antibody format. Positive antibody-phage fusions usually have to be re-tested as soluble antibodies to ensure that the clone's binding properties rely solely on the antibody fragment. Thus, the present strategy aimed to create a more efficient pipeline for testing multiple clones identified through bioinformatics analysis.

The pMAL p5X system was the most functional when determining which vector (p5X or pET22B) to resynthesize the VHH candidates into. Using the anti-bLAC VHH as a control, the pET22B vector relied on a 6x HIS tag on the VHH antibodies as a means of detection which appeared to be inaccessible when the VHH was in its native form. In contrast, the p5X MBP fusions could be detected in both a denatured and native format with an antibody directed to the MBP. Whilst the MBP-VHH fusions also contained a C-terminal 6x HIS tag, this was inaccessible to detecting antibodies unless it was in a denatured state.

Four clones were identified from random colony picking and 39 VHH clones were identified by bioinformatics analysis for a total of 43 different VHH clones. Five clones were lost during the inverse PCR steps due to a lack of amplicons despite repeated attempts. The remaining 38 candidates were amplified and synthesised correctly as confirmed by Sanger sequencing. The data indicated that ~88% of clones could be cloned into an expression vector in a single PCR step. This was a considerable improvement from the scFv rescue strategy where the majority of the scFv candidates were not rescued as intended. The scFv clones that produced the correct size amplicons and subsequent transformed colonies still failed to resemble the target sequence identified by the bioinformatics strategy. Instead, the expression of the 38 VHH clones (94%) was confirmed by both whole cell and periplasmic western blot analysis.

Although the majority of the VHH candidates were expressing as intended, the clones failed to produce positive ELISA signals as whole cell extracts. This included the 4 clones identified by random colony picking that demonstrated positive binding reactions as phage fusions, and also the anti-bLAC positive control VHH. The whole cell extract ELISA for the active phytase candidates, produced stronger signals after using the anti-MBP primary antibody at a higher concentration (1:4,000 compared to 10,000). However, this concentration led to a very high background value (~1.91) which was subtracted from all of the plotted values for the active phytase candidates. The reaction also had to be stopped after 5 minutes with substrate, where it would usually have been left for 10—15 minutes. In contrast, the anti-bLAC soluble VHH ELISAs, had previously been successful using periplasmic extract and anti-MBP diluted 1:10,000 (refer to chapter 4). The anti-bLAC VHH were derived from the periplasm in that case, and the whole cell extracts may have suffered from a proportion of the VHH-fusion being misfolded due to incorrect secretion to the periplasm or possible inhibitory effects from the detergents used for extraction or cellular components present in the crude sample. The periplasmic extracts should not have encountered such problems.

Despite the higher anti-MBP concentration (1:6,000), the VHH clones and the anti-bLAC VHH clone also failed to produce positive ELISA signals against their target antigens as periplasmic extracts. The anti-bLAC ELISA had shown that periplasmic extracts produced near equal absorbance values as nickel purified fractions so it was unlikely that the purity of the extracts was a problem in these cases. The clones were expressed overnight at 30°C and it may be likely that expressing the clones for such a long duration had led to considerable degradation of the fusion protein by proteases. Although fusions were present in most cases this degradation may indicate sub-optimal expression and stability of the fusion proteins that may have caused low levels of VHH-MBP being present in the ELISAs and may even have caused the presence of free VHH that could have interfered with the ELISA. A dominant MBP-only band was also present for the anti-bLAC clone (figure 5.40 & 5.45) when detected with an anti-MBP antibody and the sample failed to produce a positive functional ELISA signal. In the single case where the anti-bLAC clone had produced a positive functional ELISA signal (refer to chapter 4), it was again expressed overnight at 30°C, but when detected on

a western blot using an anti-HIS antibody (figures 5.40 & 5.41) this sample clearly had a dominant fusion protein at much higher levels than cleaved VHH. This data may indicate that the expression levels of MBP-VHH fusion require considerable optimisation to ensure that fusion protein is dominant in the samples and at relatively high levels. As purified anti-bLAC fusion-protein also failed to bind in functional ELISAs, purity cannot be the only issue.

The loss of binding activity from the positive control between preparations of the fusion protein raises the question of how valid the expression system is for producing functional VHH when fused to MBP. In addition, the loss of binding activity in the 4 anti-phytase clones identified by random colony picking is also unexpected. All 4 candidates showed repeatable functional binding as phage fusions although they were not phytase-conformation specific. The other 34 candidates had shown such enrichment within their experiments as indicated by NGS analysis, so that it is unlikely that every single one was in fact a false positive, and that the highly stringent bioinformatics analysis failed to filter them out. The study by Ravn et al. (2010a) was discussed in chapter 4, detailing how multiple clones only bound reliably to their antigens as phage fusions. In some cases, it is possible that the phage entity becomes an essential component of the antibody-antigen interaction however antibodies such as these are undesirable reagents for downstream processes. In the case of Ravn et al. (2010a), the soluble antibodies were detected by their 6x HIS tag. The present study has already shown the temperamental nature of the 6x HIS tag detection system in protein fusion systems, and how the presentation of the complex might have an adverse effect on its detection. In the case of the VHH antibodies, the pET22B and p5X vectors resulted in constructs where the StrepII-6x HIS tags were exposed at the C terminus. In the case of p5X, the MBP fusion albeit a large protein remains at the N terminus, leaving the 6x HIS tag as exposed for all of the subsequent VHH candidates that were resynthesized into that vector. It is possible that the presence of the StrepII-polyHis tag at the C-terminus of the VHH constructs forms a conformation that not only obstructs the polyHis tag but also the CDR3 region of the VHH, thereby preventing binding to the antigen. When the pIII protein is present after these VHH tags this may alter the conformation allowing CDR3 accessibility for binding.

This is not the case for the scFv clones (C7 and E6) from chapter 4, where they were digested with restriction enzymes NCOI and NOTI just before their 6x HIS tag, thus it was not carried over to the new vector. An alternative argument could suggest that the MBP protein itself causes interference with the antibodies ability to bind to the targets due to steric hindrance. This was not the case for one of the preparations of anti-bLAC VHH clone but this may have been a rare exception.

A recent study considered three commonly used fusion tags for the expression of VHH (Liu et al., 2018). The study investigated the use of 6x HIS tag, MBP and Small Ubiquitin-like Modifier (SUMO) tags to determine which tag was optimal in terms of VHH expression and purification. The VHH sequence was a known binder for the anti-Cap protein of porcine circovirus type 2 (PCV2) and was transferred between different expression vectors (pCold I: 6xHIS; c2X: MBP; pE-SUMO: SUMO) so as to contain one of the three tags. The study concluded that the MBP fusion has the lowest antigen binding activity in comparison to the VHH containing the other tags where the authors also suggest the likelihood of steric hindrance interfering with protein folding and antigen interaction. Another study by Salema and Fernandez (2013) demonstrated the use of a MBP-VHH fusion as a means of increasing the overall yield of soluble antibodies extracted from the periplasm. In this case the MBP fusion result in a higher yield (≥12 mg/L) but also retained its ability to bind to its antigen. But it's ability to bind to the target was reduced considerably in comparison to its non-fusion form which continued to bind and remain detectable at a 3-fold lower concentration than the MBP fusion. The authors also attribute the lower affinity for the target antigen to possible steric hindrance caused by the large MBP protein.

To date, only a handful of phage display studies (Miyazaki et al., 2015, Turner et al., 2016) have combined the VHH antibody format with next generation sequencing as a means of identifying new target binders. The present study identified enriched potential binders using Z score rankings and common motif analyses.

In each case, the studies made use of immune libraries rather than the naïve/synthetic libraries used within this study. This introduced a bias towards the production of a specific type of antibody tailored against the target antigen. In the present study, immune libraries were not an option but

the bioinformatics approach may be improved by considering published strategies even though the libraries were already more biased to begin with. The study by Miyazaki et al. (2015) used an immune VHH library and only conducted one round of panning before introducing a sequencing (Illumina) and bioinformatics approach. The study took into account both the original library and panning round sequence diversity. The authors looked for VHH sequences that had become enriched through panning by comparing the frequencies of clones before and after biopanning to determine the clones that had become enriched by panning. The study then compared 100 of the most frequent sequences and found that 75% of them closely resembled the sequences of 13 previously identified binders from the colony picking approach. Three sequences identified by the high-throughput sequencing approach, that did not resemble the previously identified binders, were tested for antigen binding activity and two of them successfully bound to the target antigen. The study highlighted the usefulness of a bioinformatics approach for isolating less diverse ligands that might have otherwise gone undetected by colony picking. The study by Turner et al. (2016) also used an immune VHH library but took a slightly different approach to the bioinformatics analysis. Instead, the authors only sequenced the initial immune library using an Illumina platform without conducting any panning rounds. The library sequencing data was interrogated using a known binder sequence previously identified by colony picking. The authors carried out a BLAST search to identify clones with similar sequences. A series of VHH candidates were tested by surface plasmon resonance (SPR), including two with unnaturally high copy numbers (4,100—1,100) within the library. The study discovered 9 new binders using these approaches however the clones with unnaturally high copy numbers, despite being within an immune library, failed to bind to the target antigen.

The bioinformatics strategies described by others for VHH libraries (Miyazaki et al., 2015, Turner et al., 2016) involved much simpler approaches to that of the present study. In both cases, known binder sequences were used to interrogate the sequencing data of immune libraries to identify similar or novel sequences, the former are likely to possess similar binding properties with varying degrees of affinity for the target antigen. The present study used a very stringent bioinformatics approach as a means of filtering large datasets

of sequence diversity. The strategy involved generating Z scores (Zhang et al., 2011), Naqid et al. (2016) by comparing enriched target antigen, HCDR3 sequences to a negative control dataset. In addition, a reversed analysis was conducted which considered the Z scores generated when the negative control replicates were compared against the pool of target antigen replicates. The translated HCDR3 sequences were also used in a common motif analysis to determine the most common residue combinations amongst the most enriched candidates. Despite the authors best efforts, the strategies did not identify target binders that would continue to bind as soluble antibodies. The study's antigens were particularly difficult targets in terms of being able to pinpoint the specific epitopes required to distinguish between closely related proteins, or target a hapten with limited epitope presentation. The scFv clones identified in chapter 4 appeared to be conformation specific when part of a phage fusion, and 4 VHH clones appeared to bind consistently to the two phytase conformations although they failed to distinguish between conformations as intended. Only the VHH candidates were synthesised successfully after identification by NGS, but it remains unclear as to why the clones were all enriched despite a lack of interaction with the target antigen, or whether there was a fundamental fault with the assay or clones themselves. As stated above, given the data produced with the control anti-bLAC VHH it is likely to be the latter. Nevertheless, some target antigens are more challenging than others and even antibodies that bind to the target may have low affinity and still be unsuitable for downstream assays. Evidence in the literature would suggest that the coupling of NGS with phage display strategies can still offer researchers an alternative to laborious colony picking, and in many cases enable the rapid identification of a multitude of potential binders against the target (Miyazaki et al., 2015, Turner et al., 2016, Lovgren et al., 2016, Yang et al., 2017).

Chapter 6: Assessment and optimisation of bioinformatics analysis methods

This chapter aimed to re-evaluate the bioinformatics analysis strategies for the NGS data used within the study, whilst also considering the strategies used in the next generation phage display field as a whole, regardless of the ligand being displayed.

First, the NGS data from chapter 4 (scFv) was interrogated. As the library was based on naturally occurring scFv sequences with diversity even within the framework regions, the analysis considered the effect of HCDR3 flanking region changes on the selection of the top 20 scFv candidates that were identified.

The VHH library diversity was also considered and how this changes through panning experiments.

Finally, the VHH sequencing data from chapter 5 was re-analysed using a less stringent, published method. The only validated positive VHH candidates identified within the present study were found by colony picking methods. As the same candidates could not be produced as functional soluble antibodies, it is likely that for the VHH format there was a fundamental problem with the ELISA assay itself. Several studies have successfully used the NGS approach to identify a multitude of binders against a range of target antigens. The VHH datasets obtained within the study and other datasets produced elsewhere were therefore analysed further to assess whether the NGS strategies used were too stringent, or whether it was perhaps more likely that an assay problem had prevented the discovery of novel antibodies.

6.1 scFv flanking region changes

The scFv rescue strategy was a major setback for demonstrating the benefits of using NGS with phage display, however an accumulation of mutations such as truncations or out of frame sequences within the library itself also hindered its usefulness. Another challenge with the scFv library is that the flanking regions of HCDR3 are not fixed and so the degeneracy of the flanking sequence allowed when 'capturing' HCDR3 sequences could influence the datasets analysed and therefore the scFvs selected. Here, the

scFv dataset from chapter 4 was re-analysed to look at this effect of degeneracy allowances within the HCDR3 flanking regions on scFv clone selection.

The normal NGS analysis pipeline allowed 1-change in each of the flanking regions (VYYCAR and WGQGT). The following analyses selected HCDR3 sequences but allowed 0 or 2 changes per flanking sequence; this was conducted for all three target's scFv NGS datasets (active and denatured phytase and phytic acid).

The analysis aimed to see where the original 1-change scFv candidates then fell within the data, whether they would still be considered viable candidates or whether they had been replaced by antibodies that contained fewer changes within their flanking regions.

The scFvs identified by the original '1-change allowed' analysis were selected using Z score thresholds of $Z \geq 10$, 5 or 2 and filtering candidates that were seen in at least 4 positive replicates as described in methods section. Following this, clones seen as being enriched in unrelated panning experiments and then those with high Z scores after the reversed analysis were removed. A resulting 17 scFvs against denatured phytase, 12 for active phytase and 12 for phytic acid (table 6.1) were shortlisted. Within the '0-changes allowed' dataset, 6 of the original 41 candidates were completely absent from the data (AEAAD, GDPS, GGASPLHN, RVFLLFD for denatured phytase and ASCS, GWFFV for active phytase) (table 6.1), suggesting that they had at least 1 change within their flanking regions and had been omitted by the search criteria.

The final list of top 20 candidates from the 0 and 2 changes datasets were compared against the top candidates originally selected for rescue from the 1 changes dataset to see whether the original candidates still made the top 20 (table 6.1).

Table 6.1: Original list of rescue candidates for panning against active and denatured phytase and phytic acid based on 1 change allowed within each flanking sequence compared to 0 or 2 changes.

Original Rescue [1]	0 changes [2]	2 changes [2]	Still a top 20 candidate? [3]	
			0 changes	2 changes
AEAAD	No	Yes	N/A	Yes
AETAD	Yes	Yes	No (excluded by reversed analysis)	No (excluded by reversed analysis)
ALNLR	Yes	Yes	Yes	No
ASMSPVH	Yes	Yes	Yes	No
CQIEW	Yes	Yes	Yes	No
GAMEYG	Yes	Yes	No	No
GAYEDYVQ	Yes	Yes	Yes	No
GDPS	No	Yes	N/A	No
GGASPLHN	No	Yes	N/A	No (duplication, seen in both phytases)
GTYLY	Yes	Yes	Yes	No
IGATNPREISSL	Yes	Yes	Yes	No
KPGPLFASS	Yes	Yes	Yes	No
LLDGQG	Yes	Yes	Yes	Yes
PEPQPFDY	Yes	Yes	Yes	Yes
RVFLLFDY	No	Yes	N/A	No (duplication, seen in both phytases)
VLGDRH	Yes	Yes	No	Yes
WGPGSE	Yes	Yes	Yes	No (duplication, seen in all three targets)
AFFIQTPD	Yes	Yes	No	Yes
AVDL	Yes	Yes	No	No
CGSGL	Yes	Yes	No	No
GLKHFYRPVWID	Yes	Yes	Yes	No (duplication, seen in all three targets)
IHRMVYL	Yes	Yes	No	Yes
KIYTH	Yes	Yes	No	No
LNPA	Yes	Yes	No	No
LTPSQG	Yes	Yes	No	No
QGGGW	Yes	Yes	No	No
QWSYSKDFSPN	Yes	Yes	Yes	No (duplication, seen in all three targets)
REICR	Yes	Yes	Yes	No (duplication, seen in all three targets)

Original Rescue [1]	0 changes [2]	2 changes [2]	Still a top 20 candidate? [3]	
			0 changes	2 changes
TRSPK	Yes	Yes	No	No
ASCS	No	Yes	N/A	No (excluded by reversed analysis)
DSEFDTK	Yes	Yes	Yes	Yes
GWFFV	No	Yes	N/A	No
HMLFTS	Yes	Yes	Yes	No
IKPACNTMA	Yes	Yes	No	Yes
LMSQQ	Yes	Yes	Yes	No (duplication, seen in all three targets)
PDDLIG	Yes	Yes	Yes	Yes
PPSVHFDY	Yes	Yes	Yes	Yes
STQLP	Yes	Yes	Yes	No (duplication, seen In both phytases)
TSLLGVNR	Yes	Yes	Yes	No (duplication, seen in all three targets)
VGESQ	Yes	Yes	Yes	No (duplication, seen In both phytases)
YLAEFG	Yes	Yes	Yes	No (duplication, seen In both phytases)

[1] The amino acid sequence of the original HCDR3 being rescued based on the 1-change allowed dataset. Key: Blue text = Denatured phytase candidate. Green text = Phytic acid candidate. Orange text = Active phytase candidate

[2] The data was re-analysed permitting either 0 or 2 changes within each flanking sequence. The data was searched to see whether the original rescue candidates were still present within the data.

[3] Indicating that the scFv was still ranked within the top 20 in the new 0 or 2 flanking region datasets, scFvs also seen in other datasets were removed (No (Duplication, seen in "..") and candidates were also removed if they were also seen in the reversed analysis (No (excluded by reversed analysis).

Referring to table 6.1, when no changes were allowed in the flanking regions, only 23 of the 41 original rescues (1 change allowed), would be ranked in the top 20 candidates due to higher ranking of other HCDR3 sequences. A further HCDR3 was removed during reversed analysis. So overall only 53% of the original target clones would have been selected. After allowing up to 2 changes in the flanking sequences, only 10 of the original candidates (24%) were still ranked in the top 20 candidates and were not excluded due to reversed analysis (also being seen enriched in the negative target) or due to also being enriched in the other panning experiments (indicating non-specificity). The lists of top 20 candidates for 0 and 2 changes

therefore identified potential new rescue candidates. The 2 changes allowed datasets for all three antigens, were also searched for the presence of the new, 0 changes allowed candidates. Only 8 out of the 60 candidates were present in both datasets. This data shows that the majority of the selected candidates are unlikely to be antigen specific and were selected only because other 'versions' of the same HCDR3 were omitted from the data due to multiple changes in the flanking sequences (as highlighted with the 2 changes dataset analysis). Whether this is due to natural diversity or due to errors in PCR/sequencing is unknown. The data may highlight the difficulties of analysing NGS data from naïve or semi-synthetic antibody libraries where flanking regions contain variable sequences.

Predicting error rates caused by amplification and sequencing

The approximate error rates caused by amplification mutations and incorrect base calls during sequencing were predicted using the conserved VHH library. The total number of reads within each target antigen's panning round files (rounds 1—4) were determined after permitting 0—2 changes within the non-degenerate HCDR3 flanking regions (table 6.2). Using the 0 change read counts as a baseline, the percentage increase in total read counts were calculated after allowing 1—2 changes (table 6.3).

The additional reads represented data that was omitted due to errors within the flanking region calls. The percentage increase averaged at 12% across the 12 datasets after allowing 1 change in the flanking region. After allowing 2 changes, the average percentage increase in reads was much higher at 83.59%. These additional reads represent a small fraction of what error rates might be expected in the HCDR3 regions which can be 2—4X the length of the flanking regions.

Table 6.2: Total reads in each round of panning, for each target antigen, after 0—2 flanking region changes

Total Reads[1]	Active				Denatured				Phytic Acid			
Changes Permitted[2]	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4
0	127732	138061	157881	236973	144883	229184	165899	78064	380888	169310	1126718	222584
1	144080	151911	172202	254155	163283	256513	187095	85985	428027	196564	1293342	259525
2	244528	233712	251146	356032	273084	411242	304748	131196	735167	343177	2290239	477412

[1] The total number of reads in each panning round file (rounds 1—4 per target antigen)
[2] The read count data was analysed by permitting 0—2 changes within each flanking sequence.

Table 6.3: Percentage increase in total reads for each round of panning after comparison to 0 flanking region changes

Error Rate[1]	Active				Denatured				Phytic Acid			
Changes Permitted[2]	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4
1	12.80	10.03	9.07	7.25	12.70	11.92	12.78	10.15	12.38	16.10	14.79	16.60
2	91.44	69.28	59.07	50.24	88.49	79.44	83.69	68.06	93.01	102.69	103.27	114.49

[1] The error rate caused by amplification and sequencing errors, as determined as a percentage increase in the total number of reads when compared to the 0 changes total read counts.
[2] The read count data was analysed by permitting 1—2 changes within each flanking sequence.

Griffin library HCDR3 capture

The Griffin library is comprised of both natural and synthetic HCDR3 sequences with natural variation within the framework 3 and framework 4 sequences and therefore flanking sequences. By allowing multiple changes within the already degenerate flanking sequences, a greater number of HCDR3 sequences can be identified. The naïve Griffin library was analysed before and after allowing 1—2 changes within the flanking regions and the number of HCDR3 sequences captured after each criterion was calculated. The total number of HCDR3 sequences captured were compared to that of the 0 changes allowed total and the percentage increase in read counts were calculated (table 6.4).

Table 6.4: Percentage of Griffin library HCDR3 sequences captured after allowing 0—2 changes in the flanking regions

READS	0 changes	1 changes	2 changes
Griffin library	515194	600329	636113
% increase	0%	16.5%	23.4%

By allowing 1 change in the flanking regions, 600,329 HCDR3 sequences were identified which equated to a 16.5% increase in the number of HCDR3 sequences identified in comparison to 0 changes allowed. By allowing 2 changes in the flanking regions, 636,113 HCDR were identified; a 23.4% increase in HCDR3 sequences in comparison to the fixed set of flanking sequences.

6.2 VHH & Griffin library diversity

The next stage of assessing bioinformatic analysis tools considered the camelid VHH and Griffin scFv library diversity. The three VHH sub-libraries (12mer, 16mer and 21mer) and the Griffin library were miniprepped twice from their bacterial glycerol stocks. They were analysed by NGS in duplicate. The two datasets for each sub-library were then compared using a diversity estimate perl script.

The two Griffin library barcodes each had 122,267 and 76,824 reads with 8,844 HCDR3 sequences seen within both datasets. The estimated diversity for the Griffin library was $1x10^6$.

For the VHH libraries, the two 12mer barcodes each had 142,326 and 30,997 distinct reads with 80 HCDR3 seen in both datasets. The estimated diversity (Lincoln) for the 12mer VHH library was $5x10^7$. The two 16mer barcodes each had 197,836 and 167,943 distinct reads with 1,895 HCDR3 sequences present within both datasets. The estimated diversity (Lincoln) for the 16mer VHH library was $1x10^7$. The two 21mer barcodes each had 141,139 and 162,597 distinct reads with 2,588 HCDR3 seen in both datasets. The estimated diversity for the 21mer VHH library was $8x10^6$. The total diversity for the VHH library was $7x10^7$.

6.3 Copy number as an estimate of enrichment through panning

The 12mer, 16mer and 21mer libraries were analysed separately. The number of times each HCDR3 sequence was seen once, twice, three times, up to 15 times were documented. The counts were expressed as a percentage of the total reads (tables 6.5—6.7). The VHH library template with the HCDR3 sequence "VRGYFMRLPSSHNFRY" was for a β-lactamase binder (Conrath et al., 2001) and was counted 212 times within the 12mer library, 166 times within the 16mer library and 143 times within the 21mer library. The presence of this clone was higher in the library due to its use as a template during the library's production.

The majority of the HCDR3 sequences had only one copy within the library sequencing data, accounting for 85%, 83% and 88% of the distinct HCDR3 sequences within the 12mer, 16mer and 21mer sub-libraries respectively.

Table 6.5: Sequence (HCDR3) copy number and a percentage of the total reads in the 12mer VHH library.

Copies [1]	HCDR3 count [2]	% of total [3]
1	737,782	85.11314
2	102,314	11.80331
3	19,875	2.29285
4	4,514	0.520751
5	1,351	0.155856
6	539	0.062181
7	217	0.025034
8	105	0.012113
9	57	0.006576
10	25	0.002884
11	15	0.00173
12	6	0.000692
13	12	0.001384
14	6	0.000692
>15	7	0.000808

[1]The number of times a HCDR3 sequence was seen within the VHH library.
[2] The number of HCDR3 seen with each copy number ([1]).
[3] The number of HCDR3 seen with each copy number ([1]) as a percentage of the total distinct HCDR3.

Table 6.6: Sequence (HCDR3) copy number and a percentage of the total reads in the 16mer VHH library.

Copies [1]	HCDR3 count [2]	% of total [3]
1	570,364	83.04236
2	93,186	13.56745
3	17,147	2.496524
4	3,966	0.577431
5	1,312	0.191021
6	438	0.063771
7	194	0.028246
8	102	0.014851
9	63	0.009173
10	25	0.00364
11	15	0.002184
12	5	0.000728
13	6	0.000874
14	6	0.000874
>15	6	0.000874

[1] The number of times a HCDR3 sequence was seen within the VHH library.
[2] The number of HCDR3 seen with each copy number ([1]).
[3] The number of HCDR3 seen with each copy number ([1]) as a percentage of the total distinct HCDR3.

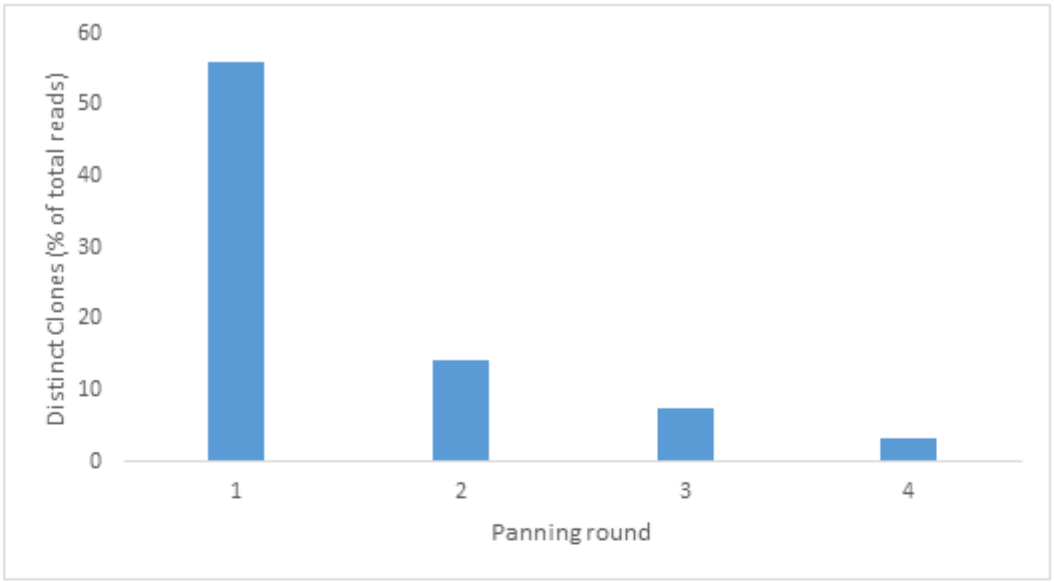
Table 6.7: Sequence (HCDR3) copy number and a percentage of the total reads in the 12mer VHH library.

Copies [1]	HCDR3 count [2]	% of total [3]
1	468,700	88.25346
2	50,359	9.482304
3	8,592	1.617823
4	2,066	0.389016
5	703	0.132371
6	309	0.058183
7	143	0.026926
8	79	0.014875
9	38	0.007155
10	27	0.005084
11	20	0.003766
12	14	0.002636
13	8	0.001506
14	11	0.002071
>15	15	0.002824

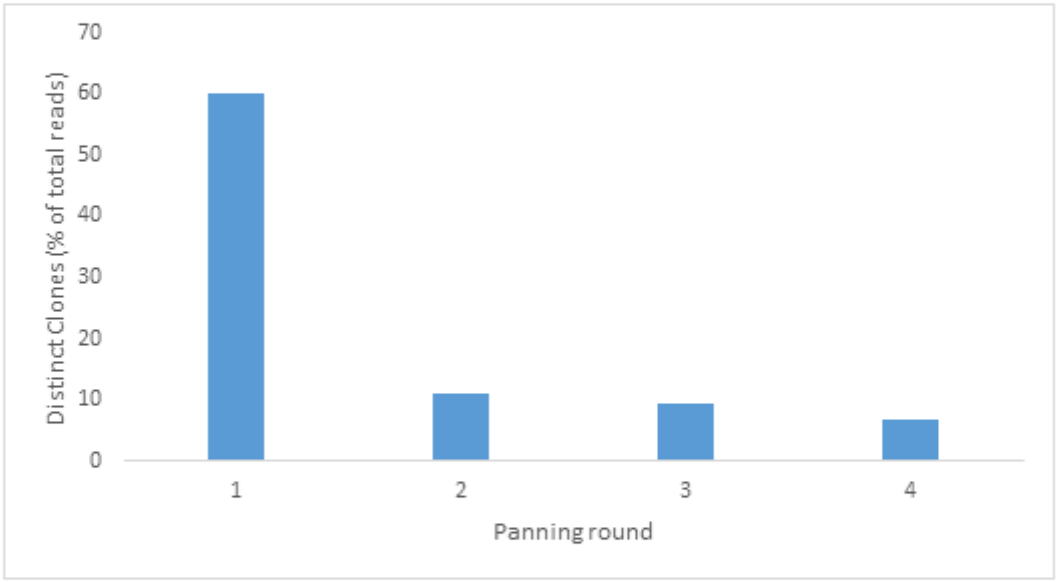
[1] The number of times a HCDR3 sequence was seen within the VHH library.
[2] The number of HCDR3 seen with each copy number ([1]).
[3] The number of HCDR3 seen with each copy number ([1]) as a percentage of the total distinct HCDR3.

The VHH panning data (glycerol stocks of sub-libraries) from chapter 5 were revisited. Each panning round's total read count and number of distinct reads (diversity) was taken into account (figures 6.4—6.6) for all three target antigens. This data demonstrated what had been selected by each round of panning against a target antigen.

A


Distinct Clones (% of total reads)
60
50
40
30
20
10
0
1
2
3
4
Panning round

B


Distinct Clones (% of total reads)
70
60
50
40
30
20
10
0
1
2
3
4
Panning round

C

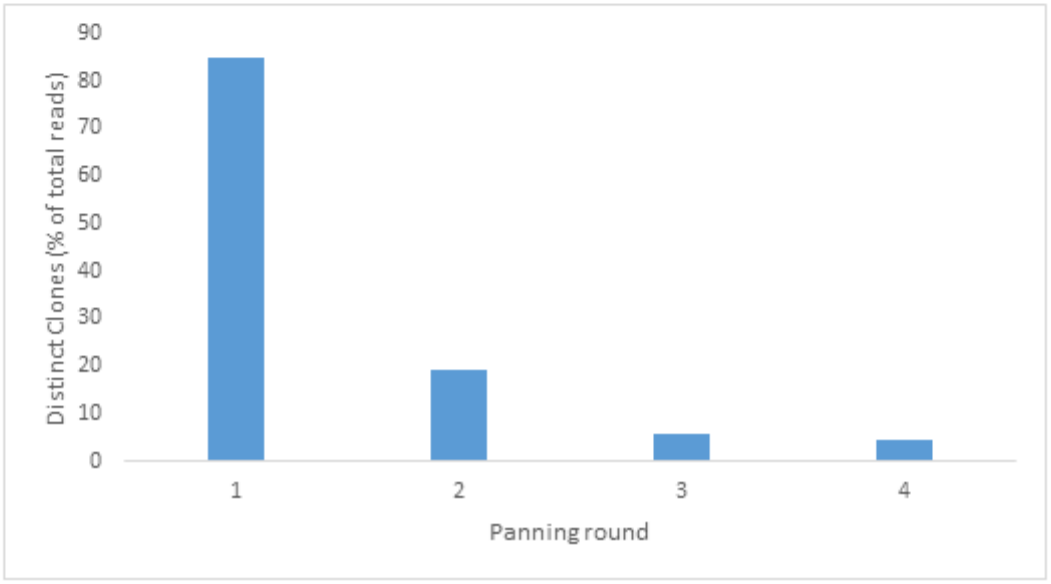

Distinct Clones (% of total reads)
90
80
70
60
50
40
30
20
10
0
1
2
3
4
Panning round

Figure 6.1: The number of distinct HCDR3 sequences captured during each panning round for active and denatured phytase and phytic acid. The bacterial glycerol stocks for each active (A) and denatured phytase (B) and phytic acid (C) panning round using the camelid VHH library were miniprepped and amplified in preparation for NGS. The barcoded files were analysed and the HCDR3 sequences were identified based on known, flanking regions. The number of distinct HCDR3 sequences (as percentages of the barcode's total reads) were compared between panning rounds 1—4.

The diversity dropped throughout the rounds of panning and phage output samples as expected when specific clones are being enriched for.

6.4 Reanalysing the VHH data using alternative published methods

The bioinformatics analysis method used within this study was quite complicated and had very strict selection criteria to try and select for hapten specific antibodies (phytic acid) and even conformation specific binders (phytase). However, many NGS phage display publications to date based their candidate selection on overall frequency, and by using common amino acid clustering tools. They often do not generate a ranking system similar to the Z score, or compare against sequencing data covering a negative panning control using the same library. In combination with the highly stringent panning techniques, it is possible that this NGS analysis strategy was too strict and thus failed to identify true target-specific binders, rather than the chapter 5 soluble ELISAs having a fundamental fault.

To understand whether alternative bioinformatics analysis would have changed the candidate binders that were identified, the VHH NGS datasets for active and denatured phytase, and phytic acid, were reanalysed using an alternative published method. The new data analysis strategy was based on the VHH selection paper published by Miyazaki et al. (2015). This study filtered the NGS data based on the overall frequency of each HCDR3 candidate before and after panning and did not compare unrelated datasets. The positive panning replicate's HCDR3 sequences, from the round 4 screen, were pooled into a single list. The same was done for the negative panning replicates. The HCDR3 sequences were ranked by their read counts, displayed as a percentage of the total reads. Their read counts within the negative pool was also calculated, but for comparison purposes only – the study by Miyazaki

et al. (2015) did not use a negative control pool. The top 20 frequency candidates are shown in tables 6.8—6.10 and any of the original VHH candidates identified by this study's Z score or AA motif analysis were highlighted. The frequencies of each HCDR3 sequences were also determined in each positive panning replicate's barcode individually, and expressed as a % of that barcode's total reads (tables 6.11—6.13).

Table 6.8: The top 20 frequency candidates for the active phytase VHH dataset.

Rank [1]	Sequence [2]	% of total reads (positive) [3]	% of total reads (negatives) [4]
1	qHRPRPHRPTKRRKSY	30.49499684	33.16868584
2	qNSKKPRRNSKVQ	7.262791153	7.831573142
3	qSARCSKNCRVRKMSH	5.167309903	4.592766886
4	qRQKRPLQSRSTKKRL	4.968968132	4.846146023
5	qKRHRRDGTPTRTKKR	2.837537617	2.025576754
6	qAKSNKAQQRPRKRKR	2.260674622	2.33980669
7	qGKKSGRGPNTKLRRK	2.191807956	1.790280523
8	SRSRKqLLSRRSPKSR	1.594270524	1.990794466
9	qKGLLKRPTKHPSRIR	1.307369397	1.276289087
10	AqSLRKLNREKKNK	0.830211474	0.945748126
11	LRHKQNCYAPVRSRRG	0.826919734	0.745792448
12	qRTMKKTKPPSRGARK	0.778150007	0.694529229
13	YKVAGRQPVYPKKPRK	0.69351764	0.765914232
14	qKTKKVTSIKSSSSRR	0.584572596	0.487388933
15	qRRKYGKRRRYI	0.570539389	0.487437477
16	RRqKNLRTHTFLNWKK	0.57019289	0.587026959
17	TKKHPLGGIEYPMRRPKTYKR	0.494309621	0.458650462
18	GRMKLQqVQVLRRWKK	0.480045414	0.549841708
19	RRHKSRKPqRTKKPPL	0.476118426	0.480519856
20	KKRTHDSPRPRHPRKW	0.449928881	0.369595183

[1] Frequency ranking of each HCDR3 candidate.
[2] The VHH candidate's raised against active phytase HCDR3 sequence. Candidates from the original NGS analysis were highlighted in pink. The clones identified originally by random colony picking were highlighted in green.
[3] The positive pool read count of the HCDR3 candidate expressed as a % of the total reads.
[4] The negative pool read count of the HCDR3 candidate expressed as a % of the total reads

Table 6.9: The top 20 frequency candidates for the denatured phytase VHH dataset.

Rank [1]	Sequence [2]	% of total reads (positive) [3]	% of total reads (negatives) [4]
1	qFSKPKSVKPLKQNKPMKRRY	8.886648761	6.197692017
2	CSqRRPSKKKCKRRAP	5.909104258	5.887812943
3	ARANKLKRWRTLSEPH	4.294279704	7.400373214
4	RKKFARAPNHNETKWR	3.694236733	4.249825427
5	AYPVCTCSKKGAPRALRqRSV	2.595446042	1.041607083
6	IKAKHLTTVRFKKVKT	2.495644974	1.946291496
7	qRFTRRTSQGNKRWKK	2.076114673	1.824539471
8	qRRMKPTKTDWPRRTT	2.062126036	1.840096982
9	RqGTQRKPPRTRHARL	2.059910524	1.46160459
10	qLKHKPTTHTRRRKRH	1.393685586	1.461494057
11	FKNIAqGLKKHKKPRK	1.294013327	1.18988701
12	RKSqIKTLTHLMRFRR	1.226800889	1.090573171
13	STRqLKTLPKKGPKKR	1.087996513	0.741123307
14	WRTRRKQMqPSLRPRR	1.03387105	0.725703964
15	qASKFKRPPKKKHVRL	0.99878353	1.232469822
16	qRERKRCSRKRCSRRS	0.918664448	0.878764441
17	MKHRRFINLDDMKPRK	0.906788276	1.130862425
18	STHPqLKKPPKRRGYM	0.88916723	0.759637574
19	RRGRLKqAPRTRSKPK	0.825535677	1.388514689
20	KIMLNqRLSKKRKSKI	0.816390252	0.602155779

[1] Frequency ranking of each HCDR3 candidate.
[2] The VHH candidate's raised against denatured phytase HCDR3 sequence. Candidates from the original NGS analysis were highlighted in blue. The clones identified originally by random colony picking were highlighted in green.
[3] The positive pool read count of the HCDR3 candidate expressed as a % of the total reads.
[4] The negative pool read count of the HCDR3 candidate expressed as a % of the total reads

Table 6.10: The top 20 frequency candidates for the phytic acid VHH dataset.

Rank [1]	Sequence [2]	% of total reads (positive) [3]	% of total reads (negatives) [4]
1	IYWWHLNSECCLSTTSPDVPN	0.502477023	0.518440233
2	SPLqFS	0.470291162	0.21734989
3	qPCFASqPRTTSQNPLRPKLV	0.458393782	0.177626446
4	AHTMTHTTGRPq	0.436756732	0.196995203
5	GTEFPYYLDRqF	0.394741298	0.558545327
6	TDqHSMSLMFERARST	0.382574204	0.390014881
7	LPqRHLQKSSHREqRR	0.361416645	0.129729324
8	FLPWTIANqSANPKSHqPRAT	0.347811077	0.279336272
9	qPIIEKSPYLSHSqGV	0.343885241	0.30538391
10	qTLLqSPILLSL	0.333576176	0.226000632
11	THAqHSGDTPNTNHAq	0.3081931	0.182969552
12	qPWTqRWENQTVSTPS	0.303877677	0.123114051
13	qYPPWMKFPNSAqMMK	0.303787772	0.22288382
14	SqQLNYqADIRM	0.299622191	0.45311441
15	CRCAKHSSGPRTTVAKDPDTV	0.299322509	0.14000208
16	PEEPLLYKLQTVqPKQ	0.298902954	0.010336364
17	HLKSqYPHNTVAPPHA	0.290362013	0.282898342
18	TIMPqCEFVCqI	0.289882522	0.028878212
19	qCLLVPALAMIQPTRT	0.269863756	0.143500542
20	qRKRPTRYTEAqAGKL	0.264529414	0.236909472

[1] Frequency ranking of each HCDR3 candidate.
[2] The VHH candidate's raised against phytic acid HCDR3 sequence. Candidates from the original NGS analysis are not present.
[3] The positive pool read count of the HCDR3 candidate expressed as a % of the total reads.
[4] The negative pool read count of the HCDR3 candidate expressed as a % of the total reads

Table 6.11: The frequencies of the top 20 VHH candidates (frequency analysis) in each active phytase, positive replicate barcode.

		Read count (% of ttl) distribution across positive barcodes (round 4 screening) for active phytase [3]											
Rank [1]	**Sequence** [2]	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**St Dev** [4]	**CV** [4]
1	qHRPRPHRPTKRRKSY	30.77	29.54	28.52	32.04	29.73	31.12	29.55	37.65	30.24	31.29	2.54	8.18
2	qNSKKPRRNSKVQ	7.50	7.40	7.15	8.14	7.15	6.63	6.52	6.93	7.10	7.11	0.46	6.41
3	qSARCSKNCRVRKMSH	5.86	6.71	4.58	4.90	4.60	5.20	3.46	5.33	4.92	6.06	0.91	17.58
4	qRQKRPLQSRSTKKRL	5.16	4.75	5.47	4.82	5.03	4.77	4.71	3.19	4.91	4.94	0.60	12.63
5	qKRHRRDGTPTRTKKR	2.54	2.65	2.35	1.94	3.33	3.55	3.19	2.54	3.49	2.95	0.54	18.75
6	qAKSNKAQQRPRKRKR	2.26	3.28	2.45	2.43	2.12	1.80	1.92	2.41	1.95	2.29	0.42	18.15
7	qGKKSGRGPNTKLRRK	2.07	2.24	2.31	1.90	2.54	2.30	2.35	2.20	2.18	2.10	0.18	7.90
8	SRSRKqLLSRRSPKSR	1.74	1.26	1.45	2.00	1.53	1.55	1.49	1.64	1.55	1.47	0.20	12.57
9	qKGLLKRPTKHPSRIR	1.48	1.36	1.43	1.30	1.19	1.17	1.51	1.31	1.18	1.30	0.12	9.08
10	AqSLRKLNREKKNK	0.92	0.58	0.93	0.87	0.72	0.76	0.85	0.50	0.85	0.87	0.14	18.49
11	LRHKQNCYAPVRSRRG	0.67	0.60	0.86	0.92	0.82	0.75	0.81	0.69	1.03	0.74	0.13	16.19
12	qRTMKKTKPPSRGARK	0.91	0.71	0.76	0.73	0.75	0.75	0.70	0.75	0.81	0.81	0.06	7.95
13	YKVAGRQPVYPKKPRK	0.58	0.53	0.69	0.91	0.67	0.70	0.52	0.66	0.69	0.69	0.11	16.51
14	qKTKKVTSIKSSSSRR	0.61	0.61	0.64	0.52	0.59	0.61	0.57	0.40	0.54	0.62	0.07	12.15
15	qRRKYGKRRRYI	0.71	0.50	0.63	0.47	0.54	0.50	0.58	0.58	0.61	0.56	0.07	12.47
16	RRqKNLRTHTFLNWKK	0.50	0.44	0.61	0.62	0.60	0.57	0.54	0.54	0.64	0.48	0.07	12.00
17	TKKHPLGGIEYPMRRPKTYKR	0.40	0.47	0.47	0.41	0.58	0.55	0.57	0.50	0.55	0.51	0.06	12.31
18	GRMKLQqVQVLRRWKK	0.44	0.33	0.43	0.63	0.54	0.44	0.50	0.37	0.48	0.44	0.08	18.32
19	RRHKSRKPqRTKKPPL	0.45	0.70	0.37	0.50	0.53	0.34	0.37	0.40	0.51	0.53	0.11	22.74
20	KKRTHDSPRPRHPRKW	0.37	0.36	0.56	0.32	0.51	0.60	0.41	0.49	0.49	0.38	0.09	20.67

[1] Frequency ranking of each HCDR3 candidate.
[2] The active phytase VHH candidate's HCDR3 sequence. Candidates from the original NGS analysis and random colony picks were highlighted.
[3] The positive replicates read count for the HCDR3 candidate expressed as a % of the total reads, for round 4 screening.
[4] The standard deviation (SD) and coefficient of variation (CV) for the read count % as a group.

Table 6.12: The frequencies of the top 20 VHH candidates (frequency analysis) in each denatured phytase, positive replicate barcode.

		Read count distribution across positive barcodes (Round 4 screening) for denatured phytase [3]											
Rank [1]	**Sequence** [2]	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**	**St Dev** [4]	**CV** [4]
1	qFSKPKSVKPLKQNKPMKRRY	7.38	11.49	10.08	10.67	11.86	8.35	8.70	9.01	7.73	8.78	1.55	16.43
2	CSqRRPSKKKCKRRAP	5.23	7.75	5.65	5.93	6.84	6.20	6.22	6.29	5.71	5.38	0.74	12.15
3	ARANKLKRWRTLSEPH	5.50	7.81	4.30	4.31	3.73	4.01	4.10	4.40	3.92	3.90	1.23	26.79
4	RKKFARAPNHNETKWR	3.25	6.17	3.23	4.86	4.05	4.27	3.60	3.56	3.43	3.69	0.91	22.65
5	AYPVCTCSKKGAPRALRqRSV	2.16	3.27	2.55	2.79	3.01	2.83	2.80	2.54	2.95	2.20	0.35	12.98
6	IKAKHLTTVRFKKVKT	2.47	3.19	2.52	2.72	2.63	2.33	2.73	2.72	2.30	2.34	0.27	10.32
7	qRFTRRTSQGNKRWKK	2.10	3.44	2.11	2.09	2.18	2.03	2.20	2.03	1.79	1.96	0.46	20.78
8	qRRMKPTKTDWPRRTT	1.92	3.11	1.96	1.55	1.95	2.10	1.92	1.95	2.90	1.88	0.49	22.98
9	RqGTQRKPPRTRHARL	2.64	2.77	1.80	1.76	1.87	1.98	1.94	1.99	2.56	2.00	0.38	17.65
10	qLKHKPTTHTRRRKRH	1.76	1.93	1.33	1.16	1.29	1.25	1.17	1.24	1.24	2.10	0.35	23.96
11	FKNIAqGLKKHKKPRK	1.64	1.99	1.20	1.08	1.25	1.06	1.30	1.11	1.57	1.23	0.30	22.28
12	RKSqIKTLTHLMRFRR	1.19	1.68	1.17	1.25	1.29	1.13	1.32	1.22	1.04	1.42	0.18	14.16
13	STRqLKTLPKKGPKKR	1.26	2.01	1.03	1.10	0.94	1.15	1.12	0.98	1.15	0.93	0.31	26.84
14	WRTRRKQMqPSLRPRR	0.89	1.45	1.05	1.02	0.99	1.09	1.17	1.12	0.93	1.01	0.16	14.71
15	qASKFKRPPKKKHVRL	0.88	1.40	1.00	0.99	1.02	1.02	1.08	1.05	0.93	1.00	0.14	13.48
16	qRERKRCSRKRCSRRS	0.73	1.24	0.96	1.03	0.69	0.94	1.02	1.20	0.91	0.82	0.18	18.73
17	MKHRRFINLDDMKPRK	0.89	1.38	0.96	0.83	1.04	0.85	0.79	0.83	0.92	1.03	0.17	18.23
18	STHPqLKKPPKRRGYM	0.79	1.75	0.80	0.82	0.81	0.77	0.80	0.73	1.35	0.82	0.33	35.39
19	RRGRLKqAPRTRSKPK	0.94	1.18	1.09	0.83	0.61	0.75	0.89	0.81	0.69	0.95	0.18	20.16
20	KIMLNqRLSKKRKSKI	0.87	1.22	0.82	0.72	0.81	0.72	0.86	0.89	0.73	0.78	0.15	17.41

[1] Frequency ranking of each HCDR3 candidate.
[2] The denatured phytase VHH candidate's HCDR3 sequence. Candidates from the original NGS analysis and random colony picks were highlighted.
[3] The positive replicates read count for the HCDR3 candidate expressed as a % of the total reads, for round 4 screening.
[4] The standard deviation (SD) and coefficient of variation (CV) for the read count % as a group.

Table 6.13: The frequencies of the top 20 VHH candidates (frequency analysis) in each phytic acid, positive replicate barcode.

		Read count distribution across positive barcodes (Round 4 screening) for phytic acid [3]													
Rank [1]	**Sequence** [2]	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**St Dev** [4]	**CV** [4]
1	IYWWHLNSECCLSTTSPDVPN	1.1	0.1	0.2	1.2	0.0	0.8	0.7	0.4	0.4	0.4	0.4	0.5	0.377957	71.3666
2	SPLqFS	1.0	0.1	0.5	1.3	0.5	0.4	0.5	0.0	0.6	0.1	0.9	0.2	0.401777	77.74356
3	qPCFASqPRTTSQNPLRPKLV	0.2	0.7	0.6	0.2	0.5	0.2	1.0	0.4	0.0	0.5	0.9	0.2	0.295988	65.05599
4	AHTMTHTTGRPq	0.0	0.7	1.3	0.6	0.2	0.0	0.0	0.1	0.2	0.5	0.5	1.2	0.443495	97.30145
5	GTEFPYYLDRqF	0.6	0.2	0.4	0.5	0.1	0.0	0.7	0.5	0.5	0.7	0.4	0.0	0.260635	64.76943
6	TDqHSMSLMFERARST	0.0	0.5	0.2	0.4	0.3	0.5	0.7	0.5	0.5	0.4	0.2	0.0	0.227124	65.52672
7	LPqRHLQKSSHREqRR	0.6	0.3	1.5	0.0	0.0	0.4	0.0	0.1	0.7	0.2	0.2	0.5	0.43877	109.9622
8	FLPWTIANqSANPKSHqPRAT	0.6	0.0	0.3	0.2	0.2	0.0	0.1	1.6	0.1	0.3	0.6	0.7	0.451734	106.678
9	qPIIEKSPYLSHSqGV	0.5	0.3	0.3	0.3	0.7	0.7	0.2	0.0	0.3	0.2	0.4	0.3	0.20094	58.19463
10	qTLLqSPILLSL	0.6	0.5	0.5	0.0	0.0	0.3	0.0	0.7	0.7	0.1	0.0	0.9	0.329532	89.4852
11	THAqHSGDTPNTNHAq	0.1	0.3	0.0	0.4	0.2	0.0	1.8	0.6	0.0	0.0	0.0	0.0	0.503786	162.6481
12	qPWTqRWENQTVSTPS	0.7	0.0	0.4	0.0	0.1	0.0	0.9	0.0	0.7	0.4	0.2	0.6	0.314342	95.29888
13	qYPPWMKFPNSAqMMK	0.0	1.1	0.0	0.2	0.0	0.3	0.3	0.5	0.5	0.1	0.2	0.2	0.294062	100.0526
14	SqQLNYqADIRM	0.0	0.0	0.1	0.6	0.0	0.3	0.3	0.0	0.2	0.3	0.0	1.6	0.448298	144.7065
15	CRCAKHSSGPRTTVAKDPDTV	1.7	0.1	0.1	0.3	0.4	0.2	0.2	0.2	0.2	0.7	0.1	0.2	0.442147	114.9053
16	PEEPLLYKLQTVqPKQ	0.5	0.0	0.5	0.3	0.0	0.0	0.1	0.0	0.0	0.5	0.0	1.9	0.556799	155.6059
17	HLKSqYPHNTVAPPHA	0.0	0.2	1.1	0.4	0.2	0.0	0.7	0.4	0.0	0.0	0.0	0.2	0.348195	117.3145
18	TIMPqCEFVCqI	0.0	0.1	0.6	0.0	0.0	0.8	0.4	0.9	0.0	0.3	0.0	0.1	0.342861	119.382
19	qCLLVPALAMIQPTRT	0.0	1.3	0.0	0.0	0.7	0.0	0.3	0.0	0.1	0.1	0.3	0.2	0.400944	147.0688
20	qRKRPTRYTEAqAGKL	0.3	0.8	0.0	0.0	0.0	0.5	0.5	0.0	0.3	0.4	0.0	0.0	0.273018	107.5187

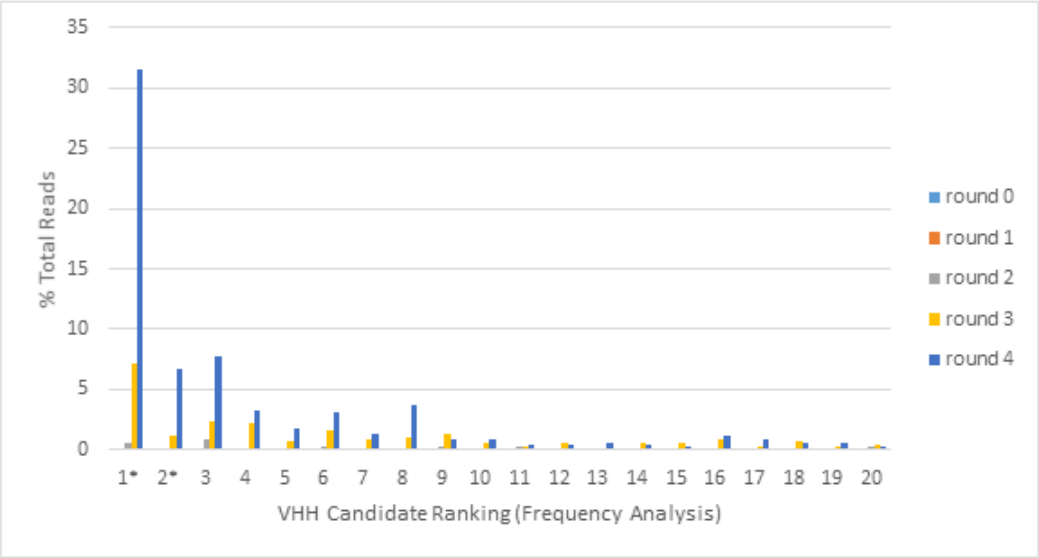
[1] Frequency ranking of each HCDR3 candidate.
[2] The phytic acid VHH candidate's HCDR3 sequence. Candidates from the original NGS analysis and random colony picks were highlighted.
[3] The positive replicates read count for the HCDR3 candidate expressed as a % of the total reads, for round 4 screening.
[4] The standard deviation (SD) and coefficient of variation (CV) for the read count % as a group.

Three of the eleven, original active phytase VHH candidates suggested by the Z score analysis were identified within the top 20 for the frequency analysis. Two of denatured phytase original candidates were also identified, but none of phytic acids original candidates were present in the top 20 for the frequency analysis. As expected, the clones identified by colony picking ("qHRPRPHRPTKRRKSY", "qNSKKPRRNSKVQ" and "RKKFARAPNHNETKWR") were found in the list of top 20 frequency candidates for their respective targets. However, the fourth clone identified by colony picking "RTTKSHQHSSRKRFKH" against denatured phytase was not seen in the top 20, and instead ranked #703.

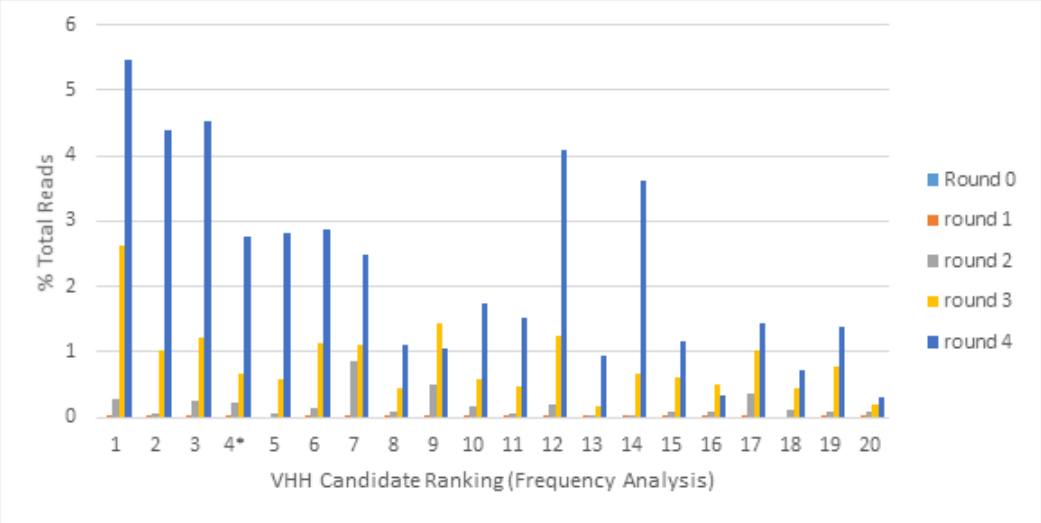
The majority of the candidates identified against phytase conformations had a consistent distribution of frequencies within their positive barcodes, with coefficient of variation (CV) values across 10 replicates usually <20%. The high frequencies and consistently low CV values (for phytases) suggest persistent binders however the frequencies for the phytic acid candidates were very low (0.5%) and highly variable (>60% CV values) which may suggest that the selection experiment was not optimal.

The study by Miyazaki et al. (2015) also considered the candidates frequency (%) within the original library as well as after panning. The authors only selected the sequences whose frequencies increased after panning as an indication of functionality and binding specificity. Following this method, the top 20 frequency candidates read counts isolated in the present study were checked within the libraries (12mer, 16mer, 21mer), and within the panning round glycerol samples against their target antigen. Their read counts were then expressed as percentages of the total read count within their individual panning glycerol samples and plotted for comparison (figures 6.2). The phytase candidates (active and denatured) mostly demonstrated a clear enrichment process through the panning rounds however they were all undetectable within their original VHH libraries. The majority of the phytic acid candidates seemed to show round 3 as the maximal level of enrichment, with a considerable reduction in representation in round 4. Only the rank 8 candidate 'FLPWTIANqSANPKSHqPRAT' was detected within its original VHH library and had a relatively high copy number of 7.

A


35
30
25
20
15
10
5
0
% Total Reads
1* 2* 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
VHH Candidate Ranking (Frequency Analysis)
round 0
round 1
round 2
round 3
round 4

B


6
5
4
3
2
1
0
% Total Reads
1 2 3 4* 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
VHH Candidate Ranking (Frequency Analysis)
Round 0
round 1
round 2
round 3
round 4

C

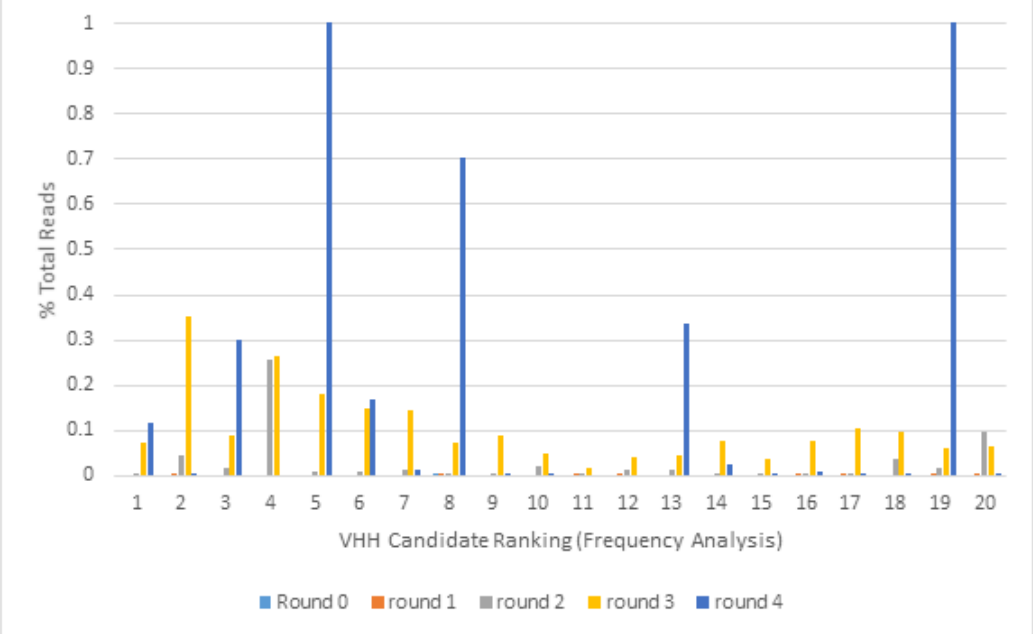

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
% Total Reads
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
VHH Candidate Ranking (Frequency Analysis)
Round 0
round 1
round 2
round 3
round 4

Figure 6.2: The frequency of the top 20 active and denatured phytase and phytic acid candidates during panning (frequency analysis). The read counts for the top 20 candidates (frequency analysis) for active (A) and denatured phytase (B) and phytic acid (C) were expressed as percentages of the total reads within each panning round. Round 0 referred to the relevant VHH library (12mer, 16mer or 21mer). The candidates were displayed by their rank number based on the frequency analysis. * denotes VHH candidates identified by random colony picks that are confirmed binders.

6.5 Growth and selection patterns

The next analysis aimed to assess whether positive clones could be identified based on completely novel criterion, their frequency during growth and selection phases of panning. Firstly, 7 of the top VHH candidates were chosen based on the criteria described in table 6.14.

Table 6.14: The top 20 frequency candidates for the denatured phytase VHH dataset.

Identifier [1]	VHH HCDR3 [2]	Target [3]	Top candidate selection criteria [4]
1	qHRPRPHRPTKRRKSY	Active	Colony Picked - positive phage ELISA, Frequency Analysis
2	qNSKKPRRNSKVQ	Active	Colony Picked - positive phage ELISA, Frequency Analysis
3	RKKFARAPNHNETKWR	Denatured	Colony Picked - positive phage ELISA
4	RTTKSHqHSSRKRFKH	Denatured	Colony Picked - positive phage ELISA
5	qKTKKVTSIKSSSSRR	Active	9 of 10 (Z≥2), Acid eluted 5 of 10 (Z≥2), Frequency Analysis
6	qSARCSKNCRVRKMSH	Active	10 of 10 (Z≥10), MEME, Frequency Analysis
7	AYPVCTCSKKGAPRALRqRSV	Denatured	10 of 10 (Z≥10), MEME, Frequency Analysis

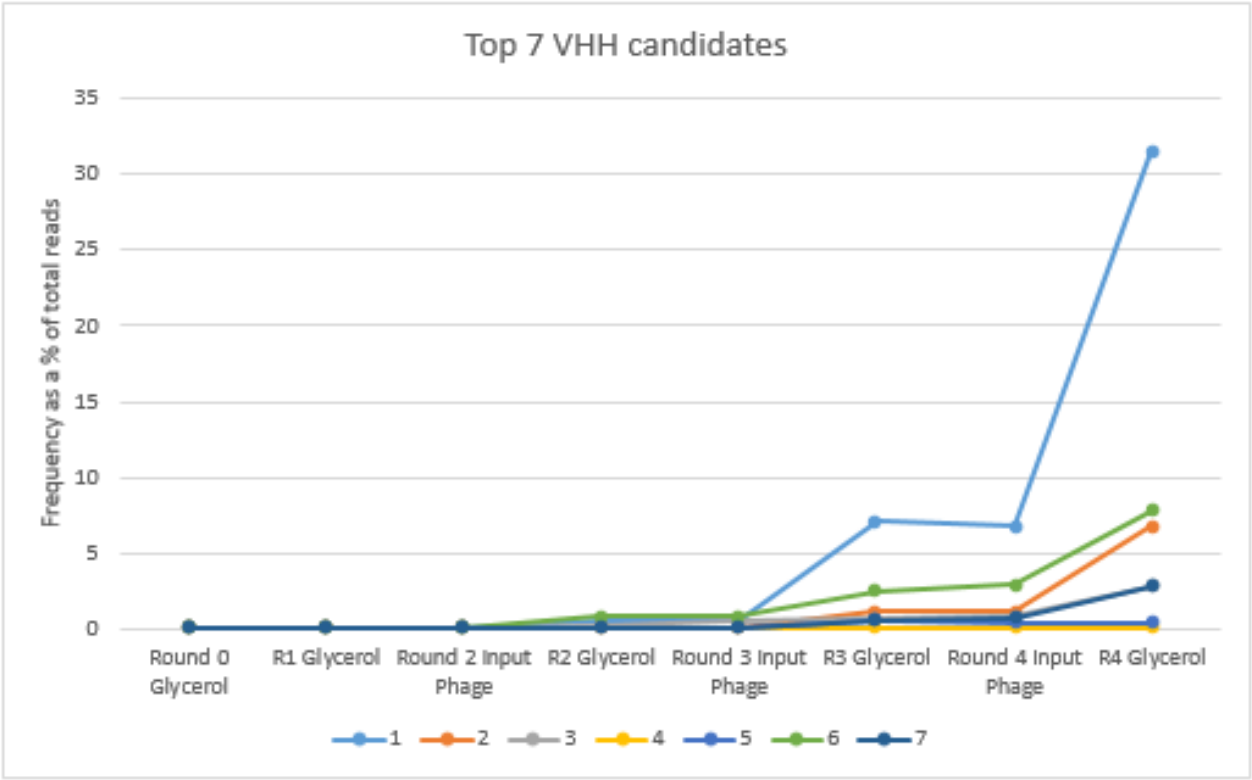
[1] Identifier number assigned to each VHH HCDR3 sequence for the purpose of this analysis only.
[2] The VHH candidate's HCDR3 translated amino acid sequence.
[3] The target antigen that the VHH candidate was isolated against.
[4] Selection reasoning for each VHH candidate.

The VHH candidates (I.D 1—4) were selected as they were isolated by random colony picking and demonstrated consistent binding activity via multiple phage ELISAs. The VHH candidates (I.D 5—7) were selected because they were ranked amongst the top 20 candidates for the Z score and frequency analyses. In addition, I.D #6 & 7 were also selected as they were identified as some of the most common motifs by the 'MEME' analysis method. In the case of the clone qKTKKVTSIKSSSSRR (I.D #5), the HCDR3 sequence was also ranked within the top 20 (rank #14) candidates for the 2-

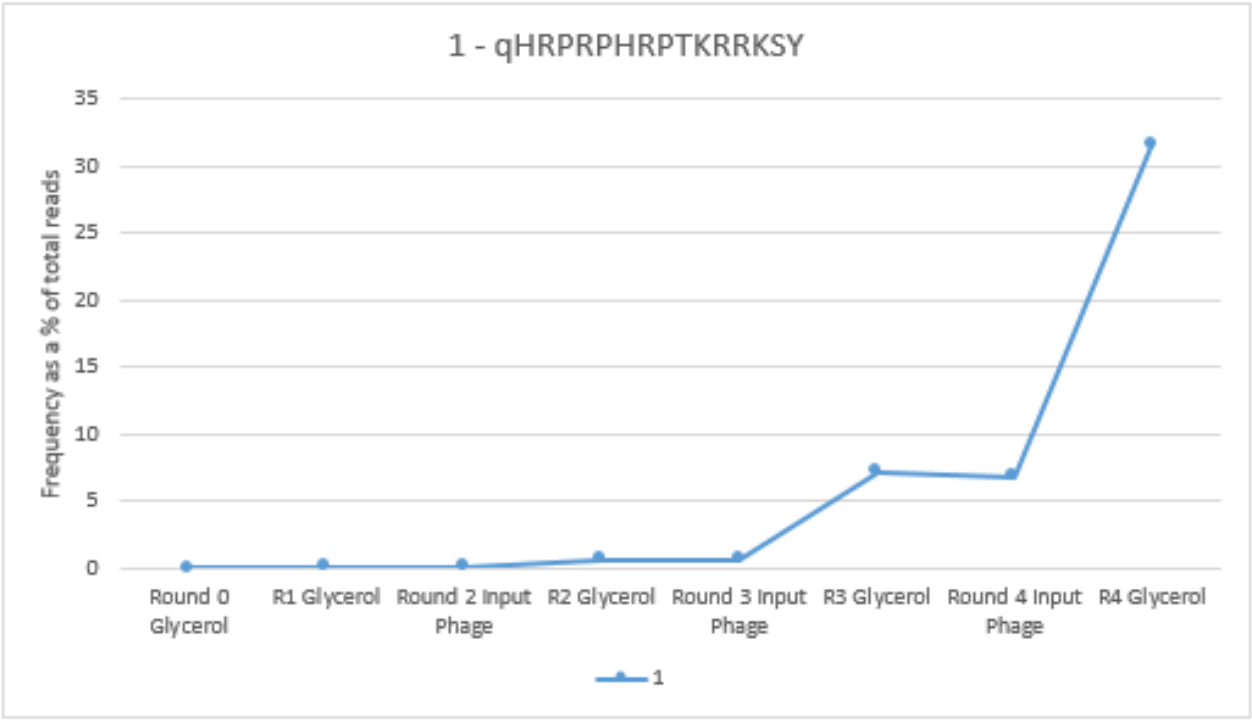
panning round, acid eluted dataset for active phytase. The analysis determined the HCDR3 frequency, as a percentage of the total reads within the relevant barcode, for both the panning round bacterial glycerol stock and phage input sequencing files. Transition from panning round glycerol samples to input phage phases represented growth stages only, whereas the transition between input phage to the next set of panning round bacterial glycerol represented both selection and growth stages. The frequencies for each of the 7 VHH candidates described in table 6.14 were plotted (Figure 6.3).

For qHRPRPHRPTKRRKSY (I.D #1), the growth and selection phase between round 4 input phage and round 4 glycerol demonstrated the largest difference in total frequency of 24.78%, followed by the growth and selection phase between round 3 input phage and round 3 glycerol with a frequency difference of 6.59%. Interestingly, the growth only phase between round 3 glycerol and round 4 input phage saw a reduction in total frequency whereas there was only a small increase (0—0.08%), if any, difference within the other growth only phases. The candidates qNSKKPRRNSKVQI, qSARCSKNCRVRKMSH and AYPVCTCSKKGAPRALRqRSV (I.D 2, 6 and 7 respectively) demonstrated similar trends with total frequency increases occurring mostly between input phage to glycerol phases (0.01-4.84%). Candidate RKKFARAPNHNETKWR (I.D #3) demonstrated a consistent increase in total frequency throughout the growth and selection phases with the largest increase during the 4 input phage and round 4 glycerol (selection+growth phase). Given that clones 1—4 are validated binders, this preferential enrichment during the selection phages could indicate true binders. Candidate qKTKKVTSIKSSSSRR's (I.D #5) total frequency also increased during growth/selection phases but decreased during growth phases which may indicate some toxicity but it being a true binder. Candidate RTTKSHqHSSRKRFKH (I.D #4) accounted for very few (0—0.02%) of the total reads throughout the panning experiment.

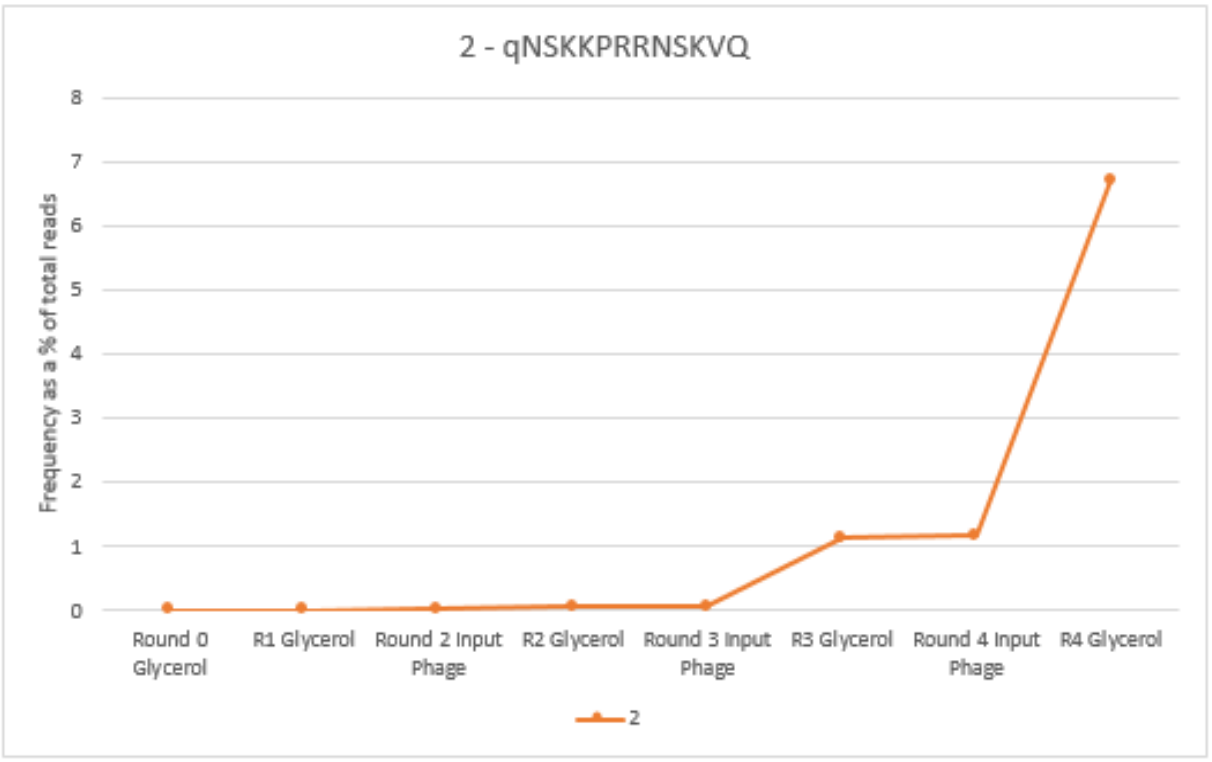
A


Top 7 VHH candidates
Frequency as a % of total reads
35
30
25
20
15
10
5
0
Round 0 Glycerol
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
Round 4 Input Phage
R4 Glycerol
1
2
3
4
5
6
7

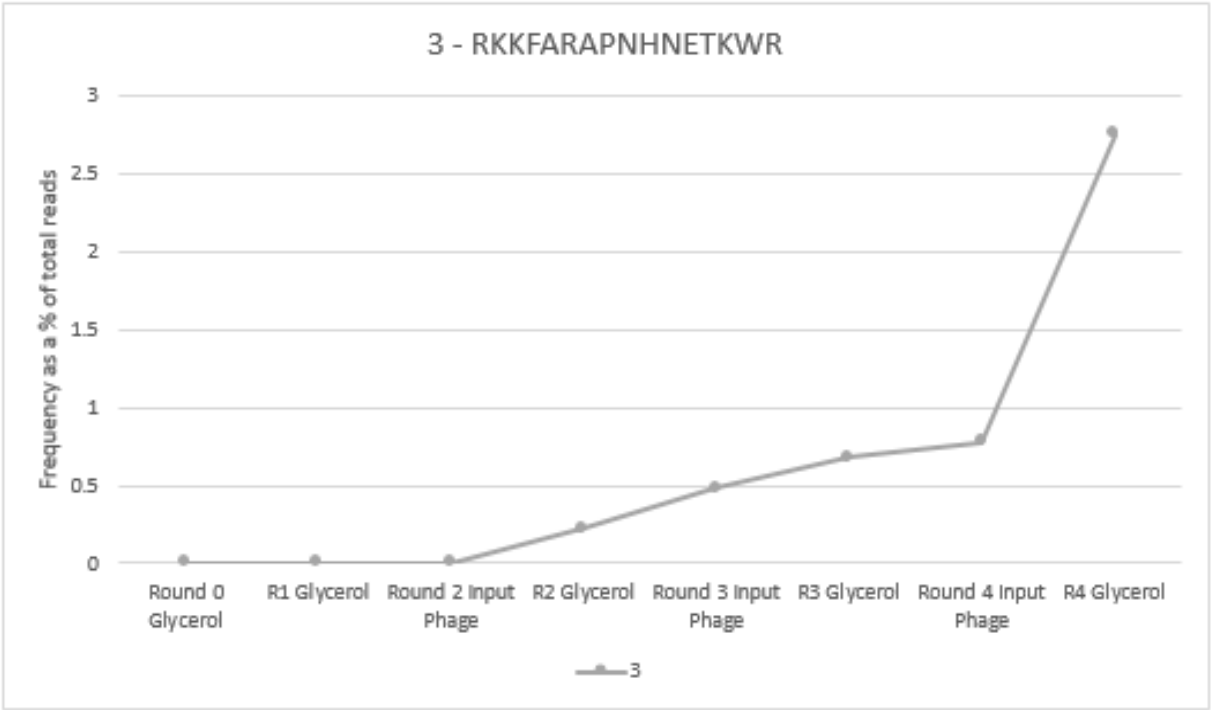
B


1 - qHRPRPHRPTKRRKSY
Frequency as a % of total reads
35
30
25
20
15
10
5
0
Round 0 Glycerol
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
Round 4 Input Phage
R4 Glycerol
1

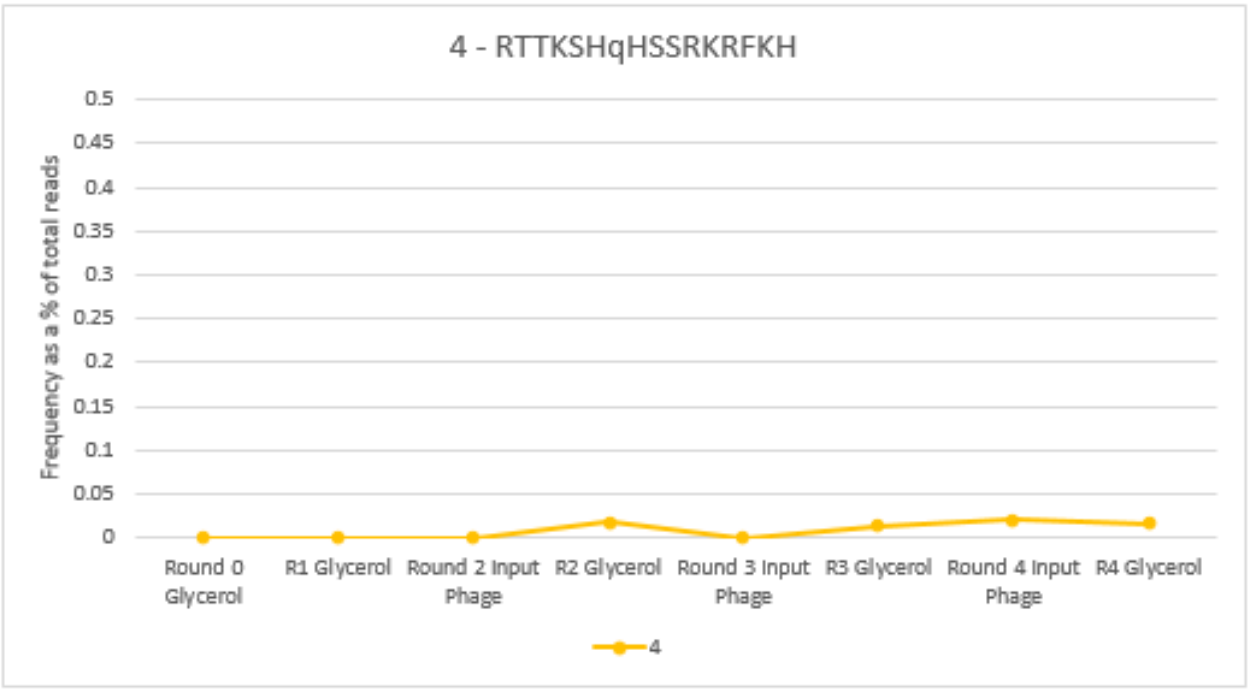
C


2 - qNSKKPRRNSKVQ
Frequency as a % of total reads
8
7
6
5
4
3
2
1
0
Round 0 Glycerol
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
Round 4 Input Phage
R4 Glycerol
2

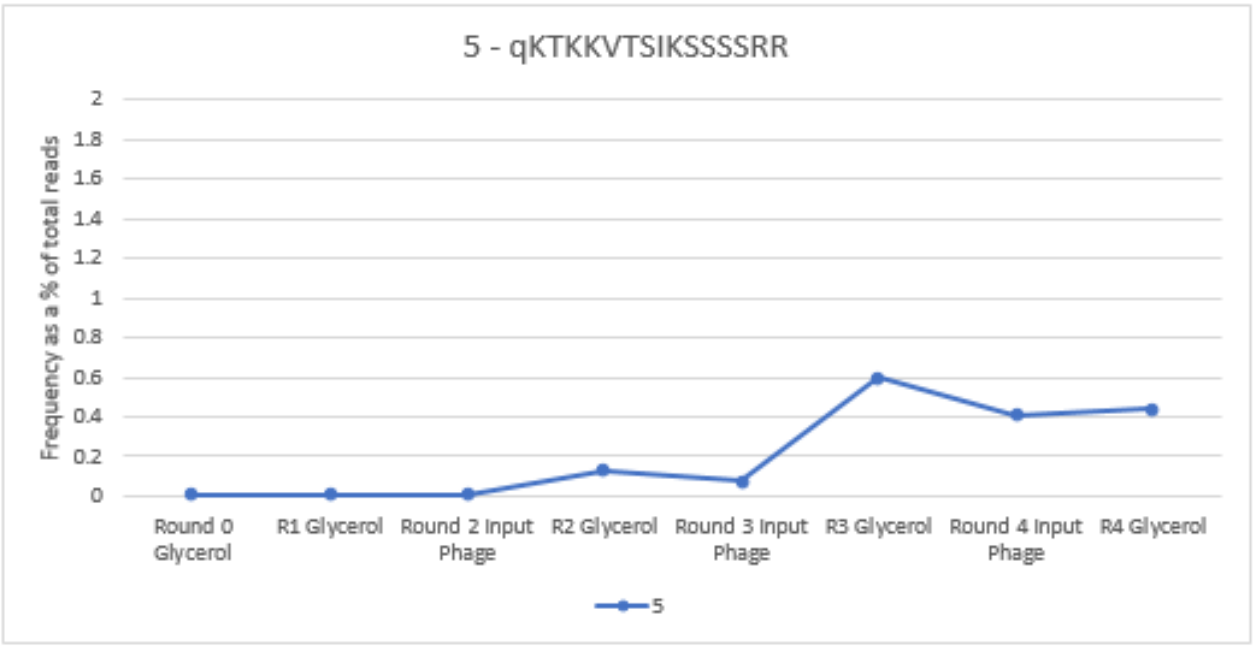
D


3 - RKKFARAPNHNETKWR
Frequency as a % of total reads
3
2.5
2
1.5
1
0.5
0
Round 0 Glycerol
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
Round 4 Input Phage
R4 Glycerol
3

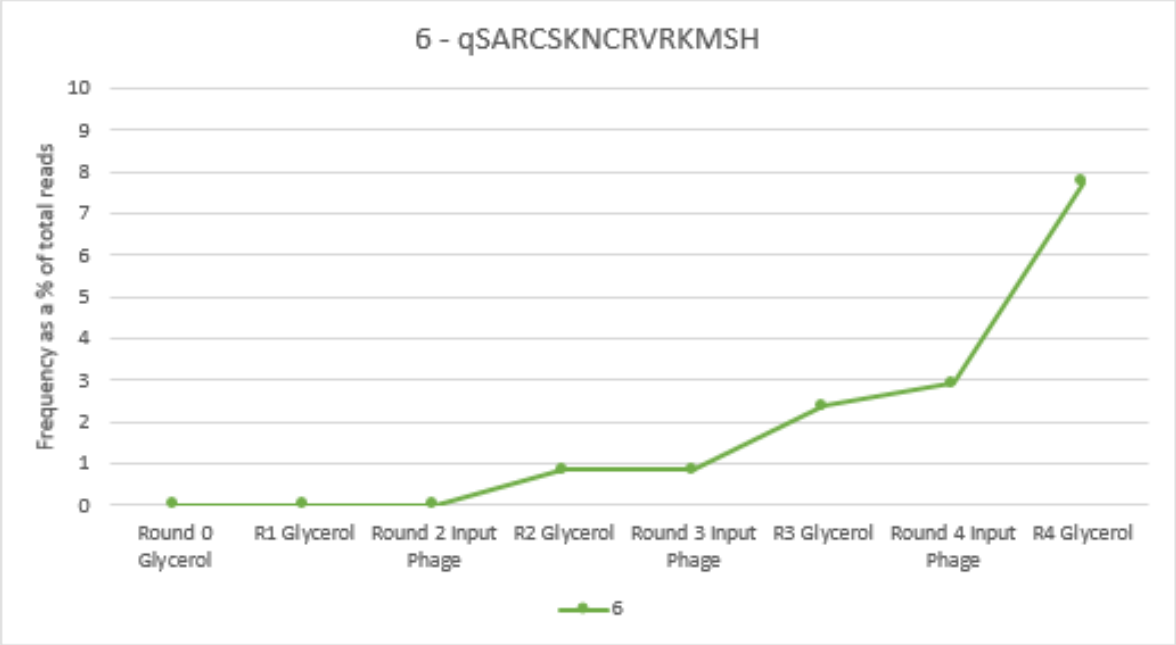
E


4 - RTTKSHqHSSRKRFKH
Frequency as a % of total reads
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Round 0 Glycerol
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
Round 4 Input Phage
R4 Glycerol
4

F


5 - qKTKKVTSIKSSSSRR
Frequency as a % of total reads
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Round 0 Glycerol
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
Round 4 Input Phage
R4 Glycerol
5

G



H

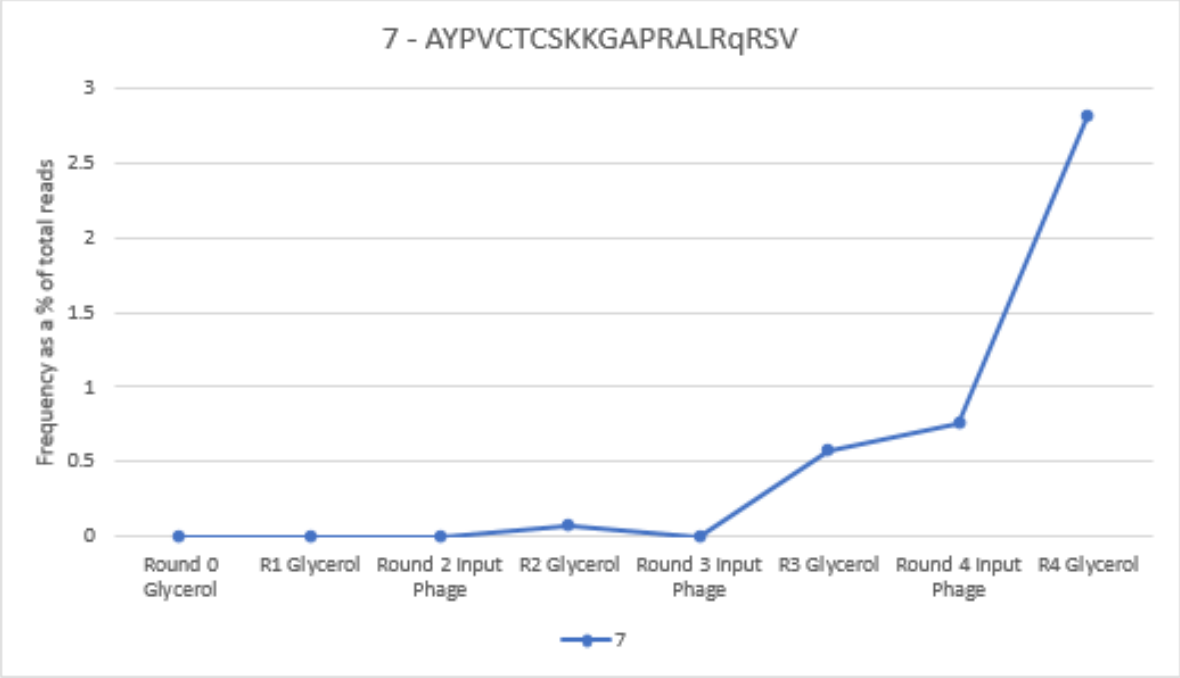


Figure 6.3: Growth and selection phase frequencies for top 7 VHH candidates. The read counts for the top 7 (I.Ds 1—7) VHH candidates (colony picks and bioinformatics analysis (Z score, MEME motif, frequency) for the phytases were expressed as percentages of the total reads within each panning round or phage input barcode file.

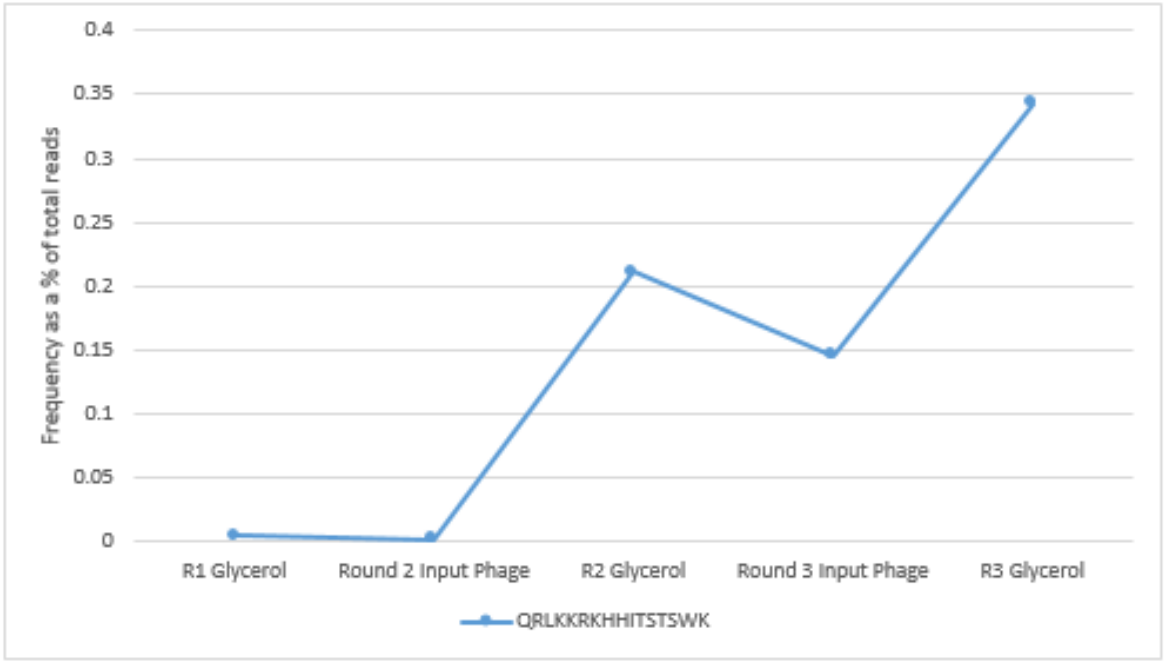
To further test whether true binders could be identified by their high enrichment during selection phases of the panning, the same growth and selection phase analysis was applied to the three known VHH elastin binders identified from the same VHH library (QRLKKRKHHITSTSWK, SKERKRNKHPKHI and TKPTKPRSPPKKTRLH) (Dr Alison Gray, personal communication) (Figure 6.4). Their binding activity was based on phage ELISAs only (data not shown).

The three anti-elastin binders all demonstrated high enrichment (an increase in total frequency) during the growth and selection phases, and TKPTKPRSPPKKTRLH between growth phases also. Similarly, to candidate #5 (figure 6.9), the anti-Elastin clone QRLKKRKHHITSTSWK decreased in frequency during growth phases which may indicate some toxicity. Overall, the data indicate that high enrichment during selection phases of panning may allow the identification of true binders in NGS datasets.

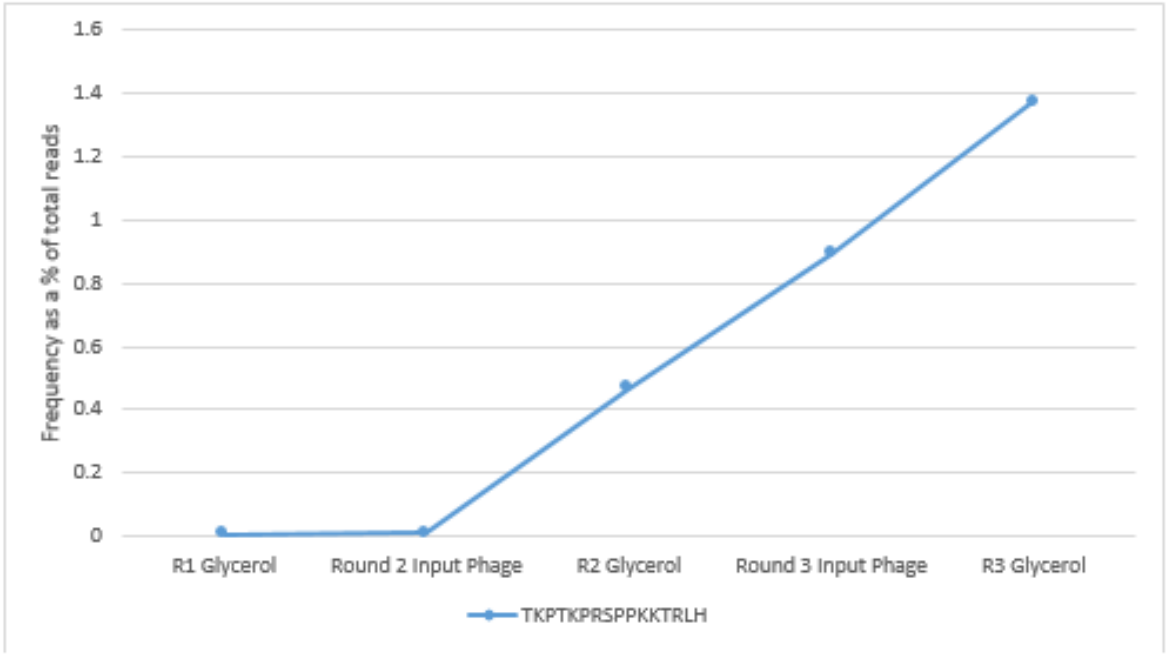
The original panning experiments against elastin were subjected to a frequency analysis to identify the most enriched clones after 3 panning rounds. The growth and selection phase analysis was then applied to the top 10 clones (Figure 6.5).

NGS data from the top 10 anti-elastin candidates (frequency analysis) demonstrated similar trends where most clones became enriched during the growth and selection phase (1, 2, 4—10). Interestingly, the clone KMPqTPNKKPHRKWKR (rank 3) had a sudden drop in frequency (-1.41%) after round 3 selection, potentially being out competed by better suited clones such as SqPFGPPMYYKL (rank 1). Alternatively, this may suggest that the clone KMPqTPNKKPHRKWKR (rank 3) was not a true binder and was out competed by those that were.

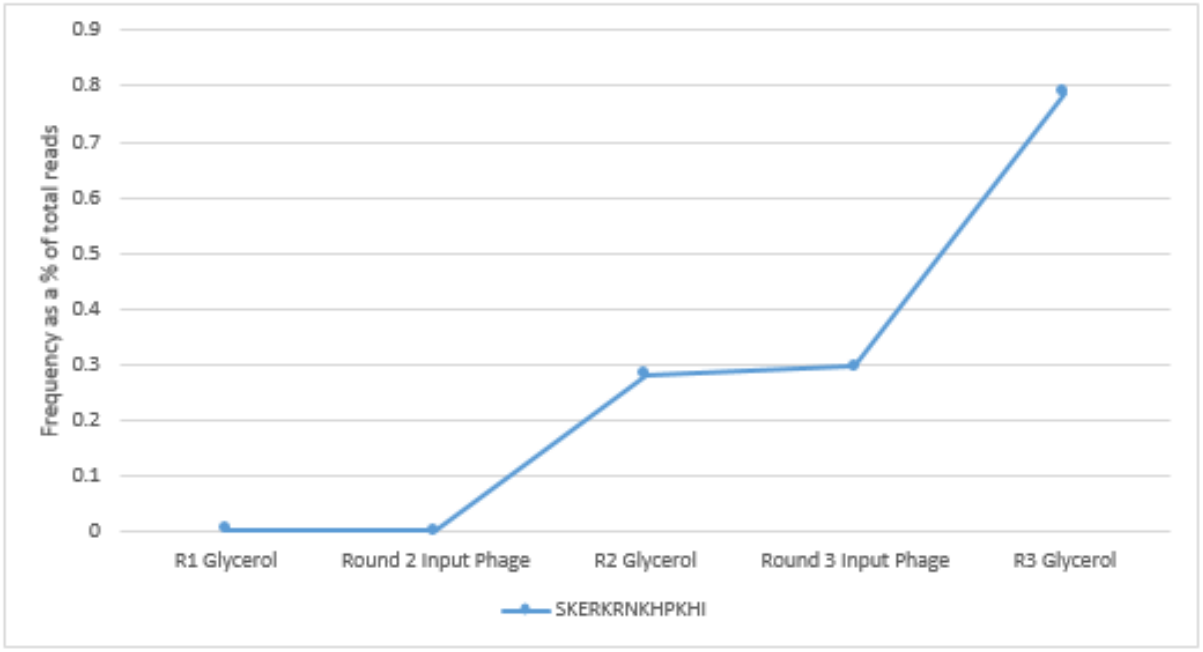
A


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
QRLKKRKHHITSTSWK

B


1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
TKPTKPRSPPKKTRLH

C



D

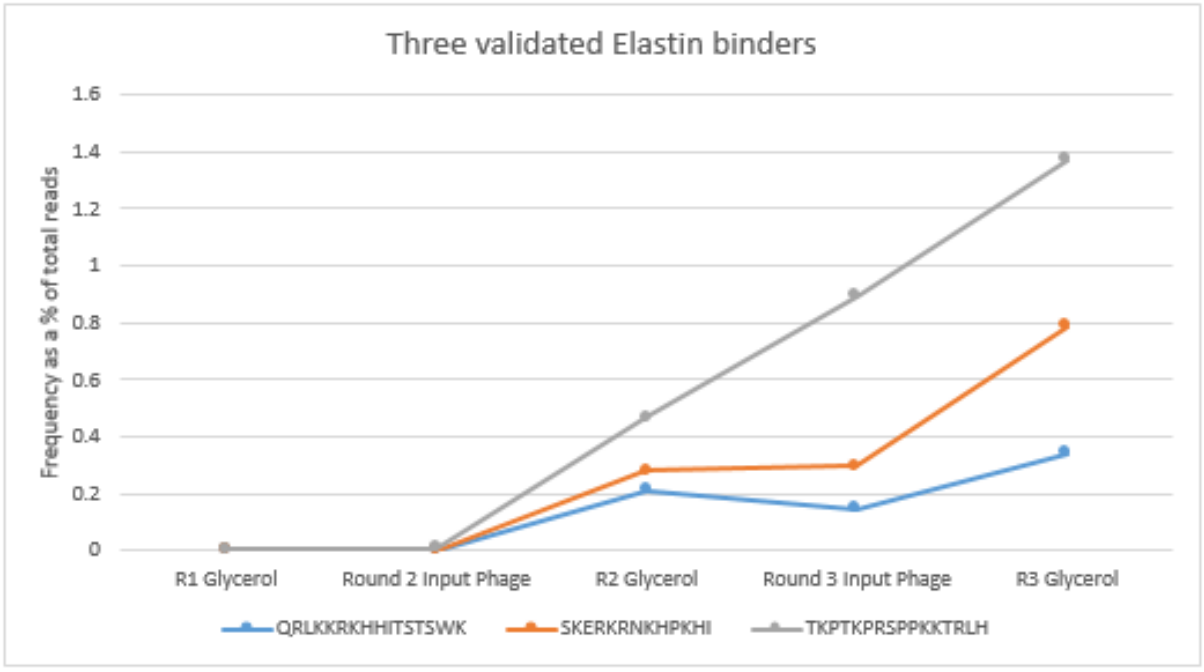
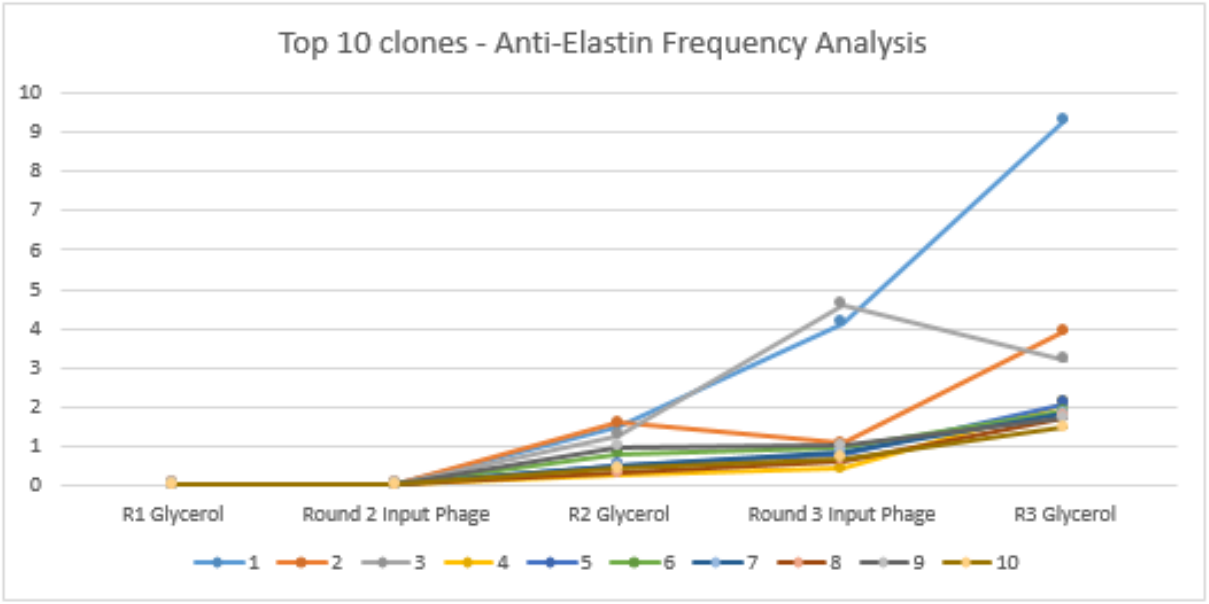
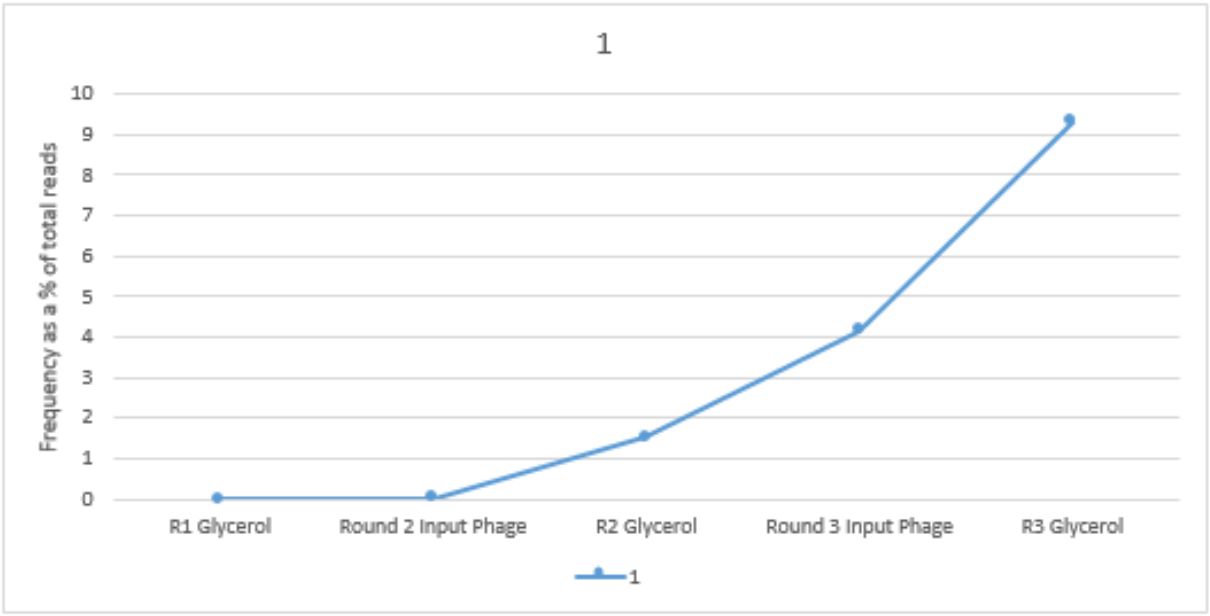


Figure 6.4: Growth and selection phase frequencies for 3 known elastin binders. The read counts for the three known Elastin binders were expressed as percentages of the total reads within each panning round or phage input barcode file.

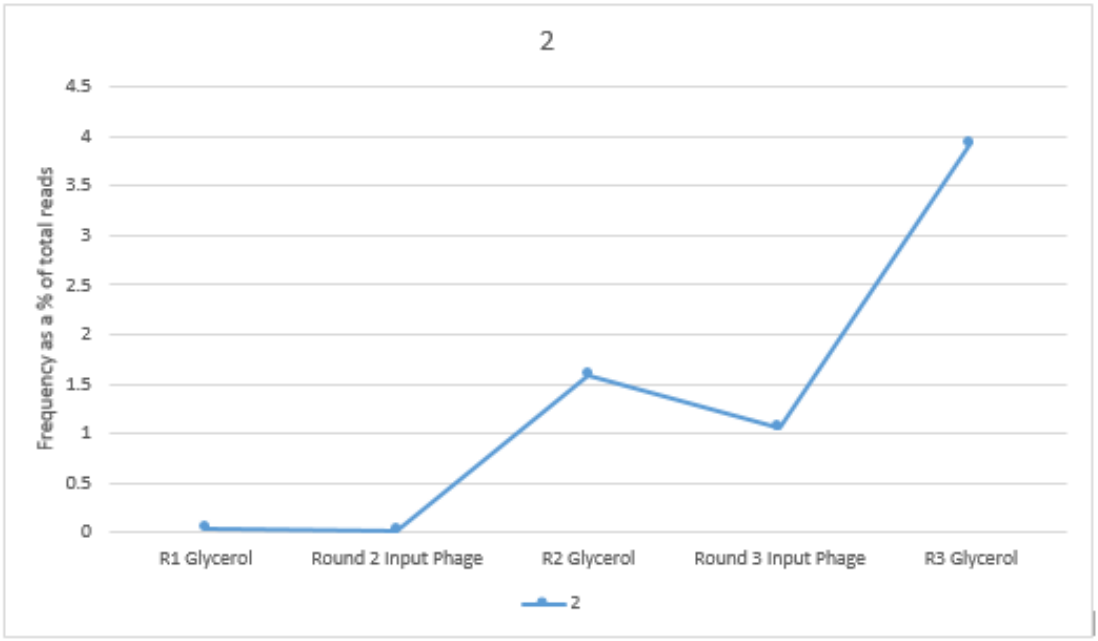
A


Top 10 clones - Anti-Elastin Frequency Analysis
10
9
8
7
6
5
4
3
2
1
0
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
1
2
3
4
5
6
7
8
9
10

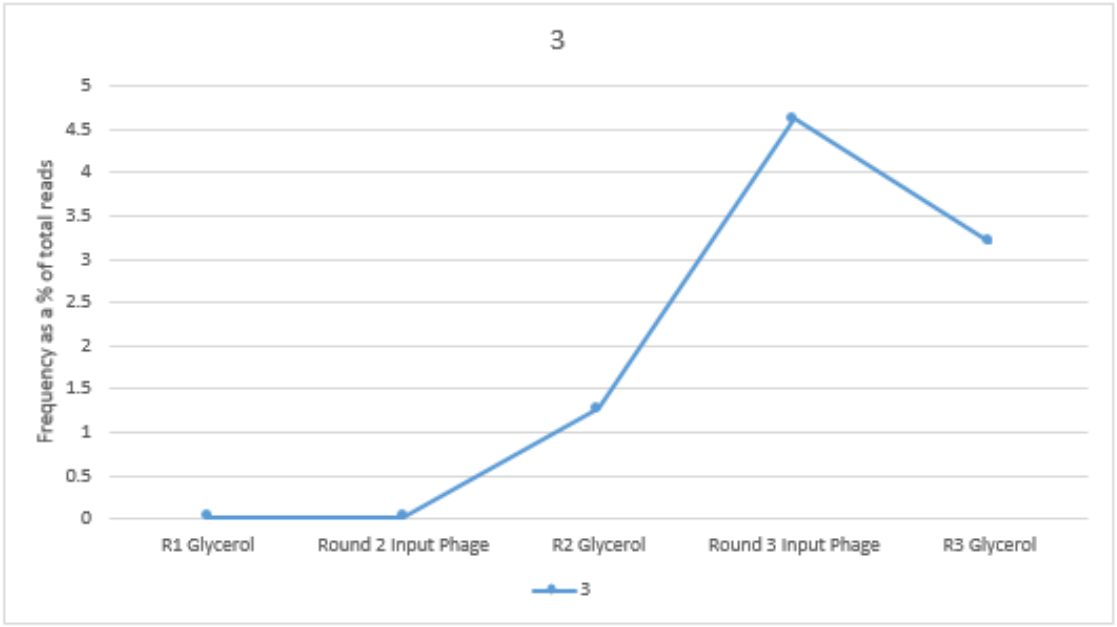
B


1
Frequency as a % of total reads
10
9
8
7
6
5
4
3
2
1
0
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
1

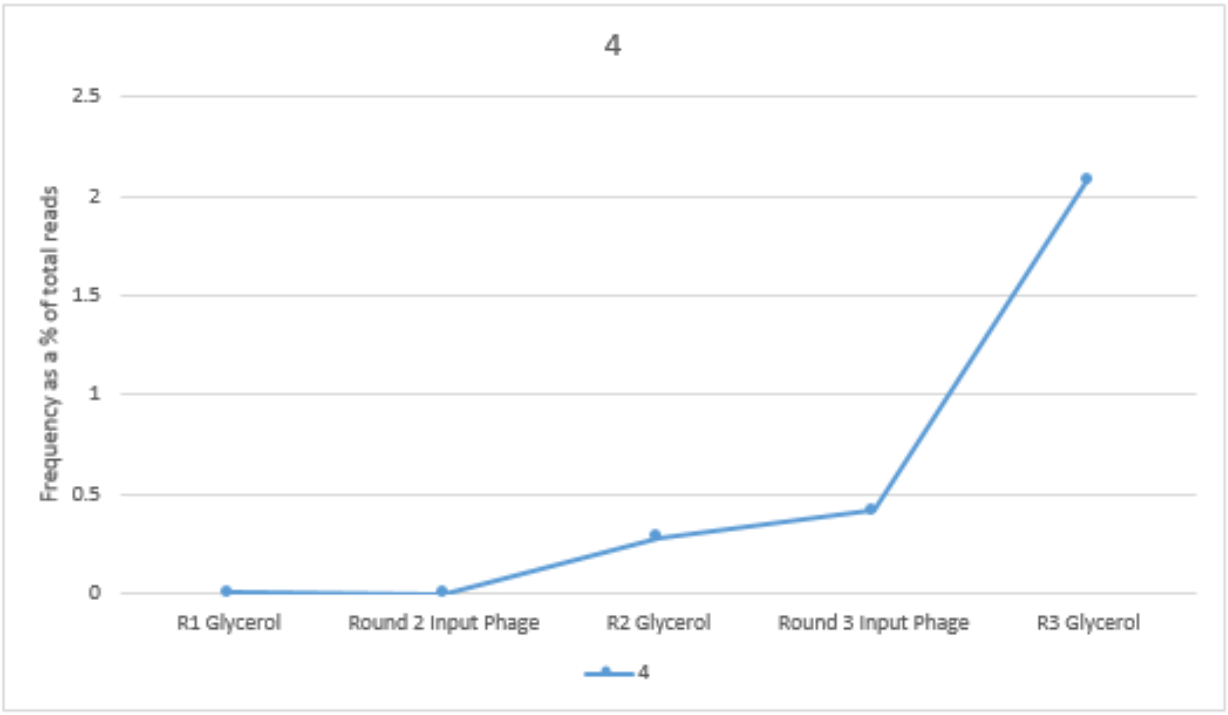
C


2
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
2

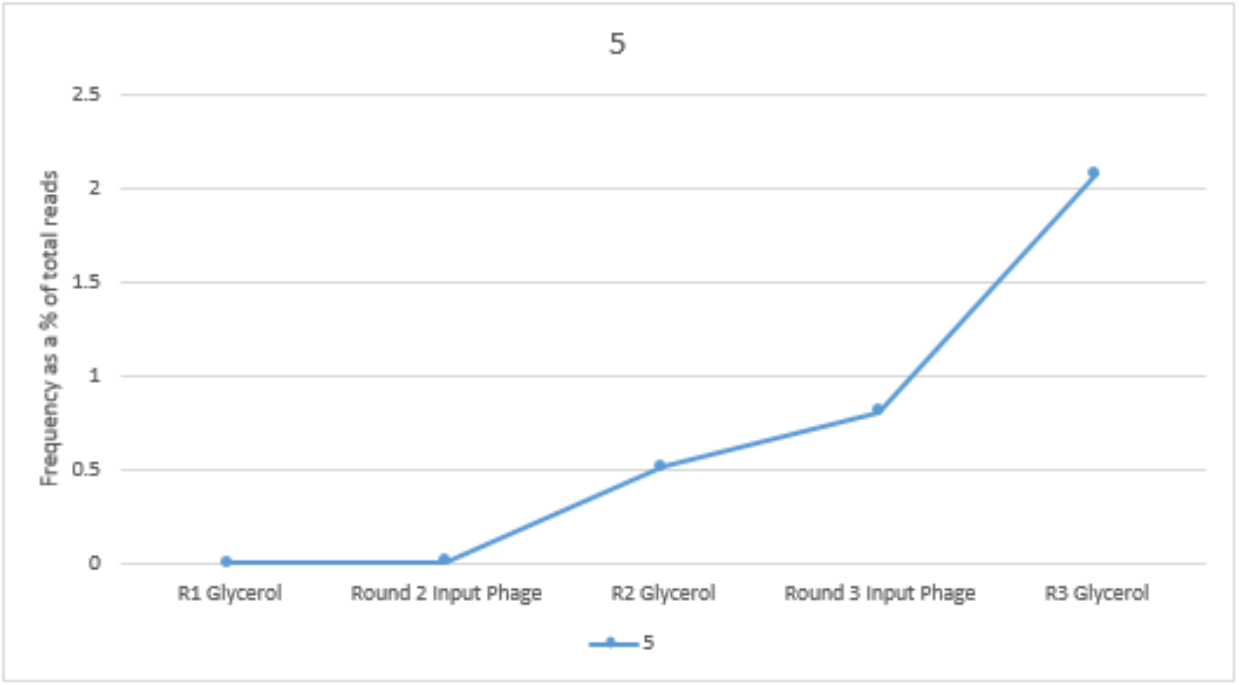
D


3
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
3

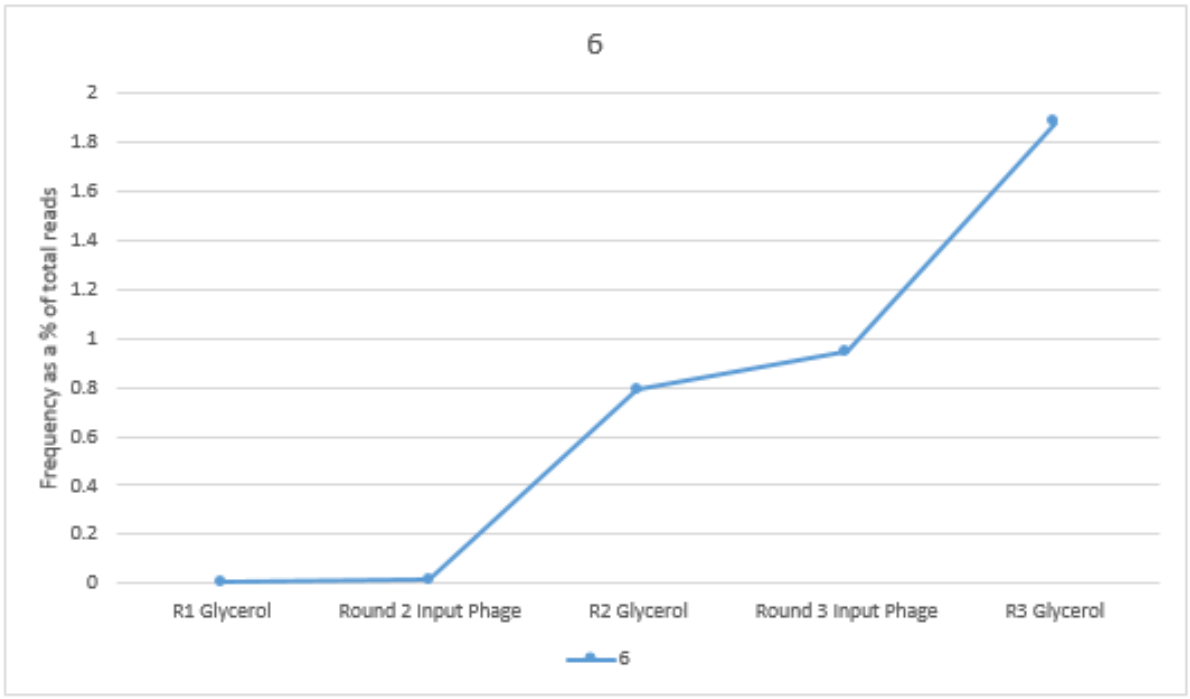
E


4
2.5
2
1.5
1
0.5
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
4

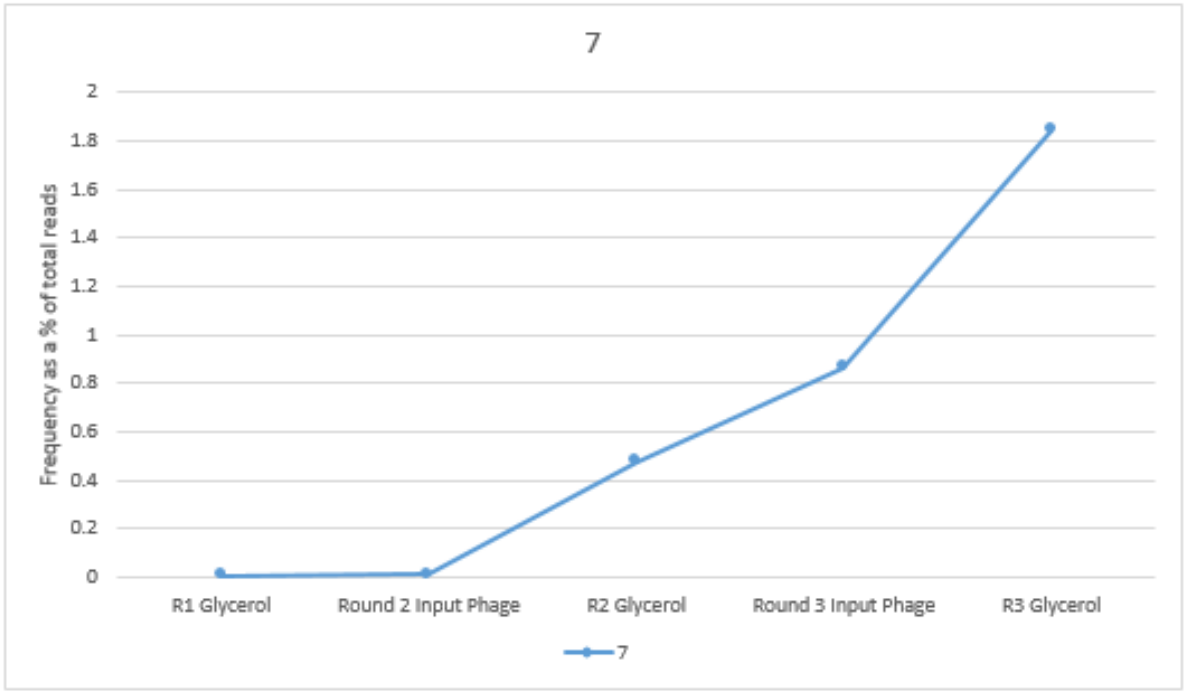
F


5
2.5
2
1.5
1
0.5
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
5

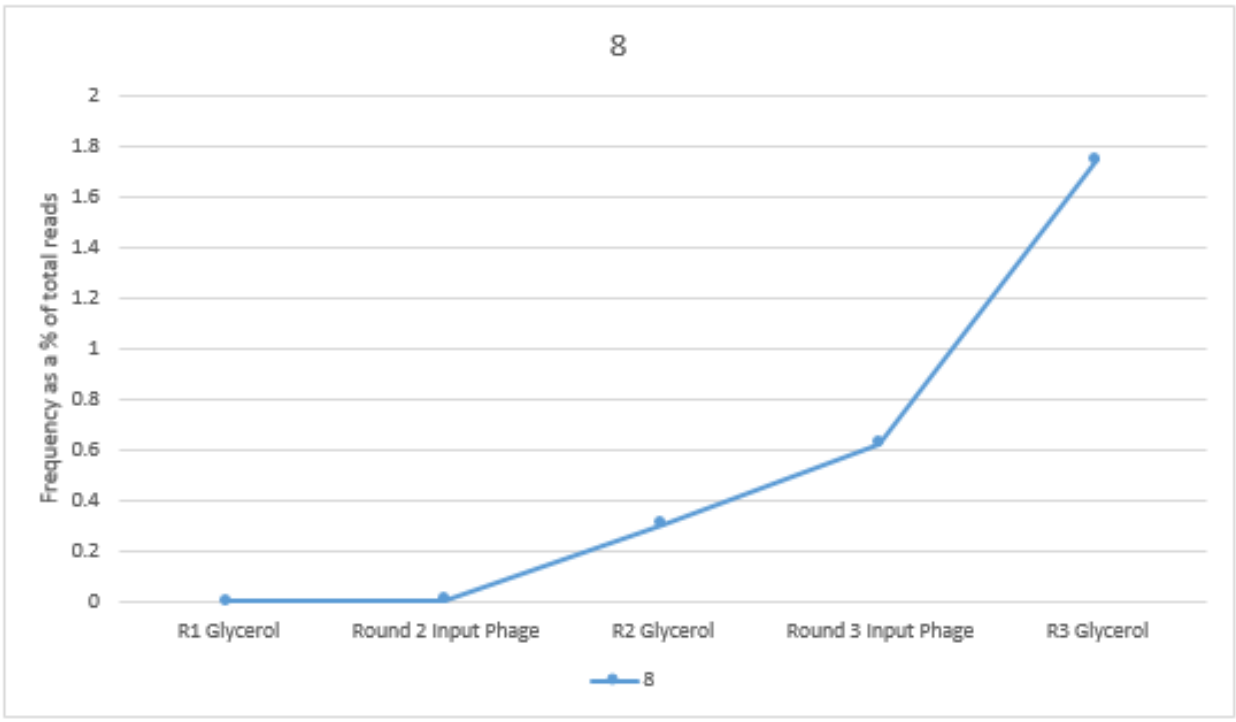
G


6
Frequency as a % of total reads
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
6

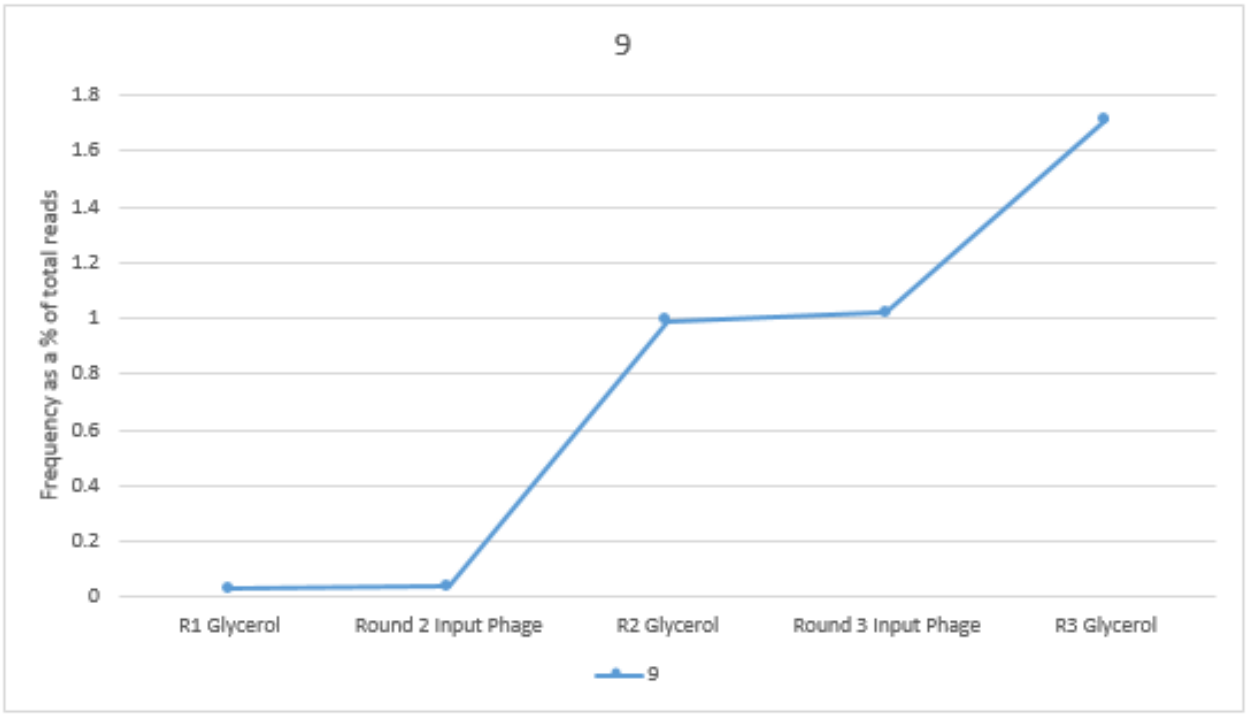
H


7
Frequency as a % of total reads
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
7

I


8
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
8

J


9
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Frequency as a % of total reads
R1 Glycerol
Round 2 Input Phage
R2 Glycerol
Round 3 Input Phage
R3 Glycerol
9

K

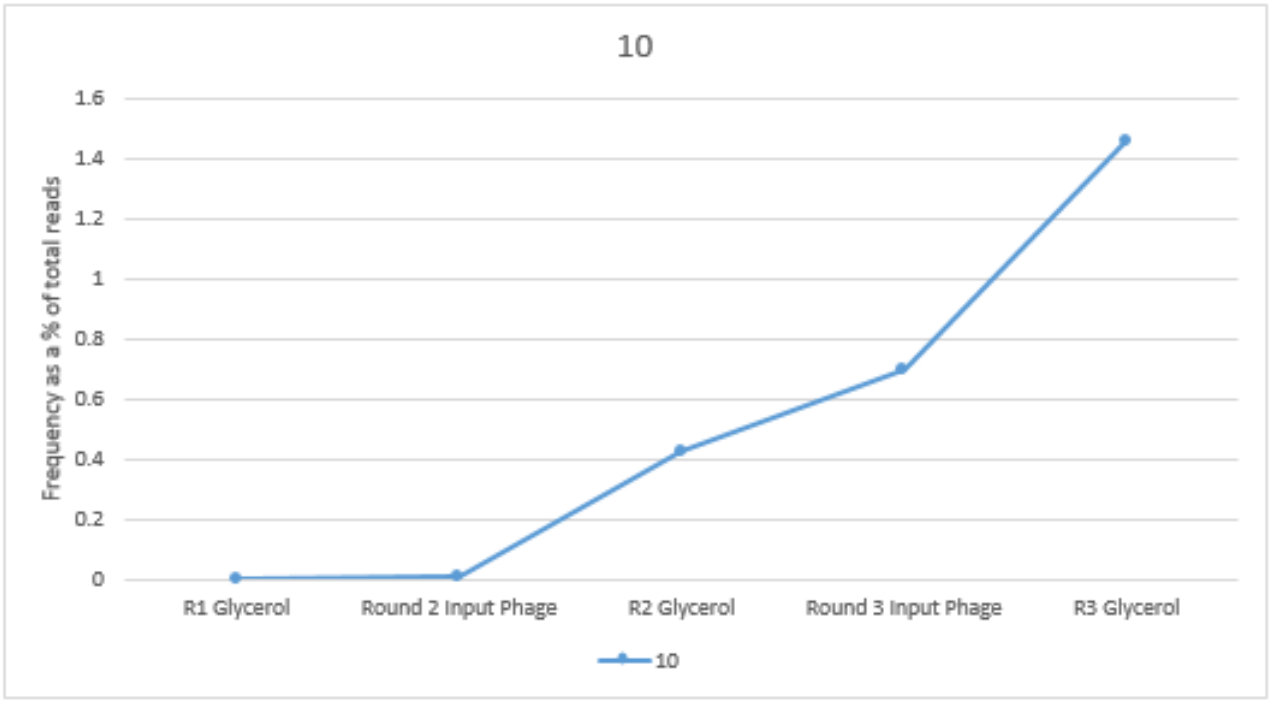


Figure 6.5: Growth and selection phase frequencies for the top 10 anti-elastin clones (Frequency Analysis). The read counts for the top 10 frequency analysis, Anti-Elastin clones were expressed as percentages of the total reads within each panning round or phage input barcode file.

6.6 Revisiting the colony picks; comparing unrelated datasets

This particular study had the benefit of running three different panning experiments in parallel, using the same phage display library. This allowed the identification of potentially non-specific binders, i.e clones occurring in multiple unrelated datasets as they either have growth advantages or bind to commonly used blockers or solid supports. Although this was a potentially powerful tool, it is not known whether this is possibly too stringent a deselection process.

The challenge here was to either enforce a complete absence of clones from unrelated datasets, resulting in the exclusion of potentially positive clones as they occur by chance at very low levels in other panning experiments, or make some sort of allowance for low levels of enrichment for such background clones. The anti-elastin binder (VHH) with the HCDR3 sequence 'QRLKKRKHHITSTSWK' (described in chapter 5) was used to determine the carry-through of non-specific binders during panning. Its specificity for elastin should have prevented it from being able to bind to either phytase conformation or phytic acid or commonly used blockers/plastics throughout their panning experiments. In theory, the clone should have been either absent or minimally enriched within the phytic acid and phytase datasets. The clone's frequency in the unrelated panning experiments would give an indication as to whether a complete absence of non-specific clones could be an acceptable threshold for filtering the data, or whether some level of background was to be expected and perhaps some frequency, albeit low, was inevitable.

The three NGS datasets frames.LR.fa files were searched for the presence of the anti-elastin HCDR3 sequence using the standard grep command in Linux terminal. The frames.LR.fa files contained all of the HCDR3 sequences present in the files before ranking them based on Z score and thus the search was much more thorough than filtering the data using the normal NGS pipeline described in this study.

The anti-elastin clone's HCDR3 sequence "QRLKKRKHHITSTSWK'" could not be found in any of the frames.LR.fa files for the phytase targets including panning glycerol stocks and phage output files and screening rounds. The anti-elastin clone did have 5 reads within the negative pool for

the round 3 screen against phytic acid. The two other anti-elastin clones, as described in section 6.5, were also subjected to searches within the three unrelated data sets generated here. Only the TKPTKPRSPPKKTRLH clone was found, and again only had very low read counts: three reads within a single active phytase positive replicate. The anti-elastin clones demonstrated a minimal presence within the unrelated datasets. This data supports the original method used within the study, that is, the removal of candidates that had become enriched enough to make the minimal threshold (presence in 4 or more replicates with Z scores of two or greater) in multiple unrelated panning experiments.

6.7 Discussion

6.7.1 The application of NGS to phage display panning

The coupling of phage display and next generation sequencing has the potential to revolutionise the way new ligands are discovered. Different sequencing platforms offer different ranges of read depth and accuracy. During NGAPD experiments, antibody gene data is often limited in the sense that, even as fragments, the genes are too lengthy to sequence in full by high-throughput sequencing platforms. Illumina's paired end MiSeq (2x300bp) or Life Technologies' Ion Torrent (400bp) platforms are the most frequently used with antibody phage display experiments. Ion Torrent sequencing, as used within the present study, can only capture a segment of the scFv gene and published studies always focus on the HCDR3 domain as it is widely believed to contribute the most to antigen binding (Xu and Davis, 2000). The MiSeq platform offers the ability to capture segments of both the VH and VL chains however not as one full length amplicon. This limitation means that the sequencing technique cannot accurately determine the VH and VL pairing as they are found in the library. NGAPD publications such as Saggy et al. (2012) have managed to get around this by pairing the most abundant VH with the most abundant VL sequences after bioinformatics analysis, and successfully identified strong target binders. However, an important difference was that the authors identified the NGS candidates by data mining an immune phage display library rather than by panning against the target antigen. The immune library had already gone through in vivo maturation and binders were already enriched.

Generally speaking, platforms that offer the longest reads, also have the least read depth. The Ion Torrent and MiSeq platforms have shorter read lengths, but they can produce millions of reads with a high level of accuracy (Glanville et al., 2015). Platforms offering longer sequence reads covering ~800bp, and thus both VH and VL chains of an scFv, include technologies such as Roche's 454 pyrosequencing and Pacific Bioscience's single-molecule real-time (SMRT) sequencing. Although these platforms offer read lengths long enough to cover most of the scFv it comes at the expense of read depth and in the case of SMRT, the accuracy of base calls. The Ion Torrent, MiSeq (2x300bp) and 454 platforms maintain base call accuracies of >99% whereas

the SMRT platform only achieves 84—85% (Lee et al., 2013). The sequencing platforms are an important consideration when conducting phage display experiments. Chapter 4 provided interesting results where truncated, light chain only, antibodies demonstrated specific binding properties to the target antigen. If they were truly binding to the target without the need of a heavy chain, the NGS strategies focusing solely on the heavy chain diversity are currently missing important data that could facilitate the identification of novel binders to their targets. This argument also assumes that all of the binding properties were originating from the scFv and not the scFv-phage fusion as binding properties could not be confirmed using soluble antibodies. A more practical consideration for NGPD experiments is the costs associated with sequencing technologies, for which Ion Torrent and MiSeq remain the cheapest (Glanville et al., 2015). Thus most publications focus on the CDR3 domains of the VH and VL chains.

6.7.2 Effects of flanking region identity on antibody selection

Within the present study, the human scFv HCDR3 sequence was typically flanked by VYYCAR HCDR3 WGQGT. There was however degeneracy within the flanking sequences as not all of the amino acids are conserved within the flanking region or between antibodies and their VH families. This chapter re-analysed the NGS data from chapter 3's scFv isolation and allowed 0 or 2 changes within the HCDR3 flanking regions. The original top candidates identified by allowing only one change were looked for within the new datasets to determine whether they would still be considered viable candidates. When no changes were permitted in the flanking regions, only 53% of the original clones were now ranked in the top 20 candidates and met specificity criteria. When two changes were allowed within the flanking regions only 24% of the clones remained as selected candidates. Often, the candidates that appeared to be antigen specific when 1 change was allowed in each flanking region were also found in the datasets of the other antigens when up to 2 changes were permitted, and so would not have been selected. Although the HCDR3 sequence remained the same, the more stringent flanking sequence criteria had prevented their detection in the other antigen datasets. The HCDR3 sequence's presence in multi-antigen datasets suggests that they were not specific to the intended target, or in the case of the

phytase, were not binding to conformational epitopes. This may well be a considerable limitation in the methods used to select candidate binders.

Ideally, in future experiments the HCDR3 sequences would be determined using the least stringent flanking region criteria to maximise HCDR3 identification. The Kabat numbering system identifies the CDR and the framework regions, however, some residues will vary thus the flanking sequences VYYCAR and WGQGT are not the only combination naturally present within an immune library (Elvin A. Kabat, 1991). In the context of next generation sequencing experiments using natural repertoires, it is impossible to know which combinations of flanking regions are truly present in the samples, and which combinations arose due to genetic mutations during growth phases or as amplification errors during PCR and sequencing. Sequencing platforms can be susceptible to errors such as mis-amplification and incorrect base calling with frequencies ranging from 10^{-2} (PacBio and Ion Torrent) to 10^{-3} (Illumina) (Rouet et al., 2018). To address this, the approximate error rates caused by this study's amplification and sequencing steps were predicted using the VHH library. The VHH library used a conserved VHH scaffold where the only variation between clones was restricted to the HCDR3. Therefore, any additional reads identified after allowing 1-2 changes within the flanking regions, which should have been conserved between clones, were likely present due to either amplification or sequencing errors. The percentage increase in reads, or error rate, were the highest after allowing 2 changes in the flanking sequences. As expected, more reads are identified as the selection criteria became less stringent.

Suprisingly, the VHH HCDR3 numbers increased by 83% when allowing 2 'errors' in the flanking regions and for the scFv library this was 23%. Indicating that the PCR and sequencing errors contribute to flanking region diversity more than natural diversity found in the scFv library.

In NGS studies scFv HCDR3 sequences are often identified based on either the International Immunogenetics database (IMGT) residue numbering system (Lefranc, 2001) i.e (Ravn et al., 2010b), (Saggy et al., 2012), (Hemadou et al., 2017) or specific flanking sequences provided during the analysis pipeline (Lovgren et al., 2016). To the author's knowledge, no NGS study has explicitly considered the effect of allowing degeneracy within the flanking regions. A concern is that allowing too much degeneracy within the

sequences will lead to the false identification of HCDR3 sequences, or if the light chain data is also present in the data, the identification of the wrong CDR3 due to their flanking sequence similarities (typical flanking sequences with fixed residues in bold: 5' VYY**C**AR-HCDR3-**WG**Q**G**T 3' and 5' VYY**C**-LCDR3-**FG**Q**G**T5 '3). For the present study, the 'contamination' of datasets with light chain sequences was not a problem as only the heavy chain was amplified. As sequencing techniques improve, high-throughput read lengths will inevitably become longer, therefore allowing the entire scFv gene to be covered. It is at this point that the pipeline will need to be adjusted to account for the similarities in the HCDR3 and LCDR3 flanking regions to prevent false identification. The number of changes allowed in each flanking region are rarely discussed by published papers, and for those that did, they allowed 1 change per sequence (Lovgren et al., 2016, Spiliotopoulos et al., 2015). The study by Yang et al. (2017) allowed up to two changes using known flanking sequences however this was done using an IgY library and didn't attempt different levels of flanking sequence stringency. By allowing multiple changes within the flanking sequences, a greater number of HCDR3 sequences can be captured as demonstrated here, with a 16.52% increase after allowing 1 change and a 23.47% increase after allowing 2. When the scFv samples were prepared for sequencing the primers that were used to amplify the HCDR3 fragment were degenerate in order to capture as many different HCDR3 as possible. The data presented within this chapter demonstrated the importance of combining degenerate primers intended to amplify HCDR3 fragments containing natural diversity from within a sublibrary, with a bioinformatics strategy that maximises the identification of HCDR3 once they have actually been sequenced.

6.7.3 The estimation of library diversity using NGS

NGS can be used to determine the diversity and bias of naïve antibody libraries. The analysis of very large naïve libraries represents a considerable challenge, where the diversity of the antibody library will most often exceed the read depth capabilities of most NGS platforms. Illumina's HiSeq platform offers up to ~3 billion reads with a base call accuracy of ~99% (Lee et al., 2013). However, the diversity of the Griffin scFv library was reported to be ~ $1x10^9$ (Griffiths et al., 1994); however most platforms will only capture a

fraction. Platforms such as MiSeq capture 17 million reads and Ion Torrent platform with a 318 chip capture up to 5 million reads both with >99% accuracy. Considering the entire diversity and individual clone count cannot be captured by a single sequencing run, the diversity of libraries can be estimated instead. The present study used the Lincoln Index statistical measure to estimate the diversity using two individual, Ion Torrent sequencing runs of the Griffin library and initial VHH libraries (12mer, 16mer, 21mer). The files had low read counts (76824—122267 for the Griffin scFv and 30997—197836 for the camelid VHH) which had a negative effect on the final diversity predictions (10^6—10^7); an additive total of $7x10^7$. The VHH libraries had previously been estimated at ~10^8—10^9 by Sanger sequencing a random selection of picks and titering transformants. It is possible that the prediction using the NGS approach may be more accurate than the conventional method. However, the Griffin library was expected to have a diversity of ~ $1x10^9$, but the NGS analysis indicated it was only $1x10^6$. Some discrpency might be explained by the fact that the NGS pipeline only considered the HCDR3 diversity but the other CDR sequences also vary with a collective diversity of ~ $1x10^9$. However, the Griffin scFv library involved the same variable heavy chain repertoire as the Fab library developed by Griffiths et al. (1994) and should have contained ~ >10^8 different clones, although this is also split between the three HCDR loops and the authors estimated diversity using BstNI fingerprinting and sequencing of 24 clones.

The VHH library template (anti-bLAC) "VRGYFMRLPSSHNFRY" was counted 212 times in the 12mer library, 166 times in the 16mer library and 143 times in the 21mer library. As for the rest of the HCDR3 sequences, most only had one copy, accounting for 85%, 83% and 88% of the distinct sequences within the 12mer, 16mer and 21mer sub-libraries respectively. This was similar to other studies such as Turner et al. (2016) where single copy number candidates (VHH and peptides) accounted for 88% and 92% of the sequences respectively. The data indicates that the libraries were highly diverse and only a select few antibodies were represented by more than 1 copy. The study by Turner et al. (2016) also found that most of the sequences were only represented once when considering the copy numbers of 186 randomly selected sequences from their original library. A handful of

sequences had very high copy numbers >250 and were suspected to be artefacts of the library preparation.

The VHH copy number data does not correlate with the estimated diversity using the NGS method which suggests that both scFv and VHH clones are present >100 times (8—387) within their respective libraries. The NGS method calculates the Lincoln index statistic ((Sample 1 distinct reads x Sample 2 distinct reads)/number of sequences common to both sample sets) to predict the library diversity. A disadvantage to the Lincoln index statistic is that it requires a very large sample set and the diversity estimates were greatly influenced by the total read counts in each barcode file. It is likely that a higher read depth would be required to accurately estimate library diversity using this method.

6.7.4 Following antibody enrichment through panning

NGS can also be used to investigate the panning process. The diversity dropped throughout sequential rounds of panning and phage output samples as expected. Based on the frequencies of the top 20 candidates (frequency analysis) for both phytases, the individual clones were obviously enriched whereas the frequencies for the phytic acid candidates were much lower despite the expected drop in diversity during panning. The low frequencies suggest that the selected anti-phytic acid clones may not be true binders or that they are particularly toxic to their hosts and inhibit bacterial growth.

6.7.5 Alternative methods for defining enriched clones within NGS data

The present study demonstrated some of the limitations associated with naïve phage display libraries. The large diversity within a naïve library has the potential to facilitate the identification of novel antibodies against almost any given target, however filtering out the background and parasitic clones remains a considerable challenge. To try and address this issue, phage display experiments are increasingly coupled with next generation sequencing platforms to help scrutinise and sort the ligands more efficiently. Ultimately the sequencing data reflects the library diversity at various stages of panning, implicating the most enriched clone as the most likely candidates with the desired binding properties. The methods by which the libraries are analysed vary considerably between publications due to the nature of the

experimental design, and the resources available to generate and interrogate the data.

Here, the bioinformatics analysis focused primarily on generating Z scores (Zhang et al., 2011) for each sample's HCDR3 sequence. The HCDR3 sequences were then ranked based on their Z score and how many positive replicates they were seen in. In chapter 3, the three target antigens were run in parallel and their NGS datasets could be compared to help omit HCDR3 sequences common to unrelated panning experiments. The present study's 4th and 5th results chapters brought in additional analysis strategies such as common motif analysis (MEME) and also a reversed analysis to omit non-specific clones (clones seen as being enriched against a non-target control). The Z score strategy has successfully identified target specific binders in other publications such as Zhang et al. (2011) and Naqid et al. (2016) . The peptide phage display study by Naqid et al. (2016) used a naïve peptide library for 2 panning rounds. The study aimed to map antibody responses to bacterial infection, and selection was carried out comparing binding to post-infection and pre-infection IgG samples. To identify the peptides within the data, sequences were a minimum of 4 amino acids long and had up to two mismatches in the flanking sequences used to identify the peptide sequences. Z scores were generated for each peptide sequences with a threshold of $Z \geq 2$. The authors also excluded likely target unrelated peptides (TUP) candidates that were already documented by the online peptide databases SAROTUP and Pepbank. The most enriched candidates post-analysis were synthesised and 22 of the 27 peptide candidates bound to their target.

AC't Hoen et al. (2012) , Ryvkin et al. (2012) and Ngubane et al. (2013) also used random, naïve peptide libraries but coupled with the Illumina sequencing platform. The studies analysed the sequencing data for both the naïve library and one (AC't Hoen), three (Ryvkin) or five (Ngubane) panning rounds. AC't Hoen et al. (2012) demonstrated that after only 1 round of panning, high affinity peptides could be identified from the sequencing data. A peptide sequence ranked #14 (by frequency) after the first panning round had exceptionally strong target binding abilities. Candidates that demonstrated clear enrichment throughout panning were considered top candidates; enrichment was defined as a reduction in sequence diversity as the panning rounds progressed but a gradual increase in the frequency of

enriched clones. Ngubane et al. (2013) and AC't Hoen et al. (2012) also compared the sequences to the SAROTUP database, and Pepbank as well in AC't Hoen's case. Peptide sequences were then ranked by frequency after confirming their enrichment by panning and excluding clones with naturally high copy numbers in the naïve library, such as two or more (AC't Hoen et al., 2012). Ngubane et al. (2013)'s study identified a good target binder which was synthesised after identification by the NGS approach. However, the clone could not be identified after x10 random colony picks from round 5 despite representing 80% of the round 5 sub library. Ryvkin et al. (2012) on the other hand interrogated their sequencing data using the MEME motif analysis software to identify common motifs that had become enriched within the data; again, enrichment was characterised by a drop in sequence diversity throughout panning but a gradual rise in frequency for enriched clones. The motif analysis considered the entire captured diversity for each round to determine the most common motif within the affinity-enriched candidates.

The study by (Zhang et al., 2011) used a naïve human scFv library coupled with Roche 454 pyrosequencing. The author panned against both a target and control antigen for 2 rounds and sequenced the variable heavy chain of the scFvs. Using the target and control data, each HCDR3 sequence was ranked by the Z scores generated by the comparison. Two clones were recovered and were shown to specifically bind to the target antigens in phage-fusion and full IgG formats. This is the only antibody phage display publication that has applied the NGS Z score method to identify novel antibodies capable of binding to the target.

The publications by Ravn et al. (2010b), Sasso et al. (2015), Lovgren et al. (2016) and D'Angelo et al. (2014) also used naïve or synthetically randomised human scFv libraries. Each study conducted two to four rounds of panning before applying NGS using the Illumina or Ion torrent platforms. The resulting candidates were ranked by frequency in each study. In each publication, the most abundant antibodies (VH) were rescued by 2-step, inverse PCR or simply synthesised, and each tested positive against their antigens by phage or yeast display. The study by Sasso et al. (2015) also considered each antibody sequence's enrichment throughout the panning rounds; enrichment of scFv was defined as described previously for the peptide studies by AC't Hoen et al. (2012) and Ryvkin et al. (2012). In a

further publication, Ravn et al. (2013) used 5 different human scFv libraries that had been diversified within the CDR3 regions of the VH and VL chains. Again NGS was applied using the Illumina platform after three panning rounds. Novel scFv were generated that proved to be effective binders. They were identified based on frequency and cluster analyses that looked for common motifs.

All of the studies described above used naïve libraries and each study managed to identify target specific clones by considering the frequency of clones, their enrichment throughout the panning rounds and also which motifs were most frequent. Although the Z score analysis method has been successful for the noted publications, it has only been demonstrated once by Zhang et al. (2011) in the case of antibodies. Even then, the strategy used by the authors relied entirely on the Z score ranking system used within the present study without application of negative selection. Studies by (Ngubane et al., 2013) and (Sasso et al., 2015) conducted subtractive panning steps similar to those used during the present study. The authors were targeting cell wall components specific to *M. tuberculosis* and the CLDN1 protein on the surface of HEK cells. To remove binders to other cell membrane components it was nessesary to subtractively pan against related mycobacterium and HEK cells that were not displaying CLDN1 to remove non-specific phage. This approach is common throughout the literature and may in some cases be effective enough to facilitate specific binders to the intended targets.

Overall, the data presented here show that for panning against phytase, the most enriched clones in terms of frequency were very likely to be binders and indeed those ranked #1 and #2 for active phytase and #4 for denatured phytase were (previously) demonstrated to bind phytase after being randomly picked. However, they could not be confirmed as specific for a particular conformation of the enzyme and it is likely that most binders will not discriminate between the closely related targets and so require different criteria/panning strategies to select those with this trait. It is possible that the additional steps such as the comparison of unrelated VHH panning databases or reversed analysis proved to be too stringent and removed binders specific for the targets. Could it have been more realistic to expect clones to come through in both target and control data, and unrelated experiments using the same initial library? The anti-elastin VHH clones

identified from the same library demonstrated a minimal presence within the unrelated datasets, suggesting that non-binding clones were unlikely to become enriched during panning against an entirely different target. Although this may be dictated by the individual toxicities of each clone and the effects this has on the growth and proliferation throughout panning. In terms of the presence of high frequency clones in the control datasets (reversed analysis), it seems likely that if the clone is of high frequency it is likely to also contribute a significant proportion of the background phage during the negative screen having been grown from the same sub-library as the phage used for the positive screen. This could be tested by comparing the positive screens for each phytase conformation against one another, rather than the negative control datasets that had been screened using the same sub libraries. It may be more likely that conformation specific binders would be identified using this method having an expectedly lower enrichment profile against the opposite conformation. The specific clones would not then transfer to the negative control replicates and appear enriched due to the use of the same sub-library and would be less likely to be ruled out by reversed analysis.

Other studies using immune antibody libraries have also been coupled with NGS platforms (Roche 454 and Illumina) and various bioinformatics strategies (Saggy et al., 2012, Miyazaki et al., 2015, Turner et al., 2016) (Yang et al., 2017). In these cases, the need for multiple panning rounds was reduced due to the immune status of each library and therefore the relative abundance of pre-tailored antibodies. This effect was evident by the number of panning rounds conducted by Saggy et al. (2012), Miyazaki et al. (2015) and Turner et al. (2016) who ran only 1 round. The exception here was the study by Yang et al. (2017) who still conducted 4 rounds of panning despite using an immune library. This is not to say that studies using naïve libraries cannot learn from the successes of immune libraries, when coupled with next generation sequencing. The study by Saggy et al. (2012) used the 2x300bp Illumina platform, generating sequences for both the heavy and light chains. The authors paired the most abundant VH and VL sequences together when constructing the NGS candidates for testing. This proved to be an effective way of pairing the VH and VL sequences whose natural pairing was otherwise unknown. The study by Yang et al. (2017) used both an NGS approach and colony picking in combination with an immune scFv library (IgY). The

sequencing data was generated for the initial immune library and all four rounds of panning. Candidates were ranked based on their frequency (enrichment or impoverishment) throughout the panning rounds; enrichment was characterised by an incremental population shift as the panning rounds progressed, whereas impoverishment was defined as the opposite effect. Candidates from both the NGS and colony picking method were shown to be specific for the target.

When considering VHH antibodies, only a handful of VHH studies that coupled their display strategy with an NGS platform have been published to date. The studies by Miyazaki et al. (2015) and Turner et al. (2016) both used immune VHH libraries. The studies both sequenced the initial immune library using the Illumina platform but the study by Miyazaki et al. (2015) also sequenced the subsequent panning rounds. Turner and co-workers interrogated the immune library data by using a BLAST search for a known target binder sequence. Clones with similar sequences were synthesised and were shown to have binding properties comparable to the existing clone. This was not applicable to most studies when target specific antibody sequences will not be available. Miyazaki's study was similar to Zhang et al. (2011) and Sasso et al. (2015) in the sense that the candidates were ranked based on not only their total frequency within each panning round, but also their copy number in the initial immune library. The majority of the HCDR3 sequences within the present study had only one copy within the library sequencing data. A high copy number would suggest pre-existing bias towards a clone if it is already over represented the library preparation.

For Miyazaki's study, the top 100 sequences that displayed clear enrichment throughout panning were also subjected to a phylogenetic tree analysis; enriched candidates were defined by an increase in frequency (%) after panning and 75% of the sequences had very similar amino acid sequences and were grouped within the major branch. Positive binders identified by colony picking were identified during the study and shared the most homology with the candidates in the branch representing the 75%. The authors aimed to identify VHH candidates that were functionally different to the 75% forming the primary branch of the phylogenetic tree. The authors expressed candidates from three minor subbranches and out of the three minor clusters, only VHH candidates from two of the three branches

demonstrated binding activity against the target antigen suggesting that the NGS strategy could also be used to identify rare, more diverse binders that might otherwise remain unfound by conventional methods.

In summary, most studies rank ligand sequences by frequency and may or may not take other factors into consideration depending on the experiment design. In addition to frequency, the prevalence of sequences throughout panning rounds, to demonstrate enrichment, and within the initial library sequencing data was also considered as a means of identifying sequences that potentially already had a high copy number to begin with. Another common strategy was to look at common motif clusters, suggesting that the panning strategy was selecting for a specific amino acid sequence. This chapter re-evaluated the VHH NGS data from chapter 5 that was previously ranked based on Z score and common motifs. All of the NGS and colony pick candidates failed to produce positive absorbance signals in their soluble antibody ELISAs, and thus brought the bioinformatics strategy into question. However, considering the number of candidates that were tested it is possible that there was a fundamental error with the assay itself. The panning experiments against phytase appeared to be successful due to the isolation of positive, anti-phytase VHH and scFv phage fusions. The scFv candidates appeared to be target specific as confirmed by multiple, monoclonal ELISA. Furthermore, the VHH candidates consistently bound to phytase although they were not conformation specific. The VHH candidates for phytic acid were unsuccessful as demonstrated by negative ELISAs and poor enrichment data in comparison to the phytase datasets. Here, further analysis aimed to re-evaluate the VHH NGS data using a published method, so as to compare the new top candidates against those suggested by the original Z score and motifs analysis. The bioinformatics strategy conducted by Miyazaki et al. (2015) was used as the alternative method. Candidate HCDR3 sequences were ranked based on frequency followed by their enrichment throughout the panning rounds. In the present study, only three of the original candidates against active phytase were identified again using the frequency/panning enrichment strategy. For denatured phytase only two of the original candidates were re-identified and for phytic acid none of the original candidates made the top 20 candidates. Only the clone "RTTKSHQHSSRKRFKH" of the 4 positive clones against phytase identified

through random colony picking, was absent from the top 20 candidates for denatured phytase after the new analysis method was applied. Also, the frequency of the top 20 clones was established as percentages of the total reads for each positive replicate. Interestingly, all of the candidates for panning against phytases had fairly even distributions within the positive barcodes, indicating that the candidate's frequencies were not unfairly biased by individual replicates. This repeatable selection of the high frequency clones possibly indicates that these represent true binders (and this was previously confirmed for 3 randomly picked clones in Chapter 5). This was not the case for the phytic acid panning data where clones were present at relatively low frequencies (0.5% or less) and showed high variability between replicates, possibly indicating these are not true binders.

The study by Miyazaki et al. (2015) did not have a negative control dataset to compare the frequencies against, however for the present study the frequency of the clones in the negative pool were calculated for comparison. All of the top 20 candidates had similar frequencies in the negative pool as they did in the positive suggesting that the candidates were likely binding to both the target and negative control antigens as indicated in earlier chapters by phage ELISAs. This included three clones identified by random colony picks. In their case the phage ELISAs already eluded to the fact the clones were reactive to both phytase conformations so it is unsurprising that they become enriched in both screening replicates.

The similar frequencies in both the positive and negative replicate pools would suggest that the reversed analysis applied in chapters 4 and 5 would often result in the exclusion of these new candidates suggested by the frequency analysis alone as the overall aim was to select conformation specific clones. However, as stated above, when clones are of high frequency it is likely that they will be obtained from both the positive and negative screens. Defining the level of true specific binders in panning sublibraries that are seen in negative screens is important for future application of these selection methods. Using earlier rounds of panning may overcome such issues as analysis would take place before a small number of clones become dominant.

Using the frequency analysis it was evidently not possible to identify enriched clones that did not also have a high frequency within the negative

replicates. Other studies, as previously discussed, have compared their datasets to public databases of known binders to help filter the data. However, this is carried out for peptides and the equivalent databases for antibody sequences where potentially at least 6 regions can contribute to binding would be far more complicated and are not available. Therefore, for antibody isolation it is difficult to compare datasets any other way than to carry out negative screening within the experiments. Ultimately, care must be taken to ensure that sequences are not unjustly ruled out due to experimental design or bioinformatics analysis as this will severely hinder the usefulness of the final data.

A number of the new VHH candidates demonstrated clear enrichment throughout panning and for the most part, could not be identified within the naïve library. The latter is unsurprising as it is impossible for the current sequencing platforms to capture the entire diversity of the naïve libraries used within this study. One top candidate was detected, and had an abnormally high read count of 7. If the new candidates identified by the frequency analysis had been pursued further, that particular candidate would have been excluded as it was already enriched within the library prior to panning. The study by AC't Hoen et al. (2012) made a note of excluding candidates with naïve library copy numbers of 2 or greater.

The frequency analysis demonstrated that this method produced very different candidates compared to the previous Z score analysis. Only five of the original VHH candidates panning against phytase conformations out of 26 were identified within the top 20 for the frequency analysis for each target. None of the original 13 candidates raised against phytic acid were present. It may be that frequency analysis alone is effective to identify binders but more stringent methods such as the Z score analysis is better suited for highly selective analyses such as the discrimination of antibody binding between closely related antigens. In addition to the subtractive bioinformatics analysis used (e.g excluding clones through analysis of unrelated datasets and/or reversed analyses of sublibrary binding), as mentioned previously, the criteria used to identify HCDR3 sequences in the first place (e.g flanking region degeneracy) will also dicate the clones that are identified.

As part of the study, a further analysis was conducted that took into consideration the differences in total frequency between growth and growth

and selection phases during panning. Whilst this will not Inform whether binders can distinguish between closely related antigens, it may provide an additional criteria to select true binders above background clones. Anti-phytase binder candidates and anti-Elastin VHH binders frequently demonstrated enrichment during the growth + selection phase and less in the growth only phase, as was to be expected of clones becoming enriched within the panning experiments due to binding to the target. Some candidates demonstrated an almost linear increase in total frequency throughout the experiment, regardless of growth only or growth+selection phases. This may indicate that the clones had bound the target but also had low host toxicity and therefore thrived in each phase. Out of the 7 anti-phytase candidates, the colony picked candidate #1 (qHRPRPHRPTKRRKSY vs active phytase) accounted for the highest % of total reads and demonstrated the most enrichment during the growth + selection phases and little to no enrichment during the growth stage. This was the optimal trend expected of a clone that was becoming enriched due to binding to the target antigen. This clone also demonstrated a positive binding interaction during multiple phage ELISAs as did clone I.D's 2—4; the three known anti-Elastin binders also demonstrated a similar trend. This growth and selection pattern may therefore be indicative of true binders and may serve as some additional criteria to support the selection of new clones from vast NGS datasets. To the authors knowledge, this type of analysis considering total frequency between growth and growth+selection phases is the first of its kind. The study by Sasso et al. (2015) analysed the enrichment profiles for 75 different clones and demonstrated the gradual increase in frequency throughout panning rounds. However, the authors did not consider growth and selection phases to assess exactly where the major enrichment occurred and whether it was likely due to an interaction event with the target antigen or whether it was more likely to be a clone with a growth advantage.

Relatively little has been described in the literature about optimising bioinformatics selection methods and identifying true binders from the NGS data. This would require multiple different bioinformatics analyses of many different datasets followed by testing the selected candidates. Overall, the data presented here starts this process and demonstrate that the specific bioinformatics criteria applied to NGS data will produce very different

candidate antibodies for rescuing and testing. This included optimising the flanking regions used to identify CDR sequences and also consideration of whether defining enrichment based on frequency (considering both growth and selection patterns) or more stringent Z scores, and whether to apply additional exclusion criteria (reversed analysis and comparison to unrelated datasets) is required to obtain the binding specificities aimed for.

Chapter 7: Conclusions

The aim of the study was to use NGAPD strategies to identify novel antibodies against the hapten phytic acid and against conformational epitopes on the enzyme phytase. Phage display experiments were tailored to each antigen with a particular emphasis on selection strategies using a novel bioinformatics approach. Major challenges of the study included immobilising a hapten with no chemically active groups (phytic acid), rescuing scFv candidates from their polyclonal sub-libraries and finally obtaining binders that demonstrated high specificity i.e against conformational epitopes.

Phytic acid was a particularly difficult target and at present there are still no commercially available antibodies that will bind to it. This is not only due to the size of the hapten itself and its lack of immunogenicity but the fact that the immobilisation strategies were more complicated than for proteins and many haptens. In the present study phytic acid was demonstrated to be immobilised on TiO_2 plates on the TiO_2 Titansphere beads as used by Wilson et al. (2015) to capture phytic acid extracted from tissue extracts. Here, panning against phytic acid used the TiO_2 plates enabling efficient panning in a novel 96-well format. However, the bead approach would have provided a greater surface area to which phytic acid could be presented to the phage display libraries. This format would potentially allow a larger capture of target specific antibodies in comparison to the 96-well formats and could be considered in future studies.

Future alternative panning strategies could also involve conjugating phytic acid to various carrier proteins such as BSA or keyhole limpet haemocyanin (KLH) (Sathe et al., 2011). A phytic acid-like molecule with an additional aliphatic side chain, with an amine group attached via a phosphate oxygen, had been synthesised in preparation for such an experiment. This approach had been reserved as another backup strategy. 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) conjugation chemistry could be used to facilitate the covalent bonding between the amine handle (NH3) with a carrier protein's carboxylic acid group. The EDC cross-linking reagent does not become part of the conjugation complex; this is useful to preserve the structures of the components and the intended target. This approach would also require the addition of subtraction steps against the unaltered carrier

protein to reduce non-specific binders. However, the EDC chemistry does have fairly low specificity, enabling any primary amine group to become reactive and bind to any carboxylic group. This may lead to aggregation between carrier proteins so would require careful monitoring using protein gels to check molecular weights. Alternative chemistries such as a maleimide tether could be used to react with a cysteine's sulphydryl sidechain on the protein carrier. Unlike EDC, this does become part of the conjugated complex, requiring a more complicated subtraction step against the carrier protein with a reactive malemide tether. The addition of a carrier protein would enable phytic acid to become immobilised via the carrier, so by any support able to adsorb proteins such as a Nunc MaxiSorp® plates. The potential drawback is that antibodies may become reliant on the carrier-hapten complex in order to react and bind sufficiently.

A major setback for the bioinformatics approach was the rescue process for scFv candidates identified from the data, whereas the novel synthesis strategy for the VHH candidates in chapter 5 was very efficient and could be applied to any expression vector. A similar process for scFv could not be developed due to the presence of diversity throughout the gene.

Producing functional soluble VHH antibodies was also particularly challenging where only the VHH positive control (anti-bLAC) consistently demonstrated a true binding event. Clones identified by random colony picks had already demonstrated consistent binding events via phage ELISA for both the scFv and the VHH candidates. However, their soluble antibody formats for both scFv and VHH failed to produce positive ELISA signals. Future experiments could be designed to test whether the scFv and VHH clones identified by the bioinformatics approach could have bound to their target antigens as a phage fusion as this was the format that was selected during panning. The antibody candidates would be amplified from their panning sub-libraries (scFv) or synthesised by inverse PCR (VHH) in their respective libraries' phagemid vector and expressed as recombinant phage. Although this format is far from ideal and limits their usefulness as a reagent, it may have at least shed light on the bioinformatics strategies used, and also whether there was a potential problem with the soluble ELISA assay. Positive phage ELISA data may also elude to whether the MBP protein was leading to steric hindrance of antibody binding, or whether the C-terminal display of the

StreptagII-6xHIS tags on the VHH clones were interfering with antibody-antigen interactions.

Future experiments might consider cleaving the MBP tag from the VHH antibodies via the factor Xa cleavage site and purifying them using an amylose column that would retain the cleaved MBP proteins and the antibodies would be eluted during column wash steps. In addition, experiments designed to block the 6X HIS tag with a metal such as nickel could also serve as possible remedies to any tag interference. Alternatively, antibodies could be subcloned into new expression vectors without their StreptagII-6xHIS tags or expressed by a completely different system entirely for comparison. The scFv clones from chapter 4 (E6 and C7) were subcloned into the pMAL vector p5X but were digested from the pHEN2 vector just before the 6X HIS tag. This ensured that the MBP protein was the only system that could be used to identify and purify the clones. To remove the possibility of Streptagg II - HIS tag interferences this could be done for the VHH clones as well and they could be detected via the MBP. Strategies such as those described above could be tested and optimised using positive controls such as the anti-bLAC or anti-elastin clones before applying the most suitable method to the novel candidates isolated during each panning experiment. The most promising VHH candidates identified by this study are shown in table 7.1. There were many scFv and VHH clones that appeared to be enriched based on phage ELISA data, Z scores, frequency analyses, MEME analyses and enrichment profiles. However, the unsuccessful rescue attempts for the scFv candidates and potentially flawed soluble antibody expression methods meant that the candidates could not be tested effectively.

Table 7.1: The most promising VHH candidates identified during the study.

VHH HCDR3 [1]	Target [2]	Selection criteria [3]
qHRPRPHRPTKRRKSY	Active	Colony Picked - positive phage ELISA, Frequency Analysis, Enrichment Profile
qNSKKPRRNSKVQ	Active	Colony Picked - positive phage ELISA, Frequency Analysis, Enrichment Profile
RKKFARAPNHNETKWR	Denatured	Colony Picked - positive phage ELISA
RTTKSHqHSSRKRFKH	Denatured	Colony Picked - positive phage ELISA
qKTKKVTSIKSSSSRR	Active	9 of 10 (Z≥2), Acid eluted 5 of 10 (Z≥2), Frequency Analysis
qSARCSKNCRVRKMSH	Active	10 of 10 (Z≥10), MEME, Frequency Analysis, Enrichment Profile
AYPVCTCSKKGAPRALRqRSV	Denatured	10 of 10 (Z≥10), MEME, Frequency Analysis, Enrichment Profile

[1] The HCDR3 sequence for the VHH candidate.
[2] The target antigen for each VHH candidate.
3 The selection criteria used to determine the strongest VHH candidate. Colony picked clones were isolated by randomly picking colonies from their respective panning rounds. Positive phage ELISA refers to the positive monoclonal phage ELISA results shown in earlier chapters. Frequency analysis indicates that the candidate was within the list of top 20 candidates identified by the frequency analysis conducted in chapter 6, and likewise for enrichment profiles as established in section 6.5. Candidates that were suggested by the standard bioinformatics Z score analysis used throughout the study were indicated by no. of replicates seen in followed by their best Z score threshold. MEME refers to candidates suggested by the MEME analysis method that considered the most populous motifs after panning.

This study used naïve human scFv and camelid VHH phage display libraries and it was therefore more difficult to identify potential binders compared to immune libraries due to a lack of bias towards the targets within a high level of antibody diversity. However, immune libraries are not appropriate when the aim is to isolate antibodies that distinguish a protein's conformation or when panning against non-immunogenic haptens. Despite this, positive phage fusions displaying scFv and VHH were identified within the study. One and three phage-scFv fusions were identified that were positive against denatured phytase and phytic acid respectively. Furthermore, two phage-VHH fusions per target were identified by colony picking that were positive against active and denatured phytase. As determined in chapter 5, the scFv libraries were less efficient at displaying antibodies in comparison to the VHH. This is possibly due to the more complex pairing and assembly of the VH and VL genes during cloning and the proteins within the periplasm after expression, the potential toxicity caused by the scFv or increased likelihood

of premature stop codons within the scFv libraries as suggested by the large number of truncated antibodies. The result chapters' 3 and 4 also demonstrated the difficulties of rescuing and subcloning the scFv clones identified by the bioinformatics analyses. This was successfully overcome by using VHH libraries with a single region of diversity (HCDR3) which enabled the synthesis of clones directly into any expression vector. This was achieved with great accuracy and efficiency, successfully synthesising 38 of the 43 targets (88%). The VHH synthesis method was considerably more successful than the scFv rescue strategies where only 4 of 24 targets (16%) were recovered during chapter 3 and none were recovered in chapter 4.

At the beginning of the study there were very few antibody phage display publications that involved a bioinformatics approach to identify target specific clones. During the course of the study, an increasing number of NGAPD publications were released and most commonly used a frequency analysis approach as opposed to a ranking system based on Z scores. This is likely due to the complexity of the bioinformatics screening and analysis pipeline when using Z scores to rank candidates. Furthermore, most NGAPD publications to date were not isolating antibodies against especially difficult epitopes such as those that would distinguish between conformations of the same protein, therefore a complex ranking system used to filter out candidates common to a control and experimental dataset was unnecessary. The Z score approach was designed to maximise the stringency of the analysis and identify the clones most likely to distinguish between phytase conformations.

To develop the bioinformatics approach further it would be interesting to conduct a panning experiment against a simpler protein target and compare bioinformatics analyses to determine the best selection strategies to identify true binders. Considerations when carrying out panning and bioinformatics analysis are summarised in figure 7.1.

Library

Naïve

- High diversity – good for targeting multiple antigens but requires excessive filtering of non-specific clones. Naqid et al. (2016), AC't Hoen et al. (2012), Ryvkin et al. (2012), - peptide), Zhang et al. (2011), Ravn et al. (2010b), Sasso et al. (2015), Lovgren et al. (2016), D'Angelo et al. (2014) **(Present Study)**

Immune

- Tailored diversity but requires use of animals and an immune response. Miyazaki et al. (2015), Yang et al. (2017), Turner et al. (2016)
- May not require a bioinformatics approach Saggy et al. (2012).

Panning

Immobilisation of Target

- Solid supports – Maxisorb, Maleic Anhydride, Metal Complexes
- In solution – Beads with greater surface area
- Conjugation – Use of carriers to facilitate immobilisation

Target vs Control Panning

- Panning vs a control dataset required for Z scores.
- Peptides often replace this using databases e.g SAROTUP and PepBank.

Subtractive Panning Steps

- Difficult targets may require more stringent panning experiments where phage are subjected to a control antigen prior to incubation with the target e.g for enzyme conformations, haptens fused to carriers, specific cell membrane protein. Ngubane et al. (2013), Sasso et al. (2015)

Phage Elution

- Generic methods – Bacterial infection, pH shift
- Selective methods - Competitive using antigen

Identifying Clones

Conventional Colony Picks

- Most enriched candidates identified rapidly and inexpensively.
- Multiple panning rounds and hundreds of picks required.
- Rare binders are often missed. Difficult to isolate a range of different candidates.

A combination of the two

Ravn et al. (2010b)
Lovgren et al. (2016)
Turner et al. (2016)
Yang et al. (2017)
(Present Study)

Bioinformatics

- Intelligent filtering of data.
- Chance of reducing panning rounds.
- Identify rare binders with specific phenotype.

Zhang et al. (2011), Naqid et al. (2016), AC't Hoen et al. (2012) D'Angelo et al. (2014), Miyazaki et al. (2015), **(Present Study)**

Candidate Analyses

Clone Characterisation

- ELISA – Poly or monoclonal specificity
- Soluble antibody production and expression levels
- SPR – Affinity
- pH & thermostability

Antibody Formats

- Phage Fusion
- Soluble Antibody fragment
- Full IgG

Modifications

- Optimal codon usage for expression
- Purification tags and chaperones
- Optimal biophysical scaffolds

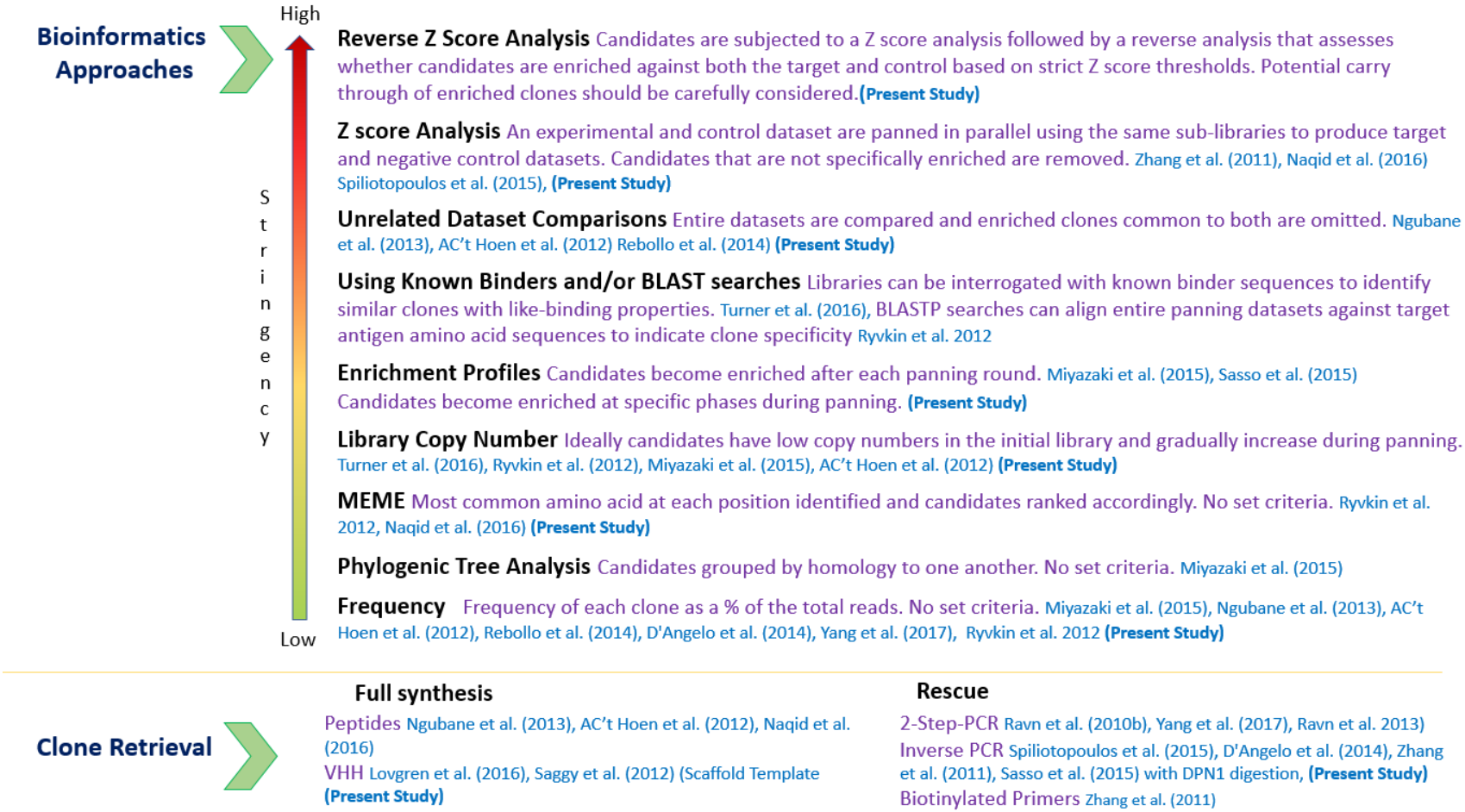


Figure 7.1 A summary of phage display approaches relevant to the present study.

Bioinformatic analyses has potential to be employed to identify true binders and then with increased stringency to identify binders that can differentiate very closely related targets.

This study began to evaluate the successes of existing strategies within the most recent NGS phage display publications, and also investigated novel approaches to data analysis. Chapter 6 results demonstrated that novel strategies such as considering enrichment profiles during growth and selection phases might be used to identify positive binders. Ideally these profiles would be compared against that of negative controls (non-binding phage antibodies that came through the paning experiment) included within the same panning experiment. Future experiments would require empirical testing of the clones suggested by this method. Within this study, known phage-fusion binders followed a similar profile to one another and suggested a promising, novel strategy for identifying new target binders from the diverse phage display sub-libraries. The study demonstrated the use of different bioinformatics analysis methods that can be used to filter the sequencing data with the intention of identifying true binders. The method used should depend on what type of binders are required. Frequency analysis and enrichment profiles may be efficient to identify ligand sequences that have become enriched from the initial library copy number; ideally this would occur predominantly during the selection phase. MEME analyses can also identify common motifs within a sub-library at any given stage of panning. Additionally, a motif search can be ideal when an existing binder has already been isolated, and the user is looking to identify similar motifs that may harbour greater affinity for the target. Alternatively, sequencing data may have to be filtered or ranked in order to identify very specific binders such as those that might distinguish between different epitopes on the same target. Methods such as the Z score ranking system considers the frequencies of different sequences in an experimental dataset and compare them to a control. Sequences are then ranked based on how often they appear in each dataset; candidates are ranked higher if they were enriched in the experimental data but not in the control. A Z score strategy enables a more stringent filtration of the sequencing data which can be especially useful when using a naïve phage display library and the diversity and carry through of phage fusions can be quite high. Finally, analysis to detect selected clones in unrelated datasets and also in reversed analysis datasets may provide even higher stringency to filter for binders with very high selectivity. In addition, the data presented here also demonstrates that the criteria used to identify

CDR regions for this inclusion in the datasets is of high importance, and less stringent criteria will produce larger datasets that may lead to more accurate selection of antibodies with the desired binding properties.

In conclusion, high specificity binders against the study's challenging targets could not be obtained. However, the study identified phage-antibody fusions positive against phytase, identified immobilisation and novel panning methods for phytic acid, and described a streamlined and highly efficient subcloning strategy for the camelid VHH library used within this study. In addition, novel bioinformatics analysis tools are described to enchance CDR identification, and to select antibody candidates.

Chapter 8: Future Approaches

Different approaches may be considered to improve the likelihood of success of the project. The soluble antibodies identified during this study lost their binding ability after being expressed as MBP fusions. This effect has been descrbied in the literature and may be due to steric hinderance caused by MBP. It may be beneficial to first optimise the growth conditions and expression time in order to possibly prevent degredation of the antibodies and the loss of function. Furthermore, the MBP could be cleaved from the antibody to reduce the likelihood of interfearance. Alternatively, the antibodies may be expressed as viral fusions to assess whether the phage particle forms a vital part of the binding complex.

Future approaches may benefit by panning against the targets in solution, rather than in a 96-well format. This will increase the surface area available to the target antigens so that they can be immobilised. The phytic acid molecule may be immobilised on TiO_2 beads and the phytases may be engineered with a biotin tag so that they can be captured by streptavidin coated magnetic beads.

A simpler approach to identifying an antibody against the active conformation of phytase could involve an inhibitory enzyme activity assay. An inhibitory assay would demonstrate whether any of the clones picked from a panning experiment against active phytase were binding specifically to the active site of the enzyme, an epitope only available when the enzyme was in an active form.

Phytic acid may also be conjugated to a carrier protein that could enable the target to be immobilised more accessibly than it was on the metal supports used during this study. This has been done successfully in the literature for other hapten targets and may be a promising alternative.

Finally, this study focused on the development of the bioinformatics approach in terms of how it might benefit phage display and isolating novel antibodies from naïve libraries. At present, there is little discussion in the literature as to what analysis is suitable for different targets, ligand types and phage libraries. A thourough comparison of different data analysis strategies followed by empirical testing of the top candidates will help to shed light on the most effective and efficient methods.

References

'T HOEN, P. A. C., JIRKA, S. M. G., TEN BROEKE, B. R., SCHULTES, E. A., AGUILERA, B., PANG, K. H., HEEMSKERK, H., AARTSMA-RUS, A., VAN OMMEN, G. J. & DEN DUNNEN, J. T. 2012. Phage display screening without repetitious selection rounds. *Analytical Biochemistry,* 421**,** 622-631.

AC'T HOEN, P., JIRKA, S. M., BRADLEY, R., SCHULTES, E. A., AGUILERA, B., PANG, K. H., HEEMSKERK, H., AARTSMA-RUS, A., VAN OMMEN, G. J. & DEN DUNNEN, J. T. 2012. Phage display screening without repetitious selection rounds. *Analytical biochemistry,* 421**,** 622-631.

AL QARAGHULI, M. M., PALLIYIL, S., BROADBENT, G., CULLEN, D. C., CHARLTON, K. A. & PORTER, A. J. 2015. Defining the complementarities between antibodies and haptens to refine our understanding and aid the prediction of a successful binding interaction. *BMC Biotechnology,* 15**,** 1-14.

ARNDT, K. M., MÜLLER, K. M. & PLÜCKTHUN, A. 2001. Helix-stabilized fv (hsfv) antibody fragments: substituting the constant domains of a fab fragment for a heterodimeric coiled-coil domain11Edited by I. A. Wilson. *Journal of Molecular Biology,* 312**,** 221-228.

AUGSPURGER, N. R. & BAKER, D. H. 2004. High dietary phytase levels maximize phytate-phosphorus utilization but do not affect protein utilization in chicks fed phosphorus- or amino acid-deficient diets. *J Anim Sci,* 82**,** 1100-7.

BAEK, H., SUK, K.-H., KIM, Y.-H. & CHA, S. 2002. An improved helper phage system for efficient isolation of specific antibody molecules in phage display. *Nucleic Acids Research,* 30**,** e18.

BAILEY, T. L. & ELKAN, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol,* 2**,** 28-36.

BERTANI, G. & WEIGLE, J. J. 1953. Host controlled variation in bacterial viruses. *J Bacteriol,* 65**,** 113-21.

BHAVSAR, K. & KHIRE, J. M. 2014. Current research and future perspectives of phytase bioprocessing. *RSC Advances,* 4**,** 26677-26691.

BITAR, K. & REINHOLD, J. G. 1972. Phytase and alkaline phosphatase activities in intestinal mucosae of rat, chicken, calf, and man. *Biochimica et Biophysica Acta (BBA)-Enzymology,* 268**,** 442-452.

BODER, E. T. & WITTRUP, K. D. 1997. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol,* 15**,** 553-7.

BOEKE, J. D., MODEL, P. & ZINDER, N. D. 1982. Effects of bacteriophage f1 gene III protein on the host cell membrane. *Molecular and General Genetics MGG,* 186**,** 185-192.

BREITLING, F., DUBEL, S., SEEHAUS, T., KLEWINGHAUS, I. & LITTLE, M. 1991. A surface expression vector for antibody screening. *Gene,* 104**,** 147-53.

CERINO BADONE, F., AMELOTTI, M., CASSANI, E. & PILU, R. 2012. Study of Low Phytic Acid1-7 (lpa1-7), a New ZmMRP4 Mutation in Maize. *Journal of Heredity,* 103**,** 598-605.

CHARLTON, K., HARRIS, W. J. & PORTER, A. J. 2001. The isolation of super-sensitive anti-hapten antibodies from combinatorial antibody libraries derived from sheep. *Biosensors Bioelectron,* 16.
CHARLTON, K. A. & PORTER, A. J. 2002. Isolation of anti-hapten specific antibody fragments from combinatorial libraries. *Methods Mol Biol,* 178.
CHEN, Q. C. & LI, B. W. 2003. Separation of phytic acid and other related inositol phosphates by high-performance ion chromatography and its applications. *J Chromatogr A,* 1018**,** 41-52.
CHUNG, W.-J., LEE, D.-Y. & YOO, S. Y. 2014. Chemical modulation of M13 bacteriophage and its functional opportunities for nanomedicine. *International Journal of Nanomedicine,* 9**,** 5825-5836.
CONRATH, K. E., LAUWEREYS, M., GALLENI, M., MATAGNE, A., FRERE, J. M., KINNE, J., WYNS, L. & MUYLDERMANS, S. 2001. Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae. *Antimicrob Agents Chemother,* 45**,** 2807-12.
COSGROVE, D., IRVING, G. & BROMFIELD, S. 1970. Inositol phosphate phosphatases of microbiological origin. The isolation of soil bacteria having inositol phosphate phosphatase activity. *Australian Journal of Biological Sciences,* 23**,** 339-344.
COSGROVE, D. J. 1966. *Reviews of Pure and Applied Chemistry,* 16**,** 209-252.
COSGROVE, D. J. 1980. The determination of myo-inositol hexakisphosphate (phytate). *Journal of the Science of Food and Agriculture,* 31**,** 1253-1256.
COSTELLO, A. J. R., T. GLONEK, AND T. C. MYERS 1976. 31P nuclearmagnetic resonance-pH titrations of myo-inositol hexaphosphate. *Carbohydrate Research,* 46**,** 159-171.
COWIESON, A., WILCOCK, P. & BEDFORD, M. 2011. Super-dosing effects of phytase in poultry and other monogastrics. *World's Poultry Science Journal,* 67**,** 225.
COWIESON, A. J., ACAMOVIC, T. & BEDFORD, M. R. 2004. The effects of phytase and phytic acid on the loss of endogenous amino acids and minerals from broiler chickens. *Br Poult Sci,* 45**,** 101-8.
COWIESON, A. J., ACAMOVIC, T. & BEDFORD, M. R. 2006. Phytic Acid and Phytase: Implications for Protein Utilization by Poultry. *Poultry Science,* 85**,** 878-885.
COWIESON, A. J., BEDFORD, M. R., SELLE, P. H. & RAVINDRAN, V. 2009. Phytate and microbial phytase: implications for endogenous nitrogen losses and nutrient availability. *World's Poultry Science Journal,* 65**,** 401-418.
COWIESON, A. J., RAVINDRAN, V. & SELLE, P. H. 2008. Influence of dietary phytic acid and source of microbial phytase on ileal endogenous amino acid flows in broiler chickens. *Poult Sci,* 87**,** 2287-99.
CRICK, F. H., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. 1961. General nature of the genetic code for proteins. *Nature,* 192**,** 1227-32.
CROZE, M. L. & SOULAGE, C. O. 2013. Potential role and therapeutic interests of myo-inositol in metabolic diseases. *Biochimie,* 95**,** 1811-27.

D'ANGELO, S., KUMAR, S., NARANJO, L., FERRARA, F., KISS, C. & BRADBURY, A. R. M. 2014. From deep sequencing to actual clones. *Protein Engineering Design and Selection,* 27**,** 301-307.
DERSJANT-LI, Y., AWATI, A., SCHULZE, H. & PARTRIDGE, G. 2015. Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. *Journal of the Science of Food and Agriculture,* 95**,** 878-896.
DOST, K. & TOKUL, O. 2006. Determination of phytic acid in wheat and wheat products by reverse phase high performance liquid chromatography. *Analytica Chimica Acta,* 558**,** 22-27.
ELVIN A. KABAT, T. T. W., CARL FOELLER, HAROLD M. PERRY, KAY S. GOTTESMAN 1991. *Sequences of Proteins of Immunological Interest*, DIANE Publishing.
ERPEL, F., RESTOVIC, F. & ARCE-JOHNSON, P. 2016a. Development of phytase-expressing chlamydomonas reinhardtii for monogastric animal nutrition. *BMC Biotechnol,* 16**,** 29.
ERPEL, F., RESTOVIC, F. & ARCE-JOHNSON, P. 2016b. Development of phytase-expressing chlamydomonas reinhardtii for monogastric animal nutrition. *BMC Biotechnology,* 16**,** 29.
FAGHIHIAN, H. & FARSANI, S. N. 2013. Modification of polyacrylamide–β–zeolite composite by phytic acid for the removal of lead from aqueous solutions. *Polish Journal of Chemical Technology,* 15**,** 1-6.
FISKE, C. H. & SUBBAROW, Y. 1925. THE COLORIMETRIC DETERMINATION OF PHOSPHORUS. *Journal of Biological Chemistry,* 66**,** 375-400.
GIBSON, D. M. 1987. Production of extracellular phytase fromAspergillus ficuum on starch media. *Biotechnology letters,* 9**,** 305-310.
GLANVILLE, J., D'ANGELO, S., KHAN, T. A., REDDY, S. T., NARANJO, L., FERRARA, F. & BRADBURY, A. R. 2015. Deep sequencing in library selection projects: what insight does it bring? *Curr Opin Struct Biol,* 33**,** 146-60.
GLANVILLE, J., ZHAI, W., BERKA, J., TELMAN, D., HUERTA, G., MEHTA, G. R., NI, I., MEI, L., SUNDAR, P. D., DAY, G. M. R., COX, D., RAJPAL, A. & PONS, J. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 20216-20221.
GOLOVAN, S. P., MEIDINGER, R. G., AJAKAIYE, A., COTTRILL, M., WIEDERKEHR, M. Z., BARNEY, D. J., PLANTE, C., POLLARD, J. W., FAN, M. Z., HAYES, M. A., LAURSEN, J., HJORTH, J. P., HACKER, R. R., PHILLIPS, J. P. & FORSBERG, C. W. 2001. Pigs expressing salivary phytase produce low-phosphorus manure. *Nat Biotechnol,* 19**,** 741-5.
GREINER, R. & FAROUK, A.-E. 2007. Purification and Characterization of a Bacterial Phytase Whose Properties Make it Exceptionally Useful as a Feed Supplement. *The Protein Journal,* 26**,** 467-474.
GREINER, R., KONIETZNY, U. & JANY, K.-D. 1993. Purification and characterization of two phytases from Escherichia coli. *Archives of biochemistry and biophysics,* 303**,** 107-113.
GRIFFITHS, A. D. & DUNCAN, A. R. 1998. Strategies for selection of antibodies by phage display. *Curr Opin Biotechnol,* 9**,** 102-8.
GRIFFITHS, A. D., WILLIAMS, S. C., HARTLEY, O., TOMLINSON, I. M., WATERHOUSE, P., CROSBY, W. L., KONTERMANN, R. E., JONES, P. T., LOW, N. M., ALLISON, T. J. & ET AL. 1994. Isolation of high

affinity human antibodies directly from large synthetic repertoires. *Embo j,* 13**,** 3245-60.
GUSE, A. H., GOLDWICH, A., WEBER, K. & MAYR, G. W. 1995. Non-radioactive, isomer-specific inositol phosphate mass determinations: high-performance liquid chromatography-micro-metal-dye detection strongly improves speed and sensitivity of analyses from cells and micro-enzyme assays. *J Chromatogr B Biomed Appl,* 672**,** 189-98.
HAMMERS, C. M. & STANLEY, J. R. 2014. Antibody phage display: technique and applications. *The Journal of investigative dermatology,* 134**,** e17.
HANES, J. & PLUCKTHUN, A. 1997. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A,* 94**,** 4937-42.
HEMADOU, A., GIUDICELLI, V., SMITH, M. L., LEFRANC, M. P., DUROUX, P., KOSSIDA, S., HEINER, C., HEPLER, N. L., KUIJPERS, J., GROPPI, A., KORLACH, J., MONDON, P., OTTONES, F., JACOBIN-VALAT, M. J., LAROCHE-TRAINEAU, J. & CLOFENT-SANCHEZ, G. 2017. Pacific Biosciences Sequencing and IMGT/HighV-QUEST Analysis of Full-Length Single Chain Fragment Variable from an In Vivo Selected Phage-Display Combinatorial Library. *Front Immunol,* 8**,** 1796.
HOLLIGER, P. & RIECHMANN, L. 1997. A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. *Structure,* 5**,** 265-275.
HOOGENBOOM, H. R., GRIFFITHS, A. D., JOHNSON, K. S., CHISWELL, D. J., HUDSON, P. & WINTER, G. 1991. Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res,* 19**,** 4133-7.
HUSE, W. D., SASTRY, L., IVERSON, S. A., KANG, A. S., ALTING-MEES, M., BURTON, D. R., BENKOVIC, S. J. & LERNER, R. A. 1989. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science,* 246**,** 1275-81.
IGBASAN, F. A., MANNER, K., MIKSCH, G., BORRISS, R., FAROUK, A. & SIMON, O. 2000. Comparative studies on the in vitro properties of phytases from various microbial origins. *Arch Tierernahr,* 53**,** 353-73.
KALSI, H. K., SINGH, R., DHALIWAL, H. S. & KUMAR, V. 2016. Phytases from Enterobacter and Serratia species with desirable characteristics for food and feed applications. *3 Biotech,* 6**,** 64.
KIES, A. K., KEMME, P. A., SEBEK, L. B. J., DIEPEN, V. J. T. M. & JONGBLOED, A. W. 2006. Effect of graded doses and a high dose of microbial phytase on the digestibility of various minerals in weaner pigs. *Journal of Animal Science,* 84**,** 1169 - 1175.
KIM, Y.-O., KIM, H.-K., BAE, K.-S., YU, J.-H. & OH, T.-K. 1998. Purification and properties of a thermostable phytase from Bacillus sp. DS11. *Enzyme and Microbial Technology,* 22**,** 2-7.
KING, D. J., BOWERS, P. M., KEHRY, M. R. & HORLICK, R. A. 2014. Mammalian cell display and somatic hypermutation in vitro for human antibody discovery. *Curr Drug Discov Technol,* 11**,** 56-64.
KIRK, O., BORCHERT, T. V. & FUGLSANG, C. C. 2002. Industrial enzyme applications. *Curr Opin Biotechnol,* 13**,** 345-51.
KOHLER, G. & MILSTEIN, C. 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature,* 256**,** 495-497.

KOVALENKO, O. V., OLLAND, A., PICHE-NICHOLAS, N., GODBOLE, A., KING, D., SVENSON, K., CALABRO, V., MULLER, M. R., BARELLE, C. J., SOMERS, W., GILL, D. S., MOSYAK, L. & TCHISTIAKOVA, L. 2013. Atypical antigen recognition mode of a shark immunoglobulin new antigen receptor (IgNAR) variable domain characterized by humanization and structural analysis. *J Biol Chem,* 288**,** 17408-19.
KOVALEVA, M., FERGUSON, L., STEVEN, J., PORTER, A. & BARELLE, C. 2014. Shark variable new antigen receptor biologics - a novel technology platform for therapeutic drug development. *Expert Opin Biol Ther,* 14**,** 1527-39.
KOVALEVA, M., JOHNSON, K., STEVEN, J., BARELLE, C. J. & PORTER, A. 2017. Therapeutic Potential of Shark Anti-ICOSL VNAR Domains is Exemplified in a Murine Model of Autoimmune Non-Infectious Uveitis. *Front Immunol,* 8**,** 1121.
KRAMER, R. A., COX, F., VAN DER HORST, M., VAN DEN OUDENRIJN, S., RES, P. C. M., BIA, J., LOGTENBERG, T. & DE KRUIF, J. 2003. A novel helper phage that improves phage display selection efficiency by preventing the amplification of phages without recombinant protein. *Nucleic Acids Research,* 31**,** e59.
KREBBER, A., BURMESTER, J. & PLÜCKTHUN, A. 1996. Inclusion of an upstream transcriptional terminator in phage display vectors abolishes background expression of toxic fusions with coat protein g3p. *Gene,* 178**,** 71-74.
KUMAR, V., SINHA, A. K., MAKKAR, H. P. S. & BECKER, K. 2010. Dietary roles of phytate and phytase in human nutrition: A review. *Food Chemistry,* 120**,** 945-959.
KWEON, H. K. & HÅKANSSON, K. 2006. Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. *Analytical Chemistry,* 78**,** 1743-1749.
LEE, C.-Y., CHIU, Y.-C., WANG, L.-B., KUO, Y.-L., CHUANG, E. Y., LAI, L.-C. & TSAI, M.-H. 2013. Common applications of next-generation sequencing technologies in genomic research. *Translational Cancer Research,* 2**,** 33-45.
LEFRANC, M. P. 2001. IMGT, the international ImMunoGeneTics database. *Nucleic acids research,* 29**,** 207-209.
LEI, X. G., WEAVER, J. D., MULLANEY, E., ULLAH, A. H. & AZAIN, M. J. 2013. Phytase, a new life for an "old" enzyme. *Annu. Rev. Anim. Biosci.,* 1**,** 283-309.
LEMTIRI-CHLIEH, F., MACROBBIE, E. A., WEBB, A. A., MANISON, N. F., BROWNLEE, C., SKEPPER, J. N., CHEN, J., PRESTWICH, G. D. & BREARLEY, C. A. 2003. Inositol hexakisphosphate mobilizes an endomembrane store of calcium in guard cells. *Proc Natl Acad Sci U S A,* 100**,** 10091-5.
LI, Y., COCKBURN, W., KILPATRICK, J. & WHITELAM, G. C. 1999. Selection of Rabbit Single-chain Fv Fragments against the Herbicide Atrazine using a New Phage Display System. *Food and Agricultural Immunology,* 11**,** 5-17.
LIU, Z.-H., HUANG, D., FU, X.-J., CHENG, P. & DU, E.-Q. 2018. Comparison of three commonly used fusion tags for the expression of nanobodies in the cytoplasm of Escherichia coli. *Biotechnology & Biotechnological Equipment,* 32**,** 462-469.

LOEWUS, F. A. & LOEWUS, M. W. 1983. myo-Inositol:Its Biosynthesis and Metabolism. *Annual Review of Plant Physiology,* 34**,** 137-161.
LOLAS, G. M. & MARKAKIS, P. 1975. Phytic acid and other phosphorus compounds of beans (Phaseolus vulgaris L.). *Journal of Agricultural and Food Chemistry,* 23**,** 13-15.
LOVGREN, J., PURSIHEIMO, J. P., PYYKKO, M., SALMI, J. & LAMMINMAKI, U. 2016. Next generation sequencing of all variable loops of synthetic single framework scFv-Application in anti-HDL antibody selections. *N Biotechnol,* 33**,** 790-796.
LURIA, S. E. & HUMAN, M. L. 1952. A nonhereditary, host-induced variation of bacterial viruses. *J Bacteriol,* 64**,** 557-69.
MATOCHKO, W. L., NG, S., JAFARI, M. R., ROMANIUK, J., TANG, S. K. Y. & DERDA, R. 2012. Uniform amplification of phage display libraries in monodisperse emulsions. *Methods,* 58**,** 18-27.
MCCAFFERTY, J., GRIFFITHS, A. D., WINTER, G. & CHISWELL, D. J. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature,* 348**,** 552-4.
MIYAZAKI, N., KIYOSE, N., AKAZAWA, Y., TAKASHIMA, M., HAGIHARA, Y., INOUE, N., MATSUDA, T., OGAWA, R., INOUE, S. & ITO, Y. 2015. Isolation and characterization of antigen-specific alpaca (Lama pacos) VHH antibodies by biopanning followed by high-throughput sequencing. *The Journal of Biochemistry,* 158**,** 205-215.
MULLANEY, E. J. & ULLAH, A. H. 2003. The term phytase comprises several different classes of enzymes. *Biochem Biophys Res Commun,* 312**,** 179-84.
MULLER, M. R., SAUNDERS, K., GRACE, C., JIN, M., PICHE-NICHOLAS, N., STEVEN, J., O'DWYER, R., WU, L., KHETEMENEE, L., VUGMEYSTER, Y., HICKLING, T. P., TCHISTIAKOVA, L., OLLAND, S., GILL, D., JENSEN, A. & BARELLE, C. J. 2012. Improving the pharmacokinetic properties of biologics by fusion to an anti-HSA shark VNAR domain. *MAbs,* 4**,** 673-85.
MUYLDERMANS, S. 2013. Nanobodies: natural single-domain antibodies. *Annu Rev Biochem,* 82**,** 775-97.
NAQID, I. A., OWEN, J. P., MADDISON, B. C., SPILIOTOPOULOS, A., EMES, R. D., WARRY, A., TCHORZEWSKA, M. A., MARTELLI, F., GOSLING, R. J., DAVIES, R. H., LA RAGIONE, R. M. & GOUGH, K. C. 2016. Mapping polyclonal antibody responses to bacterial infection using next generation phage display. *Sci Rep,* 6**,** 24232.
NGUBANE, N. A. C., GRESH, L., IOERGER, T. R., SACCHETTINI, J. C., ZHANG, Y. J., RUBIN, E. J., PYM, A. & KHATI, M. 2013. High-Throughput Sequencing Enhanced Phage Display Identifies Peptides That Bind Mycobacteria. *PLoS ONE,* 8**,** e77844.
NISSIM, A., HOOGENBOOM, H. R., TOMLINSON, I. M., FLYNN, G., MIDGLEY, C., LANE, D. & WINTER, G. 1994. Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *Embo j,* 13**,** 692-8.
PANDE, R. & MISHRA, H. N. 2015. Fourier Transform Near-Infrared Spectroscopy for rapid and simple determination of phytic acid content in green gram seeds (Vigna radiata). *Food Chemistry,* 172**,** 880-884.
PATLEWICZ, G., CASATI, S., BASKETTER, D. A., ASTURIOL, D., ROBERTS, D. W., LEPOITTEVIN, J. P., WORTH, A. P. & ASCHBERGER, K. 2016.

Can currently available non-animal methods detect pre and pro-haptens relevant for skin sensitization? *Regul Toxicol Pharmacol*.
PATWARDHAN, V. N. 1937. The occurrence of a phytin-splitting enzyme in the intestines of albino rats. *Biochemical Journal,* 31**,** 560-564.
PHILLIPPY, B. Q. 1999. Susceptibility of wheat and Aspergillus niger phytases to inactivation by gastrointestinal enzymes. *J Agric Food Chem,* 47**,** 1385-8.
POTCHANAKORN, M. & POTTER, L. M. 1987. Biological values of phosphorus from various sources for young turkeys. *Poult Sci,* 66**,** 505-13.
POWAR, V. K. & JAGANNATHAN, V. 1982. Purification and properties of phytate-specific phosphatase from Bacillus subtilis. *Journal of Bacteriology,* 151**,** 1102-1108.
POWELL, S., BIDNER, T. D. & SOUTHERN, L. L. 2011. Phytase supplementation improved growth performance and bone characteristics in broilers fed varying levels of dietary calcium. *Poult Sci,* 90**,** 604-8.
R. E. SOUTHWOOD, T. & HENDERSON, P. 2000. *Ecological Methods 3rd edition*.
RAPP, C., LANTZSCH, H. J. & DROCHNER, W. 2001. Hydrolysis of phytic acid by intrinsic plant and supplemented microbial phytase (Aspergillus niger) in the stomach and small intestine of minipigs fitted with re-entrant cannulas. 3. Hydrolysis of phytic acid (IP6) and occurrence of hydrolysis products (IP5, IP4, IP3 and IP2). *J Anim Physiol Anim Nutr (Berl),* 85**,** 420-30.
RATHNAYAKE, U. A., SENAPATHI, T., SANDARUWAN, C., GUNAWARDENE, S., KARUNARATNE, V. & KOTTEGODA, N. 2018. Rice bran nanofiber composites for stabilization of phytase. *Chem Cent J,* 12**,** 28.
RAVINDRAN, V., COWIESON, A. J. & SELLE, P. H. 2008. Influence of dietary electrolyte balance and microbial phytase on growth performance, nutrient utilization, and excreta quality of broiler chickens. *Poult Sci,* 87**,** 677-88.
RAVN, P., DANIELCZYK, A., JENSEN, K. B., KRISTENSEN, P., CHRISTENSEN, P. A., LARSEN, M., KARSTEN, U. & GOLETZ, S. 2004. Multivalent scFv display of phagemid repertoires for the selection of carbohydrate-specific antibodies and its application to the Thomsen-Friedenreich antigen. *J Mol Biol,* 343**,** 985-96.
RAVN, U., DIDELOT, G., VENET, S., NG, K.-T., GUENEAU, F., ROUSSEAU, F., CALLOUD, S., KOSCO-VILBOIS, M. & FISCHER, N. 2013. Deep sequencing of phage display libraries to support antibody discovery. *Methods,* 60**,** 99-110.
RAVN, U., GUENEAU, F., BAERLOCHER, L., OSTERAS, M., DESMURS, M., MALINGE, P., MAGISTRELLI, G., FARINELLI, L., KOSCO-VILBOIS, M. H. & FISCHER, N. 2010a. By-passing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Research,* 38**,** e193-e193.
RAVN, U., GUENEAU, F., BAERLOCHER, L., OSTERAS, M., DESMURS, M., MALINGE, P., MAGISTRELLI, G., FARINELLI, L., KOSCO-VILBOIS, M. H. & FISCHER, N. 2010b. By-passing in vitro screening - Next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Research,* 38.
REBOLLO, I. R., SABISZ, M., BAERISWYL, V. & HEINIS, C. 2014. Identification of target-binding peptide motifs by high-throughput

sequencing of phage-selected peptides. *Nucleic acids research,* 42**,** e169-e169.
REDDY, S. T., GE, X., MIKLOS, A. E., HUGHES, R. A., KANG, S. H., HOI, K. H., CHRYSOSTOMOU, C., HUNICKE-SMITH, S. P., IVERSON, B. L., TUCKER, P. W., ELLINGTON, A. D. & GEORGIOU, G. 2010. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nature Biotechnology,* 28**,** 965.
RIECHMANN, L. & HOLLIGER, P. 1997. The C-Terminal Domain of TolA Is the Coreceptor for Filamentous Phage Infection of E. coli. *Cell,* 90**,** 351-360.
RODRIGUEZ, E., PORRES, J. M., HAN, Y. & LEI, X. G. 1999. Different sensitivity of recombinant Aspergillus niger phytase (r-PhyA) and Escherichia coli pH 2.5 acid phosphatase (r-AppA) to trypsin and pepsin in vitro. *Arch Biochem Biophys,* 365**,** 262-7.
RONDOT, S., KOCH, J., BREITLING, F. & DUBEL, S. 2001. A helper phage to improve single-chain antibody presentation in phage display. *Nat Biotechnol,* 19**,** 75-8.
ROUET, R., JACKSON, K. J. L., LANGLEY, D. B. & CHRIST, D. 2018. Next-Generation Sequencing of Antibody Display Repertoires. *Front Immunol,* 9**,** 118.
RYVKIN, A., ASHKENAZY, H., SMELYANSKI, L., KAPLAN, G., PENN, O., WEISS-OTTOLENGHI, Y., PRIVMAN, E., NGAM, P. B., WOODWARD, J. E., MAY, G. D., BELL, C., PUPKO, T. & GERSHONI, J. M. 2012. Deep Panning: steps towards probing the IgOme. *PLoS One,* 7**,** e41469.
SACHDEV S. SIDHU, C. R. G. 2017 *Phage Display In Biotechnology and Drug Discovery*, CRC Press.
SAGGY, I., WINE, Y., SHEFET-CARASSO, L., NAHARY, L., GEORGIOU, G. & BENHAR, I. 2012. Antibody isolation from immunized animals: comparison of phage display and antibody discovery via V gene repertoire mining. *Protein Eng Des Sel,* 25**,** 539-49.
SALEMA, V. & FERNANDEZ, L. A. 2013. High yield purification of nanobodies from the periplasm of E. coli as fusions with the maltose binding protein. *Protein Expr Purif,* 91**,** 42-8.
SASIREKHA, B., BEDASHREE, T. & CHAMPA, K. 2012. Statistical optimization of medium components for improved phytase production by Pseudomonas aeruginosa. *Int. J. Chem. Tech Res,* 4**,** 891-895.
SASSO, E., PACIELLO, R., D'AURIA, F., RICCIO, G., FROECHLICH, G., CORTESE, R., NICOSIA, A., DE LORENZO, C. & ZAMBRANO, N. 2015. One-Step Recovery of scFv Clones from High-Throughput Sequencing-Based Screening of Phage Display Libraries Challenged to Cells Expressing Native Claudin-1. *Biomed Res Int,* 2015**,** 703213.
SATHE, M., DERVENI, M., BROADBENT, G., BODLENNER, A., CHARLTON, K., RAVI, B., ROHMER, M., SIMS, M. R. & CULLEN, D. C. 2011. Synthesis and characterisation of immunogens for the production of antibodies against small hydrophobic molecules with biosignature properties. *Anal Chim Acta,* 708**,** 97-106.
SELLE, P. H. & RAVINDRAN, V. 2007. Microbial phytase in poultry nutrition. *Animal Feed Science and Technology,* 135**,** 1-41.
SHI, B., WANG, H., GUO, S., XU, Y., LI, Z. & GU, J. 2007. Protein III-based single-chain antibody phage display using bacterial cells bearing an additional genome of a gene-III-lacking helper phage. *Biotechniques,* 42**,** 760-5.

SHI, P., HUANG, H., WANG, Y., LUO, H., WU, B., MENG, K., YANG, P. & YAO, B. 2008. A novel phytase gene appA from Buttiauxella sp. GC21 isolated from grass carp intestine. *Aquaculture,* 275**,** 70-75.
SHIEH, T. & WARE, J. 1968. Survey of microorganisms for the production of extracellular phytase. *Applied Microbiology,* 16**,** 1348-1351.
SHIM, J. H. & OH, B. C. 2012. Characterization and application of calcium-dependent beta-propeller phytase from Bacillus amyloliquefaciens DS11. *J Agric Food Chem,* 60**,** 7532-7.
SHIMIZU, M. 1992. Purification and Characterization of Phytase from Bacillus suhtilis (natto) N–77. *Bioscience, Biotechnology, and Biochemistry,* 56**,** 1266-1269.
SHIRLEY, R. & EDWARDS, H. 2003. Graded levels of phytase past industry standards improves broiler performance. *Poultry Science,* 82**,** 671-680.
SIMONS, P. C. M., VERSTEEGH, H. A. J., JONGBLOED, A. W., KEMME, P. A., SLUMP, P., BOS, K. D., WOLTERS, M. G. E., BEUDEKER, R. F. & VERSCHOOR, G. J. 1990. Improvement of phosphorus availability by microbial phytase in broilers and pigs. *British Journal of Nutrition,* 64**,** 525-540.
SKERRA, A. & PLUCKTHUN, A. 1988. Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science,* 240**,** 1038-41.
SMITH, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science,* 228**,** 1315-7.
SOLTES, G., BARKER, H., MARMAI, K., PUN, E., YUEN, A. & WIERSMA, E. J. 2003. A new helper phage and phagemid vector system improves viral display of antibody Fab fragments and avoids propagation of insert-less virions. *J Immunol Methods,* 274**,** 233-44.
SPILIOTOPOULOS, A., OWEN, J. P., MADDISON, B. C., DREVENY, I., REES, H. C. & GOUGH, K. C. 2015. Sensitive recovery of recombinant antibody clones after their in silico identification within NGS datasets. *Journal of Immunological Methods,* 420**,** 50-55.
SULAKVELIDZE, A., ALAVIDZE, Z. & MORRIS, J. G. 2001. Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy,* 45**,** 649-659.
TAHIR, M., SHIM, M. Y., WARD, N. E., WESTERHAUS, M. O. & PESTI, G. M. 2012. Evaluation of near-infrared reflectance spectroscopy (NIRS) techniques for total and phytate phosphorus of common poultry feed ingredients. *Poultry Science,* 91**,** 2540-2547.
TALAMOND, P., DOULBEAU, S., ROCHETTE, I. & GUYOT, J. P. 2000. Anion-exchange high-performance liquid chromatography with conductivity detection for the analysis of phytic acid in food. *J Chromatogr A,* 871**,** 7-12.
TSAO, G. T., ZHENG, Y., LU, J. & GONG, C. S. 1997. Adsorption of heavy metal ions by immobilized phytic acid. *Appl Biochem Biotechnol,* 63-65**,** 731-41.
TSUMOTO, K., NAKAOKI, Y., UEDA, Y., OGASAHARA, K., YUTANI, K., WATANABE, K. & KUMAGAI, I. 1994. *Effect of the Order of Antibody Variable Regions on the Expression of the Single-Chain HyHEL10 Fv Fragment in Escherichia coli and the Thermodynamic Analysis of Its Antigen-Binding Properties*.
TURNER, K. B., NACIRI, J., LIU, J. L., ANDERSON, G. P., GOLDMAN, E. R. & ZABETAKIS, D. 2016. Next-Generation Sequencing of a Single

Domain Antibody Repertoire Reveals Quality of Phage Display Selected Candidates. *PLoS ONE,* 11**,** e0149393.
UBAH, O. C., STEVEN, J., KOVALEVA, M., FERGUSON, L., BARELLE, C., PORTER, A. J. R. & BARELLE, C. J. 2017. Novel, Anti-hTNF-alpha Variable New Antigen Receptor Formats with Enhanced Neutralizing Potency and Multifunctionality, Generated for Therapeutic Development. *Front Immunol,* 8**,** 1780.
ULUSOY, U. & SIMSEK, S. 2005. Lead removal by polyacrylamide-bentonite and zeolite composites: effect of phytic acid immobilization. *J Hazard Mater,* 127**,** 163-71.
ULUSOY, U., ŞIMŞEK, S. & CEYHAN, Ö. Investigations for Modification of Polyacrylamide-Bentonite by Phytic Acid and its Usability in Fe3+, Zn2+ and UO22+ Adsorption. *Adsorption,* 9**,** 165-175.
VATS, P. & BANERJEE, U. C. 2004. Production studies and catalytic properties of phytases (myo-inositolhexakisphosphate phosphohydrolases): an overview. *Enzyme and Microbial Technology,* 35**,** 3-14.
VIEIRA, J. & MESSING, J. 1987a. [1] Production of single-stranded plasmid DNA. *Methods in Enzymology.* Academic Press.
VIEIRA, J. & MESSING, J. 1987b. Production of single-stranded plasmid DNA. *Methods Enzymol,* 153**,** 3-11.
WILSON, M. S. C., BULLEY, S. J., PISANI, F., IRVINE, R. F. & SAIARDI, A. 2015. A novel method for the purification of inositol phosphates from biological samples reveals that no phytate is present in human plasma or urine. *Open Biology,* 5.
WINTER, G., GRIFFITHS, A. D., HAWKINS, R. E. & HOOGENBOOM, H. R. 1994. Making antibodies by phage display technology. *Annu Rev Immunol,* 12**,** 433-55.
WU, B. P., XIAO, B., WAN, T. M., ZHANG, Y. L., ZHANG, Z. S., ZHOU, D. Y., LAI, Z. S. & GAO, C. F. 2001. Construction and selection of the natural immune Fab antibody phage display library from patients with colorectal cancer. *World J Gastroenterol,* 7**,** 811-5.
WYSS, M., BRUGGER, R., KRONENBERGER, A., RÉMY, R., FIMBEL, R., OESTERHELT, G., LEHMANN, M. & VAN LOON, A. P. G. M. 1999. Biochemical Characterization of Fungal Phytases (myo-Inositol Hexakisphosphate Phosphohydrolases): Catalytic Properties. *Applied and Environmental Microbiology,* 65**,** 367-373.
XU, J. L. & DAVIS, M. M. 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity,* 13**,** 37-45.
XU, Y., LIU, X.-H. & PRESTWICH, G. D. 2005. Synthesis of phosphatase-resistant analogues of phytic acid (InsP6). *Tetrahedron Letters,* 46**,** 8311-8314.
YAMADA, K., MINODA, Y. & YAMAMOTO, S. 1968. Phytase from Aspergillus terreus: Part I. Production, Purification and Some General Properties of the Enzyme*. *Agricultural and Biological Chemistry,* 32**,** 1275-1282.
YAN, J., LI, G., HU, Y., OU, W. & WAN, Y. 2014. Construction of a synthetic phage-displayed Nanobody library with CDR3 regions randomized by trinucleotide cassettes for diagnostic applications. *Journal of Translational Medicine,* 12**,** 343.

YANG, L., LIU, H. & HU, N. 2007. Assembly of electroactive layer-by-layer films of myoglobin and small-molecular phytic acid. *Electrochemistry Communications,* 9**,** 1057-1061.
YANG, W., YOON, A., LEE, S., KIM, S., HAN, J. & CHUNG, J. 2017. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Exp Mol Med,* 49**,** e308.
YOON, S. J., CHOI, Y. J., MIN, H. K., CHO, K. K., KIM, J. W., LEE, S. C. & JUNG, Y. H. 1996. Isolation and identification of phytase-producing bacterium, Enterobacter sp. 4, and enzymatic properties of phytase enzyme. *Enzyme and microbial technology,* 18**,** 449-454.
YUN, Y.-H., ETESHOLA, E., BHATTACHARYA, A., DONG, Z., SHIM, J., CONFORTI, L., KIM, D., J SCHULZ, M., AHN, C. & WATTS, N. 2009. Tiny Medicine: Nanomaterial-Based Biosensors. *Sensors (Basel, Switzerland),* 9**,** 9275-99.
ZENG, Z. K., WANG, D., PIAO, X. S., LI, P. F., ZHANG, H. Y., SHI, C. X. & YU, S. K. 2014. Effects of Adding Super Dose Phytase to the Phosphorus-deficient Diets of Young Pigs on Growth Performance, Bone Quality, Minerals and Amino Acids Digestibilities. *Asian-Australasian Journal of Animal Sciences,* 27**,** 237-246.
ZHANG, H., TORKAMANI, A., JONES, T. M., RUIZ, D. I., PONS, J. & LERNER, R. A. 2011. Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems. *Proceedings of the National Academy of Sciences,* 108**,** 13456-13461.
ZHANG, Z. B., KORNEGAY, E. T., RADCLIFFE, J. S., DENBOW, D. M., VEIT, H. P. & LARSEN, C. T. 2000. Comparison of genetically engineered microbial and plant phytase for young broilers. *Poultry Science,* 79**,** 709-717.
ZHUANG, G., KATAKURA, Y., FURUTA, T., OMASA, T., KISHIMOTO, M. & SUGA, K. 2001. A kinetic model for a biopanning process considering antigen desorption and effective antigen concentration on a solid phase. *J Biosci Bioeng,* 91**,** 474-81.