Using Offset Recombinant Polymerase Chain Reaction To Identify Functional Determinants in a Common Family of Bacterial Albumin Binding Domains†

David A. Rozak,* Patrick A. Alexander, Yanan He, Yihong Chen, John Orban, and Philip N. Bryan

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received September 22, 2005; Revised Manuscript Received December 28, 2005

ABSTRACT: The 46 amino acid GA albumin binding module is a putative virulence factor that has been identified in 16 domains from four bacterial species. Aside from their possible effects on pathogenicity and host specificity, the natural genotypic and phenotypic variations that exist among members of this module offer unique opportunities for researchers to identify and explore functional determinants within the well-defined sequence space. We used a recently developed in vitro recombination technique, known as offset recombinant PCR, to shuffle seven homologues that encode a broad range of natural GA polymorphisms. Phage display and selection were applied to probe the recombinant library for members that showed simultaneous improvements to human and guinea pig serum albumin binding. Thermodynamic data for the most common phage-selected mutant suggest that domain-stabilizing mutations substantially improved GA binding for both species of albumin.

As many as 16 albumin binding domains have been identified in six proteins and four bacterial species (*1*−*3*). This broadly encoded three-helix module might support pathogenesis by camouflaging the bacteria from the host's immune system or scavenging albumin-bound nutrients from the blood (*4*). Variations in the abilities of these domains to bind albumin from different species (*5*) may help to define the host ranges for certain bacterial pathogens. Native or engineered versions of the module could be used to support affinity purification of albumin and other fusion proteins (*6*) or increase vaccine serum stability (*7*, *8*) and immunogenicity (*9*, *10*). However, the rich array of albumin binding domains also offers opportunities for structural biologists wishing to exploit the well-defined sequence space to study the impact of select polymorphisms on protein function and structure.

The functional and structural diversity exhibited by members of the module is evident in the *Finegoldia magna* ALB8-GA and streptococcal G148-GA3 albumin binding domains, which display variable affinities for different species of albumins and significantly different backbone dynamics. Researchers have observed that ALB8-GA demonstrates a distinct preference for primate serum albumins compared to the broader range of affinities for albumin species exhibited by G148-GA3 (*5*). Furthermore, comparative hydrogen−deuterium exchange data reveal that G148-GA3 maintains a more dynamic backbone structure than ALB8-GA, a feature that the researchers suggest may be associated with the former's ability to bind a broader range of albumins (*11*).

Unfortunately, the identity and impact of functional determinants contained within the family of albumin binding domains remain largely unexplored. As is the case for many protein families, thermodynamic, kinetic, and structural data are unavailable for most members. Despite the availability of extensive biochemical data on two distinct members of the GA[1] module (*2*−*5*, *11*−*16*), including a recently published crystal structure of ALB8-GA complexed with human serum albumin (*17*), much remains unknown about the impact of module polymorphisms on domain structure and function.

One promising technique for deciphering the manners in which the GA module and other protein families encode species-specific traits involves creating a library of recombinant homologues that can be probed with phage display for variants that accommodate specific selection criteria. Analysis of the phage-selected mutants could provide significant insights into the natural mechanisms behind phenotypic diversity, permit researchers to predict the behavior of unexamined homologues, and help to guide subsequent research.

Zhao and Arnold initially demonstrated the value of recombinant libraries in evaluating the impact of specific polymorphisms on a pair of subtilisin E mutants (*18*). However, an accumulation of experimental and computational data (*19*−*22*) suggests that the traditional DNA shuffling strategy developed by Stemmer (*23*, *24*) and used by Zhao and Arnold in their landmark study becomes ineffective when applied to coding regions of decreasing size and homology. Although other strategies have been developed specifically to promote recombination among heterologous sequences

† This work was supported by NIH Grant GM62154.

* Correspondence should be addressed to this author at the United States Army Medical Research Institute for Infectious Diseases, 1425 Porter St., Fort Detrick, MD 21702-5011. Telephone: 301-619-6032. Fax: 301-619-2152. E-mail: david.rozak@amedd.army.mil.

[1] Abbreviations: CD, circular dichroism; GA, protein G-related albumin binding; GPSA, guinea pig serum albumin; HSA, human serum albumin; IPTG, isopropyl β-D-thiogalactopyranoside; ITC, isothermal titration calorimetry; NMR, nuclear magnetic resonance; OR-PCR, offset recombinant PCR; PCR, polymerase chain reaction; PSD, phage-selected domain; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; TD, template domain; UV, ultraviolet.

10.1021/bi051926s CCC: $33.50
Published on Web 02/14/2006

(*19*, *25*−*28*), few readily produce the density of crossover events needed to efficiently shuffle families of small globular proteins such as that of the GA module.

Offset recombinant PCR (OR-PCR) is a novel strategy that appears to be capable of creating recombinant libraries from compact heterologous domains. Our initial characterization of the technique, which exploits elevated recombination frequencies near template ends and the exponential accumulation of recombinant templates during PCR, suggests that OR-PCR can generate multiple recombination events among compact heterologous domains similar in size and complexity to those defined by the GA module (*29*).

This paper describes the use of OR-PCR to create a library of recombinant GA modules, which is probed by phage display in an attempt to uncover differences between GA mutants required to bind one or two distinct albumin species. The two most prominent phage-selected domains were subjected to circular dichroic, calorimetric, and limited structural analysis in order to identify structural and functional determinants within the GA module and possibly determine whether backbone dynamics do in fact contribute to the broad affinity for different albumins observed in G148-GA3.

EXPERIMENTAL PROCEDURES

Primers. P1: 5′-CGA AGG TGT TGG CGA ACA GGG AGT TGG ACA AAT ACG GCG TGT CGG ATT ATT ACA AGA ATC TGA TCA ACA ACG CC-3′. P2: 5′-CGA TTA ACG CCT TCA CGC CTT CCA CGG TTT TGG CGT TGT TGA TCA GAT TCT TGT AAT AAT CCG-3′. P3: 5′-CAA GCG ATC CTG CAG CAT ATG GAG GCC GTG GAC GCC AAC AGC CTG GCG GAG GCG AAG GTG TTG GCG AAC AGG-3′. P4: 5′-GCT CAC GGC AGT CGC GGC CGC GAA TTC CGT CGG CAA GGC CGC CAA GAT CTC GTC GAT TAA CGC CTT CAC GCC-3′. P5: 5′-TTT TTG TGA TGC TCG TCA GG-3′. P6: 5′-TTC TGA GAT GAG TTT TTG TTC TGC-3′. P7: 5′-CCG CTG GAT TGT TAT TAC TCG-3′. P8: 5′-AAA AAG GAT CCG AGC GTC GCT TAC GTT GAA GAA GAC AAA GTA TTT AAA GCG ATG ATG GAG GCG GTG GAC GC-3′. P9: 5′-ACG TTC AAG CTT GGC CGC TTA TTC CGT CGG-3′.

A002 Assembly. The A002 G148-GA3 construct used in this study was assembled in two consecutive PCRs using the contiguous primers P1−P4. The final product was purified with the QIAquick PCR purification kit (Qiagen) before and after restriction digest with *Pst*I and *Not*I for insertion into pHEN1. The pHEN1 phage display vector was described by Hoogenboom et al. (*30*). Correct assembly of pHEN1/A002 was confirmed by DNA sequencing as described below.

Template Construction. Seven variants of A002 (TD-1 through TD-7) were produced by introducing point mutations into pHEN1/A002 using the QuickChange site-directed mutagenesis kit (Stratagene) per the protocol described in Wang and Malcolm (*31*). Each construct was generated from two consecutive QuickChange mutagenesis reactions in which changes were made separately to the 5′- and 3′-ends of the GA coding region. Stretches of amino acids that correspond to the complementary forward and reverse primers used during mutagenesis are underlined in Table 1. Where necessary, primers were constructed with randomized nucleotides to produce the amino acid polymorphisms shown in the table. Since these primers can be derived from the information presented here, they are not listed above. DNA sequencing of pHEN1 variants confirmed the accurate assembly of all seven templates.

Offset Recombinant PCR. OR-PCR was performed per the protocol described in our earlier paper (*29*). An equal mixture by mass of pHEN1 vectors containing TD-1 through TD-7 was subjected to six consecutive rounds of OR-PCR. The first of these reactions consisted of 2.5 units of cloned *Pfu* polymerase (Stratagene), 200 μM each dNTP, 0.5 μmol each of P5 and P6, and 100 ng of the mixed pHEN1 templates in 50 μL of the recommended reaction buffer. Subsequent reactions substituted a 2 μL aliquot from the previous reaction for the 100 ng template mix described above. Thermocycling began with 30 s at 95 °C followed by 30 cycles of 30 s at 95 °C, 30 s at 55 °C, and 1 min at 75 °C. PCR products were purified using the QIAquick PCR purification kit and concentrations determined via ultraviolet (UV) absorption at 260 nm.

The purified product was further amplified using a standard PCR protocol to remove heteroduplexes formed during OR-PCR and generate a smaller amplicon that was conducive to cleavage and religation into pHEN1. During amplification, 100 ng of the recombinant product was added to a PCR mix similar to the one described above and thermocycled for only eight cycles. P6 and P7 were used to generate a product with ends extending slightly beyond the indicated restriction sites. After purification with the QIAquick PCR purification kit, the product was cut with *Pst*I and *Not*I for ligation into pHEN1. The recombinant plasmids were transformed into XL-10 gold supercompetent cells (Stratagene), plated on LB agar containing 100 μg/mL ampicillin, and incubated overnight at 37 °C.

Phage Production. Transformed XL-10 gold supercompetent cells were grown in 20 mL of YT with 100 μg/mL ampicillin until mid-log phase. A 1 mL aliquot of the culture was then transferred to a fresh stock of 20 mL of YT containing ampicillin and 10^7 pfu of M13KO7 helper phage (New England Biolabs). After 1 h at 37 °C, 80 μg/mL kanamycin was added to the culture. Incubation continued at 37 °C with vigorous shaking for 16 h.

Phage Precipitation. Phage were precipitated by spinning the cell culture twice at 10000*g* for 30 min and discarding the pelleted cells. Four milliliters of PEG/NaCl (20% PEG 8000, 2.5 M NaCl) was added to the supernatant, and the solution was placed on ice for 20 min before being centrifuged at 10000*g* for 30 min. The supernatant was discarded and the pellet resuspended in 1 mL of TE buffer (100 mM Tris, pH 8.0, 0.1 mM EDTA). After addition of 200 μL of PEG/NaCl the sample was returned to ice for 20 min and centrifuged at 14000*g* for 15 min. The supernatant was discarded and the pellet resuspended in 1 mL of TE buffer for storage.

Biopanning. A 50 μL solution of Dynabeads M-280 streptavidin paramagnetic beads (Dynal Biotech) was pelleted on a magnetic manifold and resuspended in TBS Tween (50 mM Tris, pH 7.4, 150 mM NaCl, 0.5% Tween 20) and 0.1% dried milk. This solution was rocked overnight at 4 °C before the beads were repelleted on a magnetic manifold. Meanwhile, 10^8 pfu of pHEN1-containing phage was mixed with 1 μL of 10 nM biotinylated albumin and 1 mL of TBS Tween

before being rocked at room temperature for 3 h. The dried, essentially fat-free, human serum albumin (HSA) and guinea pig serum albumin (GPSA) samples used in this study were obtained from Sigma. The pelleted streptavidin beads were resuspended in the 1 mL phage solution and rocked for 30 min. Afterward, the beads were returned to the magnetic manifold to remove the supernatant before being washed seven times by rocking for 5 min at room temperature in 1 mL of TBS during each wash. Finally, the phage were eluted by resuspending the beads in 200 μL of 0.1 M glycine, pH 2.1, with 1 mg/mL bovine serum albumin (Sigma) and rocked for 20 min. The beads were pelleted and discarded before 10 μL of 2 M Tris base was added to the supernatant to neutralize the acid. Titers of the selected phage were obtained by creating serial dilutions of the neutralized solution, mixing 1 μL of each dilution with 100 μL of stationary phase TG-1 cells, and plating on LB agar with 100 μg/mL ampicillin. Phagemid-containing colonies were counted after the plates were incubated overnight at 37 °C.

The remaining eluted phage were mixed with 20 mL of YT, 100 μg/mL ampicillin, and 200 μL of stationary phase TG-1 before being shaken vigorously overnight at 37 °C. 10^7 pfu of M13KO7 helper phage was added to the culture 1 h into incubation to support production of GA-labeled phage. The resulting phage were precipitated as described above, and the entire biopanning process was repeated three more times before colonies were isolated from the titer plates for DNA sequencing and analysis.

Protein Production and Purification. Two of the phage-selected mutants (PSD-1 and PSD-7) were prepared for production, purification, and analysis by PCR amplification using P8 and P9. The PCR products were cut with *Bam*HI and *Hin*dIII for ligation into the pG58 vector. P8 and P9 were used specifically to add restriction sites and an ochre stop codon to the ends of the GA coding region. The pG58 vector enables expression of a subtilisin pro domain fusion protein, which permits the fused protein to be purified and cut with subtilisin in a one-step reaction (*32*). XL-10 gold supercompetent cells were transformed with the pG58/PSD construct and grown at 37 °C in 5 L of LB with 100 μg/mL ampicillin. At log phase the cells were induced with 1 mM IPTG for 3 h before being harvested. Cells were pelleted by spinning at 8000*g* for 30 min and resuspended in 100 mL of 100 mM potassium phosphate, pH 7.0, 30 μg/mL DNase I, and 0.1 mM PMSF for sonication on ice. Cellular debris was removed by centrifugation at 10000*g* for 30 min and 100000*g* for 1 h. Purification was carried out on a 5 mL S189 column essentially as described by Ruan et al. (*32*) and confirmed by SDS−PAGE.

Extinction Coefficients. The Edelhoch method, as described by Pace (*33*), was used to precisely determine extinction coefficients at 278 nm for each of the thermodynamically characterized GA mutants and albumins. This set of empirically determined extinction coefficients was used to compute protein concentrations from UV spectra throughout the study. Twice the absorbance at 331 nm was subtracted from that obtained at 278 nm to account for the effects of light scattering.

Circular Dichroism. Circular dichroism (CD) experiments were performed on a J-720 spectropolarimeter (Jasco Spectroscopic Co., Ltd.). Spectra for 2.96 μM PSD-1 and 2.67 μM PSD-7 in 50 mM potassium phosphate, pH 7.0, were obtained by measuring the ellipticity from 250 to 200 nm of the samples in a 1.0 cm cell at 25 °C. Melting temperatures for the same protein solutions were determined by measuring the ellipticity of the sample at 222 nm as it was heated in a 1.0 cm cell from 25 to 70 °C at 0.5 deg/min.

Isothermal Titration Calorimetry. Isothermal titration calorimetry (ITC) measurements were performed on the VP-ITC microcalorimeter (MicroCal). For each experiment the selected GA mutant and albumin were dialyzed side by side into 50 mM potassium phosphate, pH 7.0, to ensure identical buffer conditions. Each run involved nineteen 15 μL injections of approximately 250 μM GA domain into a sample cell containing around 25 μM albumin. Injections lasted 30 s each and were spaced 180 s apart. The jacket temperature was maintained at 25 °C throughout. Precise protein concentrations were determined by UV spectra as described above.

DNA Sequencing. DNA samples were prepared for sequencing by growing selected colonies overnight at 37 °C in 3 mL of LB with 100 μg/mL ampicillin and extracting plasmid DNA from the cell cultures using the Wizard Plus SV minipreps (Promega Corp.). The P2 primer was used to amplify target DNA using Perkin-Elmer/Applied Biosystem's AmpliTaq-FS DNA polymerase and Big Dye terminators with dITP. Dye-terminated products were then run on an Applied Biosystems model 3100 DNA sequencer to produce sequence chromatographs.

RESULTS

Reconstructing the Native GA Sequence Space. The 16 known members of the bacterial albumin binding module describe a finite sequence space, which encodes a range of three-helix domains with varied stabilities and albumin binding potentials. We sought to identify some of the biochemical determinants that specify phenotypic variation in these domains by shuffling representatives of the natural sequence space and selecting for broad or narrow albumin binding affinities.

This effort began by assembling the 56 amino acid protein (A002), which contained the 46 amino acid streptococcal albumin binding domain, G148-GA3, surrounded by unstructured flanking sequences. For consistency, the complete A002 amino acid sequence shown in Table 1 is largely identical to that used in earlier structural and thermodynamic studies of the streptococcal domain (*2*, *13*, *16*). *Not*I and *Pst*I restriction sites were used to insert A002 into pHEN1, an amp^R-tagged phagemid whose multiple cloning site supports the display of protein or polypeptide libraries on the surface of M13 phage by fusing the cloned fragments to the N-terminus of the gIII capsid protein (*30*).

Rather than reconstructing each of the 16 homologues shown in Table 1, we used PCR site-directed mutagenesis to create seven variants of A002, which cumulatively represented much of the natural diversity found among members of the GA module. These template domains (TD) have been labeled TD-1 to TD-7 in Table 1. Randomized bases were used to further increase TD coverage for the naturally defined GA sequence space. Although there is a slight disparity between the sequence spaces represented by the seven templates and the native GA domains, the approach significantly reduced the number of primers and reactions

Table 1: Sequence Sets Defined by Native, Template, and Phage-Selected Albumin Binding Domains

		10 20 30 40 50
	A002	MEAVDANSLAEAKVLANRELDKYGV-SDYYKNLINNAKTVEGVKALIDEILAALPTE
Native Domains [a, b]	*P. magna* L3316-GA1	.KN..EE.IK..KEA.IT..L.FS...K.....E..KN...K.
	P. magna L3316-GA2	.KN..ED.IK..KEA.IS..I.FDA..K.....E..KN...K.
	P. magna L3316-GA3	.KN..EA.IK..KEA.ITAE.LF....K.....ES.KN...K.
	P. magna L3316-GA4	.KN..ED.IK..KEA.IT..I.FDA..K...I...E..KN...K.
	P. magna ALB1-GA	.KN..ED.IA..K.A.IT..F.F.A..K.....AN..KN...K.
	P. magna ALB8-GA	.KN..ED.IA..K.A.IT..F.F.A..K.....E.N..KN...K.
	P. magna ALB1B-uGA	.Q...DK.IQ.AKAN.LT.KLLLKN.E....P.SA.SFAE.LIKS
	P. magna ALB8-uGA	.KLT.EE.EKA.K.L.IT.EFIL.Q.DK.TSR..LES.VQT.KQS
	S. dysgalactiae MAG-GA1	..KLAADTDLD..VAKIIN.-.TTKVE....A.D..KIFE.--SQ
	S. dysgalactiae MAG-GA2	..K..AD.IEI.K...I-G...IK....G..A...T..K.....S
	S. equi ZAG-GA	.L...EA.IN..KQ..I-....VT...K.....N..KA...S.
	S. canis DG12-GA2	.S...EM.I....AQ..-..F...K.....V..K.L..NS
	S. canis DG12-GA1	.DQ..QA.LK.F.R...-.N.....K.....IME.QAQVVES
	S. Streptococcus G148-GA1	..K..AD.LK.FN....-.....D.QAQVVES
	S. Streptococcus G148-GA2	.....-...H.....D.QAQVVES
	S. Streptococcus G148-GA3	.....-.....
Template Domains [b, c]		AA LK G A T
	TD-1	.....KN..EE.IR.....-.....N..KA.....
		A L
	TD-2	.....ED.IEI.K...I-G...IK.....ES.KN...K.....
		A LS Q
	TD-3	.....EA.IR..KK..I-....VT.....D.QAQVVES....
		A N NFH
	TD-4	.....EM.I....AQ..-.....K.....D.QAQVVES....
		AQ Q LK LNR G A T
	TD-5	.....EK..EA.IR.F.K...-.....N..KA.....
		SSE S
	TD-6	.....KEA.ITA...F....K.....D.QAQVVES....
		S L D
	TD-7	.....KEA.IT..T.F.A..K.....ES.KN...K.....
Phage-Selected Domains [b, d]	PSD-1	.....Q..EA.IK..KQ..I-G...IK.....ES.KN...K.....
	PSD-2	.....TQ..EA.IK..KQ..I-G...IK.....ES.KN...K.....
	PSD-3	.....AD.IEI.K...I-G...IK.....ES.KN...K.....
	PSD-4	.....ED.LEI.K...I-G...IK.....ES.KN.....
	PSD-5	.....ED.LEI.K...I-G...IK.....ES.KN.....
	PSD-6	.....EA.LS..KQ..I-....VT.....ES.KN...K.....
	PSD-7	.....EM.....-.....ES.KN...K.....
	PSD-8	.....-EM.....-.....ES.KN...K.....
	PSD-9	.....QM.....R...-.....ES.KN...K.....
	PSD-10	.....V.....KEA.IT..L.FDA..K...A...N..KA.....
	A002	MEAVDANSLAEAKVLANRELDKYGV-SDYYKNLINNAKTVEGVKALIDEILAALPTE
		10 20 30 40 50

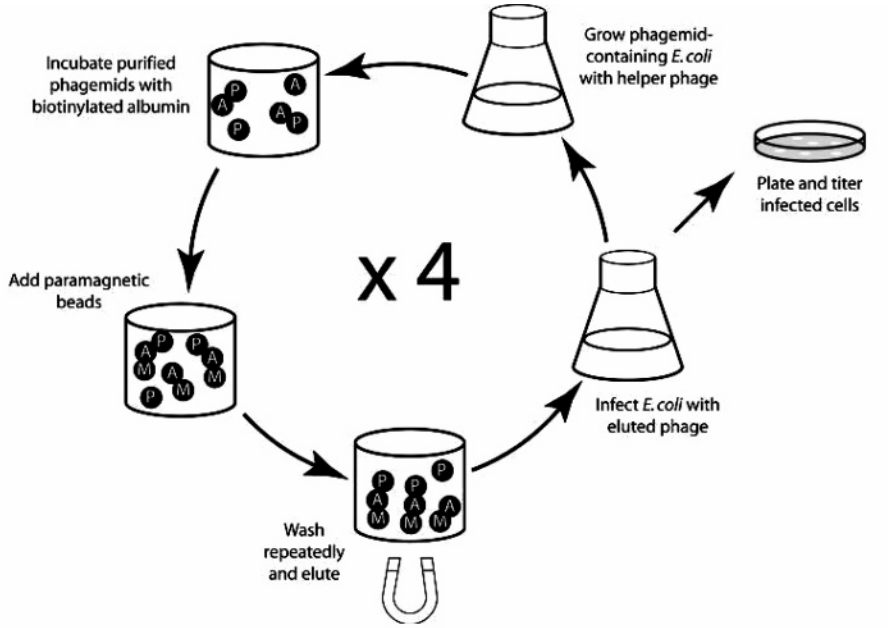
[a] Sequences of the native domain were previously compiled by Johansson et al. (*2*). [b] Gray characters are represented in only one of the three sequence spaces defined by the native, template, or phage-selected domains. [c] Underlines reveal the extent of primers used to construct the templates from A002 via site-directed mutagenesis. Stacked characters represent degeneracies caused by the use of a randomized nucleotide in the corresponding codon. [d] Gray blocks describe the regions in which crossover events likely occurred during OR-PCR.

required to produce a starting library. On average, each of the seven templates exhibits an 83% homology to one another within the 132 nt variable region subject to recombination.

Shuffling Template Domains via OR-PCR. Recent experiments suggest that OR-PCR offers an effective strategy for promoting recombination among compact heterologous domains similar in size and complexity to the members of the GA module studied here (*29*). The technique significantly elevates recombination frequency during standard PCR by locating the recombinant region near one end of the amplicon. Consequently, partially extended primers, which are terminated within the offset recombinant region, are more likely to form extendable heteroduplexes with one another during subsequent rounds of PCR rather than compete with high concentrations of unextended primers for full-length templates. Since the technique appears to benefit from the accumulation of shuffled templates in the reaction mix, recombination rates can be further increased by passing the products through multiple iterations of the OR-PCR process.

We applied OR-PCR to an equal mixture of the seven pHEN1/TD constructs by designing primers to produce a 766 bp amplicon with the start of the 177 nt GA coding region located only 24 nt from one end. The optimal elongation time for generating crossover events within the offset GA coding region was determined to be approximately 60 s, a value obtained by creating a similarly sized amplicon on the pUC19-03 and pUC19-05 constructs described previously (*29*) and measuring the impact of elongation time on *lacZ* phenotype rescue rates.

An equal mixture of the pHEN1/TD variants was shuffled via six consecutive rounds of OR-PCR. During this process, 2 μL samples from each 30-cycle reaction were transferred to fresh reaction mixes. After six rounds the final product was reintroduced into pHEN1 and cloned in *Escherichia coli* for infection with M13KO7 phage. Sequence analysis (data not shown) revealed that 12 of 15 pHEN1 samples obtained from phage-infected *E. coli* colonies had undergone recombination within the GA coding region with five of the



	Challenge Set A		Challenge Set B		Challenge Set C		Challenge Set D	
Round	Ligand	Output (pfu/mL)	Ligand	Output (pfu/mL)	Ligand	Output (pfu/mL)	Ligand	Output (pfu/mL)
1	HSA	2.0×10^5	HSA	1.2×10^5	GPSA	7.6×10^5	GPSA	8.0×10^5
2	HSA	5.3×10^6	GPSA	2.3×10^7	HSA	2.0×10^6	GPSA	2.5×10^7
3	HSA	2.1×10^6	HSA	1.2×10^6	GPSA	1.4×10^7	GPSA	1.2×10^7
4	HSA	6.1×10^7	GPSA	7.4×10^7	HSA	3.4×10^7	GPSA	7.1×10^7
Isolates	PSD-1 (x8), PSD-6, PSD-7 (x2), PSD-10		PSD-1 (x10), PSD-8		PSD-1 (x9), PSD-2, PSD-3, PSD-4		PSD-1 (x9), PSD-5, PSD-7, PSD-9	

FIGURE 1: Four rounds of selection, amplification, and purification were carried out on challenge sets A−D. During selection, purified phage (circled P) displaying GA recombinants were incubated with biotinylated albumin (circled A), exposed to streptavidin-coated paramagnetic beads (circled M), and washed repeatedly on a magnetic manifold to remove unbound phage. Infected *E. coli* were grown in the presence of helper phage to amplify selected phage and complete the cycle. Phage-infected *E. coli* were also plated to obtain the output titers reported in the accompanying table. Depending on the challenge set, albumins were either varied or maintained from one round to the next. Eleven to twelve phagemids were isolated from titer plates at the end of round 4 and sequenced to identify PSD-1−10 reported in Table 1.

sequences containing two to three crossover events. Eight of the 15 sequences exhibited frame-shift mutations that were likely to destroy the integrity of the albumin binding module. The presence of frame-shift mutations contrasted sharply with earlier OR-PCR experiments that showed no evidence of insertions, deletions, or point mutations after similar treatment (*29*). Although the frame-shift mutations reduced the number of viable species in the recombinant library by as much as half, we saw no evidence that any of the deleterious mutations were propagated through the phage selection process described below.

Selecting Phage-Displayed Mutants. It has been proposed that the dynamic G148-GA3 structure may contribute to the domain's broad affinity for albumin from different species (*11*). We sought to explore the notion that backbone dynamics were somehow tied to albumin specificity by designing a selection protocol in which phage-displayed GA recombinants were required to bind HSA, GPSA, or both albumins.

HSA and GPSA were chosen as target ligands because they appear to represent the diverse range of albumins bound by G148-GA3. Competitive binding assays show only a 10-fold difference in the abilities of the streptococcal domain to bind HSA and GPSA (*5*). At the same time, the *F. magna* ALB8-GA albumin binding domain was found to be a 1000-fold less capable at binding GPSA than HSA. Sequence analysis also recommended HSA and GPSA as ideal candidates for a study of GA binding specificities because the two albumins are on opposite ends of a phylogenetic tree depicting albumins from 11 different species (not shown here).

Four identical aliquots (challenge sets A−D) of phage displaying the GA recombinant library were each subjected to four consecutive rounds of selection, amplification, and precipitation as depicted in Figure 1. Challenge sets A and D were passed respectively over only HSA- or GPSA-labeled beads during each of the four rounds. Sets B and C alternated between the two albumins after each round in an effort to select mutants that could efficiently bind both species of ligand. Sets B and C differed only in their starting ligands in order to observe whether the initial selection criteria proved critical in determining which mutants were enriched.

After four rounds of amplification and selection challenge sets A−D showed significant signs of enrichment based upon elevated titers of phage in the eluant (Figure 1). For each of the four sets, 11−12 colonies of *E. coli* infected with phage derived from the fourth elution were isolated and sequenced. Ten distinct phage-selected domains (PSD-1 to PSD-10) were identified in the 47 sequences obtained from challenge sets A−D. More than half of these sequences, which are listed in Figure 1 and displayed in Table 1, exhibited two distinct crossover events. Two point mutations and a single codon deletion were also found among the selected mutants. Each of the biopanned challenge sets revealed a clear preference for PSD-1, which was represented by 36 of the 47 samples. The second most common mutant, PSD-7, was identified in three of the sequenced samples. All other mutants were found only once and tended to be close variants of PSD-1 or PSD-7.

Significantly, there was no discernible difference in the types or distributions of mutants appearing in each of the four challenge sets, suggesting that sequence polymorphisms in the human and guinea pig serum albumins had little effect on their respective abilities to enrich the dominant PSD-1 mutant.

Circular Dichroic Analysis of Selected Mutants. Circular dichroism was used to assess the structural and thermodynamic properties of folded PSD-1 and PSD-7. Figure 2 presents normalized CD data for both mutants alongside similar data obtained for G148-GA3 (*16*). Per Figure 2A, the CD profiles for both mutants are similar in shape to those previously observed for G148-GA3 (*13*, *15*) and indicative of peptides with a predominantly α-helical content. The G148-GA3 sample used in the referenced study uniquely contained a disordered six-histidine tag on its C-terminus, which likely contributed to the molecule's weaker signal in the normalized plot.

When mutant and wild-type domains were heated from 25 to 100 °C, they showed clean transitions from folded to unfolded states as indicated by the ellipticity at 222 nm. These transition curves were fit with upper and lower baselines to determine the ratio of unfolded to folded proteins (*K*) from the unfolded fractions depicted in Figure 2B and derive temperature-dependent values of ΔG using the equation

$$\Delta G = -RT \ln K \qquad (1)$$

where *R* is the gas constant. This transformation, which is plotted in Figure 2C, reveals that PSD-1 is more stable than G148-GA3 in the measured temperature range and has a T_m of approximately 85 °C. The opposite is the case for the less stable PSD-7, which reveals a T_m of 65 °C. The relative stabilities of the three domains are probably maintained as the solution reaches room temperature. However, the CD data do not permit us to accurately determine ΔC_p and predict the behavior of the free energy curves at lower temperatures.

Characterizing the Albumin Binding Reactions of Selected Mutants. ITC was used to determine thermodynamic state functions and binding constants for PSD-1 and PSD-7 interactions with both HSA and GPSA at 25 °C. The calorimetric technique measures heat produced when small amounts of protein bind an excess of ligand. The enthalpy involved in each binding reaction is equal to the area under the calorimetric curve. Enthalpies from successive reactions can be plotted as a function of the protein−ligand molar ratios to produce a transition curve and compute the binding constant *K*. Values for ΔG, ΔH, and ΔS are easily derived

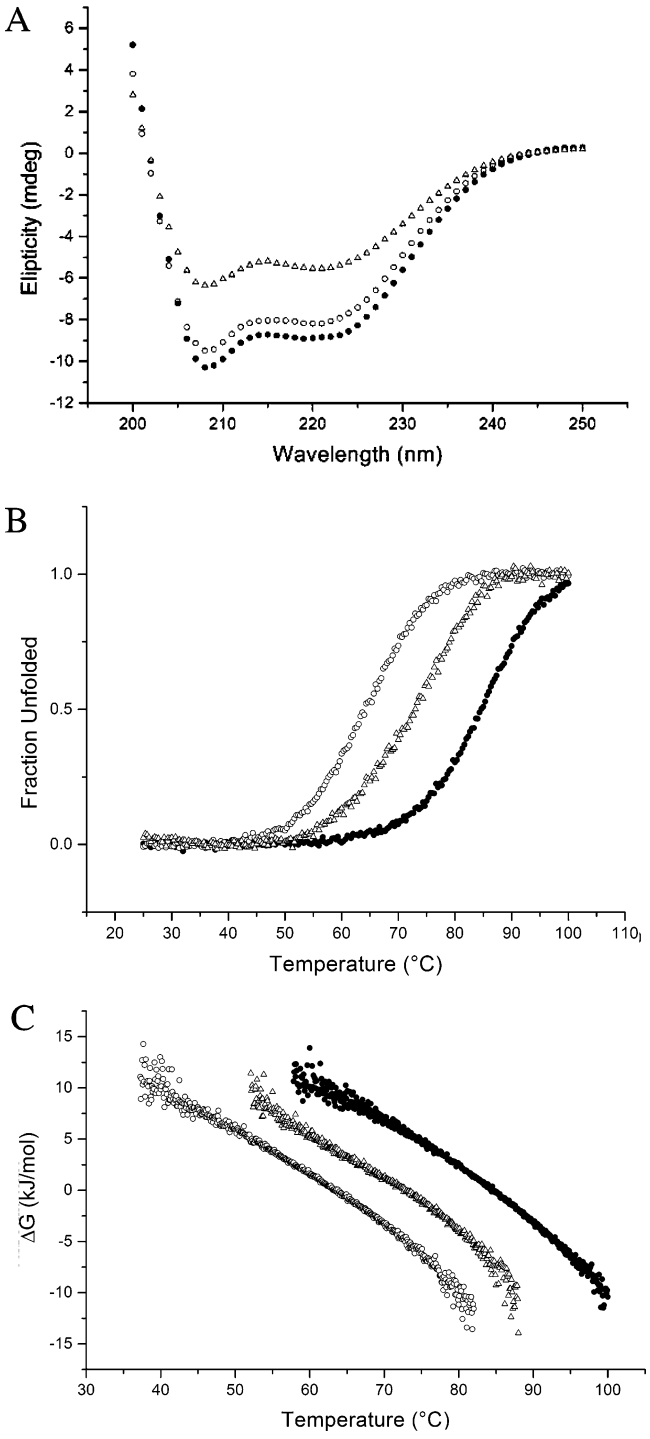


FIGURE 2: Circular dichroic measurements of PSD-1 (solid circles), PSD-7 (open circles), and G148-GA3 (open triangles) in 50−100 mM potassium phosphate, pH 7.0. G148-GA3 data were obtained from a separate study (*16*). (A) Spectral scans normalized to 1.0 μM for each mutant at 25 °C suggest that the folded proteins adopt α-helical secondary structures similar to those observed for the wild-type domain. (B) The temperature-dependent ratio (*K*) of unfolded to folded proteins was determined by fitting baselines to the superimposed 222 nm transition curves generated by heating each of the samples from 25 to 100 °C. (C) $\Delta G_{unfolding}$ was computed for each domain per eq 1.

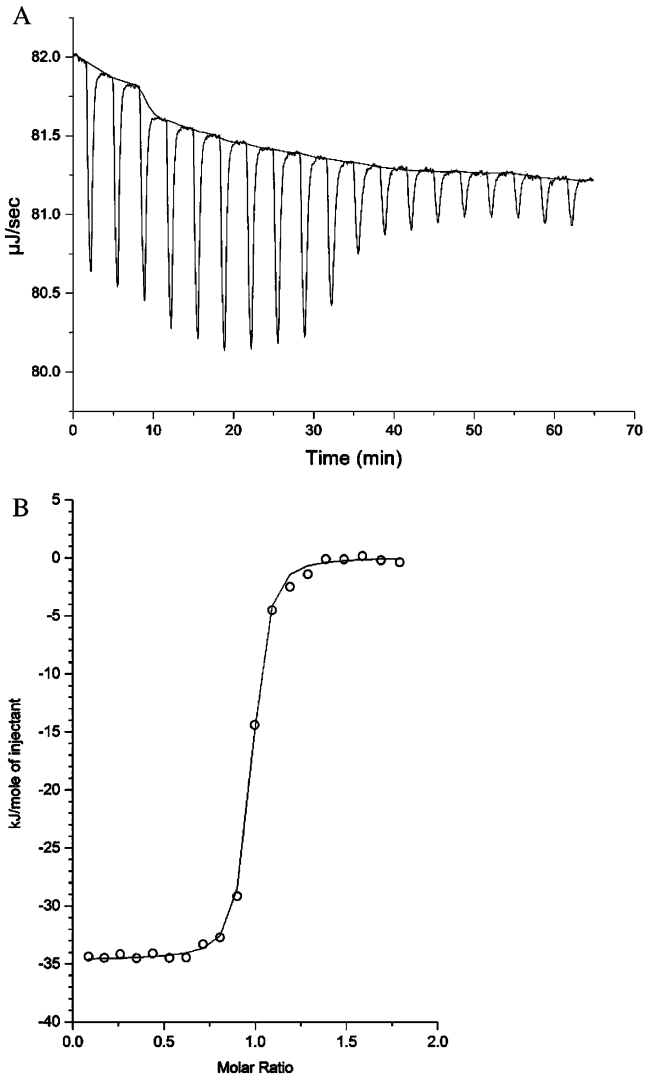


FIGURE 3: ITC analysis of the PSD-1/HSA binding at 25 °C in 50 mM potassium phosphate, pH 7.0. (A) Instantaneous heat generated by adding 1.2 nmol aliquots of PSD-1 to 12.6 nmol of HSA at 3 min intervals. (B) Total heat produced per mole of injectant as a function of the molar ratio of protein to ligand. The binding constant and van't Hoff enthalpy corresponding to the transition curve are reported in Table 2. The midpoint of the transition occurs at equivalent concentrations of protein and ligand, indicating 1 to 1 stoichiometry.

from this analysis. An example of the calorimetric data obtained for PSD-1/HSA binding at 25 °C is given in Figure 3. The midpoints for each of the binding reactions occurred when equal molar concentrations of protein and ligand were present, indicating 1 to 1 stoichiometry.

According to the ITC data reported in Table 2, the PSD-1/HSA binding constant is nearly five times that of the native streptococcal domain. Furthermore, PSD-1's GPSA binding constant is twice the value obtained for HSA. This is significant given the observation that competitive binding assays for G148-GA3 and ALB8-GA show 10- and 1000-fold decreases in their abilities to bind GPSA as compared to HSA (*5*). Remarkably, PSD-7, which is less stable than G148-GA3 at higher temperatures, yields a modest gain over the native domain in binding HSA while maintaining an equivalent affinity for GPSA.

Identifying Functional Determinants in Selected Mutants. In an effort to associate phage-selected sequence polymorphisms with observed changes in protein thermodynamics for folding and binding, the locations of PSD-1 and PSD-7 mutations were mapped onto the published structure of a wild-type GA domain. Of the two albumin binding domains represented in the Protein Data Bank, PSD-1 and PSD-7 exhibit the greatest sequence homology to G148-GA3. Therefore, we used the G148-GA3 NMR structure reported by Johansson et al. (*5*) (PDB code 1GJT) as scaffolding on which to locate and predict the possible impact of the phage-selected polymorphisms. This analysis is shown in Figure 4.

Table 2: Albumin Binding Data for G148-GA3 and Phage-Selected Mutants[a]

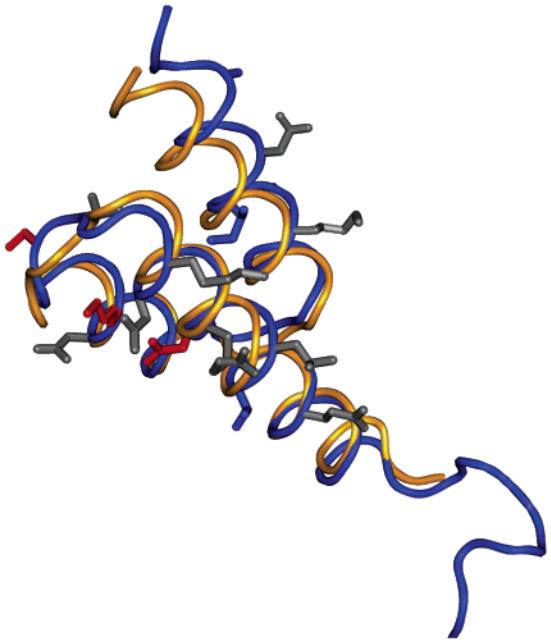
		G148-GA3[b]	PSD-1	PSD-7
HSA	K (mol^{-1})	1.2 (±0.1) × 10^7	5.4 (±0.6) × 10^7	3.5 (±0.6) × 10^7
	ΔG (kJ/mol)	−40.4 ± 0.2	−44.1 ± 0.3	−43.0 ± 0.4
	ΔH (kJ/mol)	−31.9 ± 0.1	−8.27 ± 0. 04	−11.5 ± 0.1
	ΔS (J/mol)	29 ± 1	120 ± 1	106 ± 2
GPSA	K (mol^{-1})	0.9 (±0.8) × 10^7	1.1 (±0.2) × 10^8	0.7 (±0.2) × 10^7
	ΔG (kJ/mol)	−40 ± 2	−45.9 ± 0.4	−39.1 ± 0.6
	ΔH (kJ/mol)	−3.7 ± 0.2	−6.41 ± 0.05	−3.4 ± 0.1
	ΔS (J/mol)	121 ± 6	132 ± 2	129 ± 2

[a] All values were obtained in 50 mM potassium phosphate, pH 7.0 at 25 °C. [b] Data obtained from a recently published calorimetric study (*16*).

Although the G148-GA3 NMR structure fails to pinpoint the albumin binding epitope, a recently reported crystal structure of *F. magna* ALB8-GA complexed with HSA (*17*) (PDB code 1TFO) shows that this interface occurs on the surface defined by the second and third helices of the albumin binding domain. Because the overlaid G148-GA3 and ALB8-GA structures are highly conserved and closely aligned in this region, it is reasonable to conclude that the binding epitopes are largely identical as well. The structural alignments shown in Figure 4 were made using the National Center for Biotechnology Information's (NCBI) vector alignment search tool (VAST) (*34*).

ITC analysis indicates that PSD-7 exhibits elevated and equivalent affinities for HSA and GPSA, respectively, when compared to G148-GA3, despite the recombinant's lower stability at high temperatures. Improvement to PSD-7 albumin binding is likely supported by A45S, which appears by comparison with the crystallized ALB8-GA/HSA complex to hydrogen bond with T260 on HSA (Accession Number P02768) and GPSA (Accession Number AY294645). This interaction is not supported by G148-GA3 and could give the less stable PSD-7 an added advantage in binding the two albumins. While the same mutation exists in PSD-1, its impact on albumin binding is likely mitigated by the simultaneous loss of a native G148-GA3 interaction with HSA and GPSA N342 through S27G.

Absent substantial enhancements to the binding epitope, PSD-1's superior ability to bind HSA and GPSA compared to PSD-7 and G148-GA3 might be driven by the gains in the mutant's stability observed during CD melts. The 10 polymorphisms that differentiate PSD-1 from PSD-7 are located mostly on solvent-exposed surface residues of the first two helices. Being confined to the surface of the domain, these polymorphisms are unlikely to significantly destabilize PSD-7 through steric conflicts with other residues. Further-



Position	A002 (G148-GA3)	PSD-1	PSD-7	Possible effect of selected mutation
11	E	**Q**	E	
14	V	**E**	**E**	Decreases hydrophobic surface area.
15	L	**A**	**M**	
17	N	**I**	N	Increases hydrophobic surface area.
18	R	**K**	R	
21	D	**K**	D	
22	K	**Q**	K	
25	V	**I**	V	
27	S	**G**	S	Breaks H-bond with N342 in HSA and GPSA. Increases hydrophobic surface area.
31	K	**I**	K	Increases hydrophobic surface area.
32	N	**K**	N	
44	K	**E**	**E**	
45	A	**S**	**S**	Forms H-bond with T260 in HSA and GPSA. Decreases hydrophobic surface area.
47	I	**K**	**K**	Removes bulky isoleucine from interior. Decreases hydrophobic surface area.
48	D	**N**	**N**	
52	A	**K**	**K**	Decreases hydrophobic surface area.

FIGURE 4: Structural analysis of phage-selected mutants. The graphic depicts a G148-GA3 NMR structure (blue ribbon) superimposed over a crystal structure of ALB8-GA (orange ribbon) complexed with HSA (not rendered in this figure). The N-termini for both homologues are located in the lower right-hand corner of the picture. The side chains are shown for G148-GA3 residues whose hydrophobicities increase (red), decrease (blue), or remain largely unchanged (gray) as a result of mutations in PSD-1 or PSD-7. G148-GA3, PSD-1, and PSD-7 polymorphisms are listed on the right along with a brief description of how these mutations might impact folding and binding in the phage-selected mutants.

more, these polymorphisms do not appear to encode the changes in protein stability via discernible differences in their hydrophobic natures. In fact, the three PSD-1 mutations that do impact protein hydrophobicity (N17I, S27G, and K31I) collectively increase the hydrophobic surface area of PSD-1, changes that are likely to reduce rather than enhance the overall stability of the domain by promoting thermodynamically unfavorable order in the surrounding solvent.

The alignment in Figure 4 of G148-GA3 and ALB8-GA may offer an important insight into PSD-1's enhanced stability. While closely aligned along the first and second helices, the superimposed wild-type domains appear to differ with respect to the orientations of the third helix. The out-lying G148-GA3 helix may account for a solvent-accessible gap in the G148-GA3 structure, which is absent in the ALB8-GA domain. The only difference between the highly conserved G148-GA3 and ALB8-GA core residues is an I47K polymorphism, which also exists between G148-GA3 and PSD-1. The ALB8-GA crystal structure shows that K47 extends from the domain's interior to form an ionic interaction with E19, which is also found in PSD-1. This lysine is less bulky than the similarly placed G148-GA3 isoleucine and may allow the third ALB8-GA and PSD-1 helices to nestle closer to the protein core, possibly accounting for the enhanced PSD-1 stability observed in Figure 2. NMR analysis of PSD-1 is currently underway to explore this possibility.

DISCUSSION

Families of homologous proteins such as the GA albumin binding module provide opportunities for biologists to study the mechanisms underlying observed phenotypic variations. Unfortunately, thermodynamic, kinetic, and structural data are rarely available for more than a few members of a family. Even when multiple homologues have been characterized, the diverse evolutionary pressures placed upon native proteins by their environments make it difficult to predict how polymorphisms support discrete biological functions.

Phage display and selection of recombinant libraries offer a promising strategy for unraveling the complexities of natural sequence spaces by permitting researchers to efficiently pan for functional determinants under well-defined conditions. We used the recently developed OR-PCR to shuffle a library of seven homologues defined by the natural sequence space of a medically significant family of small globular domains. Despite the occurrence of frame-shift mutations in some members of the recombinant population, OR-PCR proved to be highly effective at generating double or triple crossover events in half of the sequenced samples, probably due to the accumulation of recombinant templates and primers in the mix (*29*).

Far from revealing a distinct set of determinants for recombinants required to bind either one or two species of albumin, each of the four challenge sets exhibited an overwhelming preference for the stabilized PSD-1 mutant. These results fail to support the possibility that G148-GA3's dynamic backbone enhances its ability to efficiently bind phylogenetically diverse human and guinea pig serum albumins (*11*). However, careful analysis of PSD-1 backbone dynamics is required to support this conclusion. Furthermore, it is certainly possible that the streptococcal domain's dynamic structure supports binding to other phylogenetically diverse albumins not considered here.

PSD-1's increased stability over the wild-type streptococcal domain may be attributable to the loss of a core isoleucine at position 47, which appears to push the third G148-GA3 helix away from the protein's core. This mutation was also

among 10 PSD-1 polymorphisms that produced substantial gains over PSD-7 stability and albumin binding affinity despite the loss of hydrophilic surface residues and a native hydrogen-bonding partner in the albumin binding epitope. These results suggest a course for future research and underscore the value of OR-PCR and phage display in uncovering novel and potentially predictive insights from the complex array of natural polymorphisms found in many protein families.

ACKNOWLEDGMENT

We gratefully acknowledge the guidance and assistance we received from Biao Ruan on phage display, Kathryn Fisher regarding protein production and purification, and Fred Schwarz and Chittoor Swaminathan pertaining to protein calorimetry. We are also indebted to Fenhong Song's efforts to synthesize the many primers required for this study.

REFERENCES

1. de Château, M., and Björck, L. (1994) Protein PAB, a mosaic albumin-binding bacterial protein representing the first contemporary example of module shuffling, *J. Biol. Chem. 269*, 12147−12151.
2. Johansson, M. U., de Château, M., Wikström, M., Forsén, S., Drakenberg, T., and Björck, L. (1997) Solution structure of the albumin-binding GA module: a versatile bacterial protein domain, *J. Mol. Biol. 266*, 859−865.
3. Johansson, M. U., de Château, M., Björck, L., Forsén, S., Drakenberg, T., and Wikström, M. (1995) The GA module, a mobile albumin-binding bacterial domain, adopts a three-helix-bundle structure, *FEBS Lett. 374*, 257−261.
4. de Château, M., Holst, E., and Björck, L. (1996) Protein PAB, an albumin-binding bacterial surface protein promoting growth and virulence, *J. Biol. Chem. 271*, 26609−26615.
5. Johansson, M. U., Frick, I. M., Nilsson, H., Kraulis, P. J., Hober, S., Jonasson, P., Linhult, M., Nygren, P. A., Uhlén, M., Björck, L., Drakenberg, T., Forsén, S., and Wikström, M. (2002) Structure, specificity, and mode of interaction for bacterial albumin-binding modules, *J. Biol. Chem. 277*, 8114−8120.
6. Hammarberg, B., Nygren, P. A., Holmgren, E., Elmblad, A., Tally, M., Hellman, U., Moks, T., and Uhlén, M. (1989) Dual affinity fusion approach and its use to express recombinant human insulin-like growth factor II, *Proc. Natl. Acad. Sci. U.S.A. 86*, 4367−4371.
7. Nygren, P. A., Flodby, P., Andersson, M. U., and Wigzell, H. (1991) In vivo stabilization of a human recombinant CD4-derivate by fusion to a serum-albumin-binding receptor, *Vaccines 96*, 363−368.
8. Makrides, S. C., Nygren, P. A., Andrews, B., Ford, P. J., Evans, K. S., Hayman, E. G., Adari, H., Uhlén, M., and Toth, C. A. (1996) Extended in vivo half-life of human soluble complement receptor type 1 fused to a serum albumin-binding receptor, *J. Pharmacol. Exp. Ther. 277*, 534−542.
9. Sjölander, A., Nygren, P. A., Ståhl, S., Berzins, K., Uhlén, M., Perlmann, P., and Andersson, R. (1997) The serum albumin-binding region of streptococcal protein G: a bacterial fusion partner with carrier-related properties, *J. Immunol. Methods 201*, 115−123.
10. Libon, C., Corvaia, N., Haeuw, J. F., Nguyen, T. N., Ståhl, S., Bonnefoy, J. Y., and Andreoni, C. (1999) The serum albumin-binding region of streptococcal protein G (BB) potentiates the immunogenicity of the G130-230 RSV-A protein, *Vaccine 17*, 406−414.
11. Johansson, M. U., Nilsson, H., Evenäs, J., Forsén, S., Drakenberg, T., Björck, L., and Wikström, M. (2002) Differences in backbone dynamics of two homologous bacterial albumin-binding modules: implications for binding specificity and bacterial adaptation, *J. Mol. Biol. 316*, 1083−1099.
12. Linhult, M., Binz, H. K., Uhlén, M., and Hober, S. (2002) Mutational analysis of the interaction between albumin-binding domain from streptococcal protein G and human serum albumin, *Protein Sci. 11*, 206−213.
13. Kraulis, P. J., Jonasson, P., Nygren, P. A., Uhlén, M., Jendeberg, L., Nilsson, B., and Kördel, J. (1996) The serum albumin-binding domain of streptococcal protein G is a three-helical bundle: a heteronuclear NMR study, *FEBS Lett. 378*, 190−194.
14. Sjobring, U. (1992) Isolation and molecular characterization of a novel albumin-binding protein from group G streptococci, *Infect. Immun. 60*, 3601−3608.
15. Gulich, S., Linhult, M., Nygren, P., Uhlén, M., and Hober, S. (2000) Stability towards alkaline conditions can be engineered into a protein ligand, *J. Biotechnol. 80*, 169−178.
16. Rozak, D. A., Orban, J., and Bryan, P. N. (2005) G148-GA3: A streptococcal virulence module with atypical thermodynamics of folding optimally binds human serum albumin at physiological temperatures, *Biochim. Biophys. Acta 1753*, 226−233.
17. Lejon, S., Frick, I. M., Björck, L., Wikström, M., and Svensson, S. (2004) Crystal structure and biological implications of a bacterial albumin binding module in complex with human serum albumin, *J. Biol. Chem. 279*, 42924−42928.
18. Zhao, H., and Arnold, F. H. (1997) Functional and nonfunctional mutations distinguished by random recombination of homologous genes, *Proc. Natl. Acad. Sci. U.S.A. 94*, 7997−8000.
19. Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) Novel family shuffling methods for the in vitro evolution of enzymes, *Gene 236*, 159−167.
20. Moore, G. L., Maranas, C. D., Lutz, S., and Benkovic, S. J. (2001) Predicting crossover generation in DNA shuffling, *Proc. Natl. Acad. Sci. U.S.A. 98*, 3226−3231.
21. Moore, G. L., and Maranas, C. D. (2002) Predicting out-of-sequence reassembly in DNA shuffling, *J. Theor. Biol. 219*, 9−17.
22. Maheshri, N., and Schaffer, D. V. (2003) Computational and experimental analysis of DNA shuffling, *Proc. Natl. Acad. Sci. U.S.A. 100*, 3071−3076.
23. Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling, *Nature 370*, 389−391.
24. Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10747−10751.
25. Kikuchi, M., Ohnishi, K., and Harayama, S. (2000) An effective family shuffling method using single-stranded DNA, *Gene 243*, 133−137.
26. Ostermeier, M., Nixon, A. E., and Benkovic, S. J. (1999) Incremental truncation as a strategy in the engineering of novel biocatalysts, *Bioorg. Med. Chem. 7*, 2139−2144.
27. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D., and Benkovic, S. J. (2001) Creating multiple-crossover DNA libraries independent of sequence identity, *Proc. Natl. Acad. Sci. U.S.A. 98*, 11248−11253.
28. Gibbs, M. D., Nevalainen, K. M., and Bergquist, P. L. (2001) Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling, *Gene 271*, 13−20.
29. Rozak, D. A., and Bryan, P. N. (2005) Offset recombinant PCR: a simple but effective method for shuffling compact heterologous domains, *Nucleic Acids Res. 33*, e82.
30. Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P., and 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*, 4133−4137.
31. Wang, W., and Malcolm, B. A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis, *BioTechniques 26*, 680−682.
32. Ruan, B., Fisher, K. E., Alexander, P. A., Doroshko, V., and Bryan, P. N. (2004) Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification, *Biochemistry 43*, 14539−14546.
33. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci. 4*, 2411−2423.
34. Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) Surprising similarities in structure comparison, *Curr. Opin. Struct. Biol. 6*, 377−385.

BI051926S