Published in final edited form as:

J Mol Biol. 2009 October 9; 392(5): 1221–1231. doi:10.1016/j.jmb.2009.07.067.

Structural basis for exquisite specificity of affinity clamps, synthetic binding proteins generated through directed domain-interface evolution

Jin Huang, Koki Makabe, Matthew Biancalana, Akiko Koide, and Shohei Koide *

Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, U.S.A.

Abstract

We have recently established a new protein-engineering strategy termed “directed domain-interface evolution” that generates a binding site by linking two protein domains and then optimizing the interface between them. Employing this strategy, we have generated synthetic two-domain “affinity clamps” using PDZ and fibronectin type III (FN3) domains as the building blocks. While these affinity clamps all had significantly higher affinity toward a target peptide than the underlying PDZ domain, two distinct types of affinity clamps were found in terms of target specificity. One type conserved the specificity of the parent PDZ domain, and the other dramatically increased the specificity. Here, we characterized their specificity profiles using peptide phage-display libraries and scanning mutagenesis, which suggested a significantly enlarged recognition site of the high-specificity affinity clamps. The crystal structure of a high-specificity affinity clamp showed extensive contacts with a portion of the peptide ligand that is not recognized by the PDZ domain, thus rationalizing the affinity clamp’s improvement in specificity. A comparison with another affinity clamp structure revealed that, although both had extensive contacts between PDZ and FN3 domains, they exhibited a large offset in the relative position of the two domains. Our results indicate that linked domains could rapidly fuse and evolve as a single functional module and that the inherent plasticity of domain interfaces allows for the generation of diverse active-site topography. These attributes of directed domain-interface evolution provide facile means to generate synthetic proteins with a broad range of functions.

Keywords

PDZ; phage display; protein engineering; FN3; antibody alternative

The principles governing the relationship between protein sequences, structures and functions have traditionally been studied through characterization of natural proteins. These investigations have resulted in an impressive body of knowledge, which in conjunction with advances in molecular biology technologies has established a new branch of “synthetic protein

© 2009 Elsevier Ltd. All rights reserved.

*To whom correspondence should be addressed. skoide@uchicago.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Accession Number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 3CH8.

science.” The design, production, and analysis of synthetic proteins with novel structure and/or function rigorously tests our fundamental understanding of proteins, and also serves as a rich source of powerful tools for research and therapy. Structure-based design, computational design and directed evolution are major methodologies effective in generating new protein folds,^{1; 2; 3} binding interfaces^{4; 5} and catalysts.^{6; 7} Because, unlike natural proteins, such synthetic proteins are not constrained by the evolutionary requirements of the host organism, they are less biased platforms for determining the range of structure and function that proteins are capable of producing. However, efforts to date have focused primarily on the redesign of preexisting functional sites, such as the antigen-binding site of antibodies and enzyme active sites.^{5; 8; 9; 10} As a result, the starting scaffold architecture imposes strong restraints on the extent of new function to be implemented.

Comparative structural analysis of natural proteins suggests that dramatic changes in protein function have emerged through joining protein domains and adjusting the newly formed domain-interface.^{11; 12} Inspired by this observation, we have developed a strategy, termed “directed domain-interface evolution,” to generate leaps in protein functions.¹³ In this approach, two evolutionarily unrelated protein domains are covalently linked through a short linker, and the newly formed domain interface is subjected to directed evolution (Fig. 1a). In our proof-of-concept experiments, we have generated a set of synthetic proteins collectively termed “affinity clamps” consisting of two domains that synergistically interact with the ligand through a clamp-like architecture. In the affinity clamps, one domain serves as the capture domain that binds to a short peptide with weak affinity. The second serves as the enhancer domain that, after directed evolution, provides an optimized interface for the peptide presented by the capture domain. The resulting synergy between the two domains in “clamping” the target achieves orders of magnitude enhancement of affinity and specificity over those of the capture domain.

The first set of affinity clamps were constructed with the PDZ domain from human protein Erbin as the capture domain (Fig. 1a). Erbin-PDZ binds to the C-termini of p120-related catenins (δ -catenin and Armadillo repeat gene deleted in Velo-cardio-facial syndrome (ARVCF)) with a low-micromolar dissociation constant (K_d).¹⁴ The fibronectin type III domain of human fibronectin (FN3) was used as the enhancer domain. FN3 is a robust scaffold for producing antibody-like binding proteins with three surface loops available for creating a repertoire of binding interfaces.^{4; 15} The two domains were connected using a short linker, and the FN3 loops were diversified to generate a phage-display library. From this library, we identified a pair of affinity clamps, termed ePDZ-a and ePDZ-b, that had significantly higher affinity ($K_d = 56$ nM) than the parent PDZ (3 μ M) (Fig. 1b) to a C-terminal peptide derived from the ARVCF sequence. Affinity maturation of ePDZ-b resulted in further enhancement of affinity to the single nanomolar range (Fig. 1b).¹⁶ These results provided direct experimental support for the long-standing postulate on the role of domain combination in the evolution of protein function, and established a new branch of directed evolution-based synthetic protein science.

An intriguing aspect of these affinity clamps is that while both ePDZ-a and ePDZ-b exhibited similar levels of binding affinity to the target peptide, their specificity profiles differ dramatically. The ePDZ-b family showed higher levels of specificity by discriminating between two closely related peptides by as much as 6000-fold (Fig. 1b). Perplexingly, the affinity of ePDZ-b family to the δ -catenin peptide (Fig. 2a; sequence: PASPDSWV-COOH) was reduced relative to the starting PDZ domain, while the affinity to the ARVCF peptide (Fig. 2a; sequence: QQPVDSWV-COOH) was substantially increased. In contrast, ePDZ-a exhibited similar enhancement in affinity for both peptides.¹³ The crystal structure of ePDZ-a revealed that its FN3 surface interacts with the same region of the peptide recognized by the

PDZ domain, and thus the structure did not offer a rationale for how the higher level of specificity of the ePDZ-b family is achieved.¹³

In the current study, we aimed to elucidate the molecular mechanism responsible for the dramatically different specificity profiles exhibited by the affinity clamps. First, we used phage-displayed peptide libraries and alanine scanning mutagenesis to more comprehensively profile the binding specificity of each affinity clamp. We then determined the crystal structure of one of the high specificity affinity clamps, ePDZ-b1. The comparison of the ePDZ-b1 and ePDZ-a structures reveals significantly different peptide-binding modes, rationalizing their differences in specificity. These results demonstrate the inherent plasticity of an active site located at a domain interface, which allows for the generation of diverse interface topography and function.

Results

Binding Specificity of Affinity Clamps

Our previous measurements of binding of affinity clamps to the homologous δ -catenin and ARVCF peptides suggested that, while ePDZ-a retained essentially the same binding specificity as the parent PDZ, the ePDZ-b family exhibited much higher specificity (Fig. 1b). Because the C-terminal four residues are identical between these two peptides, these results indicate that ePDZ-b family clamps recognize one or more residues upstream of the core DSWV-COOH segment.¹³ To characterize the binding specificity of the ePDZ-b family clamps more comprehensively, we performed selection of C-terminal peptide libraries in the phage-display format. In this approach, multiple peptide sequences that bound to the ePDZ-b family clamps are identified from highly diverse libraries.¹⁴ The level of tolerance to mutations at particular positions in these peptides is used to deduce the binding specificity of the clamp.

We first performed selection using a library containing randomized amino acids at the seven C-terminal positions of the peptide. We refer to the PDZ peptides using conventional numbering, i.e., the very C-terminal residue is designated as position 0 and the remaining positions are denoted with increasingly negative numbers toward the N-terminus as -1, -2 and so on. According to this numbering, the first library contains amino acids diversified from the -6 to 0 positions. Because we already knew that the ePDZ-b family clamps preserved the underlying specificity of the parent PDZ domain with a motif (D(S/T)WV-COOH) as previously defined,^{14; 17; 18} the amino acid compositions at -2 and 0 of this library was biased toward this motif, thereby reducing the sequence space that needed to be sampled. Namely, all the 20 amino acid types were introduced to position -1 and -3 to -6, a combination of S, T, A, V, I and F to position -2, and I, L, V, A, T and P to position 0. After three rounds of library sorting, we obtained binding clones for each ePDZ-b family member (Fig. 2b and Supplementary Fig. 1a). The C-terminal four positions (-3 to 0) converged completely to an Erbin PDZ motif, DTWV-COOH, and positions -6 to -4 also showed appreciable levels of convergence. The preference for Thr over the wild-type Ser at the -2 position is consistent with that of the parent PDZ domain.^{14; 17} Ile, Leu and Met were found at the -4 position (Val in the original ARVCF target), Asn and Pro were common at the -5 position (originally Pro) and Gly and Ser were common at the -6 position (originally Gln). These results suggest that the ePDZ-b family clamps indeed specifically recognize a larger peptide segment than the recognition motif of the underlying PDZ domain, which predominantly recognizes the C-terminal four residues of the peptide.

To further characterize the sequence specificity of ePDZ-b family at the positions beyond the C-terminal D(S/T)WV segment, we constructed a second library in which the C-terminal four residues were fixed as the ARVCF sequence, DSWV, and the five preceding positions were diversified with all 20 amino acids. Specificity profiles produced with this library were similar

for ePDZ-b family members (Fig. 2c and Supplementary Fig. 1b). The motif for ePDZ-b was $X^{-8}(R)^{-7}(G/M)^{-6}X^{-5}(I/L/M)^{-4}$, that for ePDZ-b1 was $X^{-8}(R)^{-7}(G/S/\text{neutral})^{-6}X^{-5}(I/L/M)^{-4}$, and ePDZ-b2 showed further convergence at the -5 position to N and S. The apparent disagreements between results with the first and second peptide libraries (Fig. 2b and 2c) are likely due to low coverage of sequences caused by the small numbers of recovered sequences particularly from the first library (Supplementary Fig. 1). The patterns of the N-terminal three residues did not resemble the ARVCF sequence (PQPVD~~S~~WV-COOH) to which these affinity clamps were targeted, indicating that the affinity clamps could be further optimized to achieve higher binding affinity and possibly specificity. Indeed, a peptide designed to encode the most frequently observed amino acid at each position (RGSIDTWV-COOH) bound to ePDZ-b1 and ePDZ-b2 with eight-fold higher affinity relative to the ARVCF peptide (Supplementary Fig. 2b). The dissociation half-life of the new peptide from ePDZ-b1 and ePDZ-b2 was about four hours. Compared with the binding motif of Erbin PDZ ($\Phi\Phi(\text{DE})(\text{T/S})\text{WV-COOH}$; Φ refers to a hydrophobic amino acid), the specificity profiles of ePDZ-b family were clearly more stringent and distinct at positions -4, -6 and -7, indicating that the domain interface evolution strategy can significantly expand the size of the recognition site beyond that of the capture domain.

Alanine-scanning mutagenesis analysis

Although the phage-display library approach is powerful in determining general trends in sequence specificity, large-scale sequencing is necessary for quantitatively characterizing protein-interaction energetics.^{17; 19} This requirement is acute particularly when there is a strong amino acid preference and thus recovered clones are dominated by a particular pattern, as in our case above. Accordingly, to complement the library sorting experiments described above, we performed Ala-scanning experiments of the ARVCF peptide coupled with affinity measurements using SPR (Supplementary Fig. 2a). Because our goal was to understand the differences in target recognition of the affinity clamps, and because the sequence specificity of the C-terminal four residues is predominantly defined by the specificity of the parent PDZ domain, we analyzed only those positions that differed between the ARVCF and δ -catenin peptides.

The effects of alanine substitution of the ARVCF peptide on its binding to ePDZ-a and the ePDZ-b family varied greatly. Significant decreases in binding ($\Delta\Delta G > 1.0$ kcal/mol; corresponding to > 5 -fold decrease in binding affinity) were observed for substitution at each of the four positions (positions -7 to -4) for ePDZ-b family members, whereas only small effects were seen for all four positions for ePDZ-a and Erbin PDZ (Fig. 2d). The largest reduction ($\Delta\Delta G > 3$ kcal/mol) was observed at position -4 for all of the ePDZ-b family members. Also, alanine substitution primarily affected the dissociation rate of binding, and had only marginal effects on association rate, as often observed for this type of mutation.²⁰ Taken together, both the Ala substitution results and the peptide library sorting results indicate that the significantly enhanced binding specificity of the ePDZ-b family is a result of their ability to recognize a larger segment of the target peptide than the PDZ domain.

The X-ray crystal structure of ePDZ-b1

To investigate the structural basis for the dramatically enhanced affinity and specificity of the ePDZ-b family affinity clamps, we determined the X-ray crystal structure of ePDZ-b1 in complex with the ARVCF peptide at a 1.9 Å resolution. The statistics for data collection and refinement are shown in Table 1. Although we have been unable to determine the crystal structures of the other ePDZ-b family members, it is highly likely that they adopt essentially identical overall structures and peptide-binding interfaces, given the few sequence differences and similar binding specificity profiles (Fig. 1b).

The ePDZ-b1 structure has a clamp-like architecture as designed (Fig. 3a). The root-mean-square deviations of the C α atoms for the PDZ and FN3 domains (excluding the three mutated loops of the FN3 domain) are 0.973 Å and 0.752 Å relative to those of the ePDZ-a structure, respectively, indicating that the combination of the two domains and the mutations in the binding loops have not perturbed the fold of each domain. The electron density for the linker region (GGSGG) between the two domains was clearly observed, indicating that it is well ordered (Fig. 3a).

The architecture can be described as a semi-open clamshell, with the peptide positioned along a long narrow cleft created at the domain junction. The FN3 domain is slanted to one side of the peptide-binding groove of the Erbin PDZ domain. Of the three diversified loops of FN3, the FG loop contacts both the PDZ domain and the peptide, the BC loop contacts only the PDZ domain and the DE loop does not contact either the PDZ domain or the peptide (Fig. 3d). The affinity clamp buried 78% (966 Å²) of the solvent-accessible surface area of the peptide. The shape complementarity between FN3 and the PDZ/peptide complex as measured by the shape correlation value (S_c , 0.77)²¹ is quite high in comparison with other natural and engineered binding proteins.^{21; 22} Together, these data indicate that the complex possesses a large and tightly packed interface.

Peptide/ePDZ-b1 Interactions

In the ePDZ-b1 structure, the C-terminal four residues of the ARVCF peptide are inserted into the Erbin PDZ ligand-binding groove in a manner similar to the previously reported structure of a homologous peptide bound to Erbin-PDZ.¹⁷ This explains the observation that ePDZ-b1 preserved the underlying binding specificity of Erbin-PDZ. The structure indicates that the preference at position -2 for Thr over Ser (Fig. 2b) is likely to be due to more favorable interactions with Val64. Unlike the C-terminal residues that are sandwiched between the two domains, the N-terminal four residues (positions -7 to -4) of the peptide extend onto the FN3 domain, participating in an extended hydrogen bond network and making extensive hydrophobic contacts primarily with the FG loop of FN3 domain (Fig. 4c and Supplementary Fig. 3). Nearly 50% of the total surface areas of the N-terminal four residues were buried in the complex (Fig. 4a). Therefore, the recognition of a larger portion of the peptide by ePDZ-b1 is corroborated by these results, and provides a structural basis for the specificity enhancement exhibited by the ePDZ-b family.

The ePDZ-b1 structure explains the critical role of V⁻⁴ in ligand recognition, as shown in phage-display and alanine-scanning experiments, and rationalizes the puzzling observation of the reduced affinity of ePDZ-b family members to the δ -catenin peptide relative to the starting PDZ domain (Fig. 1b). V⁻⁴ is nearly completely buried in ePDZ-b1 (Fig. 4c) with two side-chain methyl groups forming hydrophobic contacts with residues H182, Y183, Y184 in the FG loop of FN3. The substitution of Val with Pro would introduce steric clashes with the nearby Y181 on the FG loop of the FN3 domain, which should result in reduced affinity. The peptide library results suggest a preference for I/L/M at this position over Val (Fig. 2 and Supplementary Fig. 1). There is a small cavity around V⁻⁴ that could accommodate a larger aliphatic side chain. Consequently, substitutions with I/L/M side chains at this position may supply additional van der Waals contacts that contribute to enhanced affinity.

The peptide library experiments revealed a strong preference for Arg at position -7 even though the original target had Pro at this position. In the structure, Pro⁻⁷ is sandwiched by the side chains of Y183 and R136 of FN3 but the packing of these residues does not appear optimal (Fig. 4c), suggesting the flexible Arg side chain at the -7 position may fit better in this cavity. The Arg side chain could also form a salt bridge with E150 on FN3 and stabilize the peptide-clamp interaction.

The ePDZ-b family members differ only in the BC loop sequence (Fig. 1b), and thus a >10 fold affinity enhancement for ePDZ-b1 and ePDZ-b2 over ePDZ-b can be attributed to the BC loop alterations. The BC loop sequence, YYDSHVS, of ePDZ-b was replaced with YRELPVS in ePDZ-b1 and FTDLPVS in ePDZ-b2 (Fig. 1b; mutations are underlined). Together these mutants differ at only five residues in the BC loop, two of which may be considered conservative substitutions (Y to F at position 127 and D to E at position 129). R128 in ePDZ-b1 (Y in ePDZ-b and T in ePDZ-b2) is located away from the domain interface. Among these differences, ePDZ-b1 and -b2 both have L130 and P131, and they are likely to be important for affinity enhancement (Fig. 1b). In the ePDZ-b1 structure, P131 is part of a turn that places the side chain of L130 into the PDZ/FN3 interface. These two residues form extensive hydrophobic contacts with residues P18, F19, T29, R30 and P43 of the PDZ domain. L130 and P131 also form hydrophobic contacts with residues H178, Y179 and Y184 on the FG loop of FN3 domain. Therefore, L130 and P131 are likely to play an important role in reducing the inter-domain movement, which should entropically favor target binding and/or stabilize a favorable FG loop conformation for peptide binding. These results further demonstrate that the FN3 loops can affect affinity and specificity by a mechanism that indirectly alters the binding interface and/or dynamic properties.

Comparison of the Two Affinity Clamp Structures

The ePDZ-b1 structure and the previously reported ePDZ-a structure¹³ offer an opportunity to understand how the combination of the same two protein domains (except for the FN3 loops) can generate a dramatic difference in the levels of binding specificity. In ePDZ-a, the FN3 loops nearly completely cover the C-terminal four residues of the peptide, but the N-terminal two residues of the peptide are largely exposed (Figs. 4c and 4d). The FN3 domain sits diagonally over the peptide bound to PDZ (Figs. 3b and 3d). The BC and FG loops interact with the peptide from the V⁰ to P⁻⁵ positions and also with adjacent PDZ residues (Fig. 3d). This mode of interaction is distinct from that found for ePDZ-b1 described above (Fig. 3d). The N-terminal four residues account for nearly 50% of total peptide surface burial for ePDZ-b1, but less than 25% for ePDZ-a (Fig. 4a).

The superposition of the PDZ domain of the two structures revealed a drastic difference in the position of the FN3 domain. The FN3 portions of the two affinity clamps are related by a 14 Å parallel translation (Fig. 3c; Supplementary Movie), which is responsible for the recognition of different portions of the peptide by FN3 of the two affinity clamps. In contrast, the interaction of the peptide with the PDZ domain is conserved, as expected (Figs. 4b and 4d). Because the two domains are loosely linked, the domain-interface is inherently plastic. This plasticity allows for the translational displacement of the FN3 domain to create a different peptide “read-out” mode and thus different levels of binding specificity.

Roles of the linker connecting the two domains

Another intriguing difference between the structures of these two affinity clamps is that the linker is well ordered in ePDZ-b1 but disordered in ePDZ-a. These observations suggest that the linker for ePDZ-b1 may play an active role in forming the clamp architecture, whereas the ePDZ-a linker may be highly flexible and simply connecting the two domains (Fig. 3b). The distance between the C-terminus of PDZ (D97) and the N-terminus of FN3 (V103) in both structures is similar (~12 Å). In ePDZ-a, the residues around the linker form a concave surface in which the linker may fluctuate (Fig. 3b). In contrast, the residues around the linker of ePDZ-b1 form a convex surface that interacts with the linker through extensive hydrogen bonds and hydrophobic contacts (Fig. 3a).

To examine the role of the linker, we constructed affinity clamps with linkers of different lengths and measured their binding affinity to the target peptide. For both ePDZ-a and ePDZ-

b, the original linker gave the highest affinity, indicating that the peptide-binding interface has been optimized in the context of a particular linker length. ePDZ-a and ePDZ-b showed significantly different levels of sensitivity to changes in linker length. ePDZ-a is not highly sensitive to either shortening or lengthening of the linker, while ePDZ-b is very sensitive to linker shortening (Table 2). The removal of one Gly reduced the binding affinity by eight fold and the removal of two Gly eliminated detectable binding to the target peptide. ePDZ-b was not as sensitive to linker extension. In the crystal structure, the linker in ePDZ-b1 appears already stretched, but the residues adjacent to the linker, i.e. the C-terminal part of PDZ domain and the N-terminal part of the FN3 domain, appear relatively flexible as judged by the crystallographic B-factor. Although the structural data suggest that these residues may be able to provide “play” that compensates for linker shortening, the results of linker mutation indicate that such adjustment was not sufficient to counteract for the two-residue deletion. The reduction of affinity by longer linkers may be due to a larger entropic loss upon binding of the peptide, as documented for other systems.^{23; 24} Together, these results indicate that the linker between the two domains of affinity clamps clearly impacts the range of accessible inter-domain geometry and thus it is an important parameter to consider in designing affinity clamps.

Discussion

Here we have presented the first detailed functional and structural analysis of affinity clamps. The comparisons of the function and structure of multiple affinity clamps revealed common features as well as distinct ones. The affinity clamps all have very high affinity to the ARVCF peptide. The two crystal structures clearly show the common molecular basis for their high affinity. By sandwiching the peptide with surfaces from two domains, the affinity clamps bury very large amounts of peptide surfaces. This mode of peptide recognition is difficult to achieve using a single, rigid scaffold such as antibodies, demonstrating an advantage of using directed interface evolution to generate high-affinity peptides-binding interfaces. The level of affinity and the mode of peptide recognition of affinity clamps are reminiscent of peptide-major histocompatibility complex (MHC) interactions.²⁵ MHC molecules achieve high affinity toward short, unstructured peptides by burying large peptide surfaces ($\sim 1000\text{-}1300\text{ \AA}^2$ for class I MHCs), which is achieved by clamping a peptide between two helices that are supported by a large β -sheet. It is interesting that peptide surface areas buried by ePDZ-b1 ($968\text{ \AA}^2$) approaches those for peptide-MHC interactions, suggesting a general benchmark for designing high-affinity peptide-protein interaction interfaces.

The crystal structures also revealed the origin of the dramatic difference in specificity between the ePDZ-a and ePDZ-b families. The FN3 portions of the two families recognize distinct segments of the target peptide as a result of a large change in the relative position of FN3 and PDZ portions. Clearly, the flexible linkage between two domains allows for the generation of diverse topography and recognition surfaces at a domain interface. The distinct difference in the binding-site topography of the two affinity clamps is most likely generated by differences in how the FN3 and PDZ domains interact, and it suggests that covalent linked domains can rapidly fuse and evolve as a single functional module. These results also suggest that, compared with the traditional “rigid scaffold” engineering (Fig. 1a),^{5; 8; 9; 10} the directed domain-interface evolution strategy is more capable of generating diverse functions.

Systematic analysis of multi-domain proteins, in particular two-domain proteins, has shown that the geometry of two connected domains within a related two-domain family (i.e. a family containing the same pair of respectively homologous domains) is well conserved.^{12; 26} Generally, when one pair of homologous proteins is superimposed, the other domain are related to each other by translation and rotation by $<5\text{ \AA}$ and 20° . By this standard, the difference of the FN3 domain location between the two clamps ($10\text{ \AA}$ translation) is quite large. As the FN3 translation appears to be limited primarily by the linker length used, the affinity clamp strategy

should be able to produce even larger range of inter-domain geometry with a longer linker and thereby expand the range of functions. It is, however, possible that the small diversity of inter-domain geometry in natural proteins is due to constraints imposed by functional requirements. A majority of multi-domain proteins whose active site is located at a domain interface act on small organic molecules (e.g. enzymes) and thus a large change in inter-domain geometry may well completely disrupt such an active site. In contrast, the affinity clamps studied here interact with a peptide segment that provides large surfaces for recognition. Indeed, large differences in inter-domain geometry (10 Å translation and 40° rotation) was found for a pair of DNA-binding proteins consisting of related domain pairs,¹² supporting this view. Taken together, large changes in inter-domain geometry can generate a broad range of functions using the same set of underlying domains, and directed domain-interface evolution is particularly suited for the generation of such functional diversity.

In conclusion, domain-interface evolution provides a rational design strategy to generate leaps in protein functions. The structure and probably dynamics of recognition interfaces can be shaped in a highly diverse manner using this strategy than with the traditional “single patch” evolution. The linker connecting the capture and enhancer domains is an integral component of affinity clamps that can significantly influence the range of functions generated. Domain-interface engineering could greatly expand the scope of synthetic protein science, both by generating new proteins with functions that are not achievable by conventional protein engineering strategies and by providing insights into structure, function and evolution of multi-domain proteins.

Methods

C-terminal peptide library construction and sorting

The peptide libraries were constructed as a fusion to the C-terminus of a M13 P8 protein mutant following the method of Laura et al.¹⁴ A gene for mutant p8 was constructed from synthesized oligonucleotides and cloned between the BamHI and HindIII site of a vector containing OmpT signal sequence originally used for phage display of a single-domain antibody.²⁷ Peptide libraries were made using the Kunkel mutagenesis method as described.^{15; 28} The hexa-peptide library contained 1×10^9 independent clones and the penta-peptide library contained 3×10^8 independent clones. The DNA sequences of the libraries were analyzed both as a mixture and as isolated individual clones, which showed no detectable bias (data not shown). The phage particles were propagated in *E.coli* XL1-blue with M13-KO7 helper phage and 10 µM IPTG following procedures described previously.¹³

To facilitate immobilization of affinity clamps in library sorting, a Cys residue was added to the C-terminus to which biotin was chemically conjugated using EZ-Link Biotin-HPDP ((N-(6-(Biotinamido)hexyl)-3'-(2'-pyridyldithio)-propionamide)following; Thermo Scientific). The C-terminus of the affinity clamp is located away from the binding site and thus the addition of the Cys residue and subsequent biotinylation did not affect their binding function.

Phage-display library sorting was performed using biotinylated affinity clamps and streptavidin-coated magnetic beads (Streptavidin MagneSphere Paramagnetic Particles; Promega). A library solution was incubated with 1 µM of biotinylated affinity clamp for 30 min at room temperature, and then the mixture was captured by streptavidin-coated magnetic beads. After washing the beads three times with TBST (150 mM NaCl, 50 mM Tris-HCl, pH7.5, and 0.5% Tween 20), the phage-streptavidin beads mixture was used to infect *E.coli* XL1-blue and the phage particles were prepared as described above. The subsequent three rounds of sorting were performed using a Kingfisher instrument (Thermo Scientific) as previously described¹³ with 400, 200 and 200 nM of ePDZ-b and 200, 100 and 100 nM of

ePDZ-b1 and ePDZ-b2, respectively. Positive clones were identified using phage ELISA and their sequences were deduced using DNA sequencing.²⁸

Protein Expression and Purification

All proteins used in this study were expressed using derivatives of the pHFT2 expression vector¹⁶ in *E.coli* BL21(DE3) cells. Protein expression was induced with 500 μ M IPTG for three to five hours at 25 °C. Proteins were purified with Ni affinity chromatography following standard protocols. When necessary, a second step of purification with Sephacryl S-100 column in PBS (50mM sodium phosphate containing 150 mM NaCl, pH 7.4) was performed to ensure that proteins were monomeric. For protein crystallization and SPR experiments, the N-terminal His₁₀ tag was cleaved with TEV protease and the cleaved protein was purified using Ni affinity chromatography.

Site-directed mutagenesis

Mutant forms of the SUMO-ARVCF peptide fusion protein¹³ and linker mutants of affinity clamps were constructed using Kunkel mutagenesis.

Affinity Measurements

SPR measurements were performed in 20 mM HEPES/NaOH buffer (pH 7.4), 150 mM NaCl, and 0.005% Tween 20 at 25°C on a BIAcore 2000 instrument. The affinity clamp or the linker mutant was immobilized on a Ni-NTA chip to the level of approximate 300 RU, and different concentrations of SUMO-peptide fusion protein were injected. The analyte concentrations were adjusted to obtain an appropriate range of kinetic or equilibrium traces. The kinetic data were analyzed using the one-to-one binding model in BIAevaluation (BIAcore). For weak interactions ($K_d > 2 \mu$ M), K_d values were determined from the saturation curve of the maximum RU values.

X-ray Crystallography

The ePDZ-b1 protein was dialyzed in 20 mM Tris-HCl buffer pH7.4 containing 100 mM NaCl, concentrated to ~15 mg/ml and mixed with the ARVCF peptide at a 1:1.2 ratio. The ePDZ-b1/ARVCF peptide complex was crystallized in 31% isopropanol, 0.1M HEPES pH7.5, 0.2M MgCl₂ by using the hanging drop vapor diffusion method at 20 °C. Crystals were cryo-protected in the mother solution containing 20% glycerol and flash-frozen in liquid nitrogen. The X-ray data were collected at the Advanced Photon Source (Argonne National Laboratory) beamline 24-ID-C by oscillation method. X-ray diffraction data were processed with HKL2000.²⁹ The structure was determined by molecular replacement with the program MOLREP in CCP4. The structures of an FN3 mutant (PDB ID code 2OBG) and Erbin PDZ domain (PDB ID code 1MFG) were used as the search models. Refmac530 was used for the structural refinement. Model building was carried out by using the program Coot.³¹ The structure of the engineered loops, linker, and ARVCF peptide were built at this stage. Data collection and refinement statistics are listed in Table 1. Molecular graphics were generated using Pymol (www.pymol.org). Solvent accessible surface areas were calculated using Areaimol in CCP4. For the analysis of MHC-peptide interfaces, the following coordinates were used: 1G6R, 1BD2, 1KJ2, 1LP9, 1MI5, 2GJ6, and 1FO0.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Wojcik and R. Gilbreth for discussion and A. Kossiakoff for access to a Biacore 2000 instrument. This work was supported in part by grants from the National Institutes of Health (R01-GM72688, R21-CA132700 and R21-DA025725 to SK) and by the University of Chicago Cancer Research Center. MB was supported by NIH grant T90-DK070076 and the Paul K. Richter and Evalyn E. Cobb Richter Memorial Fund. This work includes research conducted at the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

References

1. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–8. [PubMed: 14631033]
2. Dahiyat BI, Mayo SL. De novo protein design: fully automated sequence selection. *Science* 1997;278:82–7. [PubMed: 9311930]
3. Harbury PB, Ples JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. *Science* 1998;282:1462–7. [PubMed: 9822371]
4. Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S. High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci U S A* 2007;104:6632–7. [PubMed: 17420456]
5. Looger LL, Dwyer MA, Smith JJ, Hellinga HW. Computational design of receptor and sensor proteins with novel functions. *Nature* 2003;423:185–90. [PubMed: 12736688]
6. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, Hilvert D, Houk KN, Stoddard BL, Baker D. De novo computational design of retro-aldol enzymes. *Science* 2008;319:1387–1391. [PubMed: 18323453]
7. Kaplan J, DeGrado WF. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. U. S. A* 2004;101:11566–11570. [PubMed: 15292507]
8. Binz HK, Amstutz P, Pluckthun A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol* 2005;23:1257–68. [PubMed: 16211069]
9. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR. Making antibodies by phage display technology. *Annu. Rev. Immunol* 1994;12:433–455. [PubMed: 8011287]
10. Yoshikuni Y, Ferrin TE, Keasling JD. Designed divergent evolution of enzyme function. *Nature* 2006;440:1078–82. [PubMed: 16495946]
11. Bashton M, Chothia C. The generation of new protein functions by the combination of domains. *Structure* 2007;15:85–99. [PubMed: 17223535]
12. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol* 2004;14:208–216. [PubMed: 15093836]
13. Huang J, Koide A, Makabe K, Koide S. Design of protein function leaps by directed domain interface evolution. *Proc Natl Acad Sci U S A* 2008;105:6578–83. [PubMed: 18445649]
14. Laura RP, Witt AS, Held HA, Gerstner R, Deshayes K, Koehler MF, Kosik KS, Sidhu SS, Lasky LA. The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of delta-catenin and ARVCF. *J Biol Chem* 2002;277:12906–14. [PubMed: 11821434]
15. Koide A, Bailey CW, Huang X, Koide S. The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol* 1998;284:1141–1151. [PubMed: 9837732]
16. Huang J, Koide A, Nettle KW, Greene GL, Koide S. Conformation-specific affinity purification of proteins using engineered binding proteins: Application to the estrogen receptor. *Protein Expr Purif* 2006;47:348–354. [PubMed: 16325424]
17. Skelton NJ, Koehler MF, Zobel K, Wong WL, Yeh S, Pisabarro MT, Yin JP, Lasky LA, Sidhu SS. Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. *J Biol Chem* 2003;278:7645–54. [PubMed: 12446668]
18. Zhang Y, Yeh S, Appleton BA, Held HA, Kausalya PJ, Phua DC, Wong WL, Lasky LA, Wiesmann C, Hunziker W, Sidhu SS. Convergent and divergent ligand specificity among PDZ domains of the LAP and zonula occludens (ZO) families. *J Biol Chem* 2006;281:22299–311. [PubMed: 16737968]

19. Sidhu SS, Koide S. Phage display for engineering and analyzing protein interaction interfaces. *Curr Opin Struct Biol* 2007;17:481–7. [PubMed: 17870470]
20. Zahnd C, Wyler E, Schwenk JM, Steiner D, Lawrence MC, McKern NM, Pecorari F, Ward CW, Joos TO, Pluckthun A. A designed ankyrin repeat protein evolved to picomolar affinity to her2. *J Mol Biol* 2007;369:1015–28. [PubMed: 17466328]
21. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *J Mol Biol* 1993;234:946–50. [PubMed: 8263940]
22. Gilbreth RN, Esaki K, Koide A, Sidhu SS, Koide S. A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. *J Mol Biol* 2008;381:407–418. [PubMed: 18602117]
23. Ladurner AG, Fersht AR. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J Mol Biol* 1997;273:330–7. [PubMed: 9367765]
24. Nagi AD, Regan L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* 1997;2:67–75. [PubMed: 9080200]
25. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 1987;329:506–12. [PubMed: 3309677]
26. Han JH, Kerrison N, Chothia C, Teichmann SA. Divergence of interdomain geometry in two-domain proteins. *Structure* 2006;14:935–45. [PubMed: 16698554]
27. Koide A, Tereshko V, Uysal S, Margalef K, Kossiakoff AA, Koide S. Exploring the capacity of minimalist protein interfaces: interface energetics and affinity maturation to picomolar KD of a single-domain antibody with a flat paratope. *J Mol Biol* 2007;373:941–53. [PubMed: 17888451]
28. Koide A, Koide S. Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. *Methods Mol Biol* 2007;352:95–109. [PubMed: 17041261]
29. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–26.
30. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–55. [PubMed: 15299926]
31. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–32. [PubMed: 15572765]
32. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–90. [PubMed: 15173120]
33. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 1990;18:6097–100. [PubMed: 2172928]

Fig. 1. Directed domain-interface engineering produces new protein functions. (a) Scheme of domain-interface engineering. Two domains are connected and a new recognition site is produced at the interface of the newly connected domains. (b) Sequences and binding parameters of affinity clamps. The data were taken from Huang et al.¹³

Fig. 2.

Specificity profiles of affinity clamps. (a) Sequences of the C-terminal peptides of ARVCF and δ -catenin. (b) and (c) Weblogo^{32; 33} representations of specificity profiles for ePDZ-b family clamps. The sequence conservation at each position is represented by the overall height of the stack. The height of symbols within the stack indicates the relative frequency of each amino. The grey horizontal bars indicate positions that were kept constant in the libraries. (b) shows the results with the hepta-peptide library and (c) with the penta-peptide library. The library design is shown above the profiles. (d) Effects of Ala substitution at the indicated positions of the ARVCF peptide as measured by SPR. The values for the free energy changes

($\Delta\Delta G$) are shown on the left axis and those for the K_d ratio (K_d^{mutant}) are shown on the right axis.

Fig. 3.

X-ray crystal structures of ePDZ-b1 and ePDZ-a in complex with the ARVCF peptide. (a) and (b) Ribbon (left panel) and surface (right panel) representations of the overall structures of ePDZ-b1 (a) and ePDZ-a (b). The PDZ and FN3 portions and the peptide are shown in gray, cyan and yellow, respectively. In ePDZ-b1, the linker of is shown in orange. The weighted $2F_{\text{obs}} - F_{\text{calc}}$ electron density map of the linker segment contoured at 1.5σ is shown in the enlarged box. In ePDZ-a, the missing residues for the linker segment are indicated with dashed lines. The concave surface around the linker area is highlighted in red circle. (c) Superposition of the two structures on the PDZ portion. ePDZ-b1 is colored as in (a); ePDZ-a is colored in light brown for PDZ and dark blue for FN3. The peptide was omitted for clarity. (d) Interactions of the FN3 loops with the PDZ/peptide complex in ePDZ-b1 (left panel) and ePDZ-a (right panel). The three FN3 loops (BC, DE and FG loops) are shown as sticks in blue, cyan and red, respectively. The surface of the PDZ portion is shown in gray, and the peptide as yellow spheres. In the bottom panel, the surfaces of the PDZ and peptide portions in contact with the BC, DE and FG loops are shown in blue, cyan and red, respectively, and those in contact with both BC and FG loops are in magenta. The peptide surfaces without FN3 contact are shown in yellow, and the black lines enclose the bound peptide. The N- and C-termini of the peptide are also labeled.

Fig. 4.

Interactions of ARVCF peptide with affinity clamps. (a) Surface burial of peptide residues by affinity clamps and Erbin PDZ. (b) Comparison of the peptide binding to the underlying PDZ domain. The ePDZ-a and ePDZ-b1 structures are superposed on the PDZ portion. The PDZ portion is shown in surface representation and the peptide is shown in stick. PDZ of ePDZ-b1 is colored in gray and peptide in yellow; PDZ in ePDZ-a is colored in light brown and peptide in green. (c) Comparison of N-terminal three residues interacting with ePDZ-b1 (left panel) and ePDZ-a (right panel). The peptide and affinity clamps are colored as in Fig. 3. The orientation of the two structures are flipped by 180° to obtain a clear view of the N-terminal residues. Residues involved in H-bonds (dashed black lines) on the affinity clamps are shown in sticks with atomic coloring for nitrogen and oxygen. (d) A close view of H-bond pattern of the C-terminal five residues.

Table 1
Data collection and refinement statistics for ePDZ-b1 (PDBID: 3CH8)

<i>Data collection statistic</i>	
Space group	P3 ₂
Cell parameters	a=b=69.01
	c=46.71
	$\alpha=\beta=90$
	$\beta=120$
Beamline	APS-24ID-C
Wavelength Resolution (Å)	1.0000 Å
(highest resolution shell) ^{a)}	20-1.8 (1.86-1.80)
Completeness(%)	99.1 (97.0)
I/σ(I)	23.74 (2.712)
R _{merge} ^{b)}	0.060 (0.512)
Average redundancy	5.1 (3.9)
<i>Refinement statistics</i>	
Resolution range (Å)	20.0-1.9
Reflections used (free)	18126 (973)
R factor ^{c)}	0.201
R _{free} ^{d)}	0.260
RMS deviations	
Bonds (Å)	0.015
Angles (°)	1.620
No. protein residues	195
No. waters	89
Average B factor (Å ²)	58.92
<i>Ramachandran plot statistics</i>	
Most favored (%)	86.6
Additionally allowed (%)	12.1
Generously allowed (%)	1.3

^{a)} Highest resolution shell is shown in parenthesis.

^{b)} $R_{\text{merge}} = \frac{\sum_{\text{hkl}} \sum_i |I(\text{hkl})_i - \langle I(\text{hkl}) \rangle|}{\sum_{\text{hkl}} \sum_i I(\text{hkl})_i}$ over *i* observations of a reflection hkl.

^{c)} $R\text{-factor} = \frac{\sum ||F(\text{obs})| - |F(\text{calc})||}{\sum |F(\text{obs})|}$.

^{d)} R_{free} is R with 5% of reflections sequestered before refinement.

Table 2

The dissociation constants (in nM) of clamp linker mutants to ARVCF peptide

Linker sequence	ePDZ-a	Affinity Clamp Used
		ePDZ-b
GGSGG (wild type)	56 ± 5	56 ± 6
GGSG	117 ± 1	412 ± 5
GGG	134 ± 2	>5,000
GG	311 ± 8	>10,000
GGSGGGGSGGS	245 ± 6	247 ± 7

Errors shown were estimated from global fitting of a 1:1 binding model.