Published in final edited form as:

Curr Opin Struct Biol. 2012 August ; 22(4): 413–420. doi:10.1016/j.sbi.2012.06.001.

Structural Insights for Engineering Binding Proteins Based on Non-Antibody Scaffolds

Ryan N. Gilbreth and Shohei Koide

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, U.S.A.

Abstract

Engineered binding proteins derived from non-antibody scaffolds constitute an increasingly prominent class of reagents in both research and therapeutic applications. The growing number of crystal structures of these “alternative” scaffold-based binding proteins in complex with their targets illustrate the mechanisms of molecular recognition that are common among these systems and those unique to each. This information is useful for critically assessing and improving/expanding engineering strategies. Furthermore, the structural features of these synthetic proteins produced under tightly controlled, directed evolution deepen our understanding of the underlying principles governing molecular recognition.

Introduction

Advances in technologies and in our understanding of molecular recognition mechanisms have enabled us to create novel molecular recognition sites on otherwise inert proteins that serve as molecular scaffolds[1, 2]. Efforts over the past ~15 years have established multiple effective platforms for generating binding proteins. Such engineered binding proteins may serve as alternatives to antibodies in medical and industrial applications and have potential advantages such as superior biophysical properties, ease of production and efficient tissue penetration.

Alternative scaffold-based binding proteins are typically generated through directed evolution. In this type of approach, large combinatorial libraries are created in which the amino acid sequence of a contiguous patch on the surface of a starting scaffold is extensively diversified. Functional binding proteins are isolated from these libraries using molecular display technologies such as phage display, yeast display and ribosome/mRNA display [3]. Because the creation of a high-affinity binding surface often involves the mutation of 10–15 residues, the total number of sequences that could be encoded vastly exceeds the number that can be experimentally sampled. Thus, appropriate choices of which positions to diversify and to which amino acid types are crucial to achieving success. Three-dimensional (3D) structural data are critical in making these decisions. Clearly, a structure of the starting scaffold is essential for designing a combinatorial library, but, subsequent structural characterization of engineered binding proteins is equally important. Such structures reveal how the engineered molecules actually achieve their function, validating or

© 2012 Elsevier Ltd. All rights reserved.

Correspondence should be addressed to SK. skoid@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.

correcting hypotheses and offering insights into how library designs might be improved. Thus, combinatorial library design and structural characterization ideally form a positive feedback loop in which library designs are evaluated and incrementally improved based on structural data (Fig 1).

Of the numerous alternative scaffold systems that have been explored, there are now multiple 3D structures of binder/target complexes available from four: monobodies (derived from the tenth fibronectin type III (FN3) domain of human fibronectin), affibodies (derived from the immunoglobulin binding protein A), DARPin (based on Ankyrin repeat modules) and anticalins (derived from the lipocalins billin-binding protein and human lipocalin 2) (Table 1). The increasing number of structures available from these systems, nearly 30 now in the protein data bank (PDB), may allow us to extract meaningful information that goes beyond isolated “anecdotes”. Here we review these structures and discuss the insights they provide for protein engineering. We limit our discussion to scaffold systems with multiple available structures so that trends and tendencies may be assessed. In addition, we compare these engineered interfaces to natural interfaces, and discuss what these comparisons reveal about mechanisms of molecular recognition.

How Well Do Structures Match Designs?

Combinatorial libraries in alternative scaffold systems are designed with a particular mode of interaction with a target in mind and it is assumed that the diversified surface will mediate this interaction. Structural characterization tests this assumption and occasionally reveals unanticipated modes of interaction.

The diversified surface in DARPins is comprised by positions on a series of α -helices and well-structured loops that have been chosen because they often mediate interactions in natural ankyrin repeat proteins (Fig 2A) [4]. In all DARPin structures, this surface is used as envisioned to bind to targets. The structures of DARPin/maltose binding protein (MBP) and DARPin/BppL complexes provide examples (Fig 2A) [4, 5]. One measure of how well a structure agrees with a library design is the percentage of diversified positions that actually contact the target molecule. By this measure, among structurally characterized scaffolds, DARPin/target complexes most closely match their library design with 75% of diversified positions contacting target on average. Diversified positions in DARPins contribute an average of 68% of DARPin buried surface area and 54% of all target-contacting residues. Thus, although diversified positions typically comprise the majority of DARPin binding sites, undiversified regions, for example several positions near the center of a helix of the ankyrin repeat units, often contribute as well (Fig 2B). Conversely, there are other positions in the loops of the DARPin scaffold that are diversified in the library but do not contact target in any available crystal structures (Fig 2B). These structural data provide useful guidelines for further improving the library designs.

Anticalin libraries diversify positions in loop and sheet regions that line a pocket in the basket-like lipocalin scaffold (Fig 2A). Natural lipocalins typically use this pocket to bind small molecules. Anticalins use this surface to recognize their targets in all available structures, and thus, are also generally consistent with their library designs. However, a small number of available structures makes it impossible to assess general trends. In the anticalin/CTLA-4 complex, 63% of diversified positions contact the target. Additionally, 63% of binder buried surface and 41% of all target contacting residues are contributed by diversified positions. Alternative library designs for small molecule targets have also been constructed and resulting anticalins utilized the expected basket-like pocket for binding (Fig. 2A).

Affibody libraries essentially diversify an entire face of the helical protein A scaffold (Fig. 2A). Affibodies also generally recognize their targets as envisioned. On average, 90% of diversified positions contact target in affibody structures. These diversified positions contribute 80% of affibody buried surface and 62% of all target contacting residues. Contributions made by non-diversified positions are all from residues adjacent to diversified positions. Although these numbers suggest that affibody structures match their library design even more closely than DARPin structures do, the structure of an affibody bound to A β (1–40) peptide [6] provides a dramatic exception (Fig. 2A). This affibody unexpectedly formed a disulfide linked dimer through an engineered cysteine residue and mutations introduced into a normally helical region of the scaffold altered the structure to a β -sheet conformation which bound to the A β peptide in a hairpin conformation by forming an intermolecular β -sheet (Fig 2A) [6]. Relatedly, some affibodies are partially or completely unstructured in the unbound state [7, 8]. Thus, although the affibody system has been effective, its structural integrity seems to be often compromised, possibly because of its helical bundle architecture that, unlike β -sheets, lacks strong linkage across secondary structure elements. This low structural stability makes it difficult to predict the mode of recognition for individual affibodies. These examples highlight the importance of structural characterization in understanding how engineered proteins achieve their function.

In monobodies, three surface exposed loops at one end of the molecule are typically used to construct a binding site with an intention of mimicking antibody-like, loop-mediated interaction (Fig. 2A) [9]. Several monobodies indeed interact with their targets using these three loops, as anticipated (Fig. 2A, MBP and hSUMO1 complexes) and diversified positions contribute an average of 80% of monobody surface area burial and 68% of all contacting residues [10–12]. However, among the four systems, monobody structures show the greatest departure from their envisioned binding mode and, on average, only 51% of diversified positions contact target in monobody structures. Unexpectedly, most monobodies use a single diversified loop and positions in the undiversified β -sheet regions to form a binding site (Fig 2A, SH2 domain complex and Fig 2B) [13]. In a recent study, a monobody/yeast SUMO structure exhibiting this interaction mode was used as a guide to create a SUMO-targeted library that diversified the single-loop and contacting positions in a β -sheet. This library generated binders to other SUMO family members where a conventional monobody library did not. Mutations in the β -sheet region were crucial to this success [14]. Building on these observations, a new library design was recently reported in which this loop and a loop located on the opposite end of the molecule were diversified along with one face of the β -sheet that frequently contributed to target recognition in existing monobody structures (Fig 2A) [12]. This library was as effective as a conventional loop-based library in producing high-affinity monobodies to three distinct targets, and more effective in some cases. The crystal structure of a monobody isolated from this library validated this approach [12] (Fig. 2A). This example highlights the value of structural analyses in guiding the development of new and improved library designs.

Binding Site Topography: Are Some Scaffolds Better Than Others?

High-affinity interactions require shape and chemical complementarity. This complementarity is achieved using a variety of secondary and tertiary structural elements in natural protein-protein interactions. The structural features of binding sites in the four alternative scaffolds discussed here are similarly diverse. The success of all these systems in producing high-affinity binders to diverse targets suggests that scaffold architecture matters little as long as diverse amino acid sequences can be presented without compromising overall structure. However, these crystal structures show that each scaffold exhibits its own distinct mode(s) of recognition that may reflect their underlying architecture.

DARPin tend to recognize convex surfaces in their targets, complementary to their concave binding sites (Fig 2A). Similarly, anticalins with their basket-like architecture tend to “cradle” their targets (Fig. 2A). Affibodies have a flat binding site architecture and tend to recognize similarly flat surfaces in their targets (Fig. 2A). As discussed, monobodies use two distinct binding surfaces with different topographies. The monobody/MBP complexes exemplify cleft recognition using the protruding surface formed by three diversified loops (Fig 2A) [10, 11]. This concave epitope is distinct from a convex epitope recognized by an MBP-binding DARPin [4] (Fig. 2A). Camelid single-domain antibodies, which are similar to monobodies in structure, preferentially bind to clefts supporting a correlation between convex binding site topography and cleft binding [15]. The alternative recognition mode of monobodies utilizes a concave binding site that targets convex surfaces (Fig 2A). A very recent report published after the completion of this paper describes the structures of two FN3-based binders in complex with their targets. In one, a loop-driven binding surface is used to recognize a cleft and in the other, the concave alternative surface is used to recognize a convex epitope, again illustrating two distinct binding modes in monobodies [16]. Notably, the use of this alternative binding surface appears to dominate in cases where the target lacks a prominent cleft. Taken together, these data indicate that the topography of a scaffold’s binding site is closely correlated with the type of epitopes that it recognizes with high affinity. This observation in turn suggests that one could improve the success rate of binding protein selection through judicious choice of scaffold or library design to complement the topography of an intended epitope.

Amino Acid Composition in Binding Sites

Because of the limited coverage of sequence space in combinatorial libraries, proper choices of not only positions to be diversified but also amino acid types to be included substantially impact the odds of generating high-affinity binders. DARPin, affibody and anticalin libraries have all used largely unbiased sets of amino acids [4, 17–19] (Fig. 3A). In contrast, recent monobody libraries have utilized highly biased amino acid sets [10, 13], inspired by the enrichment of certain amino acids, particularly tyrosine, in the antigen-binding site of natural antibodies and also in other natural protein-protein interfaces [20, 21] (Fig. 3A).

In monobodies, aromatic amino acids are, not surprisingly, heavily enriched in interfaces. Aromatic residues occur, on average, at 37% of target contacting positions (41% of contacting, diversified positions) and contribute 54% of monobody surface area burial (Fig 3A). These figures are in line with highly biased library designs (Fig 3A). Notably, the overall usage of amino acid types at target contacting positions matches well with the typical monobody library design, suggesting that this composition is well-suited for recognizing diverse protein targets in this system.

Although not strongly emphasized in DARPin libraries, aromatic amino acids are enriched in DARPin binding surfaces, on average, occurring at 18% of target contacting positions (27% of contacting, diversified positions) and contributing 33% of DARPin surface area burial (Fig. 3A). When considering all target contacting positions (including non-diversified positions) DARPins also show an enrichment of aspartate and leucine (Fig. 3A). This enrichment is likely a byproduct of leucine and aspartate residues located in the non-diversified segments connecting the diversified positions in the helices and loops that are consistently in the binding interface (Fig. 2A).

There are only three affibody structures available that reflect a “typical” mode of binding, so it is difficult to draw a general conclusion for this system. However, the enrichment of aromatics is less prominent in affibodies and other amino acid types such as arginine and valine are more strongly enriched (Fig. 3A), potentially indicating different roles of amino

acids in the entirely helical context. Similarly, four of the five available anticalin structures are complexes with small molecules, which makes this type of analysis less meaningful.

Structural data reveal a consistent enrichment of aromatic amino acids across different scaffold types and even in the context of largely unbiased libraries. The unique versatility of tyrosine, in particular, in forming intermolecular contacts has been attributed to the many types of interactions that the tyrosine side chain can make (aromatic stacking, hydrogen bonding, hydrophobic interactions, cation- π interactions) [21, 22]. Interestingly, glutamine and histidine are consistently depleted at diversified positions in three scaffolds (Fig. 3A). Together, these data strongly suggest that highly biased amino acid compositions may be a generally effective strategy in library design.

Structure–Affinity Relationships in Engineered Binders

Relationships between structure/affinity have been extensively examined in the context of natural interfaces. However, simple structural parameters such as interface size, packing, and hydrophobicity have been shown to be poor predictors of affinity [20, 23]. Although all these parameters surely influence affinity, their relative contributions appear to be highly context-dependent, minimizing their individual predictive power.

Because engineered binders are produced in the context of an invariant scaffold, one might expect that the structural determinants of affinity are more consistent and thus more evident. However, obvious structural determinants of affinity are also difficult to identify in alternative scaffold binders. Neither the interface size nor shape complementarity (S_C) value, which reflects the efficiency of interface packing, is a good predictor of affinity (Fig 3B and 3C). One striking example is a DARPin with a 22 nM K_d that exhibits an S_C value of 0.48, well below that of many other protein-protein interactions (e.g. 0.67 for antibody-antigen interactions) and the average for alternative scaffold structures (0.70) [24].

Conclusions

Structural data inform us about how alternative scaffolds can be best used, how library designs might be improved, and what factors govern molecular recognition in these systems. High-affinity binders can be isolated from well-established scaffold systems, but the interface topography of individual systems dictates their epitope preferences. Structural analyses of these systems have revealed discrepancies between designed and actual usages of positions and amino acid composition for target recognition. Such knowledge will accelerate the evolution of library designs and consequently the generation of high-performance binding proteins. Systematic and in-depth analyses of the growing structural dataset of engineered binding protein/target complexes may contribute to an improved understanding of molecular recognition mechanisms.

Acknowledgments

We thank Drs. John Wojcik and Akiko Koide for critical reading of the manuscript. This work was supported by the NIH grants R01-GM072688 and R01-GM090324.

References

1. Skerra A. Alternative non-antibody scaffolds for molecular recognition. *Curr Opin Biotechnol.* 2007; 18:295–304. [PubMed: 17643280]
2. Koide, S. Design and engineering of synthetic binding proteins using nonantibody scaffolds. In: Park, S.J.; Cochran, J.R., editors. *Protein Engineering and Design*. CRC Press; 2010. p. 109-130.

3. Koide S, Koide A, Lipovsek D. Target-binding proteins based on the 10th human fibronectin type III domain ((1)Fn3). *Methods Enzymol.* 2012; 503:135–156. [PubMed: 22230568]
4. Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grutter MG, Pluckthun A. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol.* 2004; 22:575–582. [PubMed: 15097997] The authors report the first structure of a DARP in complex with its target. This structure exemplifies the characteristic binding mode of DARPs. This structure is the most prominent example of the enrichment of aromatic amino acids in DARP binding sites with 71% of interface surface contributed by aromatic amino acids.
5. Veessler D, Dreier B, Blangy S, Lichiere J, Tremblay D, Moineau S, Spinelli S, Tegoni M, Pluckthun A, Campanacci V, et al. Crystal structure and function of a DARP neutralizing inhibitor of lactococcal phage TP901-1: comparison of DARP and camelid VHH binding mode. *J Biol Chem.* 2009; 284:30718–30726. [PubMed: 19740746]
6. Hoyer W, Gronwall C, Jonsson A, Stahl S, Hard T. Stabilization of a beta-hairpin in monomeric Alzheimer's amyloid-beta peptide inhibits amyloid formation. *Proc Natl Acad Sci U S A.* 2008; 105:5099–5104. [PubMed: 18375754] The authors describe a highly unusual structure of an affibody in complex with the A β (1–40) peptide. Large conformational rearrangements and dimerization through an engineered cysteine resulted in a binding mode completely inconsistent with the combinatorial library design, emphasizing the importance of structural characterization in understanding how engineered binders actually function.
7. Wahlberg E, Lendel C, Helgstrand M, Allard P, Dincbas-Renqvist V, Hedqvist A, Berglund H, Nygren PA, Hard T. An affibody in complex with a target protein: structure and coupled folding. *Proc Natl Acad Sci U S A.* 2003; 100:3185–3190. [PubMed: 12594333]
8. Eigenbrot C, Ultsch M, Dubnovitsky A, Abrahmsen L, Hard T. Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proc Natl Acad Sci U S A.* 2010; 107:15039–15044. [PubMed: 20696930]
9. 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]
10. 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]
11. 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; 104:6632–6637. [PubMed: 17420456]
12. Koide A, Wojcik J, Gilbreth RN, Hoey RJ, Koide S. Teaching an Old Scaffold New Tricks: Monobodies Constructed Using Alternative Surfaces of the FN3 Scaffold. *J Mol Biol.* 2012; 415:393–405. [PubMed: 22198408] The authors describe a new monobody library design inspired by an unintended mode of interaction observed in monobody/target crystal structures in which large portions of the undiversified scaffold contact targets. This library effectively produced high-affinity binders and a crystal structure confirmed the designed mode of interaction.
13. Wojcik J, Hantschel O, Grebien F, Kaupe I, Bennett KL, Barkinge J, Jones RB, Koide A, Superti-Furga G, Koide S. A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol.* 2010; 17:519–527. [PubMed: 20357770]
14. Gilbreth RN, Truong K, Madu I, Koide A, Wojcik JB, Li NS, Piccirilli JA, Chen Y, Koide S. Isoform-specific monobody inhibitors of small ubiquitin-related modifiers engineered using structure-guided library design. *Proc Natl Acad Sci U S A.* 2011 The authors use a crystal structure of a monobody/yeast SUMO complex to construct a tailored library specifically for targeting members of the SUMO family. This structure-guided library successfully generated binders to members of the SUMO family that conventional monobody libraries did not.
15. De Genst E, Silence K, Decanniere K, Conrath K, Loris R, Kinne J, Muyldermans S, Wyns L. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc Natl Acad Sci U S A.* 2006; 103:4586–4591. [PubMed: 16537393]
16. Ramamurthy V, Krystek SR Jr, Bush A, Wei A, Emanuel SL, Das Gupta R, Janjua A, Cheng L, Murdock M, Abramczyk B, et al. Structures of adnectin/protein complexes reveal an expanded binding footprint. *Structure.* 2012; 20:259–269. [PubMed: 22325775] The authors report two

crystal structures that provide additional examples of the two distinct modes of interaction observed in binders derived from the FN3 scaffold.

17. Beste G, Schmidt FS, Stibora T, Skerra A. Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. *Proc Natl Acad Sci U S A*. 1999; 96:1898–1903. [PubMed: 10051566]
18. Nord K, Gunneriusson E, Ringdahl J, Stahl S, Uhlen M, Nygren PA. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol*. 1997; 15:772–777. [PubMed: 9255793]
19. Schonfeld D, Matschiner G, Chatwell L, Trentmann S, Gille H, Hulsmeier M, Brown N, Kaye PM, Schlehuber S, Hohlbaum AM, et al. An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies. *Proc Natl Acad Sci U S A*. 2009; 106:8198–8203. [PubMed: 19416843] The authors describe a structure of an anticalin bound to a protein target that validates the design of an anticalin library specifically for protein targets.
20. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol*. 1999; 285:2177–2198. [PubMed: 9925793]
21. Mian IS, Bradwell AR, Olson AJ. Structure, function and properties of antibody binding sites. *J Mol Biol*. 1991; 217:133–151. [PubMed: 1988675]
22. Koide S, Sidhu SS. The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. *ACS Chem Biol*. 2009; 4:325–334. [PubMed: 19298050]
23. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *J Mol Biol*. 1993; 234:946–950. [PubMed: 8263940]
24. Monroe N, Sennhauser G, Seeger MA, Briand C, Grutter MG. Designed ankyrin repeat protein binders for the crystallization of AcrB: plasticity of the dominant interface. *J Struct Biol*. 2011; 174:269–281. [PubMed: 21296164]
25. Schweizer A, Roschitzki-Voser H, Amstutz P, Briand C, Gulotti-Georgieva M, Prenosil E, Binz HK, Capitani G, Baici A, Pluckthun A, et al. Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism. *Structure*. 2007; 15:625–636. [PubMed: 17502107]
26. Kohl A, Amstutz P, Parizek P, Binz HK, Briand C, Capitani G, Forrer P, Pluckthun A, Grutter MG. Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure*. 2005; 13:1131–1141. [PubMed: 16084385]
27. Sennhauser G, Amstutz P, Briand C, Storchenegger O, Grutter MG. Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. *PLoS Biol*. 2007; 5:e7. [PubMed: 17194213]
28. Bandejas TM, Hillig RC, Matias PM, Eberspaecher U, Fanghanel J, Thomaz M, Miranda S, Crusius K, Putter V, Amstutz P, et al. Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin. *Acta Crystallogr D Biol Crystallogr*. 2008; 64:339–353. [PubMed: 18391401]
29. Grubisha O, Kaminska M, Duquerroy S, Fontan E, Cordier F, Haouz A, Raynal B, Chiaravalli J, Delepierre M, Israel A, et al. DARPin-assisted crystallography of the CC2-LZ domain of NEMO reveals a coupling between dimerization and ubiquitin binding. *J Mol Biol*. 2010; 395:89–104. [PubMed: 19854204]
30. Koide A, Abbatiello S, Rothgery L, Koide S. Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. *Proc Natl Acad Sci U S A*. 2002; 99:1253–1258. [PubMed: 11818562]
31. Grebien F, Hantschel O, Wojcik J, Kaupe I, Kovacic B, Wyrzucki AM, Gish GD, Cerny-Reiterer S, Koide A, Beug H, et al. Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. *Cell*. 2011; 147:306–319. [PubMed: 22000011]
32. Korndorfer IP, Beste G, Skerra A. Crystallographic analysis of an "anticalin" with tailored specificity for fluorescein reveals high structural plasticity of the lipocalin loop region. *Proteins*. 2003; 53:121–129. [PubMed: 12945055]
33. Kim HJ, Eichinger A, Skerra A. High-affinity recognition of lanthanide(III) chelate complexes by a reprogrammed human lipocalin 2. *J Am Chem Soc*. 2009; 131:3565–3576. [PubMed: 19227970]

34. Korndorfer IP, Schlehuber S, Skerra A. Structural mechanism of specific ligand recognition by a lipocalin tailored for the complexation of digoxigenin. *J Mol Biol.* 2003; 330:385–396. [PubMed: 12823976]
35. Hogbom M, Eklund M, Nygren PA, Nordlund P. Structural basis for recognition by an in vitro evolved affibody. *Proc Natl Acad Sci U S A.* 2003; 100:3191–3196. [PubMed: 12604795]
36. Lendel C, Dogan J, Hard T. Structural basis for molecular recognition in an affibody:affibody complex. *J Mol Biol.* 2006; 359:1293–1304. [PubMed: 16750222]
37. Reynolds C, Damerell D, Jones S. ProtorP: a protein-protein interaction analysis server. *Bioinformatics.* 2009; 25:413–414. [PubMed: 19001476]
38. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 1994; 50:760–763. [PubMed: 15299374]

Highlights

- A total of 28 3-D structures of non-antibody binding protein/target complexes are now published.
- The structures show scaffold-specific and universal trends in recognition mechanisms.
- The structures guide combinatorial library designs.

Figure 1. General flow of a protein-engineering project using a non-antibody scaffold
A schematic is shown depicting a positive feedback loop linking structural characterization and combinatorial library design.

Figure 2. Non-Antibody Scaffold Systems and Library Designs with Available 3D Structures
 On the left, scaffolds are depicted with positions diversified in combinatorial libraries indicated by red spheres. On the right, structures are shown that depict distinct modes of target interaction. Targets are shown as gray surface models. The target protein name, and PDB ID for each structure are indicated. In the DARPin system, scaffolds based on three (left top) or two (left bottom) diversified ankyrin repeat units have been used [4]. Example structures of binders derived from both scaffold types are shown. In the Anticalin system, libraries directed at the production of protein binders [19] (left top) or small molecule binders [17, 33] (left bottom) have been used. A representative example of each type of complex is shown. In the monobody system, two library design strategies have been used [9,

12]. (B) Structures of the Monobody and DARPin scaffolds are shown with positions colored according to the frequency with which they contact targets in the available structures. Contact is defined by burial of $> 1 \text{ \AA}^2$ of surface on complex formation. Positions diversified in typical Monobody and DARPin libraries are indicated by spheres. For monobodies only binders selected from conventional loop-based libraries were considered (PDB IDs 3QHT, 3RZW, 3K2M, 3CSB and 3T04). Also, because monobody loop lengths vary, positions from each variable loop were considered and colored as a group. For DARPins only binders based on the 3-ankyrin repeat scaffold were considered (PDB IDs: 2P2C, 2BKK, 2J8S, 1SVX and 2V5Q). Surface burial calculated using the Protorp server [37].

Figure 3. Amino Acid Usage in Target Contacts and Structural Parameters for the Target-binding Surfaces

(A) Bar graphs of the amino acid compositions in typical naïve libraries (white bars), target contacting positions in the structures of binder/target complexes (black bars), and target contacting positions that were also diversified in the library (gray bars). Contact is defined by burial of $> 1 \text{ \AA}^2$ of surface on complex formation. Naïve library compositions shown are those encoded by the mutagenic oligonucleotides used to create typical libraries. Data are shown for the Monobody, DARPins and Affibody scaffolds. In the case of DARPins, 2XZT (point mutant of 2XZD), 3NOC and 3NOG (recognize a nearly identical epitope to 2J8S) were excluded to avoid introducing artificial bias. In the case of monobodies, 3CSG was

excluded because it was isolated from a Tyr/Ser-only library. 2OTK was excluded in the case of affibodies (highly atypical structure). For 2B87, the lowest energy model (model 1) was used for analysis. Surface burial calculated using the Protorp server [37] (B) Scatter plot of the surface area buried by binders upon complex formation vs. the K_d values reported for those complexes. Surface burial calculated using the Protorp server [37] (C) Scatter plot of the surface complementarity (S_C) value for the interfaces of binder-target complexes vs. the K_d values measured for those complexes. S_C values calculated using the SC program in the CCP4 suite [23, 38]. For monobodies all structures except 2OCF were considered (no reported K_d value). For DARPin all structures except for 2XZD and 2XZT were considered (no reported K_d values). For anticalins, only the 2BX7 structure was considered since all other structures are complexes with small molecules. For affibodies, all structures were considered and for 2B87, the lowest energy model (model 1) was used for analysis.

Table 1

Structures of non-antibody binder/target complexes in the PDB.

PDB ID	Target	Resolution (Å)	R-factor (R)	Free R-factor (R _{free})	% of Residues In Allowed Regions of Ramachandran Plot	References
DARPin						
2P2C	Caspase-2	3.24	26.2	30.5	99.7	[25]
2BKK	Aminoglycoside Phosphotransferase APH (3')-IIIA	2.15	20.0	26.0	98	[26]
3NOG	Acriflavine resistance protein B (AcrB)	3.34	25.8	30.8	99.8	[24]
3NOC	AcrB	2.70	24.3	26.8	100	[24]
2J8S	AcrB	2.54	22.9	27.1	98.9	[27]
1SVX	Maltose Binding Protein (MBP)	2.24	19.5	24.9	99.6	[4]
2V5Q	Polo-like Kinase 1 (PLK-1)	2.30	18.4	22.4	99.5	[28]
2XZT	Caspase-3	2.70	19.9	23.1	100	N/A
2XZD	Caspase-3	2.10	18.7	21.8	100	N/A
3HG0	ORF49 from Lactococcal phage TP901-1	2.10	20.9	24.3	99.8	[5]
2V4H	NF-kappa-B essential modulator CC2-LZ domain (NEMO-CC2-LZ)	2.90	21.1	26.8	99.5	[29]
Monobodies						
2OCF	Estrogen Receptor α Ligand Binding Domain (ER α LB Δ D)	2.95	19.4	25.1	99.7	[30]
3QHT	Yeast Small Ubiquitin-like Modifier (ySUMO)	2.40	22.6	27.2	96.7	[14]
3RZW	Human Small Ubiquitin-like Modifier 1 (hSUMO1)	2.15	18.9	23.7	99.4	[12]
3UYO	Src Homology 2 domain of Abelson Kinase 1 (Abl1-SH2)	1.83	19.0	23.7	100	[12]
3K2M	Abl1-SH2	1.75	18.2	22.1	100	[13]
3CSB	MBP	2.00	19.9	23.6	99.1	[10]
3CSG	MBP	1.80	19.2	23.5	99.8	[10]
3T04	Abl1-SH2	2.10	19.3	25.1	100	[31]
Anticalins						
1N0S	Fluorescein	2.00	19.6	24.3	100	[32]
3BX7	Cytotoxic T-Lymphocyte Antigen 4 (CTLA-4)	2.10	20.9	22.6	99.7	[19]
3DSZ	Yttrium • diethylenetriamine pentaacetic acid (Y•DTPA)	2.00	21.3	23.0	99.1	[33]
1LKE	Digoxigenin	1.90	18.8	24.8	100	[34]

PDB ID	Target	Resolution (Å)	R-factor (R)	Free R-factor (R _{free})	% of Residues In Allowed Regions of Ramachandran Plot	References
1LNM	Digitoxigenin	1.90	20.6	24.5	100	[34]
Affibodies						
1LP1	Staphylococcal protein A, Z Domain	2.30	22.5	25.6	100	[35]
3MZW	Human Epidermal Growth Factor Receptor 2 (HER2)	2.90	20.8	27.8	99.5	[8]
2OTK	A-beta Peptide	NMR			98.9 *	[6]
2B87	Affibody	NMR			98.8 *	[36]
	Average Values	2.32	20.7	25.0	99.48	

* For NMR structures, Ramachandran statistics include all deposited models.