

Engineered OB-fold protein domains as versatile scaffolds for molecular recognition

Matthias Baake¹, John D. Steemson², Jasna Rakonjac³ and Vickery L. Arcus^{2*}

¹*School of Biological Sciences, University of Auckland, Auckland, New Zealand.*

²*Department of Biological Sciences, University of Waikato, Hamilton, New Zealand*

³*Institute for Molecular Biosciences, Massey University, Palmerston North, New Zealand*

**Corresponding author*

Corresponding author address:

Vickery L. Arcus

Department of Biological Sciences

University of Waikato

Private Bag 3105

Hamilton 3216

New Zealand

Ph: 64 7 8384679

Fax: 64 7 8384976

e-mail: varcus@waikato.ac.nz

Summary

OB-fold protein domains are found in all three kingdoms of life. They are small, five-stranded β -barrel protein domains with a common binding face. Over evolutionary time this binding face has been adapted to bind to a wide range of ligands including oligosaccharides, oligonucleotides, proteins and small molecules. In at least one case, enzymatic activity has also evolved at this common face. We have sought to mimic this adaptability *in vitro* and have developed a range of combinatorial protein libraries based on the OB-fold domain from an aspartyl-tRNA synthetase. We have characterised the protein libraries based on recombinant protein expression and solubility and have then used these libraries in phage display selection for binding back to the native ligand (asp-tRNA) as well as to a prototypic protein ligand (lysozyme). OB-fold variants obtained from selection against lysozyme have been characterised for binding and show moderate affinity (in the micromolar range). The three-dimensional structure has been determined for a single OB-fold variant in complex with lysozyme and this structure provides details of the protein-protein interface along with possible routes to tighter-binding variants. The combination of binding loops and a binding face presented by the OB-fold domain libraries (OBodies), which can be adapted for tailor-made molecular interactions, is a new avenue for exploration in seeking alternatives to antibodies in diagnostics and therapeutics.

Keywords:

Phage display, protein engineering, X-ray crystallography, binding, lysozyme, combinatorial libraries.

Introduction

Molecular recognition is central to biological processes, from high-affinity protein-ligand interactions to the more transient protein-protein recognition events of signal transduction pathways. Such events depend on the versatility of proteins, which have been adapted to new roles as organisms have evolved. We have explored this phenomenon using a widely-distributed, small protein domain, the OB-fold, originally named for the observed functions of oligonucleotide and oligosaccharide binding (OB).¹

The OB-fold is a small, stable protein domain that presents a concave β -sheet as an external binding face (Figure 1). In many contexts, this binding face is flanked by loops and in nearly all cases where this domain is present, the same face of the fold is used in ligand binding. The versatility of this domain can be seen in nature where OB-fold domains are found in a wide variety of proteins from species as diverse as humans, yeast and bacteria. In the array of proteins which contain an OB-fold, the domain binds a broad range of ligands, including proteins, oligonucleotides and oligosaccharides, leading Murzin to suggest that the OB-fold is ancient, tolerant to mutation, with a versatile binding face that is easily adapted to specifically bind a range of targets.¹ Examples of the diversity of OB-fold proteins include single stranded DNA binding in the oncogene BRCA2, tRNA anticodon binding in the aspartyl- and lysyl-tRNA synthetases, telomere end binding on chromosomes for the yeast protein Cdc13 and cell-surface oligosaccharide binding for the shiga toxin from *E. coli* and *Shigella dysenteriae*. OB-fold domains also mediate a range of protein-protein interactions in superantigens in the bacterial attack on the human immune system.^{2,3} These observations suggest that the versatility of OB-fold

proteins might be emulated *in vitro* through protein engineering for tailor-made molecular recognition.

Antibodies are the gold standard in the field of protein molecular recognition. To capture a foreign antigen, a small number of antibodies from the immune system's naïve library (which contains approximately 10^7 variants)⁴ recognise the antigen and bind to it with moderate affinity. Selection and maturation then introduces further mutations to generate the tight, highly specific binding required to eliminate the antigen. In this way a staggering array of binding modes can be grafted on to the basic antibody scaffold, to sequester targets varying from small molecules to whole cells.

This strategy has been replicated in the laboratory to produce very large libraries of antibody variants ($>10^{10}$ different clones)^{5,6} that have then been selected for binding to particular targets. Repeated cycles of amplification and selection for binding have been used to “discover” test-tube antibodies with tight and specific molecular binding characteristics. This *in vitro* approach has also been applied to other scaffolds. For example, randomisation and selection by phage display have been used to study and improve the binding of growth hormone and the growth factor heregulin to their respective receptors.^{7,8} “Affibodies” have been developed from libraries of a three-helix bundle domain from staphylococcal protein A.^{9,10} This general area has been the subject of a number of recent reviews.¹¹⁻¹³

There are two key steps in engineering proteins for tailor-made molecular recognition – the generation of diversity (library construction) and selection. For antibodies, libraries of variants have been constructed by combining the natural variation seen in humans with synthesis of randomised genes and incorporation of these genes into the library.^{14,15} For small proteins and peptides random libraries have been constructed entirely synthetically.¹⁶ For larger proteins, techniques based on the polymerase chain reaction (PCR) have been employed to generate variation. These have been variously called DNA shuffling,¹⁵ sexual PCR,^{15,17,18} staggered extension process recombination (StEP),¹⁹ chimeragenesis²⁰ and degenerate oligonucleotide gene shuffling (DOGS).²¹ Selection from a library of variants can then be used, for example, to improve binding to a target, improve enzymatic activity or improve stability.²²⁻²⁴ The single common requirement for selection is that the protein under selection be directly linked to the gene that encodes it. This allows identification of the gene which has been pulled from the original large library. Linking the gene with its encoded protein has been achieved by phage display,⁸ ribosomal display²⁵ or yeast display.²⁶ Different bacterial strains with specific properties have also been used for selection.²⁷

We have investigated the versatility of the OB-fold domain *in vitro* by constructing a range of protein libraries based on this domain and we have used phage display to select variants with selected binding specificities. Here, we report the generation of several different protein libraries based on the OB-fold domain of the aspartyl tRNA synthetase (aspRS-OB) from the hyperthermophilic chrenarchaeon, *Pyrobaculum aerophilum*. These libraries have been generated synthetically using long oligonucleotides with specific

positions randomised, followed by gene assembly using the PCR. Libraries have been tested for the rates of overexpression of their encoded proteins and estimates are made about the fraction of soluble and heat stable proteins encoded by the library. We have used phage display to isolate OB-fold domains which bind to tRNA as a nucleotide target and to lysozyme as a protein target. Further, we have determined the three dimensional structure of a complex between an engineered domain and lysozyme and this reveals the molecular details of the protein-protein interaction along with possible routes to improving the moderate affinity for the complex. To describe our engineered OB-fold domains, we have coined the term OBodies

OBodies are similar to other scaffolds which have been adopted for engineering molecular recognition (for example, single domain antibodies and affibodies) insofar as they are small, stable proteins, easily produced and randomised. However, by primarily randomising one face of the β -barrel protein scaffold, they are fundamentally different from previous scaffolds which have been studied and may provide insights into the evolution of binding at protein faces compared with binding by loops. In addition, the OB-fold is ubiquitous in all three kingdoms and thus, there is the potential to choose scaffolds to suit particular applications. In the present case we report on a thermostable scaffold with its attendant advantages. Using this approach possible future applications include a human OBody scaffold for therapeutics or a yeast scaffold for biotechnological 2-hybrid applications.

Results

The OB-fold

The OB-fold is very well represented in both the Protein Data Bank (PDB) and in Genbank. There are 124 protein structures where this fold is observed in the PDB (excluding multiple structures for the same protein) and a survey of 20 sequenced genomes puts the OB-fold at 28th in a list of the most prevalent architectures.²⁸

At the most general level, the OB-fold is a 5-stranded mixed β -barrel (Figure 1). In the majority of cases the barrel has a Greek-key topology and one end of the barrel is capped by an α -helix.^{1,3} Even at this most general description some evidence for conservation is seen between different OB-fold proteins. In a rigorous analysis of potential Greek-key topologies Zhang and Kim²⁹ identified six distinct, possible topologies for a 5-stranded Greek-key β -barrel. Only one of these is observed in nature, that seen for the OB-fold. β -barrels are uniquely described by their number of strands n , and the shear number, S .^{30,31} The shear number describes the degree to which the strands are tilted away from the axis of the barrel. There are two possibilities for the OB-fold ($S=8$ or $S=10$) and both of these are observed in nature. The scaffold used in this study is the OB-fold domain from an aspartyl tRNA synthetase which has $S=10$ (Figure 1).

Randomised residues

The OB-fold domain from the aspartyl tRNA synthetase (aspRS-OB) was cloned (the domain boundaries were identified from sequence alignments) and tested for overexpression in *E. coli* and solubility after lysis. The domain showed good expression

and heat stability, remaining well behaved (i.e. soluble) after 15 minutes at 85 °C. AspRS-OB binds to bases in the anticodon loop of tRNA and constitutes the anticodon recognition domain. The three dimensional structure of aspRS from *P. aerophilum* is not available, however there are a number of structures for homologues from other organisms including the thermophiles *Pyrococcus kodakarensis* (PDB code 1B8A) and *Thermus thermophilus* (1EFW) as well as from mesophiles, *S. cerevisiae* (1ASZ) and *E. coli* (1C0A). These structures were used to build homology models for aspRS-OB from *P. aerophilum* and from the modelled structure, 4, 9, 13 or 17 amino-acid positions were chosen for randomisation (Tables 1 & 2).

Construction of combinatorial libraries

A set of five libraries for aspRS-OB were constructed. Individual amino acids on the binding face (β -strands 1-3) were mutated separately and in combination. The loop between β -strands 4 and 5 (L4, Figure 1) was also randomised separately, as well as being incorporated into two large libraries where up to 17 codons were randomised. These libraries were named as follows: 4m – 4 randomised codons on β -strand 3; 4RL – 4 randomised codons on L4; 9m – 9 randomised codons on β -strands 1 and 2; 13m – a combination of 4m and 9m; 13mRL – a combination of 4m, 9m and 4RL (see Table 2).

Libraries were assembled using overlap extension PCR incorporating long synthetic oligonucleotides with degenerate codons at the desired positions. First, gene fragments covering the whole gene and containing overlapping regions were generated by ordinary PCR techniques. The gene was assembled by PCR using these fragments in combination

with flanking primers resulting in amplification of the full length gene incorporating the randomised codons. Positions were randomised using NNK codons where N stands for any of the four nucleotides A,C,T or G, and K stands for either T or G. This eliminates two of the three possible stop codons and reduces the frequency of truncated genes. The remaining amber stop codon is suppressed in TG1 cells (used in this study). For each library we obtained approximately 10^8 clones after ligation into an expression vector and transformation into a bacterial host. PCR screening showed that 90-95% of these clones had inserts of the correct size.

Transformation frequency places an upper limit on the diversity which can be sampled using this approach. We estimate the maximum diversity to be $\sim 10^9$ variants. Thus, for the 9m, 13m and 13mRL libraries (Table 2), the theoretical diversity is very much greater than the diversity which can be practically sampled. This is compounded by the fact that randomising 9 codons using NNK gives a $\sim 28\%$ ($9/32$) chance of incorporation of a single undesirable codon (e.g. proline in a β -strand). In the case of 17 randomised positions, this rises to a 53% ($17/32$) chance of an unwanted codon. These factors make successful screening from naïve libraries practically, and technically demanding.³²

Expression and solubility screening

Wild-type aspRS-OB expressed in *E. coli* at high levels ($10\text{-}20\text{ mg.l}^{-1}$ of culture) and was found in the soluble fraction after cell lysis. It remained soluble after heat treatment (15 min at $85\text{ }^{\circ}\text{C}$) and bound quantitatively to Ni-NTA beads. Individual members of the aspRS-OB libraries were also cloned and expressed as N-terminal His₆-tagged proteins

for screening purposes. A total of 144 colonies from each library were screened for expression, solubility and Ni-binding and the results are summarised in Table 2 (a typical SDS PAGE gel used in screening is shown in Figure 2). Approximately half of the clones from the largest aspRS-OB library (13mRL) expressed well in *E. coli* as full-length proteins. In contrast, the small '4RL' library which targets only four positions from L4 for randomisation had a very high number of expressing clones (80%). From each library 16-24 expressing clones were screened for solubility which is a crude indication that the corresponding proteins are correctly folded. For the large library (13mRL) approximately 16% of the mutants show soluble expression and bind to Ni-NTA resin whereas in the case of the small 4RL library approximately 80% are soluble (Table 2 and Figure 2). The small fraction of soluble, expressed domains is reflected in the small fraction of domains from the 13mRL library which are correctly displayed at the surface of phage, as shown in Figure 3 (inset).

To confirm that some aspRS-OB mutant domains were folded, we analysed three domains in more detail using CD and NMR. Each of three mutants were overexpressed and purified and for each mutant the CD spectra are consistent with a well-folded domain whose secondary structure is predominantly β -sheet. In each case this was supported by NMR spectra which showed the characteristic shifted resonances of folded proteins (data not shown). This provides anecdotal confirmation that mutants which are expressed and soluble are also folded. Amino acids at randomised positions for two of the 'soluble' clones are shown in Table 3B (S68, S81) and are unremarkable.

Phage display of aspRS-OB libraries

The gene for wild type aspRS-OB was cloned into the phagemid vector pRPSP2, upstream of the gIII gene generating a fusion protein with aspRS-OB at the N-terminus. This phagemid vector contains gIII under control of the phage shock promoter (psp) which is activated upon infection of the *E. coli* host by helper phage.³³ We used the helper phage VCSM13d3 (Vd3) gIII deletion mutant³⁴ which allowed multivalent display of the target protein (each of the 3-5 copies of the gIII protein will be fusions to aspRS-OB). Display of the aspRS-OB was confirmed by western blot using an antibody against the c-myc antigen sequence located between the aspRS-OB and pIII. The displayed wild type aspRS-OB is functional on the surface of the phage as it bound to immobilised aspartyl-tRNA. We also sequenced inserts from a large number of phagemid clones of the library from the randomized aspRS-OB to confirm the introduced diversity. Four representative sequences are shown in Table3A (U1-U4) and a survey of all sequences showed that amino acids have been randomly incorporated at the desired positions.

Having confirmed that our scaffold was tolerant to mutation, that a significant fraction of mutant aspRS-OB proteins were folded and that wild-type aspRS-OB is functionally displayed at the surface of phage, we then screened aspRS-OB libraries for binding to both the natural target ligand (asp-tRNA) and a foreign protein ligand (lysozyme).

For example, the assembled library 13mRL which contains 17 randomised positions was cloned into the phagemid vector pRPSP2 and the library was displayed on VCSM13 phage. Consistent with our earlier observation that ~16% of the 13mRL are expressed

and folded, only a proportion of mutants are functionally displayed. The 13mRL library showed a strong band in a Western blot at the size of the pIII coat protein and only a faint band at the size of the expected fusion protein (Figure 3, inset).³⁵

Each phage library showed enrichment after several rounds of panning against aspartyl-tRNA (Figure 3). The loop library (4RL) with four randomised positions converged to a consensus sequence which bound to tRNA more tightly than the wild-type protein (Table 3C, p12-p20). The consensus sequence of RGCR is intriguing. The presence of two positively charged residues is consistent with binding to a negatively charged RNA substrate, however the role of the cysteine residue in RNA binding is not clear but is nevertheless under strong selection. In contrast, the large 13mRL library gave enrichment (Figure 3) but no consensus sequence when panned against tRNA (Table 3D, D4-D9). Positively charged residues, typtophan and tyrosine are over-represented at the randomised sites, which is consistent with RNA binding (Table 3D, D4-D9). Six rounds of panning for 13mRL on immobilised lysozyme gave enrichment for clones which bind to the ligand with moderate affinity (Figure 3). Monoclonal phage samples were prepared and micropanning was used to assess binding. Of 22 chosen clones, 9 showed recovery numbers above the background and were sequenced. Six clones were assessed using ELISA and specificity was cross-checked by using tRNA and BSA as negative controls (Figure 4, Table 3E).

Sequences of 9 different clones indicated similarities in their composition. A few positions showed some interesting similarities. For example, position 29, which is an

acidic residue (D or E) in 6 clones out of 9, position 31 is a valine in 5 out of 9 sequences, position 35 is a positively charged residue in 4 clones, position 38 is an aromatic residue (Y, F, W) in 5 cases and finally, position 85 is glycine in 5 clones. Also, in β -strand 3 there are noticeable patterns ETET and PETE occurring in clones L16 and L34, and in β -strand 1 the consensus sequence D(V/L)(A/L) appears in, L2, L5, L6 L32, L42 and L44. Also striking is the pattern found in clones L5 and L44: β -sheet residues were identical between the two clones, whereas the loop region is different. There are no cysteines in any mutants with the exception of L6. More obvious consensus sequences could not be derived probably due to poor coverage of the very large library.

Binding to lysozyme

The aspRS-OB variants encoded by clones, L5, L6, L8 and L16 were subcloned into an expression vector and expressed and purified. Binding of each mutant to immobilised lysozyme was analysed using surface plasmon resonance (Biacore). Binding curves are shown in Figure 4C and the corresponding K_d values are L5=89 μ M, L6=139 μ M, L8=36 μ M and L16=90 mM μ M. One of these OBodies (L8) was chosen for structure determination in complex with lysozyme.

Structure of an OBody-lysozyme complex

The three-dimensional structure for the complex between the AspRS-OB variant L8 and lysozyme shows in atomic detail the specific interactions between the two proteins. The structure was solved at modest resolution (2.75 Å) and phases were determined using the lysozyme and aspRS (from *P. kodakarensis*) structures for molecular replacement (PDB

entries 193L and 1B8A). The final model has two OBody/lysozyme complexes in the asymmetric unit each with excellent geometry (88% of residues in allowed regions, 12% of residues in generously allowed regions and no residues in disallowed regions of the Ramachandran plot). For crystallographic and refinement statistics, see Table 4. For each OBody monomer the first 21 residues are not visible; these residues do not form part of the core OB-fold domain. For the full-length aspRS protein, these residues loop around the OB-fold domain and interact with the aspRS catalytic domain. Clearly, when the catalytic domain is absent, these residues are disordered. The C-terminus of the OB-fold is visible and the addition of two residues (Gly-Ala) into the loop between $\beta 2$ and $\beta 3$ to facilitate library construction extends the sequence to 111 residues.

The OBody binds to lysozyme by wrapping its binding face around the surface of the protein (Figure 5). This is accentuated by the end of β -strand 1 and its attendant short loop (residues 32-35) binding into the active site cleft of the enzyme. Of the 17 residues which were randomised to construct the library, ten are intimately involved at the protein-protein interface (Table 3, L8 residues in bold). These ten residues combine with a further eleven wild-type residues to form the interface. The binding between the OBody and lysozyme buries 841 \AA^2 per molecule which represents 14% of the surface area of the OBody (Figure 5C). Atoms at the interface are evenly divided between hydrophobic and hydrophilic character. This modest burial of surface area is enhanced by the presence of nine hydrogen bonds which include two triangular salt bridges, each involving three residues (Figure 5B). One of the salt bridges is of particular interest as it involves the lysozyme catalytic residues E35 and D52 and the OBody residue R35 that is part of the

peptide which binds into the active site cleft of lysozyme (Figure 5B). We have tested the capacity for the OBody to inhibit lysozyme and see modest inhibition which probably reflects the modest equilibrium constant for binding (36 μ M). Inspection of the sequences for lysozyme-binding variants shows that four of nine independent mutants selected for binding to lysozyme have an arginine or lysine residue at position 35 (Table 3E). This suggests that in the context of an OBody binding to lysozyme there is very significant bias towards a β 1 peptide ending in arginine or lysine which most probably binds into the active site. Given that the theoretical diversity 13mRL library vastly exceeds the practical diversity which we can screen, this sequence consensus is quite remarkable.

The OBody-lysozyme interface is commensurate with typical antibody-antigen interfaces (in terms of buried surface area, number of H-bonds and the gap-volume index) as shown in Table 5. This raises the question of the modest affinity of the OBody-lysozyme interaction in comparison to 2-3 orders of magnitude greater affinity of antibody-antigen interactions, despite similar properties of the interaction interface. Firstly, the fact that interfaces with similar properties achieve much higher affinities suggests that the OBody-lysozyme interface can also be modified to generate much tighter binding. Secondly, it points towards the presence of unfavourable interactions decreasing the affinity between the OBody and lysozyme when compared to other lysozyme complexes. To address this, we have inspected the interface for unfavourable interactions and have identified three residues that may be responsible for lowering the affinity of the complex. We will mutate these positions in the future in an effort to improve the binding of the OBody to lysozyme.

The natural bacterial inhibitor of lysozyme, YkfE, binds on the opposite face of lysozyme when compared to the OBody. However, there is a surprising intersection of the inhibitory Histidine from YkfE and the OBody Arginine which reaches into the active site of lysozyme and H-bonds with the catalytic residues Glu35 and Asp52 (Figure 6).

The binding of the camelid single domain antibody to lysozyme also has some similarities to the OBody-lysozyme complex. A long CDR3 loop from the camelid antibody binds into the active site cleft of lysozyme and presents ~70% of the binding interface. This CDR3 loop has been compared to substrate binding by NAG-NAG and there are striking similarities between the substrate and atoms in the CDR3 loop providing clear details for the mechanism of inhibition.³⁶ This is similar to the C-terminus of $\beta 1$ and the short loop between $\beta 1$ and $\beta 2$ strands of the OBody, which runs into the active site cleft and allows R35 to reach out and H-bond with the active site acidic residues of lysozyme. The relative C α positions of Y103 from the camelid antibody and Y33 from the OBody is striking as well as the similar hydrophobic interactions of Y103 and V36 (Figure 7). Whilst the OBody loop does not penetrate the active site cleft to the same degree as the camelid antibody there are nevertheless sufficient similarities between this OBody loop and the camelid loop to suggest that efficient inhibition would be possible with higher affinity binding. Indeed, the larger OBody surface area presented to lysozyme by way of the OB-fold binding face and the loop above the binding face (L4) points towards attainable tighter binding variants.

Discussion

Murzin's hypothesis is that the OB-fold is an ancient domain which is tolerant to mutation and which has been evolved in very diverse species to use one face of the protein in a broad range of biological functions.^{1,3} Examples of these biological functions include single stranded DNA and RNA binding, protein-protein interactions in bacterial pathogenesis and inorganic phosphatase enzymatic activity. This being the case, these characteristics make OB-fold proteins ideal scaffolds for the generation, through randomisation and selection *in vitro*, of new binding modes. We have coined the term OBodies for these small engineered protein domains.

We chose to focus on the tRNA anticodon binding domain of aspartatyl-tRNA synthetase (aspRS-OB) from *Pyrobaculum aerophilum* as an OB-fold scaffold to investigate OB-folds as potential carriers of diversity. Our results show that this tRNA anticodon binding domain can be converted into a specific protein binding molecule by applying a combinatorial approach.

Each mutation introduced into a protein framework can potentially affect its folding and thus its stability and solubility. To understand the tolerance towards mutations in the protein framework we generated libraries containing different sets, or combinations of mutated areas, and screened for expression and solubility of randomly picked mutants. Because we were interested in changing the binding specificity from a nucleic acid binding molecule to binding a protein target we chose to generate libraries with unrestricted diversity. Such naïve libraries contain all possible combinations of mutations

through randomisation. It is expected that a large number of mutants will not be tolerated for reasons of either stability, folding or solubility due to unfavorable combinations of amino acids in particular areas in the molecule. In a further concession to combinatorics, the theoretical diversity of our libraries containing 9, 13 or 17 randomised amino acid positions vastly exceeds the potential to practically sample this diversity. Despite this concession, one can obtain variants using combinatorial display which are useful and serve as leads for further development.

This field is quite mature and other investigators have used a similar approach to superimpose randomised regions onto well defined protein domains. Other investigators have successfully grafted new binding modes onto a range of scaffolds including single domain antibodies, camelid single domain antibodies, staphylococcal protein A, lipocalins, gamma crystalline and designed ankyrin repeat domains. In each of these cases, randomised loops (antibodies, lipocalins), helical faces (protein A) or a combination of the two (designed ankyrin repeats) are presented as binding interfaces. We now extend this repertoire by using a β -barrel binding face and adjacent loops to generate new binding interactions. This may have advantages in some contexts over other combinations of secondary structure elements. Furthermore, a protein scaffold with a different architecture from the immunoglobulin fold has potential for generating molecules with new binding properties (that may not be attainable using antibodies) for various applications in biotechnology and medicine.

Materials and Methods

Bacterial strains, growth conditions and helper phage

E. coli strain TG1 (*supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5* (rK⁻ mK⁻) [F' *traD36 proAB lacIqZΔM15*]) was utilised to construct the aspRS-OB phage display libraries. Strain XL1-Blue (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^q. lacZΔM15 Tn10* (Tet^r)]). was used for expression of His-tagged AspRS-OB variants. *E. coli* cells were incubated either in Luria-Bertani broth or in Yeast Extract Tryptone broth (2xYT) and *E. coli* transformants in LB or 2xYT supplemented with 50 µg ml⁻¹ ampicillin (Amp) were grown at 37°C with aeration. Solid medium for growth of *E. coli* transformants also contained 1.5% (w/v) agar. Helper phage VCSM13 (*gIII⁺*; Stratagene) was propagated on strain TG1. Stocks of the helper phage VCSM13d3 with deleted *gIII* were obtained by infection of complementing *E. coli* strain K1976 (TG1 transformed with plasmid pJARA112 carrying full length *gIII* under the control of phage infection-inducible promoter *psp*).³⁷

Recombinant DNA

The wild type genes for aspRS-OB (asp-tRNA synthetase from *Pyrobaculum aerophilum* IM2, bases 1-327, amino acids 1-109) was amplified by PCR from *P. aerophilum* IM2 genomic DNA. Forward and reverse primer pairs included cleavage sites suitable for cloning into either expression vector pProEx-Htb (*Bam*HI and *Eco*RI) or phage display vector pRPSP2 (*Nco*I and *Not*I). All PCR products for overexpression were digested with *Bam*HI and *Eco*RI and ligated into pProEx-Htb (Invitrogen). pProEx-Htb produces the protein as an N-terminal His₆-tagged fusion protein for purification on Ni²⁺-NTA-

sepharose. Constructs were transformed into *E. coli* XL1-Blue for plasmid preparations and expression of the corresponding gene products. For phage display cloning of the aspRS-OB gene into pRPSP2, aspRS-OB was amplified by PCR, digested using *NcoI* and *NotI*, and transformed into *E. coli* TG1. For subcloning selected mutants, genes were amplified using vector-specific primers and inserted into the donor vector pDONR221 and subsequently into pDEST15, both part of the GATEWAY[®] cloning system (Invitrogen).

AspRS-OB library construction

Libraries were constructed by incorporation of mutagenic oligonucleotides containing the codon NNK (N=A,C,G or T, K=T or G) in selected positions. AspRS-OB gene fragments carrying incorporated mutations were generated by PCR and then assembled into full-length genes. In a first PCR step, gene fragments were generated using corresponding flanking primers and incorporating the oligonucleotides randomised at selected positions. In a second step, the gene fragments were assembled into a full length gene by an overlap-extension PCR and ligated into an appropriate vector. Plasmids containing either the wild type gene or assembled libraries were transformed into *E. coli* XL1-Blue or TG1 respectively and propagated overnight at 37 °C on LB-agar plates supplemented with ampicillin (50 µg/ml).

Phage library preparation

Cultures of TG1 cells transformed with aspRS-OB gene libraries in phagemid vector pRPSP2 were diluted into pre-warmed 2xYT medium, incubated at 37 °C until OD₆₀₀ =

0.4 and then infected with VCSM13 helper phage (Stratagene) at multiplicity of infection (m.o.i.) of 50, incubated for 20 min at 37 °C without agitation and then with rotatory agitation for 1 hour. Kanamycin was then added to final concentration of 50 µg.ml⁻¹ and the culture was propagated overnight at 37 °C with aeration. Bacteria were sedimented by centrifugation and phagemid particles were precipitated from the supernatant in 5% PEG8000 0.5M NaCl solution overnight at 4 °C. Precipitated phage were pelleted by centrifugation, dissolved in 5 ml PBS, filtered through 0.45 filters and used for panning.

Expression and solubility screening

AspRS-OB libraries were cloned into pProEx and transformed bacteria were plated onto LB/Amp agar. Single colonies were randomly picked and grown overnight in 100 µl LB/Amp (50 µg.ml⁻¹) at 37 °C in a 96 deep-well plate with aeration at 1200 rpm. The cultures were diluted by adding 900 µl fresh LB/Amp and incubated for a further 60 min. Protein expression was induced by addition of isopropyl-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM followed by shaking for 4 hrs at 37 °C. Bacterial cells were collected by centrifugation and resuspended in 150 µl Tris-buffered saline (TBS, 50 mM Tris-HCl, pH 7.5, 150 mM NaCl). Cells were lysed by freeze-thaw and addition of lysozyme (50 ug/ml final concentration). The soluble fraction was obtained after sedimentation of insoluble material. A small-scale purification step was conducted by incubating the soluble protein fraction with 5 µl Ni-NTA resin (Qiagen, Germany). Ni-NTA beads were washed with TBS yielding purified protein bound to the matrix. Samples of induced bacteria, soluble protein and matrix-bound purified protein were analysed by SDS polyacrylamide gel electrophoresis.

Protein purification on a larger scale was carried out as above except that cultures were grown in 1 L flasks and purification was performed using Ni^{2+} -affinity chromatography. A second protein purification step was carried out using size exclusion chromatography.

Preparation of biotinylated RNA

Generation of biotin labeled asp-tRNA was carried out by *in vitro* transcription using the MEGAscript kit (Ambion, USA) and the Biotin RNA Labeling mix (Roche, Switzerland) containing biotin-16-UTP. The DNA template was made by PCR assembly of synthetic oligonucleotides covering the 78 bp asp-tRNA gene from *P. aerophilum* and a 150 bp DNA fragment amplified from pET28 (Invitrogen) including the T7 promotor region at the 3' end followed by GG for optimal promotor activity according to recent promotor recognition studies for T7.³⁸ This resulted in an assembled product of 230 bp which was precipitated by ethanol, dried resuspended in RNase-free H_2O and used as the template for transcription without further cloning. *In vitro* transcription was carried out following the manufacturer's manual (Ambion) and yielded ~5 μg biotinylated asp-tRNA from a 25 μl reaction.

Selection of aspRS-OB libraries

For selection against lysozyme, 4 ml Immuno tubes (Nunc, Denmark) were coated with 2.5 ml lysozyme solution ($10 \mu\text{g} \cdot \text{ml}^{-1}$ in 20 mM NaCO_3 pH 9.0) overnight at 4 °C and blocked with 4 ml 1% BSA in PBS for 1 hr at room temperature. Phage from library 13mRL were added ($\sim 2.5 \times 10^{11}$ cfu in 2.5 ml) and incubated for 2 hrs at RT with gentle

agitation and occasional inversion. Washing was performed by 8 washing steps with PBS-T, 0.1% BSA and 2 steps with PBS. Bound phage were eluted by incubation for 10 min with 2.5 ml elution buffer (0.2 M glycine-HCl pH 2.2, bromphenol blue) and immediately neutralised using 500 µl 1 M Tris-HCl pH 9.0. Eluted phagemid particles were titrated and used to infect a fresh 3 ml culture of TG1 for amplification and subsequent rounds of panning. Cultures were left for 20 min at 37 °C without agitation, incubated for one hour with shaking before addition of ampicillin and growth overnight. Overnight cultures were used to inoculate 500 ml prewarmed LB/Amp. Helper phage infection and phagemid particle production followed the same procedure as for the phage library preparation outlined above. After 4 rounds of panning of library 4RL, individual clones were analysed. After 6 rounds of selection and amplification of library 13mRL, clones were picked and analysed further as described below.

Panning of individually displayed OBody variants on lysozyme.

This panning protocol was used to investigate binding of individual OBody variants, obtained after 6 rounds of 13mRL library screening, to lysozyme. 96-well plates were coated with lysozyme or BSA and blocked in the same way as described above. Phagemid particles of individual selected library clones were produced using VCSM13 helper phage. A total of $\sim 10^{10}$ phagemid particles were added to each well and incubated for 2 hrs at RT. After 10 washes with TBST, bound phage were eluted with buffer (0.2 M glycine-HCl pH 2.2, bromphenol blue), neutralised with 1 M Tris-HCl pH 9.0 and titrated on TG1 using an appropriate dilution series.

Phagemid particle preparations

For phage ELISA, phagemid particles that display individual OBody variants at 5 copies per virion were prepared using a *gIII* deletion helper phage VCSM13d3³⁷; for the panning, phagemid particles were assembled with the help of VCSM13, which encodes for the wild-type pIII; hence most of the phagemid particles carried one copy of OBody-pIII fusion.³² To prepare stocks of phage aspRS-OB using VCSM13d3 or VCSM13, *E. coli* TG1 transformed with the corresponding OBody variant were grown in 100 ml LB-Amp to OD₆₀₀ = 0.4 and infected with VCSM13d3 or VCSM13, respectively, at an m.o.i.=50. After 20 min incubation at 37 °C without agitation, the culture was incubated for another hour with shaking. Kanamycin (50 µg.ml⁻¹ final concentration) was added and the culture was incubated overnight. Cells were sedimented by centrifugation and phagemid particles purified from the supernatant by precipitation using PEG-NaCl. Phagemid particles were resuspended in PBS and used for analysis.

Phage ELISA

ELISA experiments were performed to analyse binding of selected and individually displayed OBfold variants to lysozyme. Ninety-six-well ELISA plates were coated with 5 µg.ml⁻¹ hen egg white lysozyme, 5 µg.ml⁻¹ RNaseA or 1% BSA, in PBS (125 mM NaCl, 1.5 mM KH₂PO₄, 8 mM Na₂HPO₄ and 2.5 mM KCl, pH 7.6) at 4 °C overnight. After two washes with PBS, plates were blocked with blocking buffer (5% skim milk in PBS) for one hour at RT before phagemid particles (10⁹ well⁻¹, produced by infection with $\Delta gIII$ helper phage VCSM13d3) were added in 2.5% skim milk-PBST (PBS containing 0.1% v/v Tween 20). Plates were incubated for 2 hours at RT with agitation. After 10 washes

with PBS, mouse anti-M13 protein VIII diluted in blocking buffer was added and incubated for 1 h at RT. Plates were washed 4 times with H₂O and horseradish peroxidase (HRP)-coupled rabbit-anti-mouse immunoglobulin (Pierce) in blocking buffer was added to the wells and incubated for 1 h at RT. Wells were washed 4 times with H₂O and 50 µl substrate solution (1mg.ml⁻¹ o-phenylene-diamine in PBS 0.03% H₂O₂) was added per well. The reaction was stopped after ~ 15 min by addition of 25 µl 2.5 M H₂SO₄ and the absorbance was recorded at 492 nm. For quantification of displayed fusion protein, antibodies against c-myc tag, encoded by the vector and included into the fusions were used. Phagemid particles were used directly to coat plates. After blocking with blocking buffer, phage were detected using mouse anti-c-myc primary antibody (Zymed, Invitrogen) and an HRP-conjugated anti-mouse secondary antibody following the procedure described above.

Western blot for phage display protein detection

A phagemid particle sample of aspRS-OB was concentrated by PEG/NaCl precipitation to 1 x 10¹¹ particles.ml⁻¹ and 10 µl combined with gel loading buffer, boiled and separated on a 10% SDS-PAGE gel. After transfer onto a nitrocellulose membrane aspRS-OB-c-myc-pIII fusion protein was detected using a mouse anti c-myc primary antibody (Zymed, Invitrogen) and a HRP-linked anti-mouse secondary antibody (Amersham-Pharmacia, Sweden). Visualisation was performed using SuperSignal[®] substrate (Pierce, USA).

SPR analysis

Lysozyme was immobilised on a CM5 *Biacore* sensor chip, at 30 $\mu\text{g.mL}^{-1}$ in sodium acetate buffer at pH 4.3 by successive injections of between 10-20 μL to an adequate response. Remaining uncoupled active groups on the chip were deactivated with an injection of ethanolamine-HCl. Each of the seven L8 samples were analysed in duplicate for 1 min at 25 $\mu\text{L.min}^{-1}$, in random order, using the first flow cell as a reference. The response curves were visualised and processed using *BIAevaluation* (Biacore). Relative response at each concentration was averaged and plotted to determine R_{max} and K_d using *Sigma Plot* (Systat Software, Inc.). This protocol was repeated for three other OBody variants L5, L6 and L16 (see Figure 4).

Structure determination

The lysozyme-binding OBody L8 was cloned using Gateway (Invitrogen) into pDONR221 then subcloned into the expression vector pDEST15 which was transformed into BL21 (DE3) *E. coli* cells. These cells were inoculated into 500 mL of auto-induction media and shaken at 37 °C in 2 L baffled flasks. The fusion protein GST-L8 was purified from bacterial lysate using a GSH affinity column (GE Biosciences). The GST tag was removed using rTEV protease and separated from L8 by size exclusion chromatography.

The purified protein was combined with *Gallus gallus* egg white lysozyme (Roche) in a 1:1 molar ratio, to a final concentration of L8 at 37.5 mg.mL^{-1} and lysozyme at 42.9 mg.mL^{-1} , in TBS (25 mM TRIS, pH 7.5, 137 mM NaCl, 3 mM KCl). The complex

in solution was screened against 480 crystallisation conditions using custom screens and a sitting drop format.³⁹

A single large crystal grew from an equal mixture of protein in TBS and precipitant (7% MPEG 5K, 0.2 M HEPES pH 7.8). This crystal was then gathered in a nylon loop, coated in cyroprotectant, and frozen under a stream of cold N₂ gas (110 K). A dataset of 700 images was collected using a rotating anode X-ray generator and Mar345 detectors giving diffraction to 2.8 Å. Images were indexed using DENZO and data were scaled using Scalepack. For data collection statistics see Table 4. The structure was solved using molecular replacement (AMoRe) incorporating both lysozyme (PDB entry 193L) and the OB-fold codon recognition domain from the *Pyrococcus kodakarensis* aspartyl tRNA synthase (PDB entry 1B8A) as models. Two molecules of lysozyme were found in the asymmetric unit along with one OB-fold domain. A second OB-fold was placed by replicating the complex in the asymmetric unit based on the position of the second lysozyme molecule. The structure was iteratively built and refined using COOT, CCP4 and PHENIX. A second dataset was collected using the same crystal at the SSRL, to 2.69 Å resolution. It was indexed in the same space group and phased by molecular replacement using the complete unit cell from the previous structure. Building and refinement was done using COOT, CCP4 and PHENIX.

Acknowledgements

We would like to thank Shaun Lott for help with SPR data collection and David Goldstone for assistance with X-ray data collection. We are indebted to Ted Baker and the Maurice Wilkins Centre for Molecular Biodiscovery for the use of their X-ray facilities including robotic crystallisation screening. This project was funded by a grant from the Marsden Fund of New Zealand with additional funds from the Maurice Wilkins Centre for Molecular Biodiscovery.

References

1. Murzin, A. G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861-7.
2. Arcus, V. L., Proft, T., Sigrell, J. A., Baker, H. M., Fraser, J. D. & Baker, E. N. (2000). Conservation and variation in superantigen structure and activity highlighted by the three-dimensional structures of two new superantigens from *Streptococcus pyogenes*. *J. Mol. Biol.* **299**, 157-68.
3. Arcus, V. (2002). OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* **12**, 794-801.
4. Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* **381**, 751-758.
5. 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.
6. 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.
7. Atwell, S., Ultsch, M., De Vos, A. M. & Wells, J. A. (1997). Structural plasticity in a remodelled protein-protein interface. *Science* **278**, 1125-1128.
8. Ballinger, M. D., Jones, J. T., Lofgren, J. A., Fairbrother, W. J., Akita, R. W., Sliwkowski, M. X. & Wells, J. A. (1998). Selection of heregulin variants having

- higher affinity for the ErbB3 receptor by monovalent phage display. *J. Biol. Chem.* **273**, 11675-84.
9. Nord, K., Nilsson, J., Nilsson, B., Uhlen, M. & Nygren, P. A. (1995). A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng.* **8**, 601-8.
 10. Gunneriusson, E., Nord, K., Uhlen, M. & Nygren, P. A. (1999). Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling. *Protein Eng.* **12**, 873-878.
 11. Binz, H. K., Amstutz, P. & Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotech.* **23**, 1257-1268.
 12. Binz, H. K. & Pluckthun, A. (2005). Engineered proteins as specific binding reagents. *Curr. Opin. Biotech.* **16**, 459-469.
 13. Hosse, R. J., Rothe, A. & Power, B. E. (2006). A new generation of protein display scaffolds for molecular recognition. *Protein Sci.* **15**, 14-27.
 14. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. (2000). Phage display for selection of novel binding peptides. *Methods Enzymol.* **328**, 333-63.
 15. Stemmer, W. P. C. (1994). DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Nat. Acad. Sci. USA.* **91**, 10747-10751.
 16. Sidhu, S. S. (2000). Phage display for selection of novel binding peptides. *Methods Enzymol.* **328**, 333-363.
 17. Zhao, H. & Arnold, F. H. (1997). Optimisation of DNA shuffling for high fidelity recombination. *Nucl. Acids Res.* **25**, 1307-1308.

18. Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391.
19. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination.[see comment]. *Nat. Biotech.* **16**, 258-61.
20. Volkov, A. A. & Arnold, F. H. (2000). Methods for *in vitro* DNA recombination and random chimeragenesis. *Methods Enzymol.* **328**, 447-456.
21. Gibbs, M. D., Nevalainen, K. M. H. & Bergquist, P. L. (2001). Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene* **271**, 13-20.
22. Hemminki, A., Niemi, S., Hoffren, A. M., Hakalahti, L., Soderlund, H. & Takkinen, K. (1998). Specificity improvement of a recombinant anti-testosterone Fab fragment by CDRIII mutagenesis and phage display selection. *Protein Eng.* **11**, 311-9.
23. Coco, W. M., Levinson, W. E., Crist, M. J., Hektor, H. J., Darzins, A., Pienkos, P. T., Squires, C. H. & Monticello, D. J. (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotech.* **19**, 354-9.
24. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Pluckthun, A. (2001). Tailoring in vitro evolution for protein affinity or stability. *Proc. Nat. Acad. Sci.USA.* **98**, 75-80.
25. Hanes, J., Jermutus, L. & Pluckthun, A. (2000). Selecting and evolving functional proteins in vitro by ribosome display. *Methods Enzymol.* **328**, 404-30.

26. Boder, E. T. & Wittrup, K. D. (2000). Yeast surface display for directed evolution of protein expression, affinity and stability. *Methods Enzymol.* **328**, 430-445.
28. Qian, J., Stenger, B., Wilson, C. A., Lin, J., Jansen, R., Teichmann, S. A., Park, J., Krebs, W. G., Yu, H., Alexandrov, V., Echols, N. & Gerstein, M. (2001). PartsList: a web-based system for dynamically ranking protein folds based on disparate attributes, including whole-genome expression and interaction information. *Nucl. Acids Res.* **29**, 1750-64.
29. Zhang, C. & Kim, S. H. (2000). A comprehensive analysis of the Greek key motifs in protein beta-barrels and beta-sandwiches. *Proteins* **40**, 409-19.
30. Murzin, A. G., Lesk, A. M. & Chothia, C. (1994). Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. *J. Mol. Biol.* **236**, 1382-400.
31. Murzin, A. G., Lesk, A. M. & Chothia, C. (1994). Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **236**, 1369-81.
32. Dufner, P., Jermutus, L. & Minter, R. R. (2006). Harnessing phage and ribosome display for antibody optimisation. *Trends Biotech.* **24**, 523-529.
33. Beekwilder, J., Rakonjac, J., Jongsma, M. & Bosch, D. (1999). A phagemid vector using the E. coli phage shock promoter facilitates phage display of toxic proteins. *Gene* **228**, 23-31.
34. Rakonjac, J., Jovanovic, G. & Model, P. (1997). Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3. *Gene* **198**, 99-103.

35. Rondot, S., Koch, J., Breitling, F. & Dubel, S. (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nat. Biotech.* **19**, 75-78.
36. Transue, T. R., De Genst, E., Ghahroudi, M. A., Wyns, L. & Muyldermans, S. (1998). Camel single-domain antibody inhibits enzyme by mimicking carbohydrate substrate. *Proteins-Struct. Funct. Bioinf.* **32**, 515-522.
37. Rakonjac, J., Feng, J. & Model, P. (1999). Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J. Mol. Biol.* **289**, 1253-65.
38. Brevet, A., Chen, J., Commans, S., Lazennec, C., Blanquet, S. & Plateau, P. (2003). Anticodon recognition in evolution - Switching tRNA specificity of an aminoacyl-tRNA synthetase by site-directed peptide transplantation. *J. Biol. Chem.* **278**, 30927-30935.
39. Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N. & Lott, J. S. (2005). A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Cryst. Sect. D* **61**, 1378-1385.
40. Monchois, V., Abergel, C., Sturgis, J., Jeudy, S. & Claverie, J.-M. (2001). Escherichia coli ykfE ORFan Gene Encodes a Potent Inhibitor of C-type Lysozyme. *J. Biol. Chem.* **276**, 18437-18441.
41. Decanniere, K., Transue, T. R., Desmyter, A., Maes, D., Muyldermans, S. & Wyns, L. (2001). Degenerate interfaces in antigen-antibody complexes. *J. Mol. Biol.* **313**, 473-478.

42. Fischmann, T. O., Bentley, G. A., Bhat, T. N., Boulot, G., Mariuzza, R. A., Phillips, S. E., Tello, D. & Poljak, R. J. (1991). Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution. *J. Biol. Chem.* **266**, 12915-12920.
43. Vocadlo, D. J., Davies, G. J., Laine, R. & Withers, S. G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835-838.

Tables

Table 1. Amino acid sequence of AspRS-OB with randomised positions*

VYPKKTHWTA EITPNLHGTE VVVAGWV**WEL** **RDIGRVKEFVV** **VRDREGgaFV**
QVTLKAGKTP DHLFKVFAEL SREDVV**VIK**G IVEASK**IAKS** GVEIFPSEIW
 ILNKAKPLPI D

*β-strands are unerlined and randomised positions are in bold. The di-peptide GA (shown in lower case) was added as part of the library assembly.

Table 2. OB-fold libraries, randomised sites and expression/solubility frequencies

Library name	Randomised codons	Theoretical diversity	Expression	Solubility	Ni ²⁺ binding
4m	4 (β3)	~1.6 x 10 ⁵	86 % (144)	75 % (16)	-
4RL	4 (L4)	~1.6 x 10 ⁵	80 % (144)	100 % (11)	Yes (2)
9m	9 (β1, β2)	~5.1 x 10 ¹¹	69 % (144)	58 % (24)	Yes (4)
13m	13 (β1, β2, β3)	~8.2 x 10 ¹⁶	45% (96)	42 % (14)	Yes (5)
13mRL	17 (β1-β3, L4)	~1.3 x 10 ²²	51 % (144)	32 % (25)	Yes (5)

Table 3. Amino acids at randomised positions before and after selection*

	$\beta 1$	$\beta 2$	$\beta 3$	L4
wt	WERI	RVFVR	FQTK	IAKS
A				
U1	ECAT	IEGLV	CRHQ	GYKS
U2	GARL	LGGSF	VTNL	SRVG
U3	AAGA	DSTTN	PHQD	YRLK
U4	DPLR	LAGSR	SPTL	QRYV
B				
S68	RSVG	RVANQ	KELE	GEWS
S81	KGVM	LLGTG	SLRT	LVPQ
C				
p12				RGCA
p16				RGCR
p17				KGCR
p18				RSCR
p20				KAFR
D				
D4	SCCA	KRWYV	PARR	RAGS
D5	QRYR	KRSSV	PTCR	WNCG
D7	RKRR	PAYWR	TRRQ	RKGS
D9	WRRW	LATRK	SRTR	WWVW
E				
L4	GEAF	DMTAR	EPTL	GSTS
L5	TDVA	TWFQA	NWSV	LPSY
L6	EDVA	KTWFC	QIEV	SRAV
L8	ASGY	RVIKS	APYE	GVGR
L14	SDLA	RAYFY	EMTA	GWRD
L16	ASVG	PRWFR	ETET	GLRW
L34	VGME	ALGER	PETE	GYGS
L42	VDVL	RNLQS	RLNA	IQRS
L44	TDVA	TWFQA	NWSV	PGAA

*A. Sequences from the 13mRL library before selection. B. Sequences from the 13mRL library showing good solubility upon overexpression prior to selection. C. Variants from the 4RL library which bind to aspartyl tRNA. D. Variants from the 13mRL library which bind to aspartyl tRNA. E. Variants from the 13mRL library which bind to lysozyme. Residues in bold for L8 are those which interact with lysozyme as shown by the three-dimensional structure of the protein-protein complex.

Table 4. X-ray crystallographic and refinement statistics

<i>Crystal Properties</i>	
Space Group	$P4_12_12$
Unit Cell Axes (a, b, c)	76.76, 76.76, 166.34
Unit Cell Angles (α , β , γ)	90, 90, 90
<i>Data Collection</i>	
Resolution	34.85-2.69 (2.76-2.69)
Total Reflections	144,772
Unique Reflections	16,010
Completeness	99.2 (92.8)
Redundancy	9
R_{merge}	7.3 (54.3)
Wilson B Factor	65
Mosaicity	0.6
I/ σ I	32.5 (4.2)
<i>Refinement</i>	
Resolution	27.5-2.75 (2.82-2.75)
R	22.5 (26.5)
R_{free}	29.6 (37.6)
Protein Atoms	3384
rmsd, bond lengths	0.013
rmsd, bond angles	1.452
B factors, average	51.146

Table 5. Protein-protein interface data for proteins bound to lysozyme

	OBody, L8	Inhibitor YkfE, 1GPQ ⁴⁰	Camelid Ab 1JTP ^{36,41}	Mouse FAB 1FDH ⁴²
Buried surface area [*]	840 Å ² (14%)	796 Å ² (11%)	800 Å ² (11%)	647 Å ² (3%)
H-bonds	9	10	8	12
Salt bridge interactions	3	2	0	0
Polar:Non-polar atoms %	43:57	47:53	28:72	51:49
Gap Volume Index [‡]	2.94	3.06	2.22	3.27
K_d	36 µM	~1 nM	50 nM	22 nM

^{*}Average antibody/antigen buried surface area = 950 Å²

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

Figure Captions

Figure 1. The aspRS OB-fold and the randomised positions at the binding surface.

A cartoon representation (at left) of the aspRS OB-fold domain showing β -strands (green) and the capping α -helix (red). Loop regions are shown in grey and the N- and C-termini are labelled. The β -strands which form the five-stranded β -barrel are labelled and the randomised loop, L4, is also labelled. The aspRS OB-fold domain is displayed as a rendered surface (centre) in blue and the randomised surface is coloured maroon. At right is shown a view rotated 90° in the vertical axis. This and other molecular graphics figures were prepared using Pymol (DeLano Scientific, <http://www.pymol.org/>).

Figure 2. Expression and solubility for OB-fold variants.

An SDS polyacrylamide gel stained with coumassie blue showing the soluble and insoluble fractions of an expression test for six OB-fold variants selected from the 13mRL library. 'I' denotes the insoluble fraction and 'S' denotes the soluble fraction for each expression test. These tests were performed in a 96-well format. The arrow at right indicates the expected position of the OB-fold domain band.

Figure 3. Enrichment of naïve OBody libraries.

Enrichment ($-\log [\text{output phage} / \text{input phage}]$) of phage from multiple rounds of selection. The black line shows enrichment over four rounds of selection of the small library (4RL, four randomised positions in L4) when panned against aspartyl-tRNA. The dark grey line shows enrichment over five rounds of selection of the large library (13mRL, seventeen randomised positions) when panned against aspartyl-tRNA. The light

grey line shows enrichment over six rounds of selection of the large library when panned against lysozyme. The inset shows a western blot (using a mouse anti-c-myc antibody) of phage pIII (left lane), wild-type aspRS-OB fused to pIII displayed at the surface of phage (middle lane) and the 13mRL library fused to pIII displayed at the surface of phage (right lane).

Figure 4. Phage display and complex binding of selected clones.

A. Panning of individual AspRS-OB variants on lysozyme. Individual clones obtained after 6 rounds of selection were tested for binding (in 96-well format) to lysozyme (black bars) or BSA (white bars). Bound phagemid particles were eluted and titrated. The numbers on Y-axis indicate the number recovered phagemid particles. Clone numbers are shown on the x-axis; pIII indicates no fusion (empty vector) and OBwt, the wild-type aspRS-OB displayed as a pIII fusion protein. B. Analysis of binding of selected clones to lysozyme by ELISA. BSA (white), RNaseA (hatched) and hen egg white lysozyme (black). Experiments were performed in duplicate. C. Binding curves, using surface plasmon resonance (Biacore), to determine K_d values for binding between selected OBodies and lysozyme.

Figure 5. Structure of an OBody-lysozyme complex.

A. The protein-protein complex where the OBody is depicted as a cartoon representation (light brown) showing secondary structure elements and lysozyme is depicted as a surface (light blue). Some amino acid side chains involved at the interface are shown in yellow (OBody) and blue (lysozyme). The lysozyme surface is semi-transparent and a

α -trace is shown beneath the surface. B. Amino acids involved in hydrogen bonds at the interface. α -traces are shown for each protein as thin lines. OBody residues are shown in yellow and lysozyme residues are shown in blue. Hydrogen bonds are shown as dotted yellow lines. All residues are labelled with the exception of G102 and N103 from lysozyme (centre of picture). C. Contact surfaces for each protein are depicted by rotating the OBody 90 degrees to the left and lysozyme 90 degrees to the right.

Figure 6. Comparison between OBody binding, and YkfE binding, to lysozyme.

YkfE from *E. coli* is a potent inhibitor of lysozyme. At left the protein-protein complexes are superimposed to show the opposed binding of OBody (light brown) and YkfE (maroon) with lysozyme (blue). The inset (at right) shows the near superposition of guanido group of R35 from the OBody (in yellow) and the imidazole group of H60 from YkfE (maroon). The position of the lysozyme active site residues E35 and D52 are shown beneath the surface and D52 and N44 are labelled.

Figure 7. Comparison of active site binding.

A. The active site of lysozyme depicted as a surface (blue) showing the orientation of G31-V36 from the OBody (light brown). For comparison, an NAG-G2F covalent intermediate is shown in yellow.⁴³ B. The same active site comparing the position of the OBody amino acids to those from the CDR3 loop of the camelid antibody cAb-Lys3.³⁶ Camelid residues are labelled in dark green.

Figures

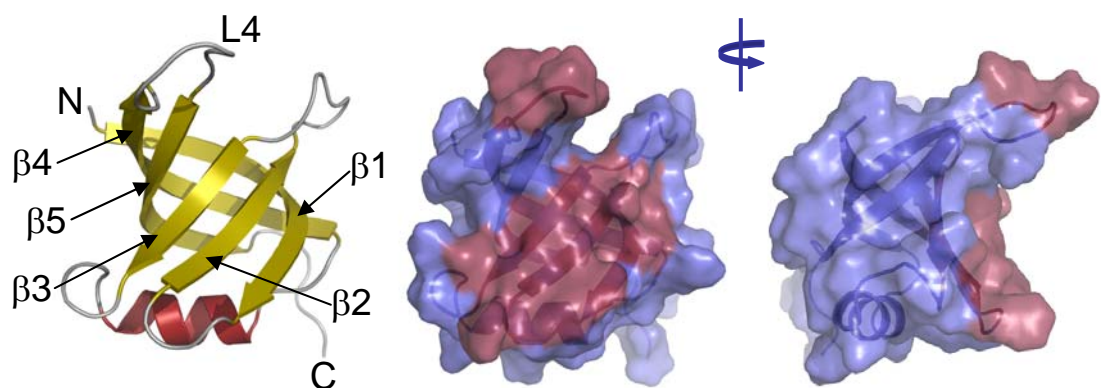


Figure 1. The aspRS OB-fold and the randomised binding surface

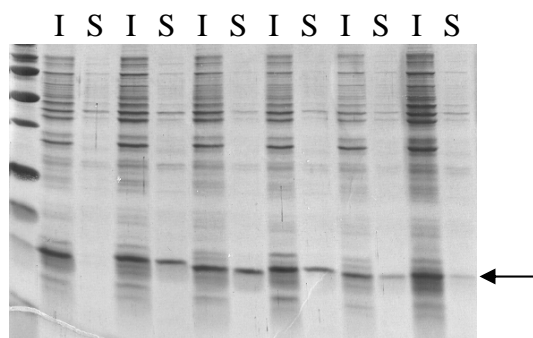


Figure 2. Expression and solubility for OB-fold variants

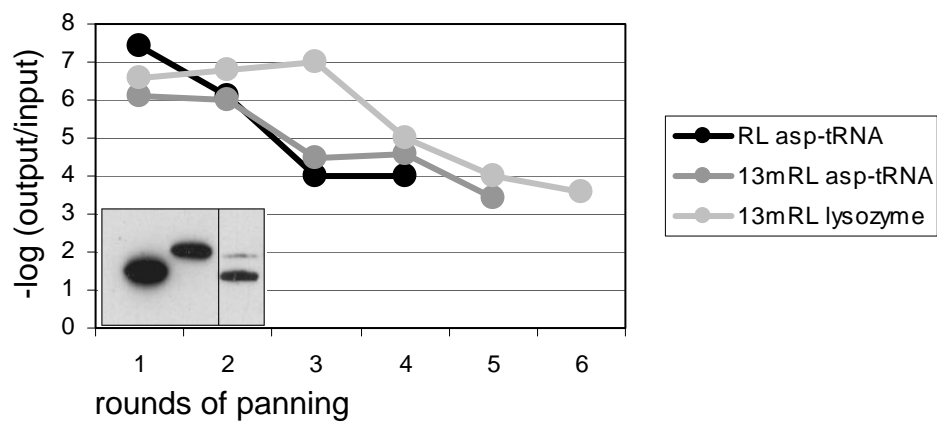


Figure 3. Enrichment of naïve OBody libraries

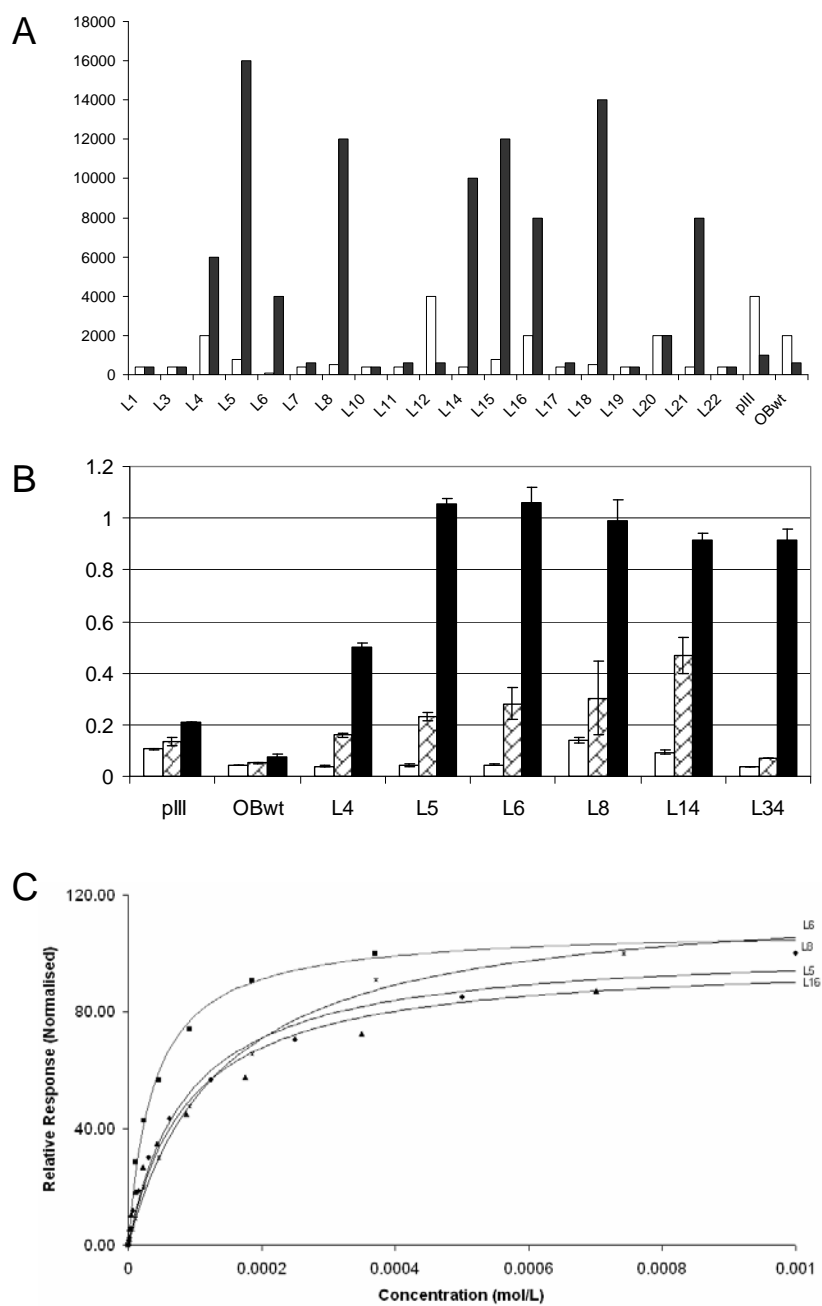


Figure 4. Phage-display binding for selected clones

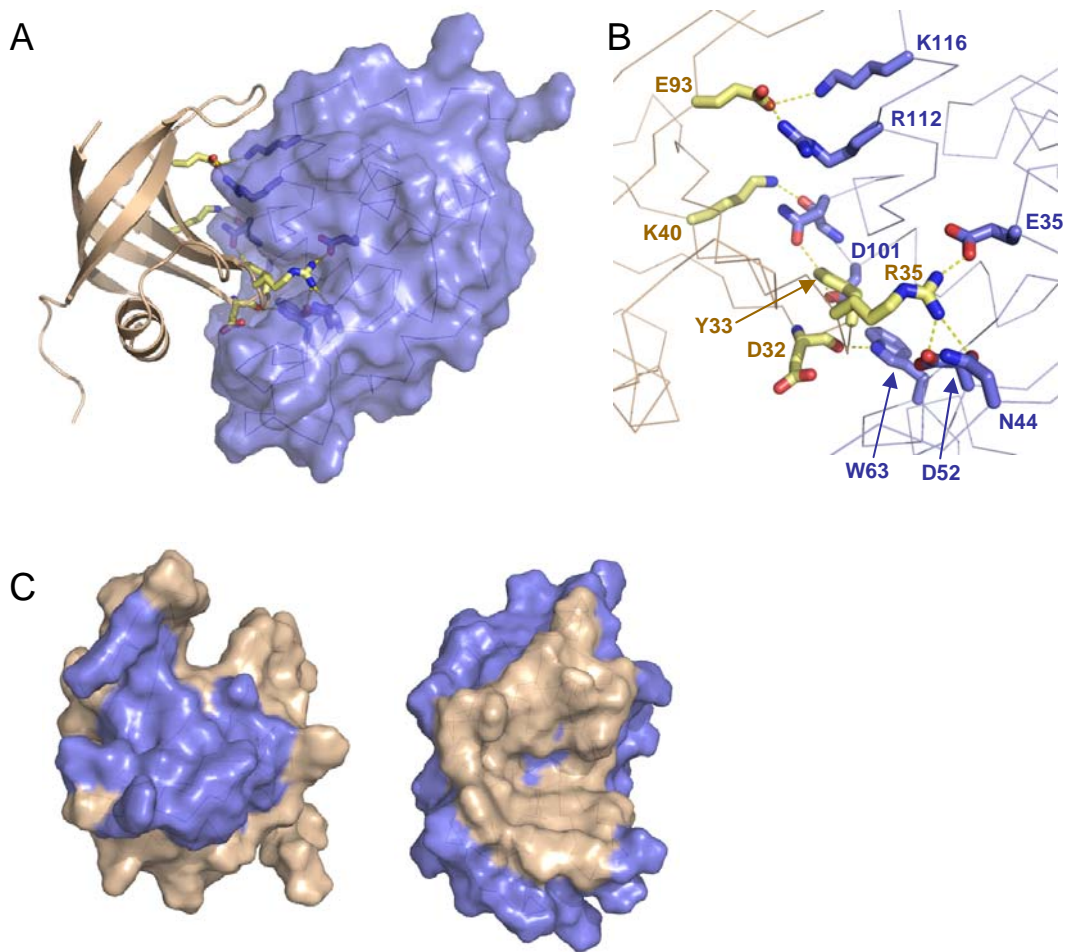


Figure 5. Structure of an OBody-lysozyme complex

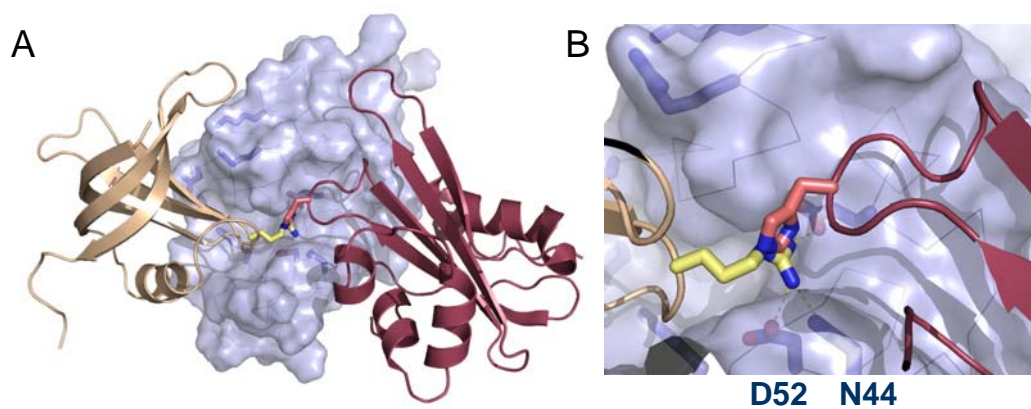


Figure 6. Comparison between OBody binding, and YkfE binding, to lysozyme

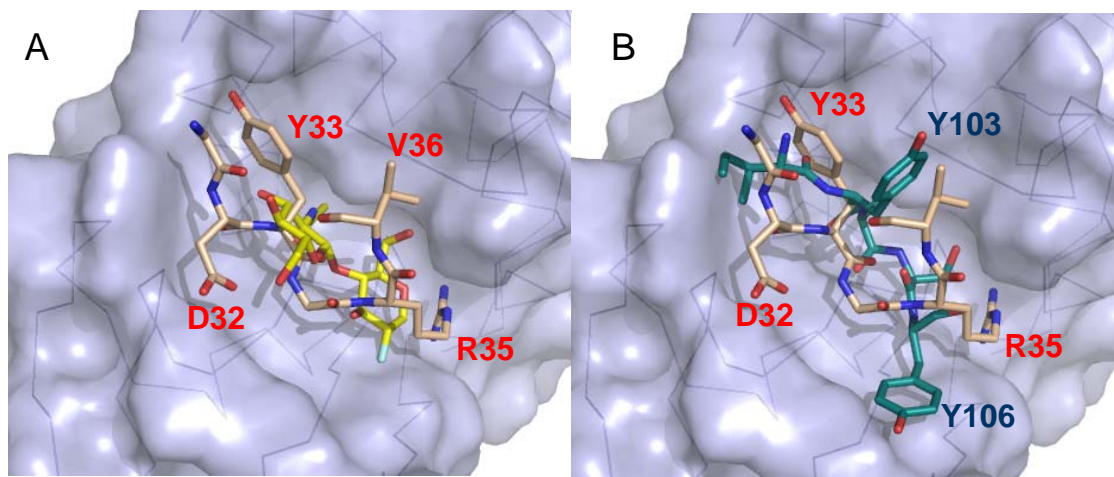


Figure 7. Comparison of active site binding