Design of protein-protein interactions via β-strand pairing

Peter Benjamin Stranges

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics.

Chapel Hill
2012

Approved by:

Brian Kuhlman Ph.D.

Nikolay Dokholyan Ph.D.

Jane Richardson

John Sondek Ph.D.

Marcey Waters Ph.D.

Abstract

PETER BENJAMIN STRANGES: Design of protein-protein interactions via β-strand pairing
(Under the direction of Brian Kuhlman Ph.D.)

The design of new protein-protein interfaces is a test of our understanding of protein interaction biophysics and can provide new tools to understand cell biology. Methods to accurately design high-affinity interactions have not been established, making it necessary to devise new strategies to facilitate the design process. Solvent exposed main chain atoms on β-strands are prone to interact with other exposed strands and could serve as the basis for the design of a novel interaction. This dissertation describes the application of β-strand pairing to design homodimeric and heterodimeric complexes. It also addresses the successes and failures in computational interface design to determine how design methods need to be improved.

One of the most common ways that proteins interact is the formation of symmetric homodimer. A way to test the hypothesis that β-strand mediated interactions can be accurately designed is to redesign a monomeric protein to form a symmetric homodimer via β-strand pairing. A computational method in Rosetta was used to find monomeric proteins with exposed β-strands then redesign them to form a symmetric homodimer by pairing exposed β-strands to form an intermolecular β-sheet. A crystal structure of one designed complex closely matches the computational model (RMSD = 1.0 Å). This work demonstrates that β-strand pairing can be used to computationally design new interactions with high accuracy

After successful design of a homodimer, β-strand pairing can be extended design to heterodimers. A computational protocol is described that identifies proteins with exposed strand capable of pairing with an exposed strand on a target protein. The interface of the identified protein is then redesigned to allow it to bind to the target protein. Experimental testing of proteins designed to bind RalA and PCSK9 show that no interaction is made. Directed evolution of the scaffold proteins could allow them to bind to their target.

Most computational protein interface designs from our laboratory and others fail to form when tested experimentally. Successful and failed protein interface designs were examined to see if they provided answers about what works in interface design. Successful designs were, in general, more hydrophobic than failed designs and had few designed hydrogen bonds buried at the interface. Rosetta designed hydrogen bonds were found not to match hydrogen bond distributions observed in high resolution crystal structures. The hydrogen bonding portion of the energy function needs to be improved to allow for design of polar interfaces similar to those found in native proteins.

To the Flying Spaghetti Monster.

Acknowledgements

First and foremost I need to thank my advisor Brian Kuhlman giving me a challenging project and sticking with it in the face of two years of disappointing experimental results. When I felt like I was hitting a wall with my research Brian provided an endless source of ideas and insight. I am also grateful that he got me into cycling.

Every member of the Kuhlman lab has contributed to my project in some way, though some deserve need additional mention here. Andrew Leaver-Fay, Steven Lewis and Doug Renfrew were instrumental in introducing me to computational protein design and making me confortable with programming. Matt O'Meara contributed the intellectual and computational basis for inspecting how Rosetta designs hydrogen bonds at protein-protein interfaces. Tim Jacobs is a much appreciated collaborator and source of ideas in the design of novel binders to PCSK9. Bryan Der has been an enormas help in developing approached to protein interface design as well as a constant motivating force.

The members of Rosetta Commons have contributed the basis for most of the computational methods described in this thesis. The progress I made was only possible because of the thousands of lines of code written by others. I especially need to thank Ingemar André for developing and answering questions about symmetric modeling and Sarel Fleishman for developing the RosettaScripts protocol.

Several faculty members at Bowdoin College piqued my interest in science and laboratory research. I especially need to thank my advisors Mike Palopoli and Amy Johnson for getting me involved in research labs and telling me I should do science instead of going to law school. My development as a competent researcher was ensured by the expert tutelage of Sasha Chervonsky at Jackson Labs and the University of Chicago.

It seems cliché to thank one's family here, but it needs to be done for several reasons. My mom and dad made sure they were the first people I saw when I woke up after bumping my head on the asphalt. My brother then took time to come to Chapel Hill and stayed with me until I had my senses back. In addition my mom and brother have been constant source of support, humor, and a reminder that it is possible to leave Chapel Hill sometimes.

There are several folks that need to be thanked for their contributions to my life outside of the laboratory. First is my initial roommate, Andrew Parsons, for demonstrating the effort required to be a successful graduate student, the joys of cycling up a 17% grade every morning, and introducing me to the little spot of heaven that is Carrburritos. My second roommate, Colin Deicth, has been a great friend and fellow political junkie since high school. I have been fortunate to find soccer and volleyball teammates that are as much fun to play with as they are to hang out with off the field/court. Finally, my classmates Jon Edwards, Aaron Hobbes, and Patrick Lackey have been some of the best friends I could hope for.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AUC	Analytical ultra centrifugation
DSSP	Database of secondary structure of proteins
FACS	Fluorescence activated cell sorting
FP	Fluorescence polarization
ΔG_{bind}	Calculated change in energy upon protein-protein binding
ITC	Isothermal titration calorimetry
MALS	Multi-angle light scattering
PDB	Protein Data Bank
REU	Rosetta energy units
RMSD	Root mean square deviation
SASA	Solvent accessible surface area
ΔSASA	change in SASA upon protein-protein binding
SEC	Size exclusion chromatography

Chapter 1

Introduction

The energetics that govern how a protein folds from an extended chain into a final conformation are understood in principle (1), but reliable techniques to accurately predict final structure based on sequence information alone have proved elusive (2). Additionally, a clear picture of how, once folded, a protein interacts with other proteins has not been elucidated. A complete understanding of how these processes work cannot be determined until the forces stabilizing the folded and interacting states are explained. *De novo* protein and protein-protein interaction design provides one way of comprehending these forces. The *de novo* design of a protein means rationally choosing a sequence that will fold into a target final structure (3). Failed and successful designs can provide valuable information about the forces involved in stabilizing a three-dimensional structure. *De novo* protein interface design extends this idea to design a sequence capable of forming an interaction between two or more protein chains. A designed interface that forms experimentally can provide valuable information about the physical chemistry of protein-protein association (4).

1.1 Protein-protein interaction design

Protein-protein interaction networks monitor the internal and external state of a cell and act as signaling circuits to transmit information and change the phenotype of the cell. Genome sequencing and proteomic databases provide a nearly complete list of the available modules for cell regulation, what is lacking is an understanding of how these parts fit together (5). Recent

high throughput techniques have allowed some of these interaction networks to be mapped (6), but a complete understanding of cell signaling remains out of reach. Like protein design, the design of new signaling circuits provides a test of our understanding of a basic biological process. An understanding of which interactions work, which do not, and why they do not can provide valuable information about natural protein-protein interactions (5). Novel methods to alter natural cellular interactions provide a wealth of information about the regulation and response of the signaling circuit (7).

The design of new protein-protein interactions has provided valuable tools to undersand the way nature constructs interactions and modify cell signaling systems. Many natural proteins exist as large complexes of small subunits that form symmetric interfaces in order to construct the larger, functional, protein (8). Grueninger *et al.* used a simple mutagenesis approach to construct ordered, symmetric oligomers out of monomeric proteins (9). This demonstrates a mechanism that nature may use to build multimeric proteins. Protein interactions that can be turned on and off with light have be used to control cell motility (10; 11) and membrane recruitment of certain proteins (12). These tools provide the ability to selectively alter protein function in different areas of the cell. Modular domain recombination of protein-protein interaction domains can create new cellular responses by providing alternative protein interaction networks (reviewed in (13) and (14)). This method recombines protein domains to form non-native interactions with a scaffold that serves to orient and control the flow of information. This method has been used to tune and improve metabolic flux (15), activate guanine exchange factors with a non-native input (16) and engineer feedback loops into MAP kinase signaling (17). These studies demonstrate that engineered protein-protein interactions can control output and tease apart natural protein pathways.

One of the most common ways to obtain a new protein-protein interaction is though directed evolution (18). Phage display (19), yeast display (20) and ribosomal display (21) have provided a wide variety of affinity reagents and protein functions used in cell biology. For example the fusion of a PDZ domain to an FNIII domain followed by directed evolution of the FNIII domain allowed the construct to recognize alternate PDZ binding peptides (22). A multiple laboratory effort engineered high affinity binders to 20 SH2 domains using a variety of

directed evolution techniques (23). It now appears that it is possible to use directed evolution to obtain an affinity reagent to any target that can be purified. A remaining challenge is obtaining binders that recognize a conformational state, post translational modification, or are specific for only one member of a protein family.

Biological engineering has made it evident that more tools are needed to control cellular machinery. Some of the primary needs are orthogonal protein interaction pathways with variable affinity, specificity (5), and computer aided design strategies to design new logic circuits(24). Metabolic pathway design requires protein interactions that can be predictably regulated and have tunable affinity for certain upstream and downstream effectors (25). Computational methods of modifying and designing new protein interactions can provide the tools to meet the needs of biological engineering (26) .

1.2 Computational protein interface design

Computational protein design attempts to search the sequence and conformation of a polypeptide chain to minimize the energy of the structure. All computational design protocols have two components, a search function to explore sequence and conformational space and an energy function that evaluates the fitness of a particular search model (27). Successful computational design of to monomeric proteins quickly suggested that these same models could be applied to engineer protein-protein interactions (28).

The idea that the molecular structure of proteins is responsible for their interaction has been recognized since the time that the first few crystal structures of protein complexes were determined (29; 30). However, the ability to accurately model how two proteins will form a complex has proved elusive (31). *De novo* computational interface design is a test of our understanding of the forces governing protein-protein interactions. A typical computational design run involves modeling the rigid body orientation of two or more protein chains in a conformation that will allow them to interact. Next, the computational protocol designs amino acids near the modeled interaction to stabilize the complex state. The design goal determines which residues near the interface are allowed to change. A common aim is to redesign residues

on a natural protein to enable it to bind to another natural protein. In this case the natural protein that is left unchanged is called the target protein and the redesigned protein is called the scaffold.

Computational interface design has helped create new interactions and functions that nature has not sampled. Recent progress in computational interaction design is reviewed in (26) and (32). Computational design provides several advantages over directed evolution and modular domain recombination. Most importantly, computational design can be used to target a specific interaction site for design and constrain the design to satisfy a particular goal. For example Reina *et al.* used a computational protocol to predictably engineer a Class I PDZ domain to recognize Class I and Class II target peptides (33). Similarly, Shifman and colleagues were able to design mutations to alter specificity of calmodulin for different target peptides (34; 35). Another group was able to find an array of substrates for a protein chaperone by investigating the sequences allowed to bind to a target cite on the chaperone (36). An antibody's affinity for its target can be improved beyond *in vivo* affinity maturation by computationally optimizing interface electrostatics (37). Computational design also allows for favoring one type of interaction (positive design) while simultaneously disfavoring an alternate interaction (negative design). This idea has been used to design heterodimers that do not interact with off target proteins (38; 39; 40).

Computational protein-protein interaction design has the potential to create proteins capable of predictably changing cell signaling. One exciting advance is the development of a computational framework to design a multi-function protein hubs capable of interacting with a variety of partners (41). To achieve this goal computational design will need to be able to repeatably engineer novel protein-protein interactions. The majority of successful computational designs addressed in the reviews above involve the redesign of an existing interaction for increased affinity or altered specificity. Most success computational novel interaction design involves idealized systems such as α-helical pairing (42; 38; 40) or hot spot grafting (43; 44). New approaches to interface design are needed to create a toolkit capable to predictably modifying cellular function.

1.2.1 Interface design using Rosetta

The Rosetta suite of macromolecular modeling software (45) was used for all protein design steps in this work. The energy function is comprised of terms for all aspects thought to be important to protein structure (46).

$$E_{protein} = W_{lj\,atr}E_{lj\,atr} + W_{lj\,rep}E_{lj\,rep} + W_{Hbond}E_{Hbond} + W_{solvation}E_{solvation} + W_{aa}E_{aa} + W_{pair}E_{pair} + W_{rama}E_{rama} + W_{rotamer}E_{rotamer} - W_{reference}E_{reference}$$

These terms include physically based potentials to capture van der Waals interactions ($W_{lj\,atr}E_{lj\,atr}$ and $W_{lj\,rep}E_{lj\,rep}$) and solvation energy ($W_{solvation}E_{solvation}$). The other, knowledge based terms, are parameterized based on high resolution structures in the Protein Data Bank (PDB). Hydrogen bond energy ($W_{Hbond}E_{Hbond}$) is calculated based on the distance and angles between the acceptor and donor atoms (47). The other terms account for electrostatic pairing ($W_{pair}E_{pair}$), torsional preferences ($W_{rama}E_{rama}$ and $W_{rotamer}E_{rotamer}$) and the contextual preferences of amino acids ($W_{aa}E_{aa}$ and $W_{reference}E_{reference}$). Rosetta's search function is Metropolis Monte Carlo (48) with simulated annealing (49), which allows the sequence and conformation of a protein to be quickly sampled during a design simulation but does not guarantee finding a global minimum.

Rosetta is well suited to design new protein-protein interactions (50). Rosetta has the ability to perform rigid body protein docking (51), backbone conformation sampling (52), and sequence design (53). The recent development of an XML scripting language allows many design and sampling modules to be easily incorporated into a single simulation. (54). Rosetta can increase the affinity of a protein interaction (55), alter the specificity of existing interactions (56; 57; 58; 59) and stabilize an interaction to serve as a biosensor (60). More recently, Rosetta has proven to be capable of designing novel protein interactions (61; 62; 63). These successes are addressed in Chapter 4. The next step is to move beyond from proof of concept designs to create novel tools, orthogonal pathways, and enzymatic function for biological engineering (26).

1.3 Complications in computational protein interface design

The design of a protein-protein interactions from scratch has proven to be a very difficult problem (reviewed in (26; 32)). Few designed interfaces have been experimentally verified to form a complex that matches the computational model. Computational interface design combines two challenging modeling goals, rigid body orientation between protein chains and the design of sequence that will allow those chains to interact (28). Most computational designs of a *de novo* interface have no measurable affinity (64). The ones that do interact often have low affinity ($K_D > 100$ μM) (4; 65) or do not form the expected complex (66). Directed evolution can help overcome low affinities by sampling additional sequences that were not in the designed model (61). Computational redesign of existing protein-protein interactions for enhanced affinity has also been plagued by similar problems. Many of the mutations predicted to increase affinity actually weaken the interaction or the redesigned proteins (67; 68).

A protein interface presents several modeling challenges that are not relevant for design of a monomeric protein. Transient protein interactions are more polar than protein cores (69) and require the desolvation of polar atoms to be offset by an interface spanning hydrogen bond (70). A transient interaction requires the proteins involved to be stable in both the bound and unbound state. Mutations that favor an interaction could cause one of the monomers to become unstable and aggregate. Monomeric proteins only need to be designed in the context of the final folded structure (53; 71). Buried water molecules can bridge hydrogen bonds between molecules and stabilize the complex (72). So far, attempts to model water mediated hydrogen bonds have not yielded encouraging results (73). Electrostatic complementation is another driving force behind protein-protein interaction (70; 74). Optimizing long range electrostatic steering can enhance the affinity of an interaction by increasing the on rate (k_{on}) (75; 76). Though useful, intensive electrostatics calculations are impractical during design because of speed and the need for a pairwise additive potential (28). A survey of successful and failed protein interface designs revealed that electrostatic complementation was not designed in most computational models (64). Recent advances in simplifying electrostatic models could allow their incorporation into design simulations (77).

Many groups have employed specialized energy functions to redesign protein-protein interactions for altered specificity or increased affinity. The requirement for specialized interface potentials should not be necessary, although different residues are preferred in interfaces versus proteins as a whole (69) the same packing density and hydrogen bond geometry is observed in monomers and across protein interfaces (78). The results of these studies have provided no consistent insight into energy function improvements. Different studies found success by increasing the weight of interchain interactions (35), down weighting the solvation terms of the energy function (68; 79), specifically favoring the electrostatic component of the energy function (37), and training the energy function on mutational data for a specific interface (67). The need for modified energy functions for protein interfaces is more likely to be indicative of a flaw in design methodology and sampling than in the energy function.

The research highlighted here suggests that one hurdle to computational interface design is the correct modeling of hydrogen bonds and electrostatics. It is imperative to devise new strategies that can overcome this and other persistent complications faced in computational protein-protein interaction design. The successes and failures of interface design can also be examined to glean any information that indicates why many designs fail and only few succeed.

1.4 Using β-strands for protein-protein interface design

"Stealing" the sequence information or local structure from natural protein-protein interfaces is a prudent way to design a new interactions or inhibit existing ones. (80). It allows us to borrow information from the billions of years life has spent engineering interactions and reapply it for our needs. Nature uses a variety of different motifs, folds, and sequences to form the basis of a protein-protein interface (81). One common interface architectural motif is pairing solvent accessible β-strands to form an intermolecular β-sheet (Figure 1.1). These paired β-strands form hydrogen bonds between the main chain atoms in a antiparallel orientation (82). Remaut and Waksman (83) surveyed heterodimeric protein interactions that result in an intermolecular β-sheet and determined that these occur through either β-strand addition (Figure 1.2A), β-strand fold completion (Figure 1.2B), or β-strand zippers. Proteins in the

Ras family form a complex with their ubiquitin-like effectors via an intermolecular β-sheet (84) (Figure 1.2C). Antiparallel cross-chain β-strand pairing also represents 8.8% of contacts observed in homodimeric proteins (Figure 1.2D) (82). Paired β-strands in homodimers are often longer than those involved in forming hetero-complexes (85).

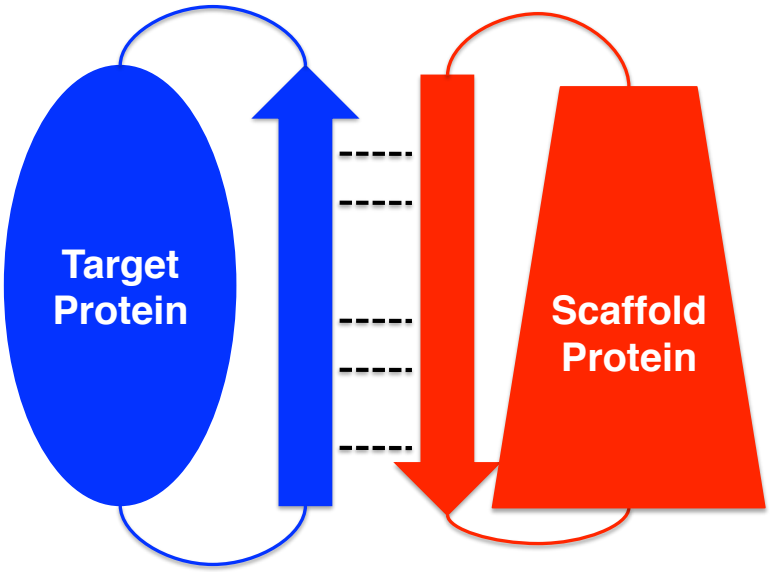

Figure 1.1: Model of heterodimer β-strand interaction. A target protein (blue) with an exposed β-strand forming a complex with another scaffold protein (red) with an exposed strand. Black dashed lines represent main chain hydrogen bonds formed across the interface.

Interface design based on β-strand pairing presents a solution to two problems in computational protein interface design: assuring orientation specificity and satisfying hydrogen bond potential. β-strand interactions are geometrically constrained by hydrogen bonding between the main chain atoms and the twist and sheer observed in β-sheets has been well documented (90; 91; 92). This specificity for a particular hydrogen bonds arrangement should prevent the formation of a complex that does not match the design model (66). Complementary strand pairing across an interface helps satisfy some of the hydrogen bond potential without the need for sequence design. Design can then proceed from well established side-chain preferences for β-sheet formation and stability (93; 94; 95; 96).

The intrinsic potential of exposed β-strands to self-assemble makes them an attractive starting point for interface design. β-strands are often called *sticky* because of their tendency

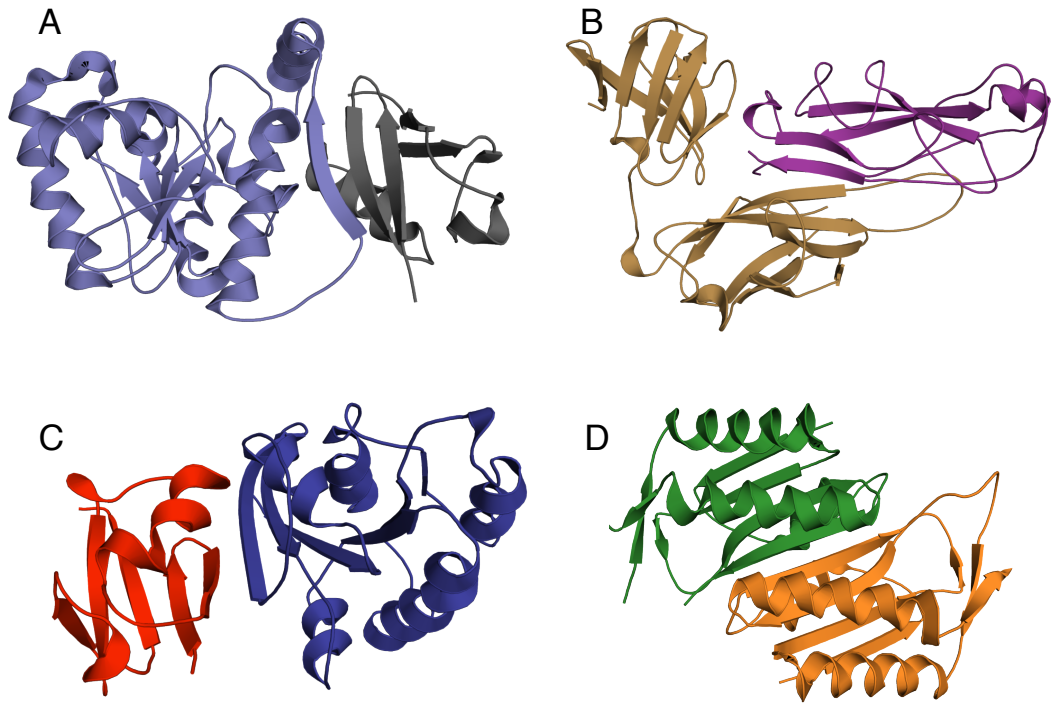

Figure 1.2: Natural β-strand mediated protein interactions. A) Crystal structure of thymine DNA glycosylase (light blue) conjugated to SUMO-1 (grey) (PDB ID: 1WYW) demonstrating β-strand addition(86). B) Crystal structure of the chaperone protein PapD (beige) in complex with PapK (purple) (PDB ID: 1PDK) demonstrating fold completion (87). C) Canonical β-strand mediated GTPase-effector interaction demonstrated by the complex of RAP1A (blue) with c-Raf1(red) (PDB ID: 1GUA) (88) D) Symmetric homodimer molybdopterin synthase (PDB ID: 1NVJ) with anti parellel β-strands at the interface between the two chains (green and orange) (89).

to interact with other strands. This is evident in β-strand pairing in crystal contacts (97; 98; 99; 53) and the propensity of β-sheet proteins to uncontrollably assemble into amyloid-like fibrils (100; 101). Exposed strands are so prone to interact with each other that nature uses negative design elements such as strand kinks, charged residues, or occlusion with a loop to prevent nonspecific β-strand interactions (102) Solvent accessible strands that interact with another protein tend to be longer and have fewer negative design elements than their counterparts (102; 85). The strength of β-strand interactions makes it possible to eliminate whole strands from the center of a β-meander motif and still maintain solubility and overall fold topology (103; 104). The potential downfall of this approach is that any mutations along an exposed

strand could cause the designed protein to uncontrollably self assemble. While this is a distinct possibility, explicit negative design steps were not required to redesign a natural β-sandwich protein (71), which suggests negative design elements exist in the backbone of exposed strands.

Several studies have shown the efficacy of using β-strands to alter protein function. Cyclic peptide inhibitors of protein-protein interactions and protease activity are thought to bind to their targets via formation of a intermolecular β-sheet (105; 106). Directed evolution experiments have generated β-strand pairing between an antibody and its antigen (107) as well as phage display targets with an exposed strand (108). Amyloid fibrillization can be inhibited by a peptide designed to form a terminating β-strand on a growing fibril (109) and an evolved β-strand presenting protein (110).

The following chapters seek to provide a solution to some of the difficulties in protein-protein interaction design. First, I introduce β-strand pairing at protein-protein interfaces as a way to ensure orientation and affinity at a *de novo* interface. I demonstrate that pairing solvent exposed β-strands, followed by sequence design at the interface, can lead to the accurate design of a symmetric homodimer in Chapter 2, beginning on page 22. Chapter 3 (page 52), extends this idea to the design of a binder to a natural protein using β-strand pairing at the targeted interface. I describe a general method to find proteins capable of forming an intermolecular β-sheet. Finally, Chapter 4 (page 77) examines successful and failed attempts to computationally design novel interfaces and finds that most successful designs have few designed polar interactions at the interface.

References

1. Dill, K. A. (1990) Dominant forces in protein folding. Biochemistry **29**, 7133–7155

2. Moult, J., Fidelis, K., Kryshtafovych, A., and Tramontano, A. (2011) Critical assessment of methods of protein structure prediction (CASP)–round IX. Proteins **79 Suppl 1**, 1–5

3. Betz, S. F., Raleigh, D. P., and DeGrado, W. F. (1993) De novo protein design: from molten globules to native-like states. Current Opinion in Structural Biology **3**, 601–610

4. Huang, P.-S., Love, J. J., and Mayo, S. L. (2007) A de novo designed protein protein interface. Protein science : a publication of the Protein Society **16**, 2770–4

5. Lim, W. a. (2010) Designing customized cell signalling circuits. Nature reviews. Molecular cell biology **11**, 393–403

6. Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., Ahn, J., Dewar-Darch, D., Reguly, T., Tang, X., Almeida, R., Qin, Z. S., Pawson, T., Gingras, A.-C., Nesvizhskii, A. I., and Tyers, M. (2010) A global protein kinase and phosphatase interaction network in yeast. Science (New York, N.Y.) **328**, 1043–6

7. Bennett, M. R., Pang, W. L., Ostroff, N. A., Baumgartner, B. L., Nayak, S., Tsimring, L. S., and Hasty, J. (2008) Metabolic gene regulation in a dynamically changing environment. Nature **454**, 1119–22

8. Janin, J. (2008) Biochemistry. Dicey assemblies. Science (New York, N.Y.) **319**, 165–6

9. Grueninger, D., Treiber, N., Ziegler, M. O. P., Koetter, J. W. A., Schulze, M.-S., and Schulz, G. E. (2008) Designed protein-protein association. Science (New York, N.Y.) **319**, 206–9

10. Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B., and Hahn, K. M. (2009) A genetically encoded photoactivatable Rac controls the motility of living cells. Nature **461**, 104–8

11. Hahn, K. M. and Kuhlman, B. (2010) Hold me tightly LOV. Nature methods **7**, 595, 597

12. Levskaya, A., Weiner, O. D., Lim, W. A., and Voigt, C. A. (2009) Spatiotemporal control of cell signalling using a light-switchable protein interaction. Nature **461**, 997–1001

13. Koide, S. (2009) Generation of new protein functions by nonhomologous combinations and rearrangements of domains and modules. Current opinion in biotechnology **20**, 398–404

14. Good, M. C., Zalatan, J. G., and Lim, W. A. (2011) Scaffold proteins: hubs for controlling the flow of cellular information. Science (New York, N.Y.) **332**, 680–6

15. Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., Prather, K. L. J., and Keasling, J. D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nature biotechnology **27**, 753–9

16. Yeh, B. J., Rutigliano, R. J., Deb, A., Bar-Sagi, D., and Lim, W. A. (2007) Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. Nature **447**, 596–600

17. Bashor, C. J., Helman, N. C., Yan, S., and Lim, W. A. (2008) Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. Science (New York, N.Y.) **319**, 1539–43

18. Jäckel, C., Kast, P., and Hilvert, D. (2008) Protein design by directed evolution. Annual review of biophysics **37**, 153–73

19. Sidhu, S. S. and Koide, S. (2007) Phage display for engineering and analyzing protein interaction interfaces. Current opinion in structural biology **17**, 481–7

20. Gai, S. A. and Wittrup, K. D. (2007) Yeast surface display for protein engineering and characterization. Current opinion in structural biology **17**, 467–73

21. Zahnd, C., Amstutz, P., and Plückthun, A. (2007) Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. Nature methods **4**, 269–79

22. Ferrer, M., Maiolo, J., Kratz, P., Jackowski, J. L., Murphy, D. J., Delagrave, S., and Inglese, J. (2005) Directed evolution of PDZ variants to generate high-affinity detection reagents. Protein engineering, design & selection : PEDS **18**, 165–73

23. Colwill, K. and Gräslund, S. (2011) A roadmap to generate renewable protein binders to the human proteome. Nature Methods **8**, 551–8

24. Clancy, K. and Voigt, C. A. (2010) Programming cells: towards an automated 'Genetic Compiler'. Current opinion in biotechnology **21**, 572–81

25. Holtz, W. J. and Keasling, J. D. (2010) Engineering static and dynamic control of synthetic pathways. Cell **140**, 19–23

26. Mandell, D. J. and Kortemme, T. (2009) Computer-aided design of functional protein interactions. Nature chemical biology **5**, 797–807

27. Street, A. G. and Mayo, S. L. (1999) Computational protein design. Structure **7**, R105–R109

28. Kortemme, T. and Baker, D. (2004) Computational design of protein-protein interactions. Current opinion in chemical biology **8**, 91–7

29. Chothia, C. and Janin, J. (1975) Principles Of Protein-Protein Recognition. Nature **256**, 705–708

30. Jones, S. and Thornton, J. M. (1996) Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America **93**, 13–20

31. Lensink, M. F. and Wodak, S. J. (2010) Docking and scoring protein interactions: CAPRI 2009. Proteins **78**, 3073–84

32. Karanicolas, J. and Kuhlman, B. (2009) Computational design of affinity and specificity at protein-protein interfaces. Current opinion in structural biology **19**, 458–63

33. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L., and Gonzalez, C. (2002) Computer-aided design of a PDZ domain to recognize new target sequences. Nature Structural Biology **9**, 621–627

34. Shifman, J. M. and Mayo, S. L. (2002) Modulating calmodulin binding specificity through computational protein design. Journal Of Molecular Biology **323**, 417–423

35. Yosef, E., Politi, R., Choi, M. H., and Shifman, J. M. (2009) Computational Design of Calmodulin Mutants with up to 900-Fold Increase in Binding Specificity. Journal Of Molecular Biology **385**, 1470

36. Kota, P., Summers, D. W., Ren, H.-Y., Cyr, D. M., and Dokholyan, N. V. (2009) Identification of a consensus motif in substrates bound by a Type I Hsp40. Proceedings of the National Academy of Sciences of the United States of America **106**, 11073–8

37. Lippow, S. M., Wittrup, K. D., and Tidor, B. (2007) Computational design of antibody-affinity improvement beyond in vivo maturation. Nature biotechnology **25**, 1171–6

38. Havranek, J. J. and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. Nature structural biology **10**, 45–52

39. Bolon, D. N., Grant, R. A., Baker, T. A., and Sauer, R. T. (2005) Specificity versus stability in computational protein design. Proceedings of the National Academy of Sciences of the United States of America **102**, 12724–9

40. Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature **458**, 859–864

41. Humphris, E. L. and Kortemme, T. (2007) Design of multi-specificity in protein interfaces. PLoS computational biology **3**, e164

42. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. Science (New York, N.Y.) **282**, 1462–7

43. Liu, S., Liu, S. Y., Zhu, X. L., Liang, H. H., Cao, A. N., Chang, Z. J., and Lai, L. H. (2007) Nonnatural protein-protein interaction-pair design by key residues grafting. Proceedings of The National Academy of Sciences **104**, 5330–5335

44. Azoitei, M. L., Ban, Y.-E. A., Julien, J.-P., Bryson, S., Schroeter, A., Kalyuzhniy, O., Porter, J. R., Adachi, Y., Baker, D., Pai, E. F., and Schief, W. R. (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. Journal of molecular biology **415**, 175–92

45. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y.-E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popović, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods in enzymology **487**, 545–74

46. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., and Baker, D. (2004) Protein structure prediction using Rosetta. Methods in enzymology **383**, 66–93

47. Kortemme, T., Morozov, A. V., and Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. Journal of molecular biology **326**, 1239–59

48. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation Of State Calculations By Fast Computing Machines. Journal Of Chemical Physics **21**, 1087–1092

49. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983) Optimization by simulated annealing. Science (New York, N.Y.) **220**, 671–80

50. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., and Meiler, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry **49**, 2987–98

51. Wang, C., Bradley, P., and Baker, D. (2007) Protein-protein docking with backbone flexibility. Journal of molecular biology **373**, 503–19

52. Smith, C. A. and Kortemme, T. (2008) Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. Journal Of Molecular Biology **380**, 742–756

53. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science (New York, N.Y.) **302**, 1364–8

54. Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., and Baker, D. (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One **6**, e20161

55. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P., and Kuhlman, B. (2007) Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. Journal of molecular biology **371**, 1392–404

56. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., and Baker, D. (2004) Computational redesign of protein-protein interaction specificity. Nature Structural & Molecular Biology **11**, 371–379

57. Joachimiak, L. A., Kortemme, T., Stoddard, B. L., and Baker, D. (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. Journal of molecular biology **361**, 195–208

58. Sammond, D. W., Eletr, Z. M., Purbeck, C., and Kuhlman, B. (2010) Computational design of second-site suppressor mutations at protein-protein interfaces. Proteins **78**, 1055–65

59. Kapp, G. T., Liu, S., Stein, A., Wong, D. T., Reményi, A., Yeh, B. J., Fraser, J. S., Taunton, J., Lim, W. A., and Kortemme, T. (2012) Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair. Proceedings of the National Academy of Sciences of the United States of America

60. Jha, R. K., Wu, Y. I., Zawistowski, J. S., MacNevin, C., Hahn, K. M., and Kuhlman, B. (2011) Redesign of the PAK1 autoinhibitory domain for enhanced stability and affinity in biosensor applications. Journal of molecular biology **413**, 513–22

61. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science **332**, 816–821

62. Sammond, D. W., Bosch, D. E., Butterfoss, G. L., Purbeck, C., Machius, M., Siderovski, D. P., and Kuhlman, B. (2011) Computational design of the sequence and structure of a protein-binding peptide. Journal of the American Chemical Society **133**, 4190–2

63. Der, B. S., Machius, M., Miley, M. J., Mills, J. L., Szyperski, T., and Kuhlman, B. (2012) Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. Journal of the American Chemical Society **134**, 375–385

64. Fleishman, S. J., Whitehead, T. A., Strauch, E. M., Corn, J. E., Qin, S., Zhou, H. X., Mitchell, J. C., Demerdash, O. N., Takeda-Shitaka, M., Terashi, G., Moal, I. H., Li, X., Bates, P. A., Zacharias, M., Park, H., Ko, J. S., Lee, H., Seok, C., Bourquard, T., Bernauer, J., Poupon, A., Aze, J., Soner, S., Ovali, S. K., Ozbek, P., Tal, N. B., Haliloglu, T., Hwang, H., Vreven, T., Pierce, B. G., Weng, Z., Perez-Cano, L., Pons, C., Fernandez-Recio, J., Jiang, F., Yang, F., Gong, X., Cao, L., Xu, X., Liu, B., Wang, P., Li, C., Wang, C., Robert, C. H., Guharoy, M., Liu, S., Huang, Y., Li, L., Guo, D., Chen, Y., Xiao, Y., London, N., Itzhaki, Z., Schueler-Furman, O., Inbar, Y., Potapov, V., Cohen, M., Schreiber, G., Tsuchiya, Y., Kanamori, E., Standley, D. M., Nakamura, H., Kinoshita, K., Driggers, C. M., Hall, R. G., Morgan, J. L., Hsu, V. L., Zhan, J., Yang, Y., Zhou, Y., Kastritis, P. L., Bonvin, A. M., Zhang, W., Camacho, C. J., Kilambi, K. P., Sircar, A., Gray, J. J., Ohue, M., Uchikoga, N., Matsuzaki, Y., Ishida, T., Akiyama, Y., Khashan, R., Bush, S., Fouches, D., Tropsha, A., Esquivel-Rodriguez, J., Kihara, D., Stranges, P. B., Jacak, R., Kuhlman, B., Huang, S. Y., Zou, X., Wodak, S. J., Janin, J., and Baker, D. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol **414**, 289–302

65. Jha, R. K., Leaver-Fay, A., Yin, S., Wu, Y., Butterfoss, G. L., Szyperski, T., Dokholyan, N. V., and Kuhlman, B. (2010) Computational design of a PAK1 binding protein. Journal Of Molecular Biology **400**, 257–270

66. Karanicolas, J., Corn, J., Chen, I., Joachimiak, L., Dym, O., Peck, S., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., T.áMontelione, G., P.áSpiegel, C., Liu, D., and Baker, D. (2011) A de novo protein binding pair by computational design and directed evolution. Molecular Cell **42**, 250–260

67. Potapov, V., Reichmann, D., Abramovich, R., Filchtinski, D., Zohar, N., Ben Halevy, D., Edelman, M., Sobolev, V., and Schreiber, G. (2008) Computational Redesign of a Protein-Protein Interface for High Affinity and Binding Specificity Using Modular Architecture and Naturally Occurring Template Fragments. Journal Of Molecular Biology **384**, 109–119

68. Filchtinski, D., Sharabi, O., Rüppel, A., Vetter, I. R., Herrmann, C., and Shifman, J. M. (2010) What makes Ras an efficient molecular switch: a computational, biophysical, and structural study of Ras-GDP interactions with mutants of Raf. Journal of molecular biology **399**, 422–35

69. Ofran, Y. and Rost, B. (2003) Analysing six types of protein-protein interfaces. Journal of molecular biology **325**, 377–87

70. Sheinerman, F. B. and Honig, B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces. Journal Of Molecular Biology **318**, 161–177

71. Hu, X., Wang, H., Ke, H., and Kuhlman, B. (2008) Computer-based redesign of a beta sandwich protein suggests that extensive negative design is not required for de novo beta sheet design. Structure (London, England : 1993) **16**, 1799–805

72. Reichmann, D., Phillip, Y., Carmi, A., and Schreiber, G. (2008) On the contribution of water-mediated interactions to protein-complex stability. Biochemistry **47**, 1051–60

73. Jiang, L., Kuhlman, B., Kortemme, T., and Baker, D. (2005) A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. Proteins **58**, 893–904

74. Schreiber, G. (2002) Kinetic studies of proteinprotein interactions. Current Opinion in Structural Biology **12**, 41–47

75. Selzer, T., Albeck, S., and Schreiber, G. (2000) Rational design of faster associating and tighter binding protein complexes. Nature structural biology **7**, 537–41

76. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G., and Herrmann, C. (2004) Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proceedings of the National Academy of Sciences of the United States of America **101**, 9223–8

77. Lippow, S. M. and Tidor, B. (2007) Progress in computational protein design. Current opinion in biotechnology **18**, 305–11

78. Cohen, M., Reichmann, D., Neuvirth, H., and Schreiber, G. (2008) Similar chemistry, but different bond preferences in inter versus intra-protein interactions. Proteins **72**, 741–53

79. Sharabi, O., Yanover, C., Dekel, A., and Shifman, J. M. (2011) Optimizing energy functions for protein-protein interface design. Journal of computational chemistry **32**, 23–32

80. London, N., Raveh, B., Movshovitz-Attias, D., and Schueler-Furman, O. (2010) Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? Proteins **78**, 3140–9

81. Lo Conte, L., Chothia, C., and Janin, J. (1999) The atomic structure of protein-protein recognition sites. Journal of molecular biology **285**, 2177–98

82. Guharoy, M. and Chakrabarti, P. (2007) Secondary structure based analysis and classification of biological interfaces: identification of binding motifs in protein-protein interactions. Bioinformatics (Oxford, England) **23**, 1909–18

83. Remaut, H. and Waksman, G. (2006) Protein-protein interaction through beta-strand addition. Trends In Biochemical Sciences **31**, 436–444

84. Kiel, C., Foglierini, M., Kuemmerer, N., Beltrao, P., and Serrano, L. (2007) A genome-wide Ras-effector interaction network. Journal of molecular biology **370**, 1020–32

85. Hoskins, J., Lovell, S., and Blundell, T. L. (2006) An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. Protein science : a publication of the Protein Society **15**, 1017–29

86. Baba, D., Maita, N., Jee, J.-G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. Nature **435**, 979–82

87. Sauer, F. G., Fütterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J., and Waksman, G. (1999) Structural basis of chaperone function and pilus biogenesis. Science (New York, N.Y.) **285**, 1058–61

88. Nassar, N., Horn, G., Herrmann, C., Block, C., Janknecht, R., and Wittinghofer, A. (1996) Ras/Rap effector specificity determined by charge reversal. Nature structural biology **3**, 723–9

89. Rudolph, M. J., Wuebbens, M. M., Turque, O., Rajagopalan, K. V., and Schindelin, H. (2003) Structural studies of molybdopterin synthase provide insights into its catalytic mechanism. The Journal of biological chemistry **278**, 14514–22

90. Chothia, C. and Janin, J. (1981) Relative Orientation Of Close-Packed Beta-Pleated Sheets In Proteins. Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **78**, 4146–4150

91. Nesloney, C. L. and Kelly, J. W. (1996) Progress towards understanding β-sheet structure. Bioorganic & medicinal chemistry **4**, 739–66

92. Ho, B. K. and Curmi, P. M. G. (2002) Twist and shear in beta-sheets and beta-ribbons. Journal of molecular biology **317**, 291–308

93. Minor, D. L. and Kim, P. S. (1994) Measurement Of The Beta-Sheet-Forming Propensities Of Amino-Acids. Nature **367**, 660–663

94. Minor, D. L. and Kim, P. S. (1994) Context Is A Major Determinant Of Beta-Sheet Propensity. Nature **371**, 264–267

95. Smith, C. K., Withka, J. M., and Regan, L. (1994) A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. Biochemistry **33**, 5510–7

96. Smith, C. K. and Regan, L. (1995) Guidelines for protein design: the energetics of beta sheet side chain interactions. Science (New York, N.Y.) **270**, 980–2

97. Tereshko, V., Uysal, S., Koide, A., Margalef, K., Koide, S., and Kossiakoff, A. A. (2008) Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. Protein Sci **17**, 1175–1187

98. Koide, S. (2009) Engineering of recombinant crystallization chaperones. Curr Opin Struct Biol **19**, 449–457

99. Nauli, S., Kuhlman, B., Le Trong, I., Stenkamp, R. E., Teller, D., and Baker, D. (2002) Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2. Protein Sci **11**, 2924–2931

100. Wang, W. and Hecht, M. H. (2002) Rationally designed mutations convert de novo amyloid-like fibrils into monomeric beta-sheet proteins. Proceedings of the National Academy of Sciences of the United States of America **99**, 2760–5

101. Yeates, T. O. and Padilla, J. E. (2002) Designing supramolecular protein assemblies. Curr Opin Struct Biol **12**, 464–470

102. Richardson, J. S. and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proceedings of the National Academy of Sciences of the United States of America **99**, 2754–9

103. Makabe, K., Yan, S., Tereshko, V., Gawlak, G., and Koide, S. (2007) β-strand flipping and slipping triggered by turn replacement reveal the opportunistic nature of beta-strand pairing. Journal of the American Chemical Society **129**, 14661–14669

104. Makabe, K. and Koide, S. (2008) The Promiscuity of Beta-Strand Pairing Allows for Rational Design of Beta-Sheet Face Inversion. Journal of the American Chemical Society **130**, 14370–14371

105. Loughlin, W. A., Tyndall, J. D. A., Glenn, M. P., and Fairlie, D. P. (2004) β-strand mimetics. Chemical Reviews **104**, 6085–6118

106. Khakshoor, O. and Nowick, J. S. (2008) Artificial β-sheets: chemical models of β-sheets. Current Opinion In Chemical Biology **12**, 722–729

107. Ni, Y. G., Di Marco, S., Condra, J. H., Peterson, L. B., Wang, W., Wang, F., Pandit, S., Hammond, H. A., Rosa, R., Cummings, R. T., Wood, D. D., Liu, X., Bottomley, M. J., Shen, X., Cubbon, R. M., Wang, S.-p., Johns, D. G., Volpari, C., Hamuro, L., Chin, J., Huang, L., Zhao, J. Z., Vitelli, S., Haytko, P., Wisniewski, D., Mitnaul, L. J., Sparrow, C. P., Hubbard, B., Carfí, A., and Sitlani, A. (2011) A PCSK9-binding antibody that structurally mimics the EGF(A) domain of LDL-receptor reduces LDL cholesterol in vivo. Journal of lipid research **52**, 78–86

108. Gilbreth, R. N., Truong, K., Madu, I., Koide, A., Wojcik, J. B., Li, N.-S., Piccirilli, J. A., Chen, Y., and Koide, S. (2011) Isoform-specific monobody inhibitors of small ubiquitin-related modifiers engineered using structure-guided library design. Proceedings of the National Academy of Sciences **108**, 7751–7756

109. Sievers, S. A., Karanicolas, J., Chang, H. W., Zhao, A., Jiang, L., Zirafi, O., Stevens, J. T., Mnch, J., Baker, D., and Eisenberg, D. (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nature **475**, 96–101

110. Smith, T. J., Stains, C. I., Meyer, S. C., and Ghosh, I. (2006) Inhibition of beta-amyloid fibrillization by directed evolution of a beta-sheet presenting miniature protein. Journal of the American Chemical Society **128**, 14456–7

Chapter 2

Computational design of a symmetric homodimer using β-strand assembly

The work in this chapter has been published in *Proceedings of the National Academy of Sciences* (2011) **108:** pp 20562-7. (1). Mike Miley helped up screening conditions for crystallography and screened crystals for diffraction. Mischa Machius collected the X-ray diffraction data and helped determine the structure. Ashutosh Tripathy designed the AUC experiments to measure affinity and size of the protein constructs.

2.1 Introduction

Protein-protein interactions and assemblies are essential for a wide array of cellular processes. The ability to rationally design unique protein interactions could provide scaffolds for functional reactions and new reagents for perturbing and monitoring cellular processes. Computational approaches for interface design have advanced rapidly in recent years and have allowed interactions to be engineered for increased affinity or altered specificity (2; 3). One long-standing goal is the creation of unique interactions. Thus far, most computational designs of new interactions have involved either the pairing of α-helices (4; 5; 6; 7) or binding of an α-helix to an open groove on a target (8; 9; 10; 11). Other methodologies have focused on grafting side-chain interactions from a known interaction onto another scaffold (8; 12; 13). There have been two examples of structurally confirmed unique computational interface designs (7; 8), however these sample a limited set of modes by which proteins can interact. New methods

of constructing an interface are necessary to mimic the ways nature forms protein-protein interactions (14).

There are many examples of naturally occurring protein heterodimers, homodimers, and larger complexes where β-strands from each chain associate to form an intermolecular β-sheet (15); β-strand pairing has also been observed in evolved antibody-antigen interactions (16) and monobody-target interfaces selected from phage display libraries (17). It has been proposed that β-strand pairing is so favorable that naturally occurring proteins often use negative design to avoid edge-to-edge association. In one study, 75 monomeric β-sheet proteins were visually examined to see if they contained structural features that would be predicted to disfavor β-sheet formation across their edge strands (18). In almost every case, one or more negative design elements were present including prolines, strategically placed charges, very short edge strands, loop coverage, and irregular edge strands. The propensity of exposed β-strands to pair is reinforced by observations of intermolecular β-sheet formation at crystal contacts of crystallization chaperones (19; 20) and designed proteins (21; 22) with exposed strands. In addition to providing affinity, β-strand interactions are geometrically constrained (23), which could provide a stable building block for designing interactions with a predetermined binding orientation. The intrinsic preference of β-strands to interact suggests that they may serve as a good anchor point for de novo interface design.

Formation of symmetric homodimers is one of the most common ways that proteins interact (24). Symmetric oligomerization provides increased stability, strict control over the number of protein units in the assembly, and low-energy structures (25). A survey of secondary structure at interfaces found that strand pairing represents 8.8% of contacts in homodimers (15). Paired β-strands at a homodimer interface are typically antiparallel (15) and longer than noninterface forming exposed strands (26). Protecting elements, typical of exposed strands in monomeric proteins, are less prevalent at β-strand mediated protein interfaces (18; 26).

There have been few successful rational designs of β-strand mediated protein interactions. Peptides that form β-strand mimetics are therapeutically used to inhibit proteases or protein-protein interactions(27; 28). One approach targeted amyloid fibrils by computationally designing a peptide to form a terminating β-strand on a growing fibril (29). Another study

took the sequence of the β-strand of a known β-strand mediated homodimer and embedded it in a cyclic peptide (30). A crystal structure of the peptide showed it formed an antiparallel β-strand paired dimer as predicted (31). However, there have been no structurally verified computational designs of a unique protein-protein interaction between two domains where the interface contains interactions between β-strands.

Here, we redesign a monomeric protein to form a symmetric homodimer via an intermolecular β-sheet. To design a β-strand mediated homodimer, we first identified structures in the protein database with exposed β-strands that could self-associate by β-strand pairing. We then used symmetric docking and sequence optimization (32) to create favorable interactions surrounding the interacting strands. Four designs were experimentally characterized and one was found to adopt the structure of the computational model.

2.2 Materials and Methods

2.2.1 Search Method for Homodimer Scaffolds

To find possible starting structures for homodimer design, we computationally scanned through a set of 5,500 high-resolution crystal structures. All computational steps were performed in the Rosetta3 suite of macromolecular protein modeling software (33). We defined a β-strand as surface-exposed if it met three criteria: (*i*) five sequential residues had β-strand secondary structure as judged by the Database of Secondary Structure of Proteins algorithm (34); (*ii*) there were no backbone-backbone hydrogen bonds formed by every other residue in the strand; and (*iii*) every other residue had fewer than 16 neighboring residues, or had 16-30 neighbors and a SASA per atom greater than 2.0 Å^2. Residues are defined as neighbors if their Cβ to Cβ distance is < 10 Å. An example command line used to find the exposed strands follows:

```
./exposed_strand_finder.<exe>
     -database <rosetta_database>
     -l <list_of_inputs>
     -ignore_unrecognized_res true
     -packing::pack_missing_sidechains
     -out::nooutput > exposed_list
```

We then created a potential homodimer. The axis of an exposed β-strand was defined as a vector from the Cα atom on the first residue of the strand to the Cα atom on the final residue of the strand. Another vector is defined at the center residue of the strand from the carbonyl carbon to carbonyl oxygen. A final vector is drawn perpendicular to the two vectors described through the Cα atom of the residue at center of the strand. The antiparallel homodimer is constructed by copying the protein and rotating the copy 180° about this axis. The copied chain is then translated away from the original by 6.0 Å to create a starting point for evaluation. The copied chain was then translated along the axis of the exposed strand in steps of 7 Å to identify alternate conformations that have no clashing backbone atoms. As a final filter, we check to make sure there are no backbone-backbone clashes between the two chains. The identification of exposed β-strands generated from the above protocol was used as input for the next step. The command line used to make potential homodimers with interacting strands of five residues was

```
./homodimer_maker.<exe>
     -database <rosetta_database>
     -s <pdb_file>
     -run::chain <chain_char>
     -sheet_start <start_residue_#>
     -sheet_stop <last_residue_#>
     -window_size 5
```

To narrow down the list of alignments, we ran short symmetric design simulations followed by side-chain and backbone minimization. These alignments were then filtered for designs that possessed an interface area of at least 850 Å^2, two or fewer polar atoms not forming hydrogen bonds at the interface, and a calculated ΔG_{bind} of less than -15.0 Rosetta energy units. The protocol used for this step is similar to the one for full design below.

2.2.2 Homodimer Design and Selection

The computational homodimer interface design strategy is similar to the Dock Design Minimize Interface protocol used previously for heterodimer design (10). Each step employed Rosettas symmetry protocols, which can perform symmetric protein-protein docking, symmetric design, and side-chain/backbone minimization (32; 35). The homodimer model generated above was used to generate the symmetry definition and starting structure for interface design. First, the protein is symmetrically docked against itself to sample rigid-body degrees of freedom. After the docking step, all residues within 8 Å of the other chain were symmetrically designed. Finally, the backbone and side chains of all interface residues were minimized. An example command line for this protocol is:

```
./homodimer_design.<exe>
     -database <rosetta_database>
     -s <pdb_file>
     -symmetry:symmetry_definition <symmdef>
     -nstruct 5000
     -pack_min_runs 4
     -make_ala_interface false
     -find_bb_hbond_E true
     -no_his_his_pairE true
     -disallow_res CGP
     -use_input_sc -ex1 -ex2
     -docking:docking_local_refine true
     -docking:sc_min true
     -docking:dock_ppk false
     -symmetry:perturb_rigid_body_dofs 3 5
     -out:file:fullatom
```

2.2.3 Evaluation of Designs

We selected which computational designs to express based on several metrics. As a first criterion, we selected designs that were in the top 10% in backbone-backbone hydrogen-bond energy across the interface, total Rosetta energy, and calculated ΔG_{bind}. We then calculated additional metrics including interface energy density ($\Delta G_{bind}/\Delta$SASA), RosettaHoles score (36), and number of buried-unsatisfied at the interface. To pick out the final designs to test experimentally, we visually inspected the designs that scored better than native interfaces in all of these metrics.

2.2.4 DNA Construct and Protein Production

DNA sequences for the wild type and all four designs were synthesized by GenScript USA and cloned into the pQE-80L vector as 6-His-maltose-binding protein (MBP) fusions as described previously (10). All proteins were expressed in BL21(DE3) pLysS cells induced with 0.3 mM IPTG overnight at 18 °C. The proteins were purified by immobilized-nickel affinity chromatography and then cleaved from 6-His-MBP with tobacco etch virus protease. The cleaved proteins were again subjected to immobilized-nickel affinity chromatography to trap the 6-His-MBP. Flow-through from the nickel column was then further purified with size-exclusion chromatography (Superdex 75) in buffer A (20 mM MES, pH 6.0, and 150 mM NaCl). Protein concentration was quantified based on absorbance and a predicted extinction coefficient (ExPASy; ProtParam) of 8;480 M^{-1} cm^{-1} for wild type and βdimer4, 13;980 M^{-1} cm^{-1} for βdimer1, 12;490 M^{1} cm^{1} for βdimer2, and 9;970 M^{-1} cm^{-1} for βdimer3.

2.2.5 Multiangle Light Scattering

Samples of βdimer1, βdimer3, and the wild-type protein were concentrated to approximately 300 μM (4 mg/mL) in buffer A and injected onto a WTC-030S5 size-exclusion column (Wyatt Technologies) connected to a multiangle light scattering instrument (DAWN HELEOS II; Wyatt Technologies) and a refractometer (OPTILAB rEX; Wyatt Technologies). Molecular mass of particles in a single elution peak was calculated based light scattering data using the ASTRA software package (Wyatt Technologies).

2.2.6 Analytical Ultracentrifugation Sedimentation Equilibrium

Sedimentation equilibrium experiments were performed using a Beckman XL-I analytical ultracentrifuge using six-sector cells and an An-50 Ti rotor. Samples of wild-type protein, βdimer1, and βdimerdimer3, at concentrations of 20, 40, and 60 μM in buffer A, were spun at 46,400 g until equilibrium was reached. Absorbance measurements at 280 nm were taken every 2 h. The absorbance offset was found by meniscus depletion after spinning the samples at 163,300 g for 6 h. The sedimentation equilibrium data were analyzed with XL-I data analysis software.

The homodimer dissociation constant was measured in a similar fashion to the method outlined above. βdimer1 was placed at in the sample cells at concentrations of 0.8, 1.5, and 2.0 μM in 20 mM KH_2PO_4, pH 7.0 and 150 mM NaCl. The absorbance was measured at 215 nm to obtain readings sufficiently above background to reliably fit the data. The data were analyzed using a monomer-dimer equilibrium model.

2.2.7 Fluorescence Polarization Assay

A variant of βdimer1 with the mutation S62C was produced for labeling with thiol reactive Bodipy (507/545)-iodoacetamide (Molecular Probes). The labeling procedure was performed as previously described (10). Buffer A supplemented with 5 mM β-mercaptoethanol was used as the binding buffer for the titrations. Bodipy-labeled βdimer1, at a concentration of 2 nM, was placed in a 1-cm path length cuvette and titrated with unlabeled protein. The change in fluorescence polarization was measured using a Jobin Yvon Horiba Spex FluoroLog-3 instrument (Jobin Yvon, Inc.). Titration amounts were calculated as described in Appendix A on page 129.The data were analyzed according to a homodimerization model (described below) and fit with Prism (GraphPad Software).

2.2.8 Homodimerization Fluorescence Polarization Fitting Procedure

We derived a homodimerization model to be used when fitting the fluorescence polarization data. This model accounts for the interaction of a protein A with itself in its labeled (A^*) and unlabeled states (A). The model was derived as follows where P is the total amount of protein in a given state:

$$A + A \rightleftharpoons AA$$

$$A^* + A \rightleftharpoons AA + A^*A^* + A^*A + AA^*$$

Writing this in terms of K_d and total protein concentration:

$$K_d = \frac{[A]^2}{[AA]}$$

$$[P_{total}] = [A^*_{total}] + [A_{total}]$$

$$[P_{monomer}] = [A^*_{monomer}] + [A_{monomer}]$$

$$[P_{dimer}] = [A^*A^*] + [A^*A] + [AA^*] + [AA]$$

$$[P_{dimer}] = \frac{[P_{total}] - [P_{monomer}]}{2}$$

Solving for the total concentration of monomeric protein gives:

$$[P_{monomer}] = \frac{-K_d + \sqrt{K_d^2 + 8[P_{total}]K_d}}{4}$$

Any change in signal seen would come from association of a labeled and unlabeled protein. We assume that the interactions between labeled and labeled is negligible because labeled protein is present in low concentrations.

$$[A^*A] = [AA^*] = [P_{dimer}]\frac{[A^*_{total}] + [A_{total}]}{[P_{total}]}$$

This model is then written to fit the change in polarization:

$$Pol_{obs} = (Pol_{max} - Pol_{min})\frac{[A^*A] + [AA^*]}{[A_{total}]} - Pol_{min}$$

The code for this fit is given in Appendix A on page 129.

2.2.9 Crystallization and Structure Refinement

Crystallization of βdimer1 was performed using the hanging-drop vapor diffusion method at 20 °C. Crystals formed after one week in a drop consisting of 2 μL of βdimer1 (7 mg/mL in buffer A) and 1 μL of well solution [100 mM sodium acetate, pH 5.0, 6% (vol/vol) isopropanol, 20% (wt/vol) PEG 8000]. Prior to data collection, crystals were cryoprotected by transferring

them into well solution supplemented with 15% (vol/vol) ethylene glycol before plunging them into liquid nitrogen. The crystals diffracted X-rays to a minimum Bragg spacing of about 1.0 Å, exhibited the symmetry of space group $P2_1$ with cell parameters of a = 50.6 Å, b = 44.3 Å, c = 53.0 Å, β= 91.91°, and contained two molecules in the asymmetric unit (solvent content, 44%). Diffraction data were collected at 100 K at a wavelength of 0.91840 Å at the Advanced Proton Source General Medicine and Cancer institutes Collaborative Access Team 23IDB beamline. The diffraction data were indexed and reduced using HKL2000 (37).

The structure of βdimer1 was determined by molecular replacement using the program Phaser (38); the computationally designed dimer of βdimer1 was used as a search model. Iterative rounds of refinement were conducted with REFMAC (39) and PHENIX (40), interspersed with manual adjustments to the model using the program Coot (41). The final model contains two molecules in the asymmetric unit with all residues defined in the electron density, except for residues 23-26 in both molecules.

2.3 Results

2.3.1 Scaffold Search Protocol

To find proteins with surface-exposed β-strands, we performed a computational search on a set of 5,500 protein crystal structures with resolution better than 2.2 Å to find proteins with a surface-exposed β-strands (Figure 2.1). A strand was defined as exposed if there was a continuous stretch of five or more residues in which every second residue did not form backbone-backbone hydrogen bonds and were not occluded from solvent (see Materials and Methods). This criterion yielded 1,500 exposed β-strands on 1,100 unique proteins. We then tested each exposed β-strand for its potential to form the basis of a homodimer interface. A copy of the entire chain of each protein with an exposed strand was rotated and translated to an ideal backbone-backbone hydrogen bonding distance to the original protein chain (Figure 2.1). The copied chain was then translated along the exposed strand in steps of 7 Å to identify alternate conformations that had no clashing backbone atoms. After this step, there were 2,800 potential alignments of 900 different proteins.

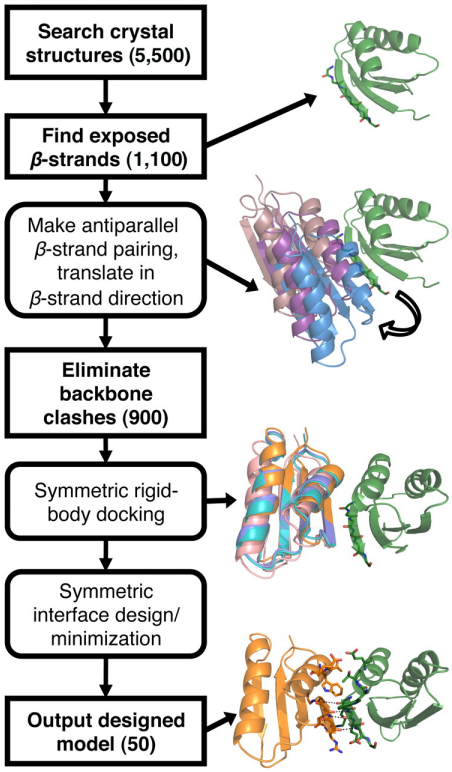

Figure 2.1: Search and design protocol for a symmetric β-strand mediated homodimer. Method used to search for, then design, scaffold proteins to create a symmetric homodimer (see full details in Materials and Methods). Numbers in parentheses represent the total number of unique input structures used in each step. Individual steps are illustrated by the structures generated during each step using the protein Atx1 (Protein Data Bank ID 1CC8).

To narrow down the list further, we performed a brief design and minimization protocol to determine which of our potential homodimers gave favorable binding energies and interface sizes after design (see Materials and Methods). This step reduced the overall number of targets to 200. From the final set of designable alignments, we removed all proteins that had not previously been expressed in *Escherichia coli*, were natural oligomers, had crystal contacts that resulted in an intermolecular β-sheet, were over 500 amino acids in length, or whose interacting β-strands were not part of globular domains. These steps generated 50 possible starting points for design of a homodimer.

2.3.2 Design Protocol

Each design simulation consisted of one round of symmetric protein docking (35) followed by five successive rounds of symmetric sequence optimization and minimization of side-chain and backbone residues at the homodimer interface (see Materials and Methods). For some protein scaffolds it was necessary to build alanine into all positions at the interface to obtain a docked structure that formed hydrogen bonds between the β-strands. We applied a stringent filter to eliminate designs that were unlikely to produce the desired experimental results. First, from each run, we selected only designs in the top 10% in total score, ΔG_{bind}, and β-strand hydrogen-bond energy. These selections were further filtered for interface energy density (ΔG_{bind}/SASA, where SASA represents the solvent-accessible surface area), number of buried polar atoms failing to form hydrogen bonds, and packing quality (RosettaHoles score) (36). We selected four homodimer designs based on the γ-adaptin appendage domain (Protein Data Bank ID 2A7B) (42). This scaffold protein was chosen because the designs of 2A7B scored favorably compared to other potential homodimers according to all metrics described above. Two of the four designs chosen had predominantly hydrophobic interfaces (βdimer1 and βdimer2), whereas the other two contained more polar interactions (βdimer3 and βdimer4) (Table 2.1), which allowed us to test our ability to design hydrogen bonding networks and hydrophobic packing interactions. All four designs exhibited a similar overall conformation and β-strand register to that of βdimer1 (Figure 2.2A). The maximum Cα rmsd from βdimer2, βdimer3, and βdimer4 to βdimer1 is 1.5 Å. All four designs have a total of six main-chain hydrogen bonds between residues 104, 106, and 108 on one chain to residues 108, 106, and 104 on the other chain, respectively. One face of the intermolecular β-sheet is exposed to solvent, whereas the other is occluded by a loop formed by residues 10-12. The crystal structure 2A7B has no crystal lattice contacts along the exposed strand, suggesting that the wild-type sequence is not prone to form an intermolecular β-sheet. Wild-type -adaptin appendage domain is likely prevented from self-association by a salt bridge between residues K10 and D107 that would be buried at the designed homodimer interface. In the designs, K10 is mutated to alanine, leucine, or serine and D107 is mutated to serine or threonine. A common feature in all four designs is charge complementation on the solvent-accessible side

of the interacting strands between residues 104 and 108 on opposite chains. For example, in βdimer1, residue 104 is a lysine and residue 108 is a glutamate. In βdimer3, residue 104 is an arginine and residue 108 is a glutamate. The buried side of the interface is dominated by either hydrophobic or polar interactions depending on the design (Figure 2.2 B-E). A search of Protein Data Bank Protein Interfaces, Surfaces, and Assemblies (43) yielded no known interfaces bearing any similarity to the designs, suggesting that the designed complexes represent a unique configuration of a protein-protein interaction.

Model	No. mutations	E_{total}	ΔG_{bind}	No. buried-unsatisfied	Polar interface area, %
Wild type	0	-561	-13	8	48
βdimer1	11	-597	-29	2	39
βdimer2	7	-593	-30	0	31
βdimer3	5	-596	-32	0	54
βdimer4	9	-593	-27	0	46

Table 2.1: Computational evaluation of designed homodimer models . Computational values used to select homodimer designs are shown compared to the wild-type protein represented as a homodimer forced into a conformation similar to the designs. No. mutations is the number of mutations to the wild type to generate the design, E_{total} is the Rosetta energy for the homodimer, ΔG_{bind} is the difference in energy between the complex and two monomers of a model ($\Delta G_{bind}= E_{AB} - E_A - E_B$), No. buried-unsatisfied is the number of polar atoms that are not solvent accessible and do not form hydrogen bonds to another atom in the protein, and Polar interface area, % represents the amount of solvent-accessible surface area of polar atoms hidden at the interface ($SASA_{polar}/SASA_{total}$).

2.3.3 Determining Oligomeric Status

We first assessed the oligomeric state of the four designs and the wild-type protein by size-exclusion chromatography. The molecular mass of a monomeric protein based on sequence is 13.5 kDa. The wild type, βdimer3, and βdimer4 eluted near the expected molecular mass for a monomeric protein, whereas βdimer1 and βdimer2 eluted close to the size expected for a dimer (Figure 2.3A and Table 2.2). We were unable to perform additional experiments with βdimer2 and βdimer4 because they did not express at sufficient levels.

To confirm the results from size-exclusion, we performed sedimentation equilibrium experiments using a Beckman XL-I analytical ultracentrifuge (AUC-SE). The wild-type protein, βdimer1, and βdimer3 were spun at 46,400 $\times g$ until equilibrium was reached. Three concentrations of protein (20, 40, and 60 μM) were used for each sample. Equilibrium absorbance

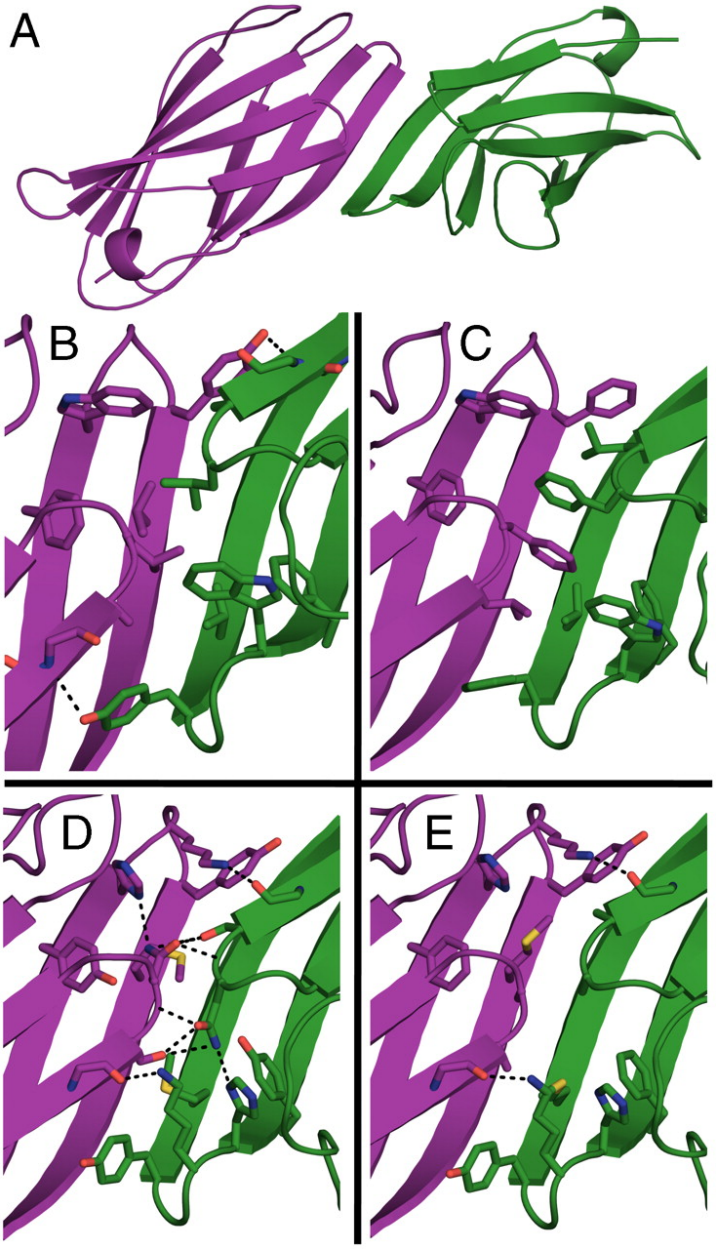

Figure 2.2: Computational designs used in experiments. (A) Overall topology of computational designs. The γ-adaptin appendage domain (Protein Data Bank ID 2A7B) is used as the scaffold for the designed interface. Coloring (purple and green) highlights the symmetric chains in the model. The solvent-excluded side of the interface is shown in detail for βdimer1 (B), βdimer2 (C), βdimer3 (D), and βdimer4 (E). Selected side chains are shown in sticks. Black dashed lines represent hydrogen bonds at the interface; the six main-chain hydrogen bonds are not shown.

profiles at 280 nm were used to determine molecular mass. The profiles for all three proteins were well fit by a single species model (Figure S2.1). The molecular mass determined from the equilibrium profile of the wild-type protein and βdimer3 were 12 and 16 kDa, respectively, close to that expected for a monomer. The molecular mass of βdimer1 was found by the same method to be 26 kDa, near that expected for a homodimer (Table 2.2).

We further tested the solution molecular mass of βdimer1, βdimer3, and the wild-type protein by size-exclusion chromatography (SEC) followed by multiangle light scattering (MALS). Each protein came off the size-exclusion column as a single peak. Light scattering and refractive index were used to determine the molecular mass of the peak (Figure 2.3B). The results were similar to the SEC experiment described above. βdimer3 and the wild-type protein were determined to have a molecular mass of 13 kDa, whereas βdimer1 had a molecular mass of 26 kDa (Table 2.2). These results further confirmed that βdimer1 forms a homodimer, but the wild type and βdimer3 do not.

2.3.4 Homodimer Binding Affinity

We used a fluorescence polarization assay to measure the dimer dissociation constant of βdimer1. Briefly, we expressed βdimer1 with the mutation S62C and labeled it with thiol reactive Bodipy. The monomer-dimer equilibrium was monitored by titrating excess unlabeled protein into dilute, Bodipy-labeled, βdimer1 protein and observing the increase in polarization from the formation of a slowly rotating dimeric species. βdimer1 was titrated with wild-type protein as a control. The change in polarization upon binding was fit to a homodimerization

	Molecular mass, kDa			
Protein	Monomer (calculated)	SEC	AUC	SEC/MALS
Wild type	13.6	11	12	13
βdimer1	13.6	26	26	26
βdimer2	13.7	21	–	–
βdimer3	13.7	10	16	13
βdimer4	13.6	12	–	–

Table 2.2: The molecular mass in solution measured for the wild-type protein and homodimer designs measured with using SEC, AUC, and MALS. Monomer (calculated) is the molecular mass expected based on sequence. The βdimer2 and βdimer4 proteins did not express in significant quantities for additional molecular mass determination.

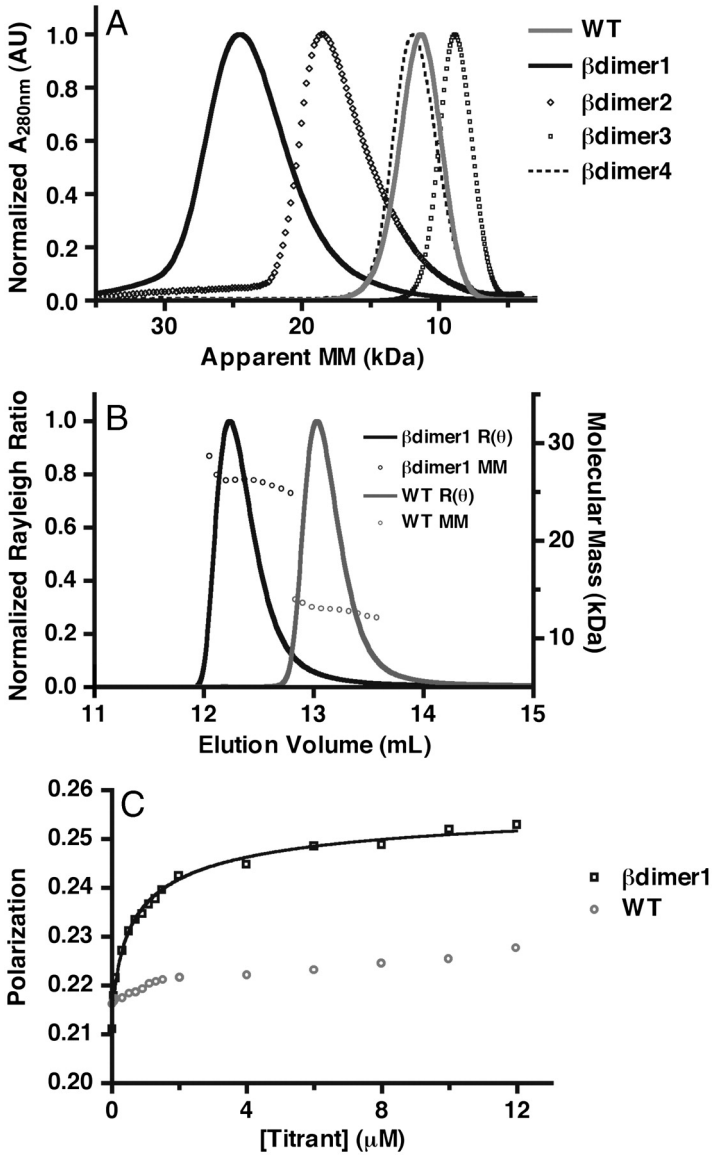

Figure 2.3: Experimental determination of molecular mass in solution. (A) Size-exclusion chromatography (Superdex 75) of the designs and wild-type protein. Absorbance has been normalized based on maximum value, the apparent molecular mass (MM) is based on a standard curve obtained from globular proteins. (B) Size-exclusion chromatography (Superdex 75) followed by multiangle light scattering of wild type (gray) and βdimer1 (black). Rayleigh ratio $[R(\theta)]$ (solid lines) has been normalized based on maximum value; MM (open circles) is calculated from light scattering and refractive index. The average molecular mass is 26 kDa for βdimer1, 14 kDa for the wild type. (C) Measurement of dimer dissociation constant of βdimer1 using a fluorescence polarization assay. Bodipy-labeled βdimer1 was titrated with unlabeled βdimer1 (black) and wild-type protein (gray), and the change in polarization was fit to a homodimerization model (see SI Materials and Methods). The calculated homodimer dissociation constant for βdimer1 is 1.0 ± 0.1 μM (SEM).

model (Materials and Methods). βdimer1 had a dimer dissociation constant of 1.0 μM, whereas the wild-type protein showed little to no interaction with βdimer1 (Figure 2.3C). To confirm this result, we performed AUC-SE with βdimer1 at concentrations of 0.8, 1.5, and 2.0 μM. The data were fit to a monomer-dimer self-association model that produced a dimer dissociation constant of 0.96 μM (Figure S2.2), which closely matches the dissociation constant determined by fluorescence polarization.

2.3.5 Crystal Structure of the βdimer1

We determined the crystal structure of βdimer1 using molecular replacement and diffraction data to a resolution of 1.09 Å (Table S2.1). The coordinates of the dimer design were used as the search model for molecular replacement. The asymmetric unit contained two molecules of βdimer1 protein, henceforth called chain A and chain B. These two chains in the crystal structure interact in a manner that is remarkably similar to the model of βdimer1 (Figure 2.4A) with an rmsd between the crystal structure and model of 1.0 Å for all backbone atoms. The intermolecular β-strand pairing found at the interface between the two chains in the crystal structure matches that of the designed model (Figure 2.4B). The conformation of the interacting strands in the crystal structure of βdimer1 show only minor differences when compared to the wild-type protein (Figure S2.3), indicating that substantial backbone rearrangement is not required for the formation of the homodimer.

The conformations of the interface residues in the crystal structure were well predicted by the computational model (Figure 2.4C). The conformations of the designed hydrophobic side chains in the crystal structure closely match those of the designed structure including L11 from one chain packing between L11 and W100 from the other chain. The computational model also accurately predicted an interface-spanning hydrogen bond between the backbone nitrogen of D9 and the hydroxyl oxygen on Y103 (Figure 2.4C). The solvent-exposed side of the paired strands presents an interesting divergence from the model. The side chains of residues E108 and Q106 appear to form hydrogen bonds from the side-chain nitrogen on Q106 to a carboxyl oxygen on E108, and from the other carboxyl oxygen on E108 to the side-chain oxygen on Q106 (Figure S2.4), suggesting the carboxyl group of E108 is protonated.

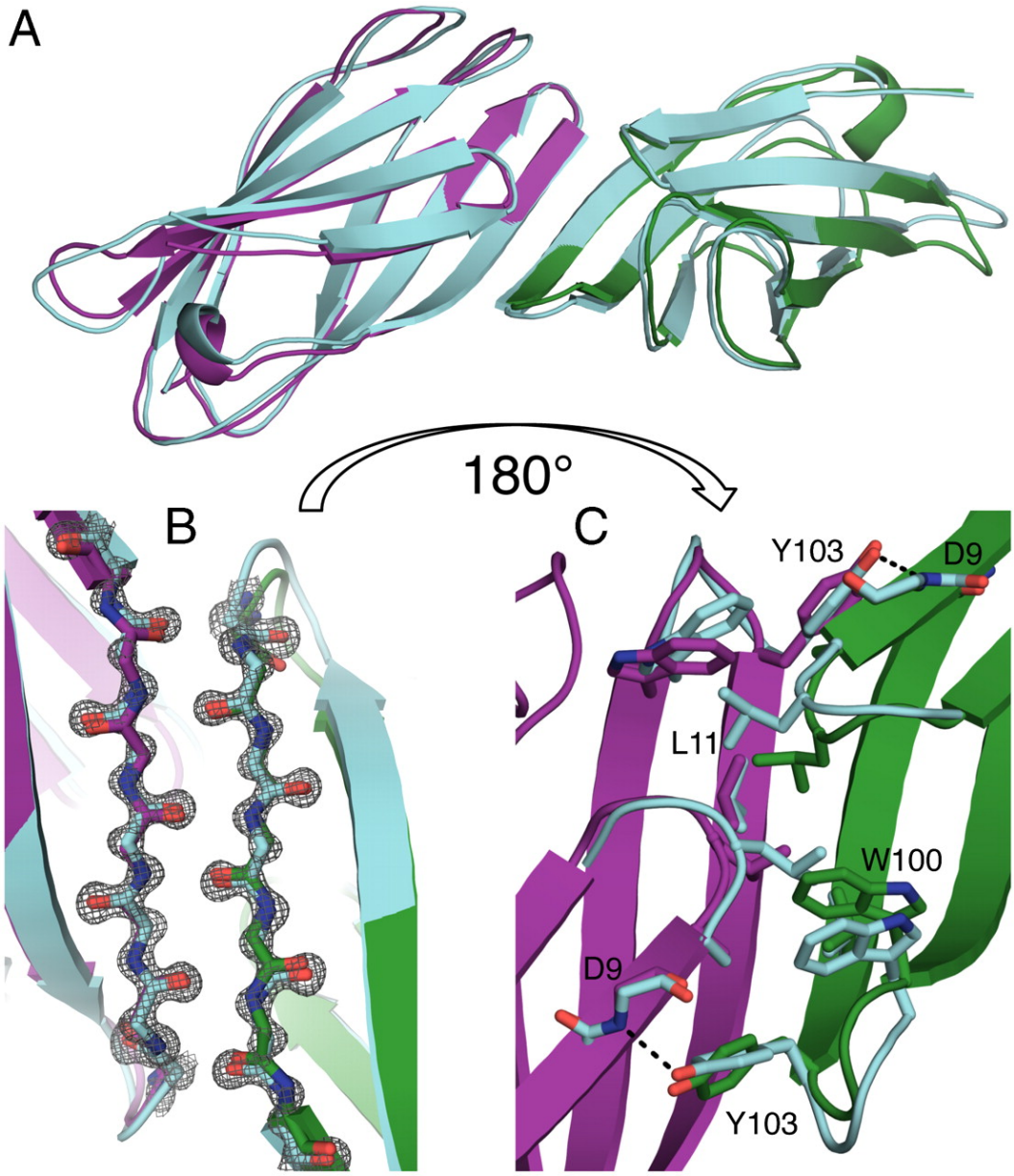

Figure 2.4: Comparison of βdimer1 computational model to crystal structure. (A) Overlay of the βdimer1 computational model (green and purple) and crystal structure (cyan). The backbone atom rmsd for the entire structure is 1.0 Å. (B) Backbone-backbone interactions between the interface-forming β-strands viewed from the solvent-accessible side of the intermolecular β-sheet. The $2F_o - F_c$ electron density (gray) is contoured to 2σ. (C) Detailed view of designed side chains forming interactions on the solvent-excluded side of interacting β-strands. A black dashed line represents the interface-spanning hydrogen-bond between D9 and Y103

.

Unmodeled protonated glutamates have also been observed forming stabilizing interactions in a computationally designed loop (44).

One possible pitfall of using β-strands to mediate an interaction is the possibility for register shifts between the paired strands. It is interesting to consider what structural elements in βdimer1 set the register between the two proteins. Docking the chains with constraints to force a register shift in either direction yielded no models with backbone-backbone hydrogen bonds at the interface. A shift in one direction introduces a clash between the side chains and backbone atoms of L11 in both chains, whereas a shift in the opposite direction creates a clash between residues Y8 and L11 on one chain with Y103 on the interacting chain (Figure S2.5).

2.4 Discussion

Our results demonstrate that protein-protein interactions can be engineered by selecting protein scaffolds with complementary, surface-exposed β-strands, then computationally designing the interface residues to form favorable interactions. One of the four homodimers we designed formed a stable dimer in solution. The crystal structure of this protein closely matches the design model.

We intentionally selected designs where the intermolecular interactions, other than the paired β-strands, were either predominantly polar or predominantly nonpolar. Although the energy and metric scores for these interfaces were similar (Table 2.1), only the designs with predominantly hydrophobic interfaces formed dimers in solution (Table 2.2). These results are unsurprising given previous observations that homodimeric interfaces are more hydrophobic than heterodimeric interfaces (45) and that new hydrogen-bond networks are difficult to design (46). A closer inspection of the computational model of βdimer3 revealed that four interface-spanning hydrogen bonds, between designed side chains and backbone atoms, were suboptimal. The N-H bond vector from the hydrogen-bond donors was more than 60° out of plane with the lone-pair electrons on the acceptor carbonyl oxygens. The N-H bond vector is typically in plane with the accepting electrons in crystal structures of natural proteins (47). This deviation is not penalized in the current implementation of the hydrogen-bond energy evaluation in Rosetta.

Some previous attempts at computational protein interface design have been plagued by problems controlling binding orientation, including a complete 180° rotation from the design model (48) or existence of multiple low-energy binding conformations (10). The β-strand pairing addresses these issues by constraining the possible geometry of the interface, as illustrated by the high similarity between the computational model and experimentally determined structure of βdimer1. One challenge of the β-strand pairing approach is that the paired strands must have complementary curvatures in order to form low-energy hydrogen bonds across the interface. This requirement limits the number of naturally occurring proteins that can be redesigned to form new homo- or heterodimers. One potential way to escape this limitation is by designing de novo scaffolds that have edge strands with the appropriate curvature for a target interaction.

Homodimerization illustrates an important step in protein evolution. Many protein-protein interfaces are built on the progression from a monomer to symmetric homodimer to asymmetric homodimer to heterodimer (25; 49). In fact, the majority of protein interfaces common across the three kingdoms of life are symmetric homodimers (24). Our results demonstrate that it is possible to make the first step in this process without disturbing the backbone conformation of the monomer. A logical next step in this path is to redesign the interface of the constructed homodimer to form a heterodimer.

The method of finding complementary, surface-exposed, β-strands presented here could be extended to other aspects of protein design. This protocol could be used to design a protein to bind a natural protein with an exposed β-strand, or build higher order oligomeric structures.

2.5 Supporting Information

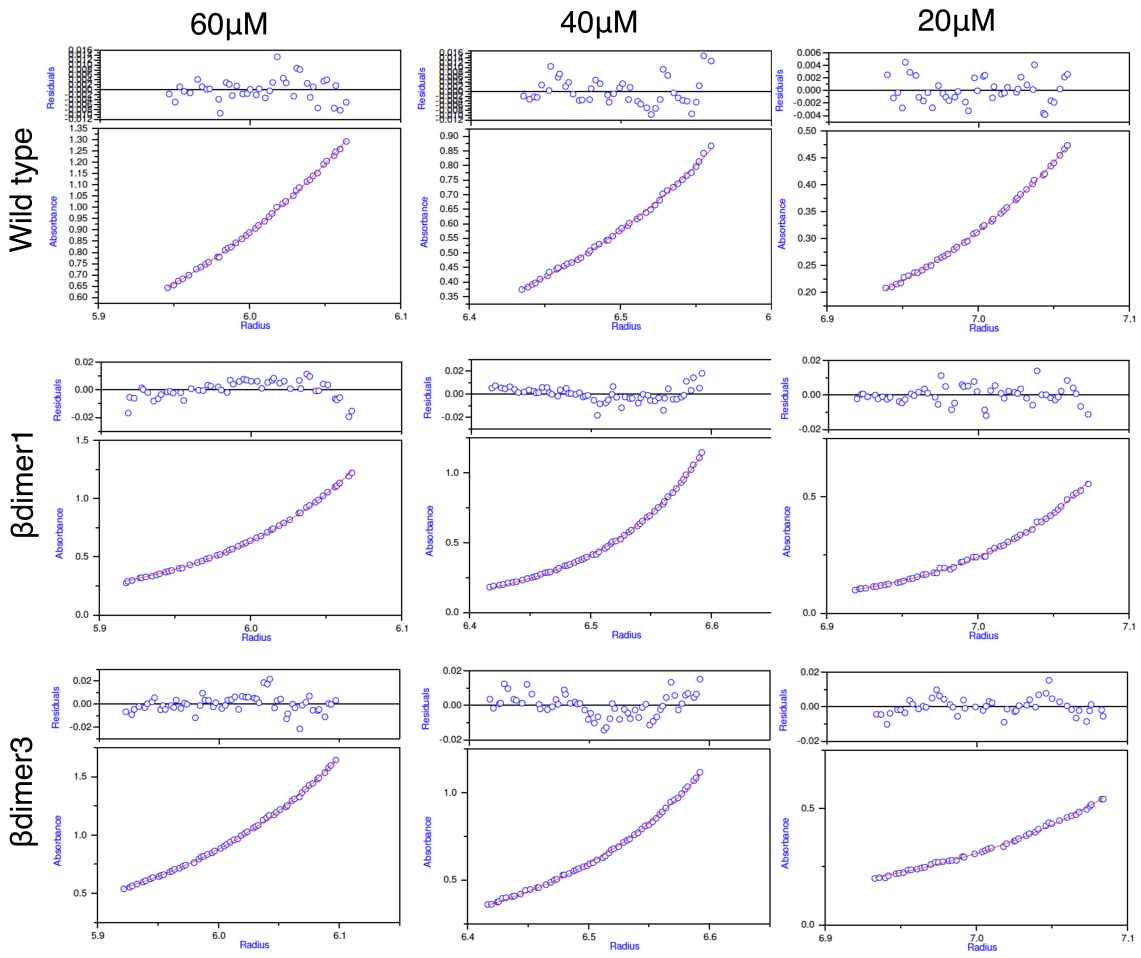

Figure S2.1: Determination of molecular mass of wild type, βdimer1, and βdimer3 by analytical ultracentrifugation sedimentation equilibrium. A single-species model was fit to the data for the wild-type protein, βdimer1, and βdimer3. Data from all three concentrations (20, 40, and 60 μM) were used in fitting to find the molecular mass. The molecular mass was 12 kDa for the wild type, 26 kDa for βdimer1, and 16 kDa for βdimer3.

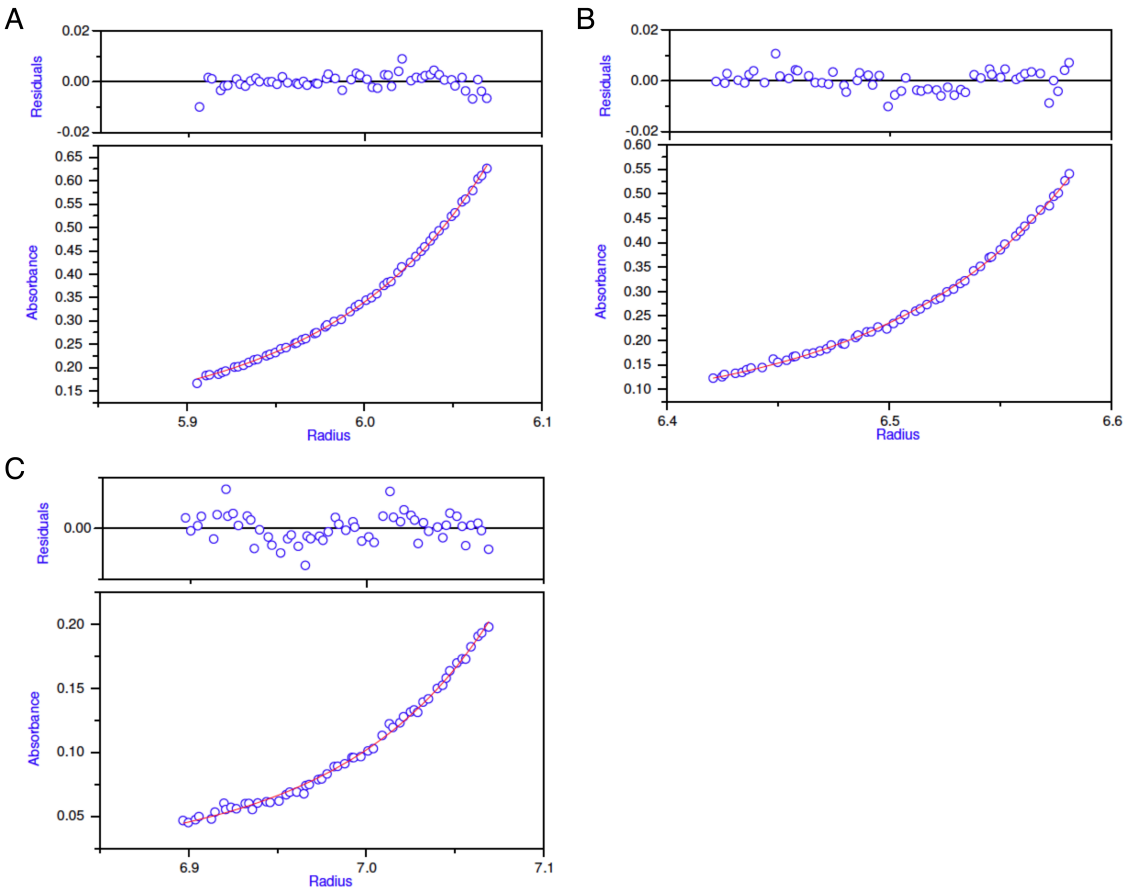

Figure S2.2: Determination of βdimer1 dimer dissociation constant by analytical ultracentrifugation sedimentation equilibrium. Absorbance was measured at 215 nm for βdimer1 at concentrations of 2.0 (A), 1.5 (B), and 0.8 μM (C). Data from all three concentrations were pooled and fit according to a self-association model with the molecular mass of the monomer assumed to be 13,628 Da. The fit produced a K_a of 9.0 in absorbance units, which can be converted to concentration units with the following equation: $K_a(M) = K_a(Abs)(\epsilon l/2)$. Where l is the path length (1.2 cm) and the extinction coefficient (ϵ) is 192,100 M^{-1} cm^{-1}. The extinction coefficient at 215 nm was found experimentally by performing serial dilutions of a known concentration βdimer1, based on absorbance at 280 nm, and measuring the absorbance at 215 nm. Using this information the dimer association constant of βdimer1 (K_a) is found to be 1.04×10^6M. The dimer dissociation constant is 0.96 μM.

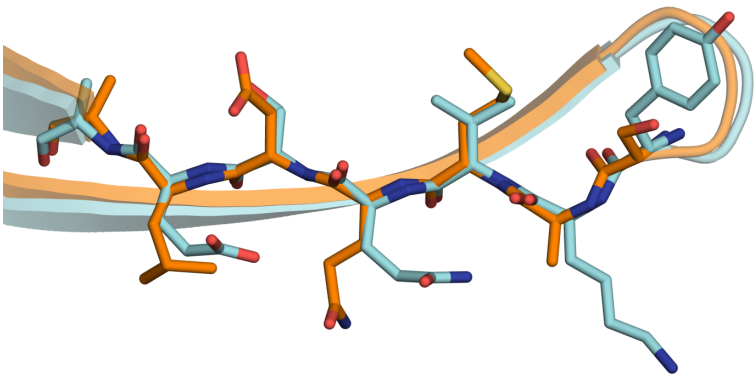

Figure S2.3: Comparison of interacting β-strands in βdimer1 chain A (cyan) to the wild-type structure (2A7B, orange). Residues between S103 (Y103 in βdimer1) to A109 are shown in sticks. Only small perturbations in the backbone are observed. The backbone atom rmsd between βdimer1 chain A and the wild-type structure is 0.3 Å for this range.

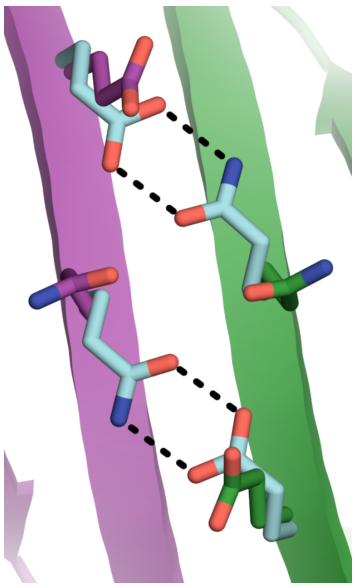

Figure S2.4: The crystal structure reveals side-chain interactions between Q106 and E108 on the solvent accessible side of the interacting β-strands in βdimer1 that were not modeled in the design. The computational model (purple and green) predicts no interaction between these side chains, whereas the crystal structure (cyan) indicates a head-on pairing of Q106 and E108. Black dashed lines represent hydrogen-bond interactions. It is likely that E108 is protonated (βdimer1 crystals were grown at pH 5.0). A protonated variant of glutamate was not considered in the computational design protocol.

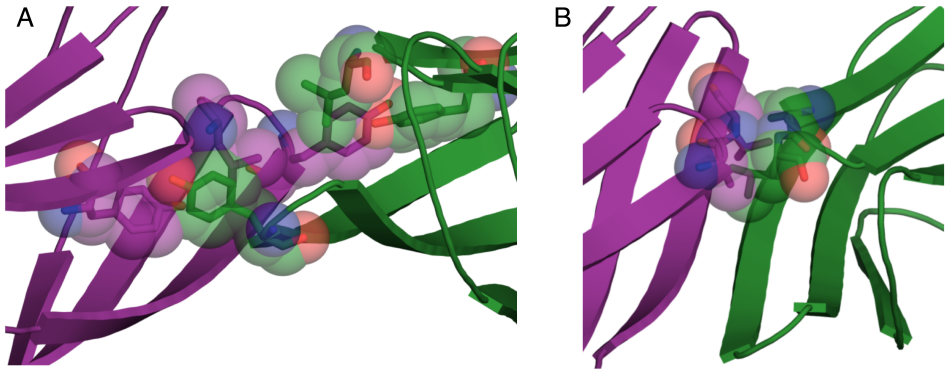

Figure S2.5: Clashes prevent register shift of βdimer1. (A) A register shift of βdimer1 to make fewer backbone-backbone contacts is prevented by the introduction of clashes of Y8 and L11 on one chain with Y103 on the matching chain. (B) A register shift in the opposite direction would be disfavored by clashes of L11 on one strand with L11 on the symmetric strand. These assemblies were made by manually moving one of the chains to the next register of backbone-backbone hydrogen-bond contacts.

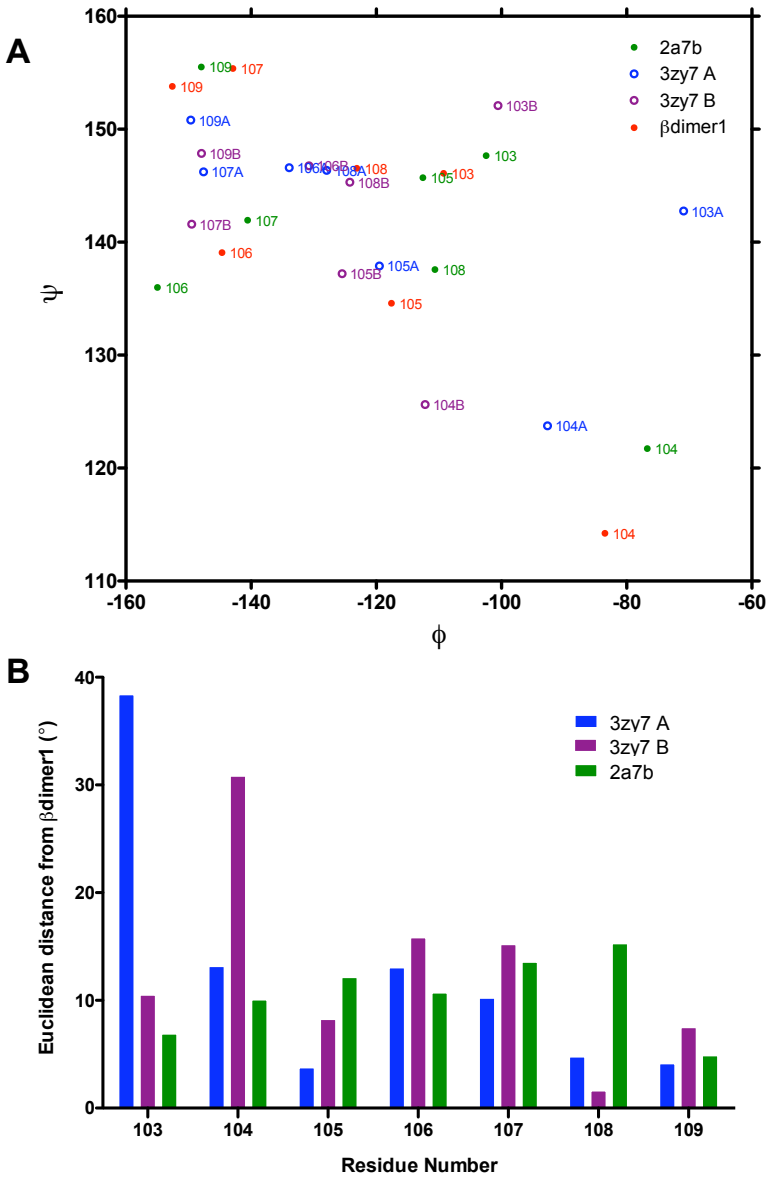

Figure S2.6: Backbone torsion angles of of interacting strands (residues 103-109) in βdimer1 design (red) the crystal structure 3a7b (chain A blue, chain B purple) and wild type crystal structure 2a7b (green). A) Ramachandran plot of residues 103-109, individual points labeled with residue number. B) Euclidian distance in $^\circ$ between the designed model for βdimer1 and the crystal structures of the wild type (2a7b) and two both chains of crystal structure of the design (3zy7 A, B).

	βdimer1
Data collection	
Space group	$P12_11$
Cell dimensions	
a, b, c, Å	50.64, 44.25, 53.02
α, β , γ, $^\circ$	90.00, 91.91, 90.00
Resolution, Å	23.14-1.09 (1.10-1.09)*
R_{merge}	0.053 (0.331)
I/σ	33.1 (1.9)
Completeness, %	92.6 (44.6)
Redundancy	3.6 (1.8)
Refinement	
Resolution, Å	23.14-1.09
No. reflections	90,300
R_{work}/R_{free}	0.159 / 0.181
No. atoms	
Protein	2,087
Ligand/ion	22
Water	314
B factors, Å^2	
Protein	14.2
Ligand/ion	23.5
Water	24.5
rms deviations	
Bond lengths, Å	0.011
Bond angles, $^\circ$	1.421
Ramachandran statistics	
Most favored, %/no.	98.2 / 220
Additionally allowed, % /no.	1.8 / 4
Generally allowed, % / no.	0 / 0
Disallowed, %/no.	0 / 0

Table S2.1: Data collection and refinement statistics. *Values in parentheses are for highest-resolution shell. X-ray coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID: 3ZY7)

References

1. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A., and Kuhlman, B. (2011) Computational design of a symmetric homodimer using β-strand assembly. Proceedings of the National Academy of Sciences of the United States of America **108**, 20562–7

2. Karanicolas, J. and Kuhlman, B. (2009) Computational design of affinity and specificity at protein-protein interfaces. Current opinion in structural biology **19**, 458–63

3. Mandell, D. J. and Kortemme, T. (2009) Computer-aided design of functional protein interactions. Nature chemical biology **5**, 797–807

4. Havranek, J. J. and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. Nature structural biology **10**, 45–52

5. Grigoryan, G., Kim, Y. H., Acharya, R., Axelrod, K., Jain, R. M., Willis, L., Drndic, M., Kikkawa, J. M., and DeGrado, W. F. (2011) Computational design of virus-like protein assemblies on carbon nanotube surfaces. Science **332**, 1071–1076

6. Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature **458**, 859–864

7. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. Science (New York, N.Y.) **282**, 1462–7

8. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science **332**, 816–821

9. Stewart, M. L., Fire, E., Keating, A. E., and Walensky, L. D. (2010) The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. Nat Chem Biol **6**, 595–601

10. Jha, R. K., Leaver-Fay, A., Yin, S., Wu, Y., Butterfoss, G. L., Szyperski, T., Dokholyan, N. V., and Kuhlman, B. (2010) Computational design of a PAK1 binding protein. Journal Of Molecular Biology **400**, 257–270

11. Sammond, D. W., Bosch, D. E., Butterfoss, G. L., Purbeck, C., Machius, M., Siderovski, D. P., and Kuhlman, B. (2011) Computational design of the sequence and structure of a protein-binding peptide. Journal of the American Chemical Society **133**, 4190–2

12. Reynolds, K. A., Hanes, M. S., Thomson, J. M., Antczak, A. J., Berger, J. M., Bonomo, R. A., Kirsch, J. F., and Handel, T. M. (2008) Computational redesign of the SHV-1 β-Lactamase/β-Lactamase inhibitor protein interface. Journal Of Molecular Biology **382**, 1265–1275

13. Liu, S., Liu, S. Y., Zhu, X. L., Liang, H. H., Cao, A. N., Chang, Z. J., and Lai, L. H. (2007) Nonnatural protein-protein interaction-pair design by key residues grafting. Proceedings of The National Academy of Sciences **104**, 5330–5335

14. Der, B. S. and Kuhlman, B. (2011) Biochemistry. From computational design to a protein that binds. Science **332**, 801–2

15. Guharoy, M. and Chakrabarti, P. (2007) Secondary structure based analysis and classification of biological interfaces: identification of binding motifs in protein-protein interactions. Bioinformatics (Oxford, England) **23**, 1909–18

16. Ni, Y. G., Di Marco, S., Condra, J. H., Peterson, L. B., Wang, W., Wang, F., Pandit, S., Hammond, H. A., Rosa, R., Cummings, R. T., Wood, D. D., Liu, X., Bottomley, M. J., Shen, X., Cubbon, R. M., Wang, S.-p., Johns, D. G., Volpari, C., Hamuro, L., Chin, J., Huang, L., Zhao, J. Z., Vitelli, S., Haytko, P., Wisniewski, D., Mitnaul, L. J., Sparrow, C. P., Hubbard, B., Carfí, A., and Sitlani, A. (2011) A PCSK9-binding antibody that structurally mimics the EGF(A) domain of LDL-receptor reduces LDL cholesterol in vivo. Journal of lipid research **52**, 78–86

17. Gilbreth, R. N., Truong, K., Madu, I., Koide, A., Wojcik, J. B., Li, N.-S., Piccirilli, J. A., Chen, Y., and Koide, S. (2011) Isoform-specific monobody inhibitors of small ubiquitin-related modifiers engineered using structure-guided library design. Proceedings of the National Academy of Sciences **108**, 7751–7756

18. Richardson, J. S. and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proceedings of the National Academy of Sciences of the United States of America **99**, 2754–9

19. Tereshko, V., Uysal, S., Koide, A., Margalef, K., Koide, S., and Kossiakoff, A. A. (2008) Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. Protein Sci **17**, 1175–1187

20. Koide, S. (2009) Engineering of recombinant crystallization chaperones. Curr Opin Struct Biol **19**, 449–457

21. Nauli, S., Kuhlman, B., Le Trong, I., Stenkamp, R. E., Teller, D., and Baker, D. (2002) Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2. Protein Sci **11**, 2924–2931

22. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. Science (New York, N.Y.) **302**, 1364–8

23. Nesloney, C. L. and Kelly, J. W. (1996) Progress towards understanding β-sheet structure. Bioorganic & medicinal chemistry **4**, 739–66

24. Kim, W. K., Henschel, A., Winter, C., and Schroeder, M. (2006) The many faces of protein-protein interactions: A compendium of interface geometry. PLoS Comput Biol **2**, e124

25. Goodsell, D. S. and Olson, A. J. (2000) Structural symmetry and protein function. Annual review of biophysics and biomolecular structure **29**, 105–53

26. Hoskins, J., Lovell, S., and Blundell, T. L. (2006) An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. Protein science : a publication of the Protein Society **15**, 1017–29

27. Loughlin, W. A., Tyndall, J. D. A., Glenn, M. P., and Fairlie, D. P. (2004) β-strand mimetics. Chemical Reviews **104**, 6085–6118

28. Khakshoor, O. and Nowick, J. S. (2008) Artificial β-sheets: chemical models of β-sheets. Current Opinion In Chemical Biology **12**, 722–729

29. Sievers, S. A., Karanicolas, J., Chang, H. W., Zhao, A., Jiang, L., Zirafi, O., Stevens, J. T., Mnch, J., Baker, D., and Eisenberg, D. (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nature **475**, 96–101

30. Khakshoor, O., Demeler, B., and Nowick, J. S. (2007) Macrocyclic β-sheet peptides that mimic protein quaternary structure through intermolecular β-sheet interactions. Journal of the American Chemical Society **129**, 5558–5569

31. Khakshoor, O., Lin, A. J., Korman, T. P., Sawaya, M. R., Tsai, S.-C., Eisenberg, D., and Nowick, J. S. (2010) X-ray Crystallographic Structure of an artificial β-sheet dimer. Journal of the American Chemical Society **132**, 11622–11628

32. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and Andre, I. (2011) Modeling Symmetric Macromolecular Structures in Rosetta3. PLoS One **6**, e20450

33. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y.-E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popović, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods in enzymology **487**, 545–74

34. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**, 2577–2637

35. Das, R., Andre, I., Shen, Y., Wu, Y., Lemak, A., Bansal, S., Arrowsmith, C. H., Szyperski, T., and Baker, D. (2009) Simultaneous prediction of protein folding and docking at high resolution. Proceedings of the National Academy of Sciences, USA **106**, 18978–18983

36. Sheffler, W. and Baker, D. (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein science : a publication of the Protein Society **18**, 229–39

37. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology **276**, 307–326

38. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. Journal of applied crystallography **40**, 658–674

39. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D-Biological Crystallography **67**, 355–367

40. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J.,

Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography **66**, 213–221

41. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography **60**, 2126–2132

42. Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A., and Weiss, M. S. (2005) On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. Acta Crystallogr D Biol Crystallogr **61**, 1263–1272

43. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. Journal Of Molecular Biology **372**, 774–797

44. Hu, X., Wang, H., Ke, H., and Kuhlman, B. (2007) High-resolution design of a protein loop. Proceedings of the National Academy of Sciences of the United States of America **104**, 17668–73

45. Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2004) A dissection of specific and non-specific protein-protein interfaces. Journal Of Molecular Biology **336**, 943–955

46. Joachimiak, L. A., Kortemme, T., Stoddard, B. L., and Baker, D. (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. Journal of molecular biology **361**, 195–208

47. Kortemme, T., Morozov, A. V., and Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. Journal of molecular biology **326**, 1239–59

48. Karanicolas, J., Corn, J., Chen, I., Joachimiak, L., Dym, O., Peck, S., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., T.áMontelione, G., P.áSpiegel, C., Liu, D., and Baker, D. (2011) A de novo protein binding pair by computational design and directed evolution. Molecular Cell **42**, 250–260

49. Ben-Shem, A., Frolow, F., and Nelson, N. (2004) Evolution of photosystem I - from symmetry through pseudosymmetry to asymmetry. Febs Letters **564**, 274–280

Chapter 3

Computational design of a β-strand mediated heterodimer

3.1 Introduction

The design of symmetric homodimers (Chapter 2) provides valuable evidence that a particular method of interface design can produce successful results (1; 2). However, the ultimate goal of protein-protein interaction design is to engineer binders to natural proteins to alter function. A fantastic example is the design of two proteins that bind influenza hemagglutinin and prevent viral entry into cells (3). Additional approaches to interface design are needed to construct proteins that can bind to the diversity of protein conformations observed in nature (4). There are no known computational designs of a heterodimer based on β-strand pairing between the two chains. A method to target exposed β-strands could provide binding partners to disrupt an array of natural β-strand mediated interactions(Figure 1.2).

We choose two proteins as targets for heterodimer design based on medical relevance and presence of X-ray crystal structures showing these protein interact with their natural partners via a β-strand. The first protein that we chose as a target for *de novo* interface design was Ras-like GTPase A (RalA). RalA is involved in the regulation of many cellular processes including cellular junction formation, endocytosis and the exocyst complex (5) and is necessary for Ras induced oncogenesis (6; 7). RalA interacts with at least two of its natural downstream effectors Sec5 and Exo84 via formation of an intermolecular β-sheet. Crystal structures of RalA in complex with the Sec5 Ral binding domain (Sec5) (8) and Exo84 Ral binding domain (Exo84)

(9) demonstrate that the same β-strand of RalA interacts with both effectors (Figure 3.1). These interfaces, like many other Ras-effector interactions, are further stabilized by hydrogen bonds and salt bridges while hydrophobic interactions are minimal (10; 11). Ras-effector interactions have been shown to be amenable to design. Mutating residues at a Ras-effector interface can alter the specificity of an interaction (12), improve affinity(13), or yield high affinity interaction to the GDP bound state(14). A protein designed to bind to RalA that is not involved in the natural signaling pathway could be a tool to further understand its function and elucidate the many downstream effects of RalA activity.

The second protein chosen as a target was Proprotein convertase subtilisin-like kexin type 9 (PCSK9). PCSK9 is a secreted protein that regulates the presence of Low Density Lipoprotein receptor (LDLr) on cell surfaces. The interaction between PCSK9 and LDLr leads to degradation of LDLr thus preventing LDLr from regulating LDL levels in the blood (15). Levels of active PCSK9 have an inverse relationship with the presence of LDLr on cell surfaces and an direct relationship with the amount of LDL in the blood (16; 17). Humans with a loss of function mutation in PCSK9 have lower LDL levels and reduced risk of heart disease (18) while gain of function mutations can lead to hypercholesterolemia (19). PCSK9 binds to EGF hand motif on LDLr, via an intermolecular β-sheet (Figure 3.2A) (17). An antibody obtained by phage display interacts with PCSK9 at the same place as LDL (Figure 3.2B). When injected into monkeys this antibody lowered blood LDL levels by as much as 50% over two weeks (20). A small protein that binds to PCSK9 in the same manner as the antibody could serve as a therapeutic protein based drug for high LDL levels.

Yeast surface display is one of the most widely used systems for *in vitro* evolution and protein engineering (21). It has been used to identify protein-protein interactions, increase solubility of a protein, increase affinity of an interaction (reviewed in (22)) and evolve bond forming enzymes (23). Yeast surface display has been shown to be an efficient method to evolve a library of protein sequences to bind to a target of interest through successive rounds of FACS and random mutagenesis (24). It is also possible to measure the affinity of individual clones (21). Recently, Fleishman *et al* used yeast surface display to identify and mature the affinity of two computational protein interface designs to a target (3). Yeast surface display was also

used to improve the stability and binding affinity of a computationally designed protein that displays a segment of HIV gp120 (25). These results suggest that yeast display is an efficient method to characterize and improve β-strand mediated heterodimer designs.

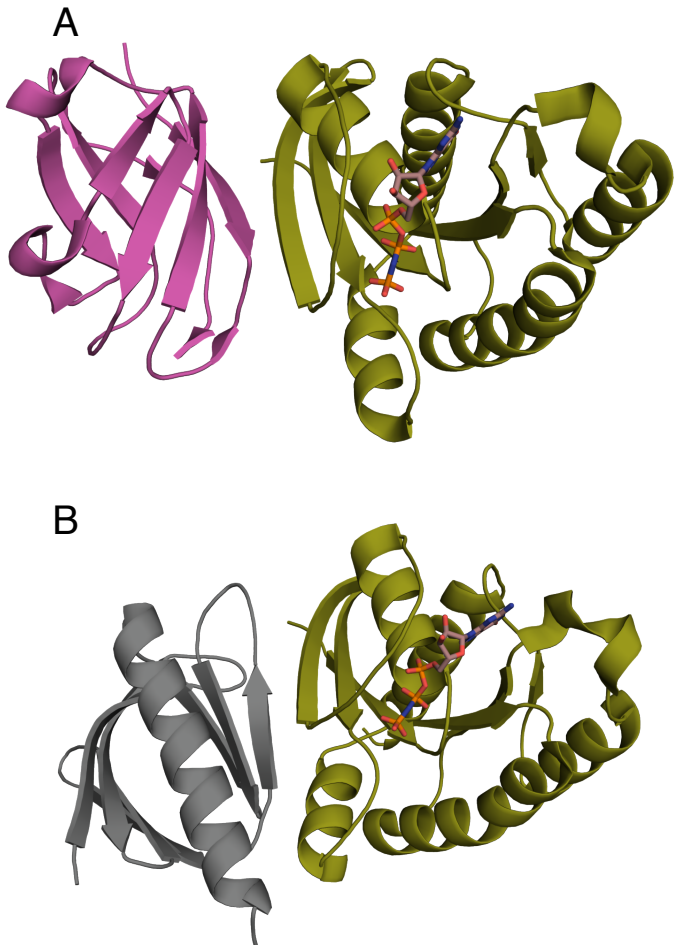

Figure 3.1: RalA forms β-strand mediated interactions to two known effectors. Crystal structure GTP bound RalA (olive) in complex with A) Sec5 (magenta) (PDB ID: 1UAD) (8) and B) Exo84 (grey) (PDB ID: 1ZC3) (9)

This chapter discribes one possible approach to finding protein scaffolds capable of binding to a target protein with an exposed β-strand and designing the scaffolds to interact with a target protein. First we identify a possible backbone conformation that could bind to a target protein then search for proteins with an exposed β-strand that matches the desired conformation. The identified scaffold proteins are then docked against the target and redesigned with Rosetta. Although initial experimental results are not encouraging, the protein scaffolds

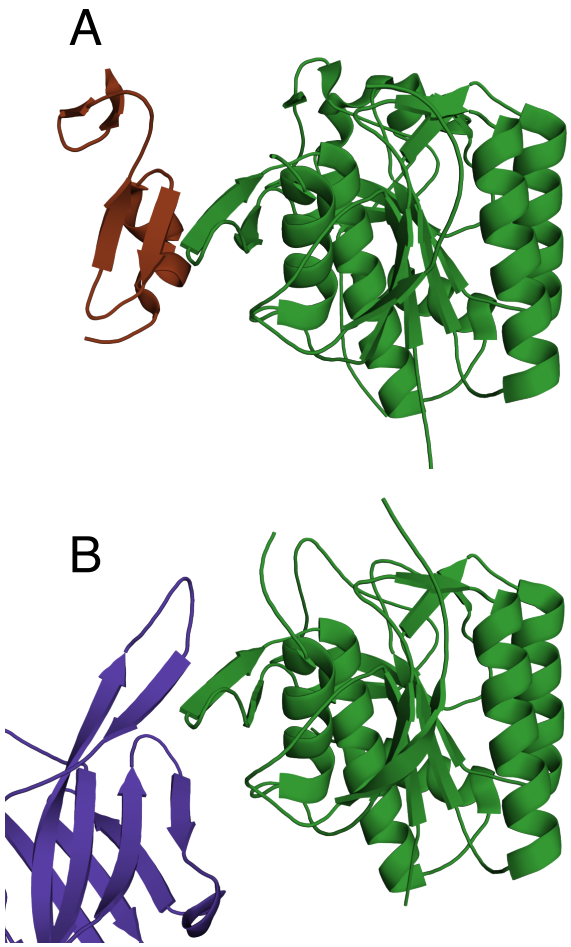

Figure 3.2: PCSK9 forms β-strand mediated interactions to two known binders. A) Crystal structure of PCSK9 (green) in complex with a domain from its natural effector LDLr (brown) (PDB ID: 3BPS) (17). B) Crystal structure of an antibody (blue) evolved to bind to and inhibit PCSK9 activity in complex with PCSK9 (green) (PDB ID: 2XTJ) (20).

found using this protocol are excellent candidates for yeast surface display followed by directed evolution to obtain a β-strand mediated heterodimer.

3.2 Materials and Methods

3.2.1 Idealized binding strand construction

To construct an idealized binding strand to RalA and PCSK9 a ten residue all alanine strand with β-strand backbone torsion angles was built along the targeted exposed β-strand of

both proteins using PyMol. The conformation of RalA was taken from the crystal structure of RalA in complex with Sec5 (PDB ID: 1UAD) (8). Two different crystal structures were used for PCSK9, one from the complex of PCSK9 with LDLr (17) and the other from a complex with an engineered antibody (20) (PDB IDs: 3BPS and 2XTJ respectively). Using Rosetta, the peptide was docked against its target protein. Conformational space of the peptide was sampled by making 200 small alterations to the backbone ϕ and ψ angles followed by energy minimization. The backbone-backbone hydrogen bond energy was upweighted in the energy function to favor strand-strand pairing. Residues at the interface on the target protein are allowed to repack to accommodate the peptide if necessary. A sample command line is as follows:

```
./peptide_minimizer.<exe>
     -s input_struct.pdb
     -database <rosetta_database>
     -ncycles 200
     -bbhbond_weight_mod 3.0
     -nstruct 100
     -jd2:ntrials 10
     -run::min_type dfpmin_armijo_nonmonotone
     -ex1
     -ex2
     -extrachi_cutoff 1
     -use_input_sc true
```

All of the above command line flags are standard for Rosetta except for `ncycles` which controls the number of backbone perturbation steps and `bbhbond_weight_mod` which upweights the backbone-backbone energy by the factor given.

3.2.2 Scaffold search procedure

Suitable scaffolds for design were found by searching a set of 5,200 high resolution ($<$ 2.0 Å) crystal structures compiled by the Richardson laboratory at Duke University. First, a potential scaffold examined for surface exposed β-strands based on three criteria (*i*) five sequential residues had β-strand secondary structure as judged by the Database of Secondary Structure of Proteins algorithm (26); (*ii*) there were no backbone-backbone hydrogen bonds formed by every other residue in the strand; and (*iii*) every other residue had fewer than

16 neighboring residues, or had 16-30 neighbors and a SASA per atom greater than 2.0 Å^2. Residues are defined as neighbors if their Cβ to Cβ distance is < 10 Å. A strand that met all these criteria was then aligned for lowest RMSD to the peptide built onto the exposed strand of RalA or PCSK9. A scaffold is accepted as a suitable starting conformation for further design if it has an RMSD to the ideal peptide of < 1.5 Å and no backbone clashes with the target protein. A backbone clash is defined a Lennard-Jones energy greater than 0.0 for cross-chain backbone atoms. A command line to perform all of these steps is given below:

```
./exposed_strand_finder.<exe>
     -database <rosetta_database>
     -l list_of_scaffolds
     -native  Ral_peptide.pdb
     -strand_span B 6 10
     -out:nooutput true
     -check_rmsd true
     -beta_length 5
     -max_E_allow 0.0
     -max_RMSD 1.5
```

There are several important command line flags here. First `native` is the target protein (RalA in the case above) with the peptide built onto it, the position on that peptide to search and align against is given by `strand_span` (in this case chain B between residues 6 and 10 in the structure. Other command line options tell the protocol to do the alignment of an exposed strand to a peptide strand (`check_rmsd`), define the length of exposed β-strand to look for (`beta_length`), give the maximum energy allowed for the backbone clash between the two chains (`max_E_allow`), and the maximum RMSD between an exposed strand on a scaffold and the idealized peptide (`max_RMSD`)

3.2.3 Interface design of selected scaffolds

All design steps were carried out using the `RosettaScripts` framework(27). An example `RosettaScripts` input file and command line to design a binder to PCSK9 are given in Section 3.5.1 on page 70. Several different design protocols were used to obtain design models depending on the scaffold and target protein. In general the steps are similar to those outlined in Section 2.2.2 on page 26. The two proteins to design are moved apart, repacked, then

docked back together to sample rigid body degrees of freedom. Once docking has been completed residues at the interface on the scaffold protein are designed while the interface residues on the target are allowed to change rotamer but not amino acid identity. After the design step, the backbone and side-chains at the interface are minimized on both proteins. Different numbers of iterations are performed depending on the target and scaffold. These steps are similar to those first used to design a protein to bind to PAK1 (28).

3.2.4 Protein expression and purification

The gene coding for RalA was obtained in the vector pGEX4T-1 and expressed as a GST fusion protein as described by Fukai *et al* (8). RalA was cleaved from GST using thrombin (Sigma) and purified to $> 95\%$ as judged by Commassie blue staining. The GDP bound to RalA after purification was exchanged for the nonhydrolyzable GTP analog Guanosine 5-[β,γ-imido] triphosphate (GppNHp) (Sigma) as described by Bauer *et al* (29). RalA-GppNHp was then dialyzed in 1x PBS pH 7.4 to remove the excess guanine nucleotide. Genes coding for the designed RalA binders, 1JL1aD1, 1JL1pD1, and 1SC0aD1, were ordered from Genscript and cloned into the pET21-b expression vector. All were purified on a Ni-affinity column followed by size exclusion chromatography (Superdex 75, Amersham). Commassie blue staining revealed that all proteins were at 95 % purity after these steps. The designs 1JL1aD1 and 1JL1pD1 eluted from the SEC column at the expected weight of a monomer, 1SC0aD1 had the expected molecular weight of a dimer. Although the protein 1SC0 appears as a monomer in the crystal structure, several homologous proteins are homodimers that use the exposed β-strand targeted for design to form part of their homodimeric interface. In retrospect it is likely that 1SC0 is a homodimer and should not have been used in this experiment (Figure 5.1B on page 119).

HEK 293 cells expressing a FLAG-tagged PCSK9 were obtained from the Horton lab at UT Southwestern. PCSK9 was expressed and purified as perviously described (16). Briefly, FLAG-tagged PCSK9 was pulled down from cellular supernatant using anti-FLAG beads (Sigma). PCSK9 was eluted from the beads using FLAG peptide. Excess FLAG peptide was removed by running the sample over a Superdex 200 SEC column. Commassie blue staining confirmed that

the single peak from the size exclusion column was comprised of 63 kDa and 10 kDa proteins which correspond to the expected sizes for the PCSK9 prodomain and catalytic domain (16).

3.2.5 Yeast expression of designs

Genes coding for the EGF-A and EGF-B domains of LDLr (residues 292-373) (17), 1N3Yd1, and 2V14d1 were ordered from GenScript and cloned into the pCTCON2b vector between the NheI and BamHI sites. The resulting vectors were transformed into EBY100 yeast cells. The two designs and LDLr were expressed using protocols and reagents described by Chao *et al* unless otherwise noted (24). Transformed yeast were grown overnight at 30 °C in 1 mL of SD-CAA media. In the morning the cells were spun down and resuspended in 200 μL of SGCAA media; 40 μL of the resuspension was added to 960 μL of SGCAA media and grown overnight at 20 °C. Protein binding and expression was checked by flow cytometry (FACSCanto, BD Biosciences). Protein expression was monitored by labeling the cells with Chicken anti-c-myc IgY (Invitrogen, catalog no. A21281) followed by Alexa Fluor 633 goat anti-chicken IgG (H+L) (Invitrogen, catalog no. A21103). Presence of FLAG tagged PCSK9 was monitored using anti-DYKDDDDK [Fitc] (GenScript catalog no. A01632).

3.2.6 ITC binding measurements

The affinity of RalA for Sec5 and the designed binders was measure by isothermal titration calorimetry (MicroCal VP-ITC). The buffer used was PBS supplemented with 5 mM $MgCl_2$. To measure the affinity of the RalA/Sec5 interaction, RalA was placed in the sample cell at 10 μM and titrated with 150 μM Sec5 in 2 μL titrations. The affinity of the designs was measured by titrating 600 μM Sec5 in 5 μL injections into the sample cell containing 60 μM RalA.

3.3 Results

3.3.1 Idealized interaction strand model

Ten residue alanine peptides were minimized to pair with the exposed β-strands on RalA and PCSK9 as described in Methods (Section 3.2.1). The structures output from in minimization protocol were chosen based on the number of backbone-backbone hydrogen bonds formed to the target and the lowest average energy per hydrogen bond. Many of the lowest energy peptides had similar backbone conformations (Figure 3.3). These peptides were later truncated to contain just the residues involved in forming backbone-backbone hydrogen bonds to the target.

3.3.2 Scaffold search protocol

The idealized stands found in the previous step were used to search for proteins with exposed β-strands that exhibit a similar backbone conformation (Figure 3.4 red boxes). Once an exposed strand on a scaffold protein is found (Figure 3.4 blue boxes) it is aligned to the idealized strand. If the backbone of the scaffold and peptide are similar (RMSD $<$ 1.5 Å) and the backbones of the scaffold and target do not clash, the structure is output for design consideration (Figure 3.4 green boxes). Of the 5,200 protein chains searched against, about 50 produced suitable matches to either RalA or PCSK9. This is a similar result to that found when scaffold proteins were screened to determine if they could form a potential β-strand mediated homodimer (Figure 2.1).

3.3.3 Selection of interface designs

The resulting structures from the scaffold search protocol (Figure 3.4) were used as input conformations for interface design. Typical interface design runs (Section 3.2.3) produced 50,000 output structures. We selected designs for expression based on presence of intermolecular β-strand mediated hydrogen bonds, favorable binding energy density ($\Delta G_{bind}/\Delta$SASA), lack of unsatisfied polar atoms at the interface, and visual inspection (Table 3.1).

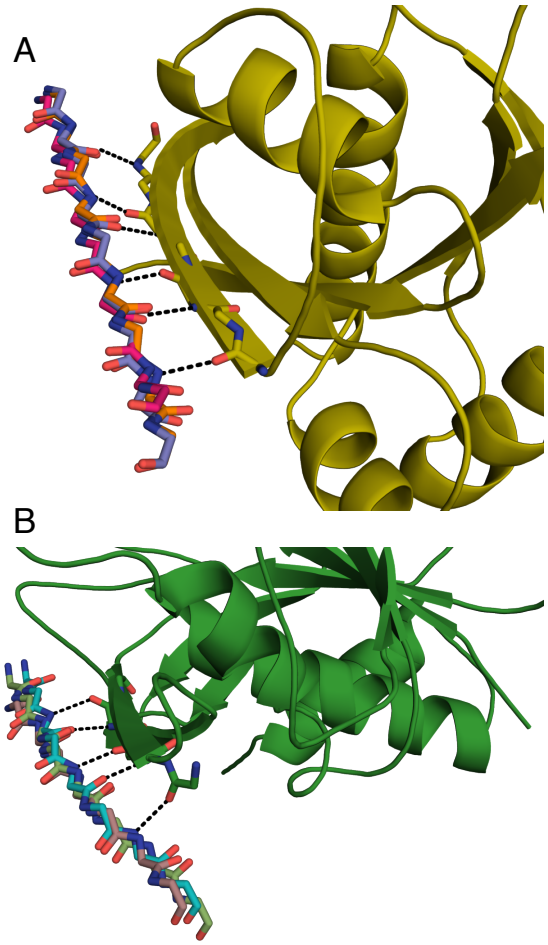

Figure 3.3: Idealized backbone conformation of three strands pairing with target proteins. Alanine peptides were minimized with Rosetta to bind to the exposed β-strands on A) RalA (olive) and B) PCSK9 (green). Black dashed lines represent main chain hydrogen bonds.

We choose three designs to test for binding to RalA (Figure 3.5). Ribonuclease H1 (PDB ID: 1JL1) (30) was designed to form a parallel (1JL1pD1; Figure 3.5B) and antiparallel (1JL1aD1; Figure 3.5A) β-strand paring with RalA. The other design forms an antiparallel intermolecular β-sheet using a scaffold protein of unknown function (PDB ID: 1SC0). This design was named 1SC0aD1 (Figure 3.5B).

Two potential binders to PCSK9 were selected based on redesigns of human α-X β2 integrin I domain (PDB ID: 1N3Y) (31) and KIF16B (PDB ID: 2V14) (32). The designs based on these domains were named 1N3Yd1 and 2V14d respectively. Each design forms an antiparallel intermolecular β-sheet with PCSK9 that includes three backbone-backbone hydrogen bonds.

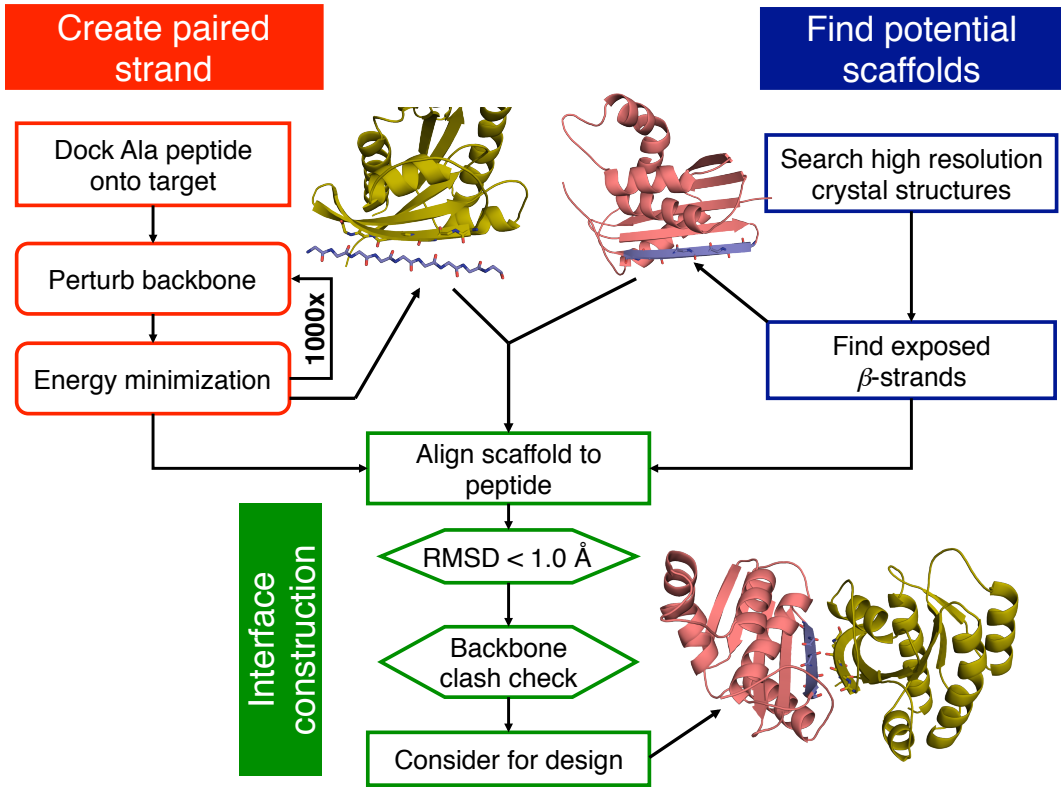

Figure 3.4: Search protocol used to find potential scaffolds to form a novel β-strand mediated interface with a target protein. Full details are outlined in Materials and Methods (Section 3.2.2). Steps shown here for matching a peptide (blue) modeled to interact with RalA (olive) to a scaffold (PDB ID: 1JL1) (pink) with an exposed strand (blue). The peptide in RalA and exposed strand on the scaffold are aligned to generate a potential starting structure for interface design. This protocol was also used to find potential partners for PCSK9.

1N3Yd1 forms an additional side-chain to backbone hydrogen bond to PCSK9 and uses an arginine to pack against PCSK9 residue F397 (Figure 3.6A), which has been shown to be critical for binding to LDLr (17). The design 2V14d1 forms more hydrophobic contacts with PCSK9 (Table 3.1) and packs against PCSK9 F297 with a proline and the main chain of an α-helix (Figure 3.6B). Only two designs were selected in order to optimize the yeast display expression system. There are several other scaffold proteins can be used if 1N3Yd1 and 2V14d1 fail to produce binders from a library.

Target	Design	ΔSASA (Å^2)	ΔG_{bind}/ΔSASA	No. buried-unsatisfied	Polar interface area (%)
RalA	Sec5 (native)	904	-0.028	1	54
RalA	1JL1aD1	1328	-0.023	0	49
RalA	1JL1pD1	1052	-0.027	2	51
RalA	1SC0aD1	1214	-0.018	2	63
PCSK9	LDLr (native)	975	-0.021	1	47
PCSK9	1N3Yd1	1087	-0.021	0	51
PCSK9	2V14d1	1038	-0.020	0	45

Table 3.1: Computational evaluation of heterodimer design models. Native interactions of RalA to Sec5 and PCSK9 to LDLr are given for comparison. ΔSASA represents the change in protein surface area upon binding, ΔG_{bind}/ΔSASA is the energy density at the interface ($\Delta G_{bind} = E_{AB} - E_A - E_B$), No. buried-unsatisfied is the number of polar atoms that are not solvent accessible and do not form hydrogen bonds to another atom in the protein, and Polar interface area, % represents the amount of solvent-accessible surface area of polar atoms hidden at the interface ($SASA_{polar}/SASA_{total}$).

3.3.4 Affinity measurements of designed binders to RalA

The binding affinity of the designs for RalA-GppNHp was measured by ITC, Sec5 was used as a positive control. Titrating Sec5 into RalA showed an exothermic interaction (Figure 3.7A). A fit of the curve revealed that Sec5 bound to RalA with $K_D = 340$ nM with a molar ratio of $n = 0.55$. These values are close to those reported previously; $K_D = 140$ nM, $n \approx 1.0$ (8). The discrepancy in the molar ratio could be caused by incomplete exchange of GDP for GppNHP in RalA. Titration of the designs into RalA produced endothermic peaks for all designs and no discernible binding interaction (Figure 3.7B, C, D). RalA is able to bind a natural effector, however our attempt to design a novel binder was unsuccessful.

3.3.5 Yeast display of proteins to bind PCSK9

Labeling of yeast cells with anti-c-myc antibody indicates that LDLr, 1N3Yd1, 2V14d1 can be displayed on yeast surface. This indicates that yeast surface display can be used as a system to evolve binding between 1N3Yd1, 2V14d1 and PCSK9. The next step was to test the designs and LDLr to determine if they bind to purified PCSK9. The binding affinity of PCSK9 to LDLr was measured as described previously (21). PCSK9 was labeled with anti-FLAG antibody.

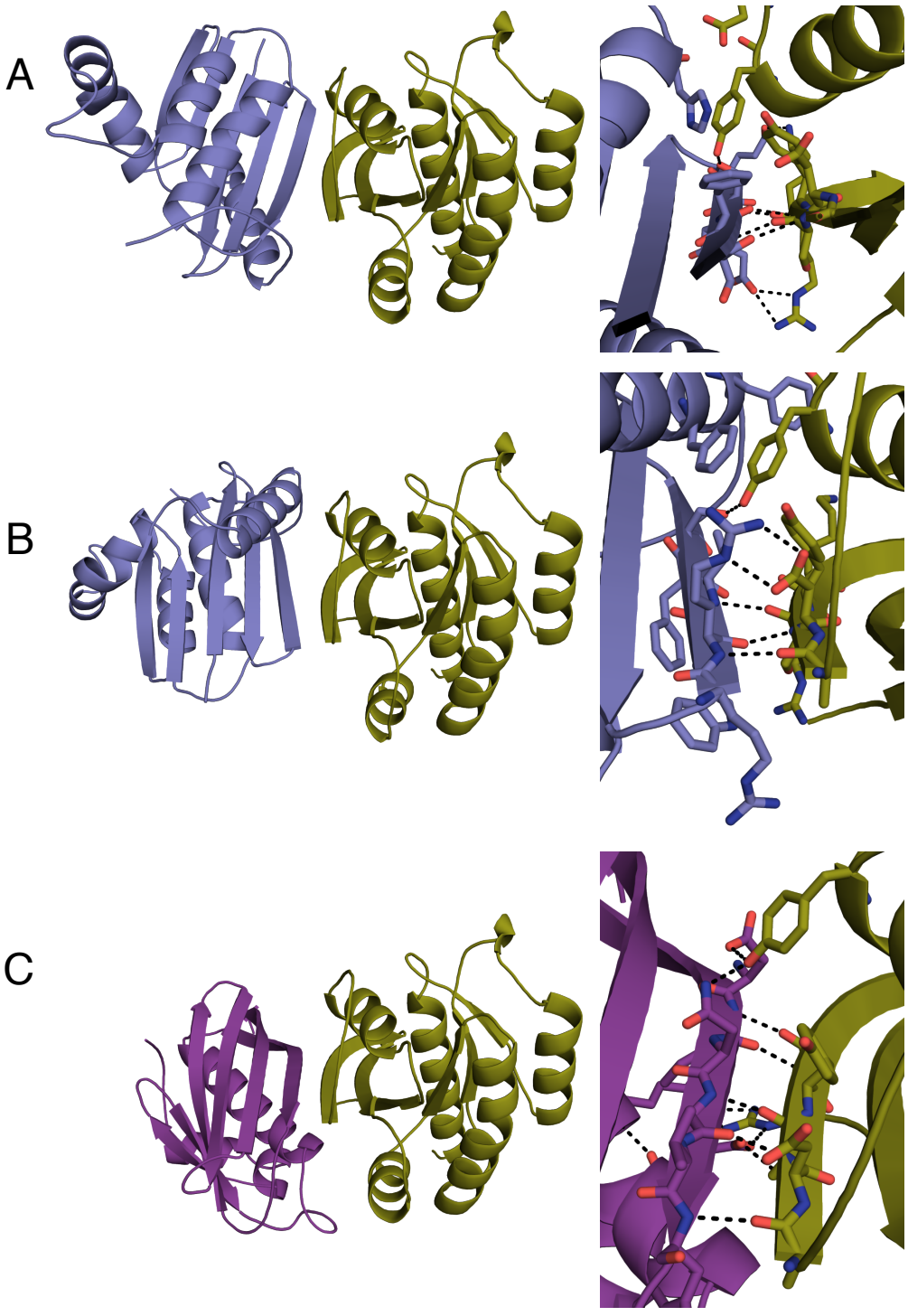

Figure 3.5: Proteins redesigned to bind to RalA (olive). The whole modeled structure of the complex formed by both chains is shown on the left with a more detailed look at interactions at the interface is shown on the right. The models are 1JL1aD1 (A, blue), 1JL1pD1 (B, blue) and 1SC0aD1 (C, purple). Black dashed lines represent hydrogen bonds across the interface.

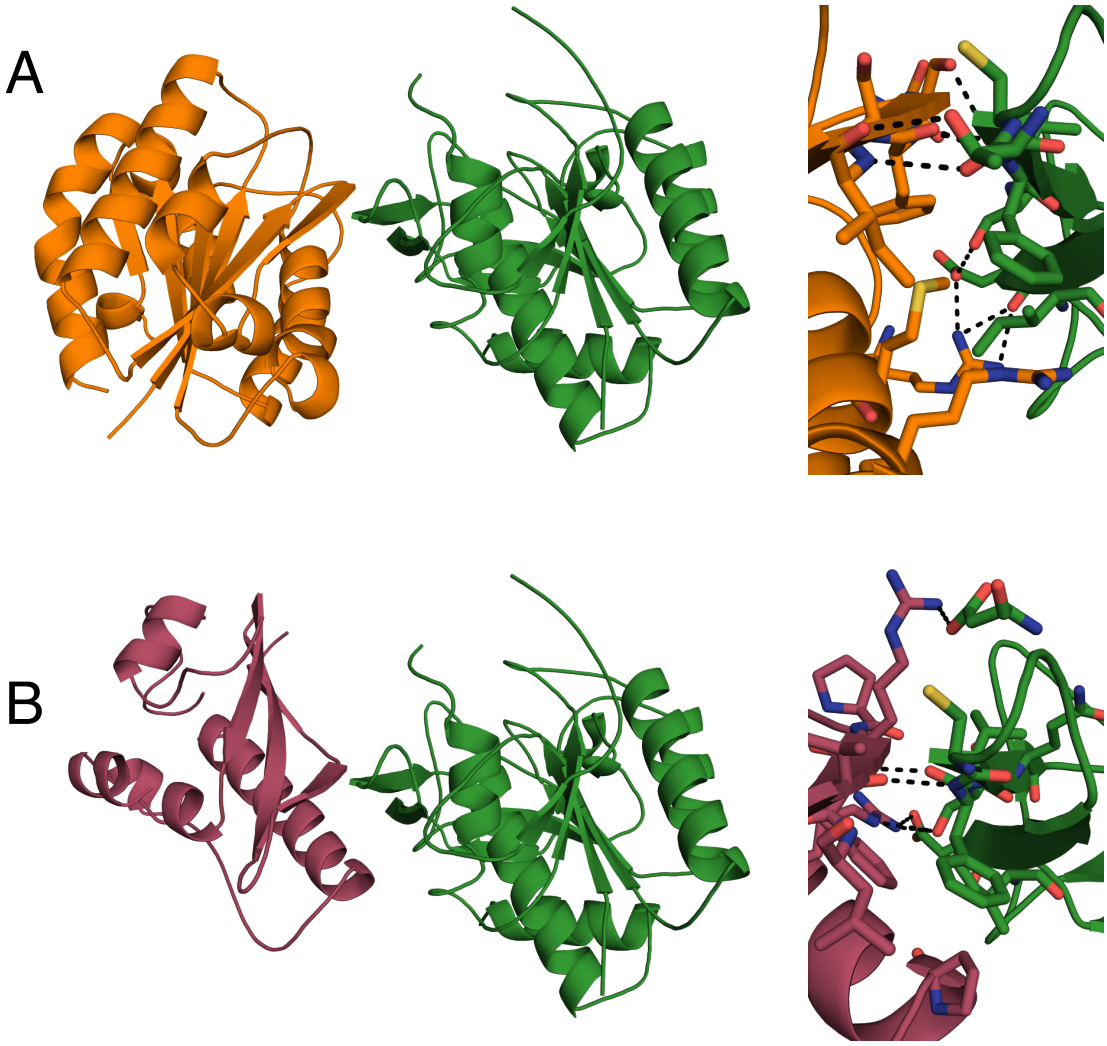

Figure 3.6: Proteins redesigned to bind to PCSK9 (green). The whole modeled structure of the complex formed by both chains is shown on the left with a more detailed look at interactions at the interface is shown on the right. The models are 1N3Yd1 (A, orange), 2V14d1 (B, maroon). Black dashed lines represent hydrogen bonds across the interface.

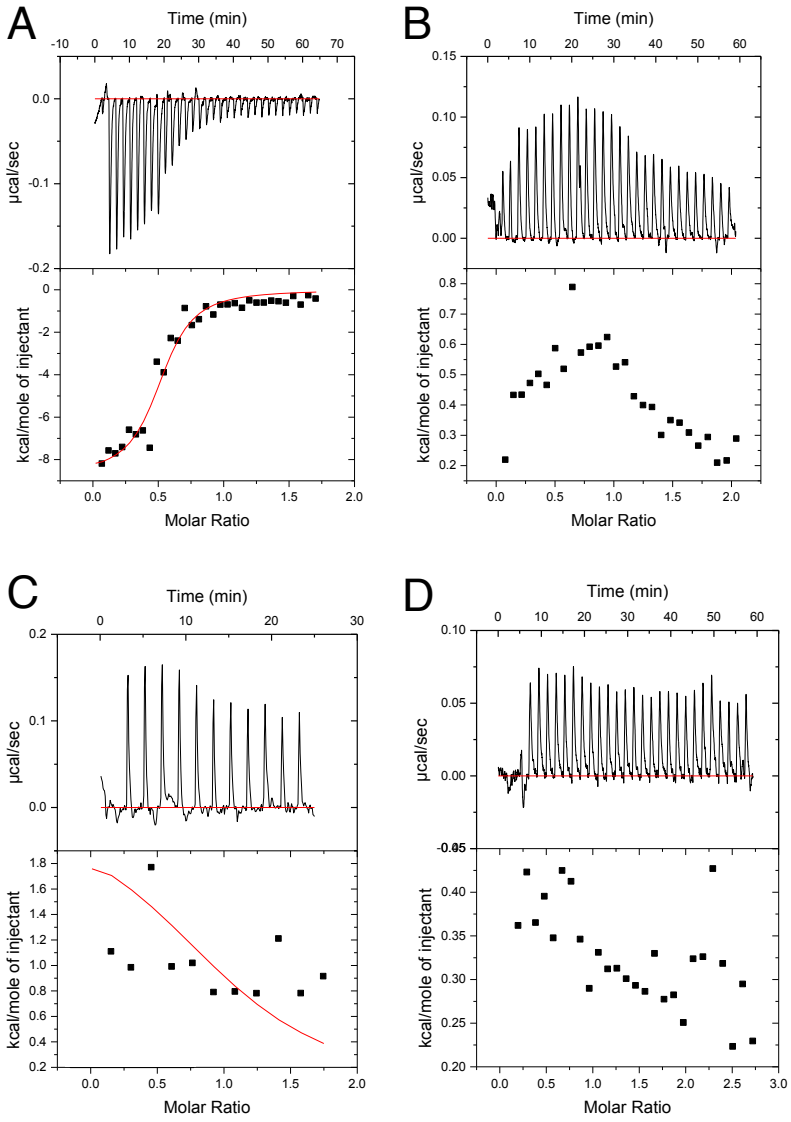

Figure 3.7: Isothermal calorimetric titration of RalA with the Sec5 and the designed scaffold proteins. Top pannel shows the raw data for each injection. Bottom panel shows the molar ratio of titrant protein to RalA vs integrated heat of addition from top pannel. A) Native interaction between RalA-GppNHp titration with Sec5 shows binding with an affinity of 340 nM with a molar ratio of 0.55. There is no detectable binding between RalA-GppNHp and the designed proteins B) 1JL1aD1, C) 1JL1pD1, and D) 1SC0aD1.

3.4 Discussion

The computational workflow described here represents a general way to find potential binders capable of forming an intermolecular β-sheet with a target protein. A peptide representing ideal β-strand pairing could be build along the exposed strand of any target protein. The search method will then find exposed strands on scaffold proteins that are a close match to the idealized strand. Thus far this protocol has only been applied to target proteins that bind to a natural effector protein along the exposed β-strand. Ribonuclease H (PDB ID: 1JL1) would be an interesting target protein because it has a long exposed β-strand and no know natural effectors that bind via this strand.

The computational results are promising. The designs chosen for expression all made at least three main chain hydrogen bonds to their target. The interface design method produced interfaces with similar energies and composition to natural binders to the target proteins 3.1. Design of four other scaffolds also yielded designs with similar interface energy. However, the experimental results have been disappointing. None of the three proteins designed to bind RalA had any measurable affinity for their target. The two designed binders to PCSK9 also showed no discernible interaction even though yeast display is capable of detecting low affinities (3). The computational models suggest that the backbone conformation of the scaffold should be able to form an intermolecular β-sheet with the target protein. An alternative method of sampling sequence space could produce proteins that are able to bind.

3.4.1 Future direction: evolving binders to PCSK9

Even though there was no detectable binding of PCSK9 to the designs, there is the potential to use directed evolution to develop binders to PCSK9. The rate of success in computational protein-protein interface design is about 1 in 20 (Chapter 4 and Table 4.1), which makes it prudent to screen a large number of interface designs. Yeast surface display provides the ability to screen large libraries of sequences (up to 10^9) and select clones that bind to the target with FACS (24). The designs were expressed as yeast surface fusions to test their viability in this system. The potential pairing between the exposed β-strand on PCSK9 and the scaffold

should serve as an anchor for the interaction and allow the incorporation of additional sequence diversity on the scaffold protein at the targeted interface. Yeast surface display was used to increase the affinity of a weakly binding computational design (3; 33). Directed evolution of proteins with exposed β-strands can enable them to interact with amyloid fibrils (34), which suggests β-strand pairing can be obtained through directed evolution

3.4.2 Computational library generation

The first step in directed evolution is the generation of a library. Rationally designed libraries that increase the frequency of useful variants are necessary as sequence space can become too large as more sites are included in the library (35). Computationally designed libraries can limit search space to amino acids that are reasonable for a protein's conformation and have been shown to produce the desired outcome better than random libraries or error prone PCR (36; 37; 38). A possible approach to computational library design is to generate an ensemble of backbone conformations in the scaffold protein followed by designing all positions at the predicted interface with PCSK9. The residues that Rosetta prefers at each position would then included in a library. Backbone flexibility is required to recapture conformation changes upon mutation (39). Degeneracy in Ramachandran space for β-strands (Figure S2.6) suggests backbone modeling will not significantly alter the ability of the β-strand to pair with PCSK9. Since sequences would only be allowed to vary near the predicted interface and β-strand paring would be available to orient the interface, it is unlikely off target binding will appear. This even could be checked for by testing the ability of LDLr to bind to PCSK9 in the presence of the designed protein. This method should provide sufficient sequence diversity without sacrificing the stability of the scaffold proteins.

Design of a novel β-strand mediated heterodimer is remains an unsolved problem. Computational results demonstrate that it is possible to find proteins with exposed stands capable of forming a heterodimer with at target, however design has failed to produce a sequence capable of interacting with the target. Alternate methods such as searching for exposed strands that match the conformation of a known interacting strand could provide additional scaffolds for

design. Another possibility is to test a greater repertoire of designs by using yeast display rapidly screen designs and evolve new binders.

3.5 Supporting Information

3.5.1 Design method for targeting PCSK9

The following is the `RosettaScripts` method used to design scaffold proteins to bind to PCSK9

```
rosetta3/rosetta_source/bin/rosetta_scripts.<EXE>
     -database <ROSETTA_DATABASE>
     -s input_struct.pdb
     -ignore_unrecognized_res
     -parser:protocol script.xml
     -nstruct 9000 -jd2:ntrials 5 -jd2:delete_old_poses
     -out:pdb_gz true
     -score12prime true -no_his_his_pairE  -no_optH false
     -docking:dock_pert 3 5
     -ex1 -ex2 -use_input_sc  -extrachi_cutoff 1
     -linmem_ig 10
     -run::min_type dfpmin_armijo_nonmonotone
     -atomic_burial_cutoff 0.01 -sasa_calculator_probe_radius 1.2
     -overwrite
```

With the following as the input to `script.xml`.

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
          <s12_prime weights="score12prime"/>
        </SCOREFXNS>
        <TASKOPERATIONS>
          <RestrictToInterfaceVector name=vectorTask chain1_num=1
              chain2_num=2 CB_dist_cutoff=10.0 nearby_atom_cutoff=5.5
              vector_angle_cutoff=65.0 vector_dist_cutoff=8.0/>
          <RestrictChainToRepacking name=repack1 chain=1/>
          <RestrictChainToRepacking name=repack2 chain=2/>
          <RestrictToRepacking name=repackonly/>
          <InitializeFromCommandline name=cmdTask/>
          <IncludeCurrent name=currentTask/>
        </TASKOPERATIONS>
        <FILTERS>
          <ScoreType name=score_filter scorefxn=score12
              score_type=total_score threshold=-1/>
          <Ddg name=ddg_filter scorefxn=score12 threshold=-10 jump=1
              repeats=3 repack=1/>
          <Sasa name=sasa_filter threshold=900 jump=1/>
          <BuriedUnsatHbonds name=unsats_filter jump_number=1 cutoff=8/>
          <Rmsd name=rms threshold=5.0 confidence=1/>
        </FILTERS>
        <MOVERS>
          <PackRotamersMover name=initial_pack scorefxn=score12
              task_operations=repackonly,cmdTask,currentTask/>
          <PackRotamersMover name=design_pack scorefxn=s12_prime
              task_operations=vectorTask,repack1,cmdTask,currentTask/>
          <MinPackMover name=design_min_pack scorefxn=score12
              task_operations=vectorTask,repack1/>
```

```
        <Prepack name=prepack scorefxn=s12_prime jump_number=1 min_bb=0
            task_operations=vectorTask,repackonly,cmdTask,currentTask/>
        <DockingProtocol name=dockprotocol docking_local_refine=0
            docking_score_high=score12 ignore_default_docking_task=1
            task_operations=vectorTask,repack1,repack2,cmdTask,currentTask/>
        <TaskAwareMinMover name=minmover scorefxn=s12_prime chi=1 bb=1
            jump=1 task_operations=cmdTask,currentTask/>
        <ParsedProtocol name=designmin mode="sequence">
          Add mover_name=design_min_pack
          <Add mover_name=design_pack/>
          <Add mover_name=minmover/>
        </ParsedProtocol>
        <build_Ala_pose name=ala_pose partner1=0 partner2=1
            interface_cutoff_distance=8.0/>
        <SaveAndRetrieveSidechains name=retrieve_sidechains allsc=0/>
        <IteratedConvergence name=designiterate mover=designmin
            filter=score_filter delta=2.0 cycles=1 maxcycles=10 />
        <InterfaceAnalyzerMover name=fullanalyze scorefxn=score12
            packstat=1 pack_input=0 jump=1 tracer=0 use_jobname=1 resfile=0
            />
        <FavorNativeResidue name=favornative bonus=0.5/>
      </MOVERS>
      <APPLY_TO_POSE>
      </APPLY_TO_POSE>
      <PROTOCOLS>
            <Add mover_name=prepack/>
            Add mover_name=ala_pose
            Add mover_name=favornative
            <Add mover_name=dockprotocol filter_name=rms/>
            Add mover_name=retrieve_sidechains
            <Add mover_name=designiterate filter_name=rms/>
            <Add filter_name=score_filter/>
            <Add filter_name=ddg_filter/>
            <Add filter_name=sasa_filter/>
            <Add filter_name=unsats_filter/>
            <Add mover_name=fullanalyze/>
      </PROTOCOLS>
</ROSETTASCRIPTS>
```

References

1. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A., and Kuhlman, B. (2011) Computational design of a symmetric homodimer using β-strand assembly. Proceedings of the National Academy of Sciences of the United States of America **108**, 20562–7

2. Der, B. S., Machius, M., Miley, M. J., Mills, J. L., Szyperski, T., and Kuhlman, B. (2012) Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. Journal of the American Chemical Society **134**, 375–385

3. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science **332**, 816–821

4. Der, B. S. and Kuhlman, B. (2011) Biochemistry. From computational design to a protein that binds. Science **332**, 801–2

5. Bodemann, B. O. and White, M. A. (2008) Ral GTPases and cancer: linchpin support of the tumorigenic platform. Nature Reviews Cancer **8**, 133–140

6. Aguirre-Ghiso, J. A., Frankel, P., Farias, E. F., Lu, Z. M., Jiang, H., Olsen, A., Feig, L. A., Joffe, E. B., and Foster, D. A. (1999) RalA requirement for v-Src- and v-Ras-induced tumorigenicity and overproduction of urokinase-type plasminogen activator: involvement of metalloproteases. Oncogene **18**, 4718–4725

7. Urano, T., Emkey, R., and Feig, L. A. (1996) Ral-GTPases mediate a distinct downstream signaling pathway from Ras that facilitates cellular transformation. Embo Journal **15**, 810–816

8. Fukai, S., Matern, H. T., Jagath, J. R., Scheller, R. H., and Brunger, A. T. (2003) Structural basis of the interaction between RalA and Sec5, a subunit of the sec6/8 complex. The EMBO journal **22**, 3267–78

9. Jin, R., Junutula, J., Matern, H., Ervin, K., Scheller, R., and Brunger, A. (2005) Exo84 and Sec5 are competitive regulatory Sec6/8 effectors to the RalA GTPase. EMBO J **24**, 2064–2074

10. Sheinerman, F. B. and Honig, B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces. Journal Of Molecular Biology **318**, 161–177

11. Kiel, C., Serrano, L., and Herrmann, C. (2004) A detailed thermodynamic analysis of Ras/effector complex interfaces. Journal Of Molecular Biology **340**, 1039–1058

12. Nassar, N., Horn, G., Herrmann, C., Block, C., Janknecht, R., and Wittinghofer, A. (1996) Ras/Rap effector specificity determined by charge reversal. Nature structural biology **3**, 723–9

13. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G., and Herrmann, C. (2004) Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proceedings of the National Academy of Sciences of the United States of America **101**, 9223–8

14. Filchtinski, D., Sharabi, O., Rüppel, A., Vetter, I. R., Herrmann, C., and Shifman, J. M. (2010) What makes Ras an efficient molecular switch: a computational, biophysical, and structural study of Ras-GDP interactions with mutants of Raf. Journal of molecular biology **399**, 422–35

15. Horton, J. D., Cohen, J. C., and Hobbs, H. H. (2007) Molecular biology of PCSK9: its role in LDL metabolism. Trends in biochemical sciences **32**, 71–7

16. Lagace, T. A., Curtis, D. E., Garuti, R., McNutt, M. C., Park, S. W., Prather, H. B., Anderson, N. N., Ho, Y. K., Hammer, R. E., and Horton, J. D. (2006) Secreted PCSK9 decreases the number of LDL receptors in hepatocytes and in livers of parabiotic mice. The Journal of clinical investigation **116**, 2995–3005

17. Kwon, H. J., Lagace, T. A., McNutt, M. C., Horton, J. D., and Deisenhofer, J. (2008) Molecular basis for LDL receptor recognition by PCSK9. Proceedings of the National Academy of Sciences of the United States of America **105**, 1820–5

18. Cohen, J. C., Boerwinkle, E., Mosley, T. H., and Hobbs, H. H. (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. The New England journal of medicine **354**, 1264–72

19. Abifadel, M., Varret, M., Rabès, J.-P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derré, A., Villéger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J.-M., Luc, G., Moulin, P., Weissenbach, J., Prat, A., Krempf, M., Junien, C., Seidah, N. G., and Boileau, C. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature

genetics **34**, 154–6

20. Ni, Y. G., Di Marco, S., Condra, J. H., Peterson, L. B., Wang, W., Wang, F., Pandit, S., Hammond, H. A., Rosa, R., Cummings, R. T., Wood, D. D., Liu, X., Bottomley, M. J., Shen, X., Cubbon, R. M., Wang, S.-p., Johns, D. G., Volpari, C., Hamuro, L., Chin, J., Huang, L., Zhao, J. Z., Vitelli, S., Haytko, P., Wisniewski, D., Mitnaul, L. J., Sparrow, C. P., Hubbard, B., Carfí, A., and Sitlani, A. (2011) A PCSK9-binding antibody that structurally mimics the EGF(A) domain of LDL-receptor reduces LDL cholesterol in vivo. Journal of lipid research **52**, 78–86

21. Orcutt, K. D., Wittrup, K. D., Kontermann, R., and Dübel, S. (2010) Antibody Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-01143-6. doi:10.1007/978-3-642-01144-3

22. Gai, S. A. and Wittrup, K. D. (2007) Yeast surface display for protein engineering and characterization. Current opinion in structural biology **17**, 467–73

23. Chen, I., Dorr, B. M., and Liu, D. R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. Proceedings of the National Academy of Sciences of the United States of America **108**, 11399–404

24. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and engineering human antibodies using yeast surface display. Nature protocols **1**, 755–68

25. Azoitei, M. L., Correia, B. E., Ban, Y.-E. A., Carrico, C., Kalyuzhniy, O., Chen, L., Schroeter, A., Huang, P.-S., McLellan, J. S., Kwong, P. D., Baker, D., Strong, R. K., and Schief, W. R. (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. Science (New York, N.Y.) **334**, 373–6

26. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**, 2577–2637

27. Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., and Baker, D. (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One **6**, e20161

28. Jha, R. K., Leaver-Fay, A., Yin, S., Wu, Y., Butterfoss, G. L., Szyperski, T., Dokholyan, N. V., and Kuhlman, B. (2010) Computational design of a PAK1 binding protein. Journal Of Molecular Biology **400**, 257–270

29. Bauer, B., Mirey, G., Vetter, I. R., García-Ranea, J. A., Valencia, A., Wittinghofer, A., Camonis, J. H., and Cool, R. H. (1999) Effector recognition by the small GTP-binding proteins Ras and Ral. The Journal of biological chemistry **274**, 17763–70

30. Goedken, E. R. and Marqusee, S. (2001) Native-state energetics of a thermostabilized variant of ribonuclease HI. Journal of molecular biology **314**, 863–71

31. Vorup-Jensen, T., Ostermeier, C., Shimaoka, M., Hommel, U., and Springer, T. A. (2003) Structure and allosteric regulation of the alpha X beta 2 integrin I domain. Proceedings of the National Academy of Sciences of the United States of America **100**, 1873–8

32. Blatner, N. R., Wilson, M. I., Lei, C., Hong, W., Murray, D., Williams, R. L., and Cho, W. (2007) The structural basis of novel endosome anchoring activity of KIF16B kinesin. The EMBO journal **26**, 3709–19

33. Karanicolas, J., Corn, J., Chen, I., Joachimiak, L., Dym, O., Peck, S., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., T.áMontelione, G., P.áSpiegel, C., Liu, D., and Baker, D. (2011) A de novo protein binding pair by computational design and directed evolution. Molecular Cell **42**, 250–260

34. Smith, T. J., Stains, C. I., Meyer, S. C., and Ghosh, I. (2006) Inhibition of beta-amyloid fibrillization by directed evolution of a beta-sheet presenting miniature protein. Journal of the American Chemical Society **128**, 14456–7

35. Jäckel, C., Kast, P., and Hilvert, D. (2008) Protein design by directed evolution. Annual review of biophysics **37**, 153–73

36. Voigt, C. A., Mayo, S. L., Arnold, F. H., and Wang, Z. G. (2001) Computational method to reduce the search space for directed protein evolution. Proceedings of the National Academy of Sciences of the United States of America **98**, 3778–83

37. Treynor, T. P., Vizcarra, C. L., Nedelcu, D., and Mayo, S. L. (2007) Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. Proceedings of the National Academy of Sciences of the United States of America **104**, 48–53

38. Guntas, G., Purbeck, C., and Kuhlman, B. (2010) Engineering a protein-protein interface using a computationally designed library. Proceedings of the National Academy of Sciences of the United States of America **107**, 19296–301

39. Smith, C. A. and Kortemme, T. (2008) Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. Journal Of Molecular Biology **380**, 742–756

Chapter 4

A comparison of successful and failed computational protein interface designs

4.1 Introduction

Computational protein design methods have progressed rapidly in recent years but several challenges remain. One of these is the design of new protein-protein interactions (1; 2). Other than notable success in designing α-helical pairing and specificity (3; 4; 5), there are no general methods for computationally designing novel interactions and successful designs are largely outnumbered by failed ones (6).

Directed evolution of protein-protein interactions has demonstrated that it is possible to generate high affinity binders to most target proteins (7; 8). Computational interface design has not developed nearly as many affinity reagents as directed evolution but can provide the ability to target specific regions of a target protein to modify function (1). Progress in computational design has been confounded by the complex nature of protein-protein interfaces including large numbers of polar interactions (9), importance of solvation loss upon binding and long range electrostatics (10), and water mediated hydrogen bonds (11). It is difficult to model these aspects in the context of a protein design simulation. It is important to correctly model polar interaction because an important determinant of an interface's ability to form is the extent to which interface spanning hydrogen bonds compensate for a desolvation penalty (12).

There have been many attempts to computationally design new protein-protein interactions, however only a few designs that have been experimentally shown to have high affinity ($K_D \leq 1$ μM) and adopt the computationally predicted structure. Experimental validation of computational designs is a necessary step not only to confirm a design matches the prediction, but to provide insight into improvements in the energy function and search methods used in a computational protocol (1). However, cloning, expression, affinity measurements, and structure determination can make experimental verification of a design time consuming. Typical interface design simulations can output hundreds of plausible models with similar energy making it imperative to apply additional metrics to select which designs to carry forward to an experimental system.

One recent study sought to improve design selection methodology by asking the computational protein docking community to determine metrics that discriminate designed proteins that were known not to bind from natural interfaces (6). Some of the best discriminating metrics showed that the designs had unfavorable solvation energy at the interface and poor electrostatic complementarity between the two proteins in the complex. However, most of the metrics failed to distinguish natural small hydrophobic interfaces from designed small hydrophobic interfaces.

Here, we examine some of the successful and failed attempts to use Rosetta to design protein-protein interactions and compare them to interfaces of natural proteins. In general we find that the designs are smaller and more hydrophobic that native protein interactions. Though most designs fail to form experimentally, the ones that do are dominated by hydrophobic packing interactions at the interface. All attempts to designing polar, hydrogen bond rich, interfaces failed to bind experimentally. We then address possible causes and solutions to the discrepancies between designed and native protein-protein interfaces.

4.2 Methods

4.2.1 Definition of an interface residue

A residue is defined as being at a protein-protein interface if it passes either of two criteria: (*i*) the distance from one of that residue's side-chain atoms to an atom on the other chain is short (less than 5.0 Å); (*ii*) the C_β of the residue in question is less than 9.0 Å from the C_β of a residue on the opposite chain and its C_α-C_β vector points at the opposite chain. This method of determining the interface is termed the interface *vector definition*. To calculate if a residue passes the first criterion, all residues nearby residues on the opposite chain(C_β to C_β distance within 10.0 Å) are found. Next the distance from all side-chain atoms in that residue to the nearby residues on the opposite chain are calculated. If any of those distances are less than 5.0 Å then the residue is considered to be at the interface. Residues that do not pass this check are then examined to determine if they meet the second criterion. Two vectors are drawn, a $\overrightarrow{C_\alpha(A)C_\beta(A)}$ vector and another vector $\overrightarrow{C_\beta(A)C_\beta(B)}$ (Figure 4.1A). If the distance from between the C_β on the different chains is short enough ($|\overrightarrow{C_\beta(A)C_\beta(B)}| < 9.0$ Å), these vectors are then normalized and the dot product between them is calculated. If either residue does not have a C_β atom then one is built onto it using the idealized position of a C_β in alanine. The residue on chain B is considered to be at the interface if the angle between $\overrightarrow{C_\alpha(A)C_\beta(A)}$ and $\overrightarrow{C_\beta(A)C_\beta(B)}$ is less than 65° (*ie.* $C_\alpha\widehat{(A)C_\beta}(A) \bullet C_\beta\widehat{(A)C_\beta}(B) > \cos 65°$). In Rosetta, residues at a protein protein interface are typically defined as all amino acids that have a C_β on one chain to C_β on the other chain distance of less than 8.0 Å. This is termed the interface *neighbor definition*. When an interface is defined in this manner, several of the residues considered to be at the interface are in fact pointing towards the core of the protein and could not be designed to be at the interface (Figure 4.1B). The interface *vector definition* avoids this problem by only considering residues to be at the interface if they are very close to the other chain or point at it (Figure 4.1B).

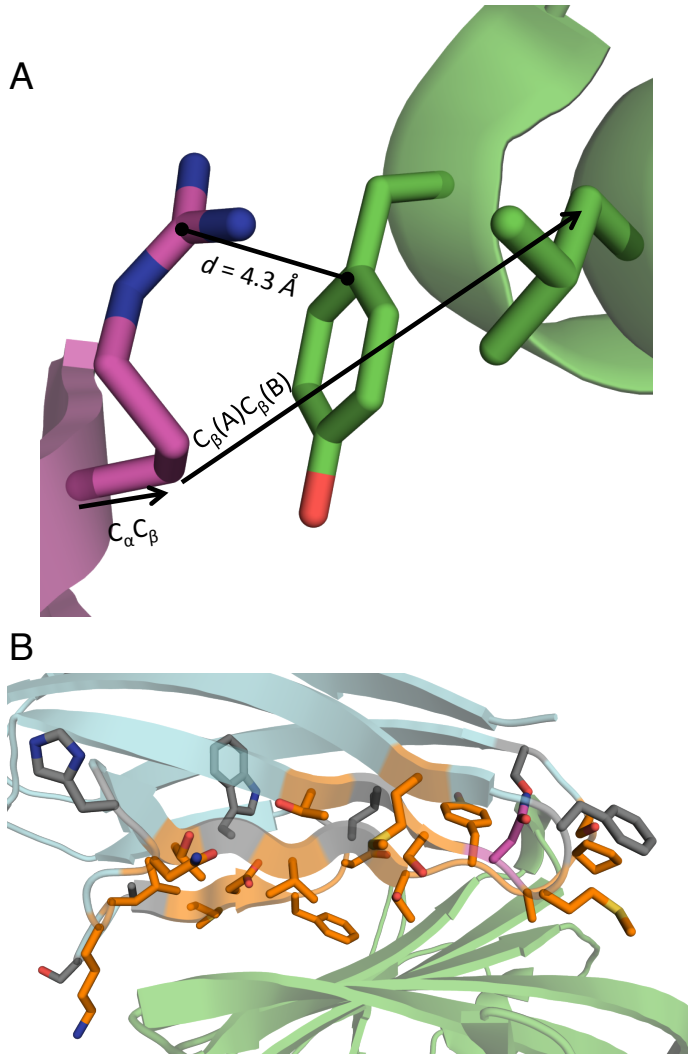

Figure 4.1: Method of finding residues at a protein-protein interface of chain A (purple) and chain B (green). A) Illustration of how the vectors and distances are drawn to determine interface residues. The highlighted arginine on chain A includes the tyrosine and leucine on chain B in two different ways. The tyrosine is included at the interface because the distance from an arginine side-chain atom to any atom on the tyrosine is less than the cutoff of 5.0 Å. The arginine on chain A includes the leucine on chain B because $|\overrightarrow{C_\beta(A)C_\beta(B)}| < 9.0$ Å and the angle between $\overrightarrow{C_\alpha(A)C_\beta(A)}$ and $\overrightarrow{C_\beta(A)C_\beta(B)}$ is less than the cutoff of 65°. B) Example interface of two chains (PDB ID: 1GL4), different chains are shown in light blue and light green. Residues defined at being at the interface on the light blue chain by either the *neighbor definition* or *vector definition* are shown in sticks. Residues shown in orange are those determined to be at the interface by both methods. Residues in grey are those defined only by the *neighbor definition*, while residues in purple are those counted only by the *vector definition*.

4.2.2 Input structure energy minimization

Many protein X-ray crystal structures have minor clashes of atomic radii, improper orientation of asparagine, glutamine, or histidine side-chains, or backbone dihedral angles outside of acceptable ranges (13). All of these problems can adversely affect the Rosetta score function. Thus every input X-ray crystal structure was repacked and minimized with Rosetta. This process has three steps. First, the side-chain χ angles are minimized to optimize local contacts, next a full rotamer packing step is done to relieve any clashes that minimization could not solve. Finally, the side-chains, backbone and rigid body orientation is minimized to obtain the structure used for further analysis. Most minimized output structures had an RMSD of less than 1.5 Å to the native crystal structure. The designed models were also run through this protocol to allow comparison between the designs and native structures. An example command line for this protocol is given below:

```
./min_pack_min.<exe>
     -database <rosetta_database>
     -l start_structs.list
     -pack_first false
     -no_rbmin false
     -min_all_jumps true
     -nstruct 50
     -score12prime true
     -out::pdb_gz true
     -ndruns 5
     -run::min_type dfpmin_armijo_nonmonotone
     -use_input_sc true -ex1 -ex2 -extrachi_cutoff 1
     -no_his_his_pairE true
     -ignore_unrecognized_res
     -no_optH false
```

4.2.3 Interface analysis protocol

The protein-protein interfaces were analyzed using the `InterfaceAnalyzerMover` (14). This protocol takes a protein-protein complex as an input structure and then creates an unbound structure. The interface energy, SASA, and other metrics are calculated based on the differences between the bound and unbound structures.

```
./rosetta_scripts.<exe>
     -l minimized_structs.list
     -parser:protocol interface_analysis.xml
     -score12prime true
     -ignore_unrecognized_res
     -no_his_his_pairE
     -out:file:score_only IA.score.sc
     -no_optH false -ex1 -ex2 -use_input_sc
     -run::min_type dfpmin_armijo_nonmonotone
     -extrachi_cutoff 1
     -linmem_ig 10
     -ignore_unrecognized_res
     -atomic_burial_cutoff 0.01
     -sasa_calculator_probe_radius 1.2
```

The script passed to `parser:protocol` is given in Supplementary Information (Section 4.5.3 on page 109)

4.2.4 Polar burial definition

Rosetta calculates solvent accessible surface area (SASA) using the Le Grand and Merz method (15) and keeps the SASA up to date as described by Leaver-Fay *et al* (16). The SASA for a polar atom is sum of the SASA for that atom, plus the SASA for any bound hydrogens. A polar atom is defined as buried if the total SASA for that atom is less than 0.1 $Å^2$. If a buried polar atom does not have a hydrogen bonding partner, as defined as having an H-bond energy of less than 0.0 REUs, then that atom is considered buried and unsatisfied. A hydrogen bond is defined as buried if the SASA for the two involved polar atoms is less than 3.0 $Å^2$. Based on distances observed from low B-factor waters to protein atoms (Reference (17) and Figure S4.1) we chose to use atomic radii from Reduce (18) and a water probe radius of 1.2 Å to find buried polar atoms and hydrogen bonds.

4.2.5 Determining hydrogen bond geometry

Geometric parameters describing interface spanning hydrogen bonds were found using the robust `Features` reporting interface in Rosetta (O'Meara unpublished results). Hydrogen bond geometry features were extracted from the minimized design models, X-ray crystal structures of the native interfaces, and set of 4400 high-resolution crystal structures (Richardson Lab

top4400 set). An example command line to extract these parameters from a list of structures (`structures.list`) is:

```
./rosetta_scripts.<exe>
     -in:file:fullatom
     -l structures.list
     -out:nooutput
     -parser:protocol features_script.xml
```

The actual features that are extracted are defined in the input XML script passed to `-parser:protocol`. The protocol used for the features described here is given in Supporting Information Section 4.5.4 (page 4.5.4). Plots of the features and hydrogen bond selection was performed using **R** scripts. Interface spanning hydrogen bonds were selected from the native dimers and the design models. Intra-chain hydrogen bonds were used from the top4400 set. Hydrogen bonds are only counted if the sequence separation of the donor and the acceptor is greater than five residues. This removes hydrogen bonds in α-helices from the distribution. Hydrogen bonds from the X-ray crystal structures were only chosen if the donor and acceptor heavy atoms had a *B-factor* < 40.

4.3 Results

4.3.1 Set of designed interfaces

The computational models used in this analysis represent a wide array of interface design goals (Table 4.1). The design models fall in two main categories: *i)* design of one protein chain to bind to a natural target (Figure 4.2A,C,E) and *ii)* design of both chains involved in an interaction to create a novel heterodimer (Figure 4.2B) or homodimer (Figure 4.2D,F). The majority of the designs, 140, fall into the first category. These predominantly consist of interfaces of a scaffold designed to bind some target of interest such as a small GTPase (Figure 4.2A, Chapter 3), PAK1 (19), proteins involved in ubiquitin transfer, and influenza hemagglutinin from Fleishman *et al* (Figure 4.2C) (20). Another 11 models represent the design of both the structure and sequence of a peptide to bind $G_{\alpha i1}$ (Figure 4.2E) (21). The second category is comprised of 18 redesigns of natural proteins to form homodimers mediated

by metal binding (Figure 4.2D) (22) or β-strands (Figure 4.2F) (23), and 11 models from Karanicolas *et al* where both interface forming chains are designed to form a new heterodimer (Figure 4.2B) (24).

Of the 59 designs from our laboratory 52 of them were successfully expressed in *E. coli*.Seventy-three of the eighty-eight proteins designed to bind HA successfully expressed using yeast surface display. All of the designed pairs from Karanicolas *et al* successfully expressed.

The interfaces used for the native dataset were taken from those chosen by Zhanhua *et al.* (25). This set is comprised of high resolution X-ray crystal structures (resolution < 2.5 Å) of 170 homodimers and 156 heterodimers. Of these, 167 homodimers and 152 heterodimers were read by Rosetta and used in this analysis (Table S4.1).

4.3.2 Definition of a successful design

For the purpose of this study, the computational interface designs were divided into three categories, strong success, weak success, and failure. A strong success is defined as a high affinity interaction ($K_D \leq 1$ μM) where the X-ray crystal structure of the complex closely matches the computational prediction. A weak success has at least a moderate affinity ($K_D \leq$ 100 μM) and either mutational or NMR chemical shift data suggesting the interface forms as designed. A failed design does not meet the previous criteria. Table 4.1 shows a summary of how many designs satisfy either definition of success. A complete list of structures used is given in Table S4.2.

There are four examples that meet the criteria for a successful protein-protein interface design. The first is the design of the structure and sequence of a peptide that binds to $G_{\alpha i1}$(GL_{helix}-4) (21). Another example is the redesign of a protein sequence to bind to HA (HB36) (20). The other two structurally verified designs involved the redesign of natural monomeric proteins to form symmetric homodimers via Zn^{+2} binding (MID1) (22) or β-strand pairing (βdimer1) (23). We chose to use the MID1 H12E mutant throughout the rest of this investigation because the crystal structure was a closer match to the design. A success that is not included in Table 4.1 is the design of a novel helical tetramer by Harbury *et al* (3). We decided not to include this design in the figures because the rest of the designs are dimers.

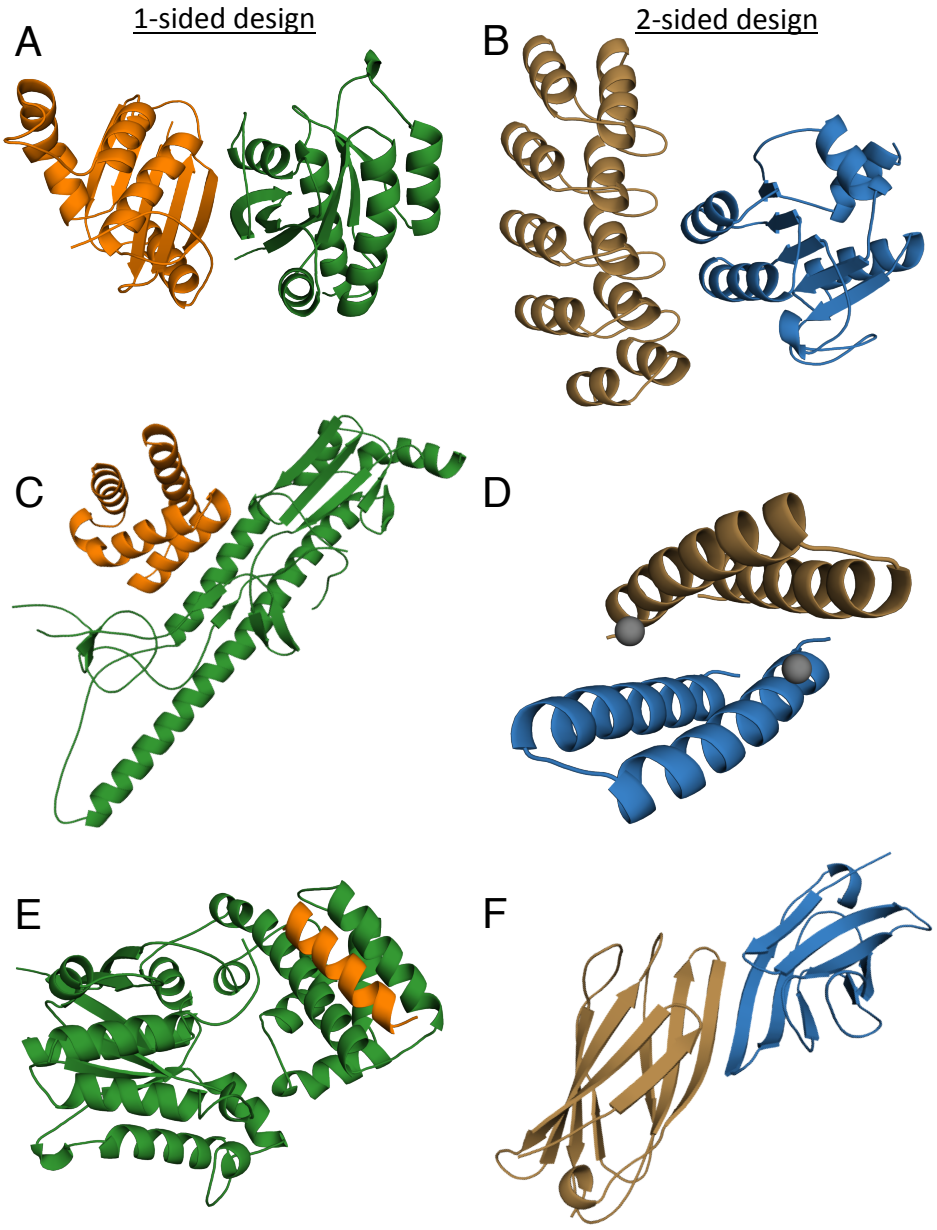

Figure 4.2: Computational interface design models of for several different targets and design goals. Natural target proteins are colored green while designed binders are colored orange (A,C,E). In the case when both sides of the interface are designed the chains are colored blue and brown (B,D,F). A) Failed design of scaffold 1JL1 to bind to RalA. B) Model of Prb/Pdar interaction. C) HB36 designed to bind HA. D) Zinc (grey spheres) mediated homodimer MID1. E) Model of the peptide GL_{helix}-4 binding to $G_{\alpha i1}$. F) Homodimer mediated by β-strand pairing βdimer1.

Four designs are classified as weak successes because there is not a crystal structure of the modeled complex. These include a high affinity binder to HA (20), a low affinity binder to PAK1 (19), a Zn^{+2} mediated heterodimer (Der unpublished results) and a β-strand mediated homodimer (βdimer2). The high affinity designed interaction between Prb and Pdar is classified as a weak success because a crystal structure of the complex is a 180° rotation from the computational model (24).

Design goal	No. tested	Express/soluble	Strong success	Weak success
PAK1 binders (19)	10	6	0	1
GTPase binders	6	6	0	0
$G_{\alpha i1}$ binding peptides (21)	11	11	1	0
Ubiquitin or UbcH7 binders	5	5	0	0
Metal mediated homodimers (22)	8	6	1	0
Metal mediated heterodimers	6	5	0	1
β-strand mediated homodimers (23)	10	10	1	1
FNIII to SH3	3	3	0	0
Flu-hemagglutinin binders (20)	88	73	1	1
Prb/Pdar (24)	11	11	0	1
Total	158	136	4	5

Table 4.1: The numbers of experimentally tested computational protein-protein interfaces examined in this work. The number of expressed/soluble proteins represents the number of total that could actually be expressed in the experimental system and did not aggregate. Strong successes are high affinity interactions ($K_D \leq 1$ μM) where an X-ray crystal structure matches the design model. Weak successes ($K_D \leq 100$ μM) have moderate to high affinity and other experimental evidence that the interface forms as designed. Citations are given when available.

4.3.3 Designed interfaces are small

Natural interfaces from the Zhanhua *et al.* set broadly sample different interface sizes ranging from 850 Å^2 to 10,000 Å^2 (up to 7,000 Å^2 for heterodimers) (Figure 4.3). The designs sample much smaller interfaces ranging from 850 Å^2 to 2,400 Å^2, with the majority of designs having an interface area between 1,000 and 1,600 Å^2 (Figure 4.3 inset). The successful designs are represented over the range of designed interface sizes and the crystal structures of the designs show a similar ΔSASA to the design models. There has been no successful dimer design where the interface area is over 1,600 Å^2 suggesting that better sampling and additional effort is required to recapitulate the sizes of native complexes.

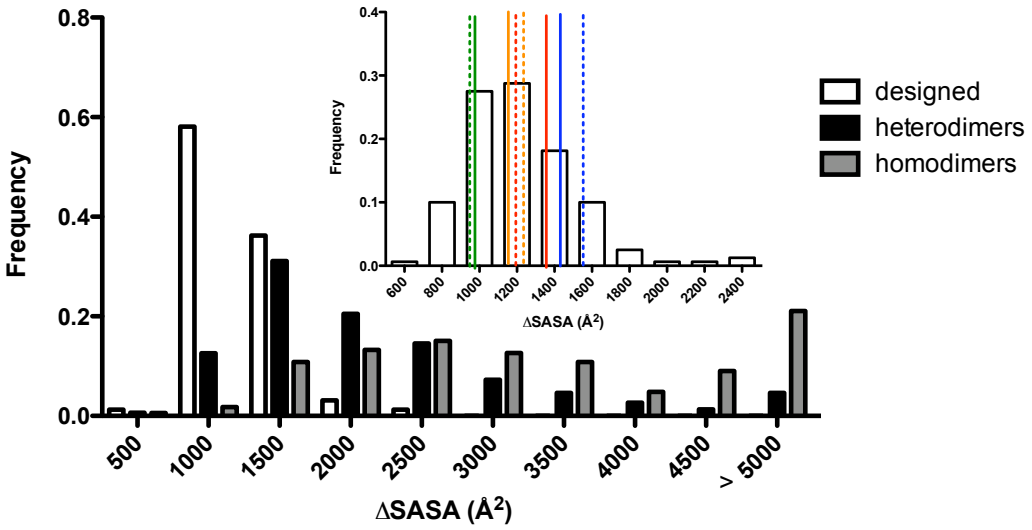

Figure 4.3: Size of designed and natural interfaces. The change in SASA upon binding for natives and designs is shown. Inset: Designs only on a smaller scale. Successful designs are highlighted by colored lines (red = GL_{helix}-4, blue = HB36, green = MID1, orange = βdimer1). Solid lines represent computational designs, dashed lines represent the crystal structure of the design. The slope of the linear regression line is not significantly different from 0 (Prism calculation) for heterodimers and homodimers while the y-intercept is -0.017 for heterodimers and -0.016 for homodimers. The slope of the regression line fit to the design data is non-zero with a y-intercept of -0.022 with a correlation value of $r = 0.29$.

Native proteins have a similar interface energy density ($\Delta G_{bind}/\Delta$SASA) across all sizes of interfaces (Figure 4.4A and B) while the designed interfaces tend to vary in energy density depending on the size of the interface. (Figure 4.4C). Larger designed interfaces tend to have a higher $\Delta G_{bind}/\Delta$SASA than smaller ones. This observation suggests that, unlike native complexes, current sampling and design strategies are unable to achieve similar energy across interface sizes. It should be noted that most of the protocols used to produce the computational designs allow rigid body motion but not substantial backbone rearrangement. Crystal structures of the successful designs maintain a similar $\Delta G_{bind}/\Delta$SASA to the computational model. The minimal change in $\Delta G_{bind}/\Delta$SASA indicates that the majority of the designed interactions were recovered experimentally. Rosetta appears to be unable to maintain a low

energy across a complex as the size of the interface increases. This could mean that more backbone sampling is needed to ensure favorable contacts across a large interaction area.

It is important to note that Zn^{+2} binding of MID1 is not accounted for by the Rosetta Energy function, thus binding energy calculations will be higher than reasonable for the metal mediated designs. If it is assumed that Zn^{+2} binding is approximately the same energy as a hydrogen bond, then adding in eight hydrogen bonds with an energy of -1.3 REUs, the $\Delta G_{bind}/\Delta$SASA for MID1 (Figure 4.4 green points) would be -0.023.

4.3.4 Successful designs have few polar interactions

Most Rosetta designed interfaces have less contribution from polar interactions at the interface than natural dimers. This was also noticed previously when comparing the proteins designed to bind HA to another set of heterodimers (6). The successful designs also tend to have fewer polar interactions at the interface than natives or the other designs. The amount of SASA that polar atoms contribute to an interface normalized by the total ΔSASAof the interface shows that the designs more closely resemble homodimers (Figure 4.5B) than heterodimers (Figure 4.5A). The successful designs have a low fraction of buried polar area even when compared to the other designed interfaces. For all of the successes, polar atoms contribute less than 40% of the total interface area (Figure 4.5). The polar atom contribution to the interface of the weak successes is also below 40% for all cases. The successful design that has the largest fraction of polar interface area, βdimer1, has six main-chain hydrogen bonds which account for a larger amount of buried polar area. The remainder of the βdimer1 interface is predominantly hydrophobic.

The design of polar residues at an interface could result in the burial of a polar atom without a hydrogen bonding partner. Native interfaces tend to have no more than two buried polar atoms per 1,000 $Å^2$ of interface (Figure 4.6A). While the designs have a similar profile to natives, three of the strong successes have no buried unsatisfied polars at the interface. HB36 has two buried unsatisfied hydrogen bonds at its interface in the X-ray crystal structure. One of these could be removed with a rotamer substitution. The other is located on the side chain of W57, a critical buried residue in the center of the interface.

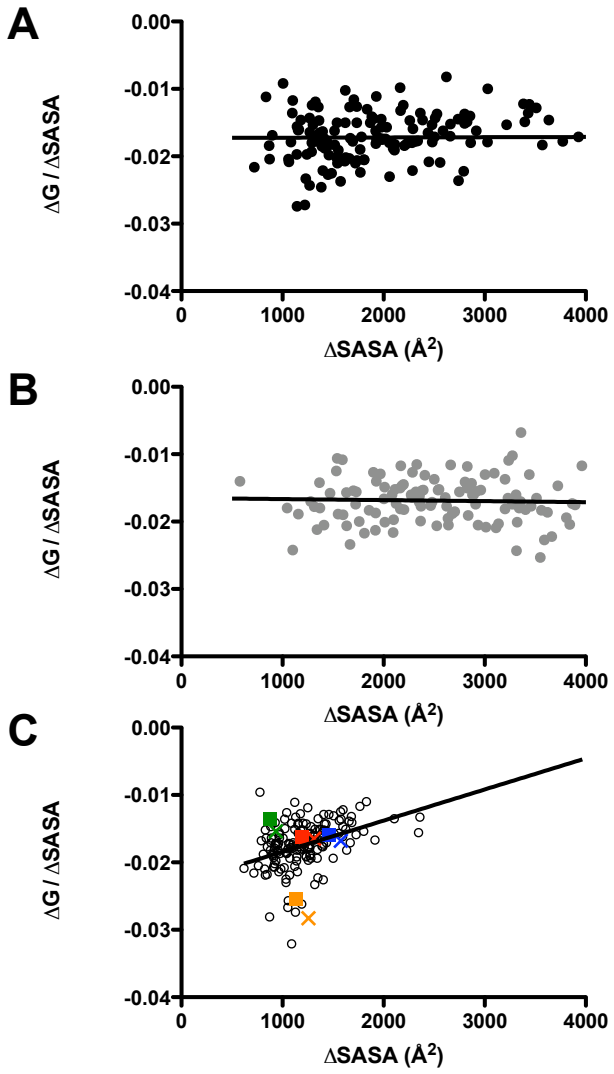

Figure 4.4: Interface energy density ($\Delta G_{bind}/\Delta$SASA) designed and natural interfaces. The change in SASA upon binding versus $\Delta G_{bind}/\Delta$SASA is shown for heterodimers (A), homodimers (B) and designs(C). Successful designs are highlighted by colored points (red = GL_{helix}-4, blue = HB36, green = MID1, orange = βdimer1). Solid points represent the computational model, $\times$'s represent the crystal structure of the design. Least-squares lines were fit to each set of interfaces. The correlation coefficient for designs is $R = 0.33$.

One of the most striking differences between designed and native complexes the amount of energy at the interface contributed by buried side-chain to side-chain or side-chain to backbone hydrogen bonds. The largest fraction of designs have near zero binding energy contributed by buried hydrogen bonds (Figure 4.6B). The successful designs all have minimal buried hydrogen bonds at the interface. MID1 and βdimer1 have none and the design models of GL_{helix}-4and HB36 each have one buried hydrogen bond across the interface. The buried hydrogen bond in GL_{helix}-4was not observed in the crystal structure. One new buried hydrogen bond was introduced to HB36 during affinity maturation of the computational design.

Examples of several designed Interfaces with buried polars are shown in Figure 4.7. These designs, and others like them, can be thought of as tests to see if it is possible to design hydrogen bonding at interfaces similar to that observed in native proteins (26). Each of these interfaces have a high value for buried $E_{H-bond}/\Delta G$ compared to the majority of the designs. Two of the designs (Figure 4.7B, C) have similar contribution from buried hydrogen bonds as native proteins but still do not interact.

This evidence suggests that design of polar interactions at a protein-protein interface is generally unsuccessful. Buried hydrogen bonds are either not present, or recovered, in most of the successful designs. Directed evolution, and not computational design, was responsible for one hydrogen bond that increased the affinity of the initial computational design of HB36. Successful designs have few buried and unsatisfied polar atoms compared to the unsuccessful designs (Figure 4.6A) indicating that if a polar atom is placed at an interface it needs a hydrogen bonding partner for an interaction to happen experimentally. However, these hydrogen bonding interactions have proven difficult to design. The difference in hydrogen bond energy between the design model of HB36 and the crystal structure is due to the gain of a interface spanning hydrogen bond during directed evolution that was not present in the design model. An interface spanning hydrogen bond in GL_{helix}-4was not present in the crystal structure. The designed helical bundle RH4 is also an example a successful design that relies predominantly on hydrophobic interactions at the interface. The polar ΔSASA fraction for RH4 is 0.23, well below that of the other successful designs observed here. It also has no buried unsatisfied polar atoms and no buried hydrogen bonds.

4.3.5 Rosetta designs suboptimal hydrogen bonds

The previous observations led us to investigate the interface spanning hydrogen bonds in the models and native proteins. One feature of Rosetta designed hydrogen bonds that stands out as being different from native proteins is the length of a hydrogen bond from a hydroxyl donor. The distance from a donor hydroxyl hydrogen to an acceptor is centered around 1.75 Å in monomers and across chains in high resolution X-ray crystal structures (Figure 4.8A). The interfaces designed by Rosetta have a narrower distribution centered at 1.9 Å. Rosetta's preference for longer hydrogen bonds could result in designs that cannot form sufficient hydrogen bonds across an interface in order to overcome the solvation penalty of burying a polar atom.

Another aspect of Rosetta designed hydrogen bonds that is different from native hydrogen bonds is a different distribution about the dihedral angle given by rotation around the acceptor-acceptor base bond (the χ angle (27)) for sp^2 acceptor atoms. A diagram illustrating the method used to generate the Lambert azimuthal equal-area projection plots is show in Figure S4.2. In high-resolution X-ray crystal structures there is a clear orientation preference for donor hydrogens around side-chain sp^2 hybridized acceptor atoms at $\chi = 180°$ with a smaller preference for $\chi = 0°$ (Figure 4.8B). Intra-chain (Figure 4.8B left) and interchain (Figure 4.8B right) show similar distributions at $\chi = 1800°$ but interfaces have fewer donors located around $\chi = 0°$. The Rosetta designed models (white points in Figure 4.8B) do not show a strong preference for χ. Many of designed hydrogen bonds are located around $\chi = 90°$ and $\chi = 270°$ which are outside of the plane containing the lone pair electrons around a sp^2 acceptor. The location of donor hydrogens around protein backbone carbonyl oxygens also shows a preference for χ that is not recapitulated in the designed interfaces (Figure S4.3). The differences between the intra-chain and interchain distributions are due to intra-chain hydrogen bonds in top4400 being dominated by β-sheets. Hydrogen bonds in the interfaces are a mix of backbone-backbone and backbone-sidechain interactions. The current Rosetta hydrogen bond model does not contain a parameter that accounts for the preference about the χ angle.

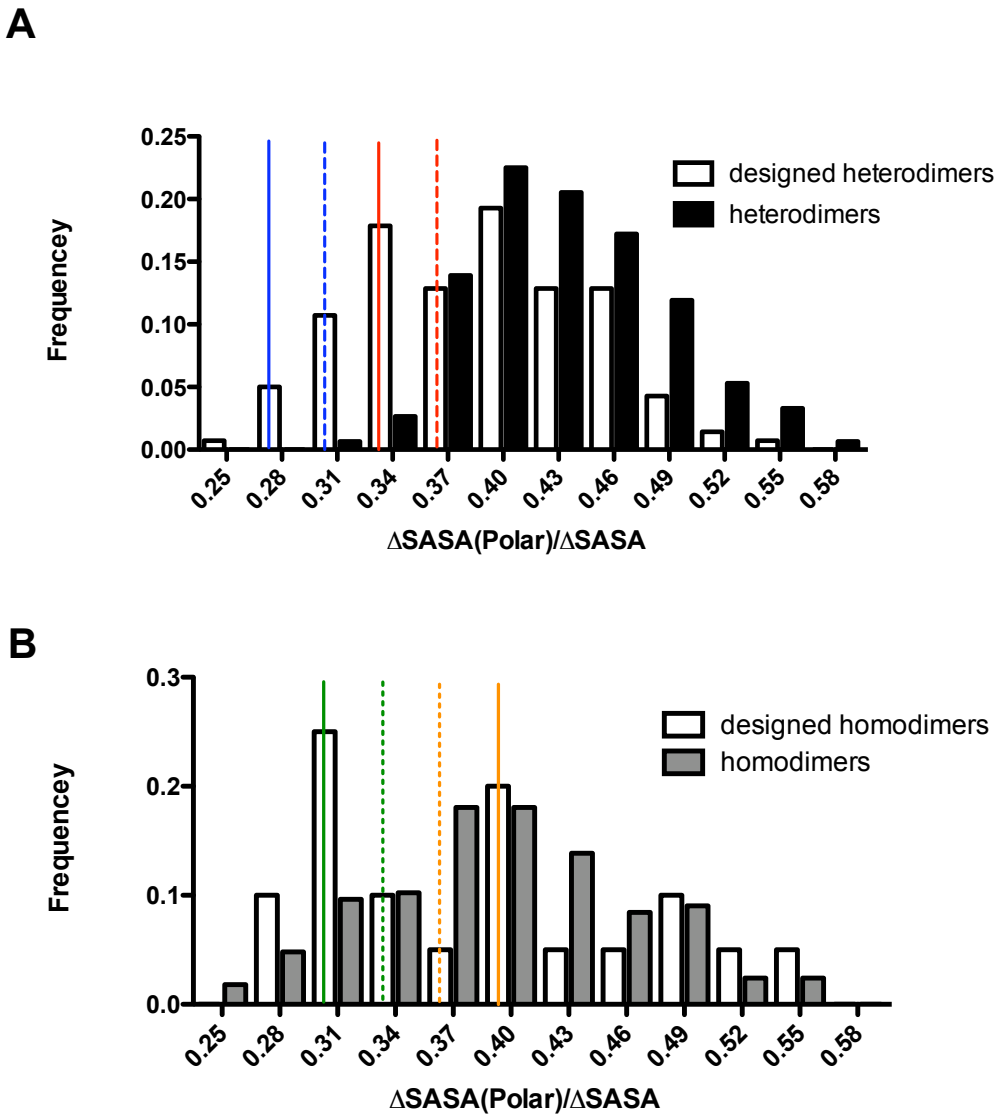

Figure 4.5: Polar content of designed and natural interfaces. The polar fraction of interface area is shown for designs versus heterodimers (A) and homodimers (B). Successful designs are highlighted by colored lines (red = GL_{helix}-4, blue = HB36, green = MID1, orange = βdimer1). Solid lines represent the computational model. Dashed lines represent the crystal structure of the design.

4.3.6 Other observations

There are two other interesting differences between native and designed interactions. The loss of solvation energy upon binding, after subtracting the energy of newly formed hydrogen

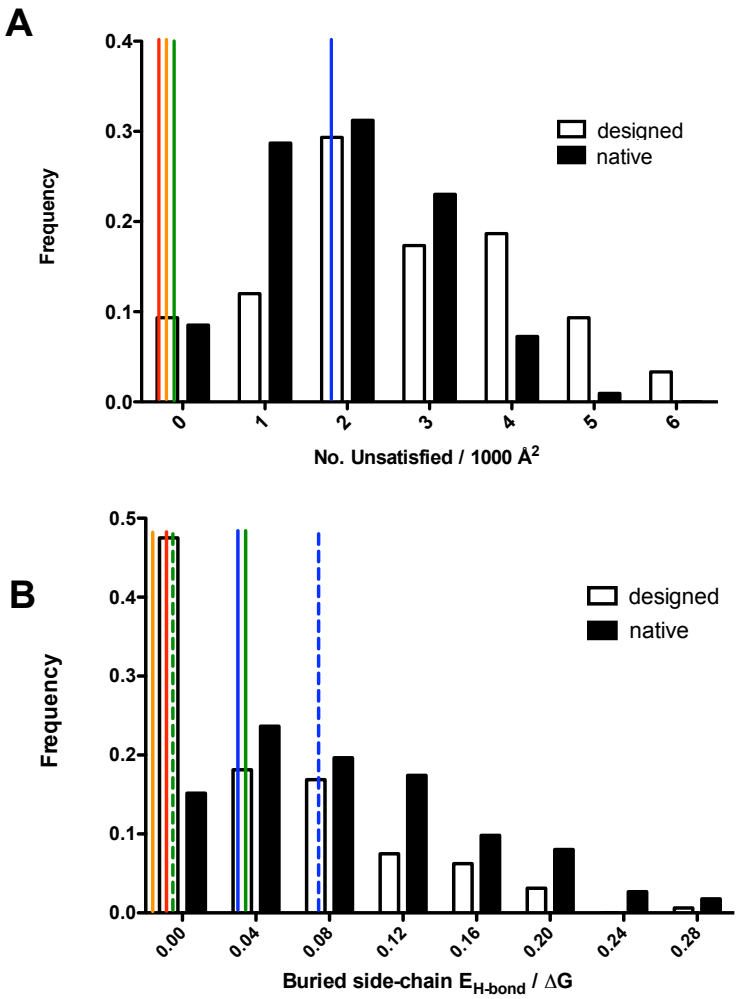

Figure 4.6: Buried polars and buried hydrogen bonds at interfaces. A) The number of buried polar atoms without a hydrogen bonding partner per 1,000 Å^2 of interface area. B) The total energy of a buried, side-chain involved hydrogen bond at the interface as a fraction of total binding energy (ΔG_{bind}). Successful designs are highlighted by solid colored lines (red = GL_{helix}-4, blue = HB36, green = MID1, orange = βdimer1). Dashed lines representing the crystal structure of the design are shown when there is a difference between the computational design and the crystal structure. The difference for HB36 in (B) is due to the addition of a hydrogen bond during directed evolution. The loss of hydrogen bond energy for GL_{helix}-4is due to the interface in the crystal structure not forming a designed, buried, hydrogen bond.

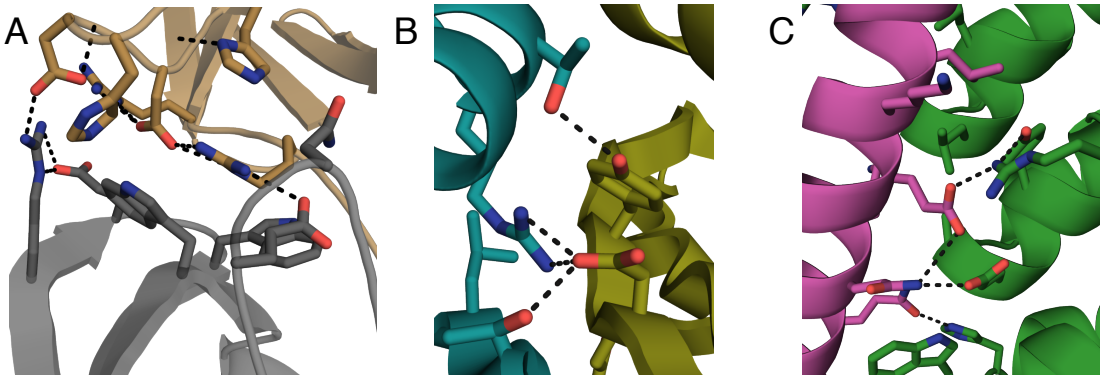

Figure 4.7: Examples of designed interfaces with buried side chain hydrogen bonds. All of these interfaces failed to bind. Interface spanning hydrogen bonds are shown in black dashed lines. A) FNIII (brown) designed to bind to βpix SH2 domain (grey), $E_{H-bond}/\Delta G = 0.262$. B) Failed design of c-Raf (aqua) to bind Rac1 (olive), buried side-chain $E_{H-bond}/\Delta G = 0.145$. C) Failed design of an α-helix (magenta) to bind to an ankyrin repeat protein (green), buried side-chain $E_{H-bond}/\Delta G = 0.115$.

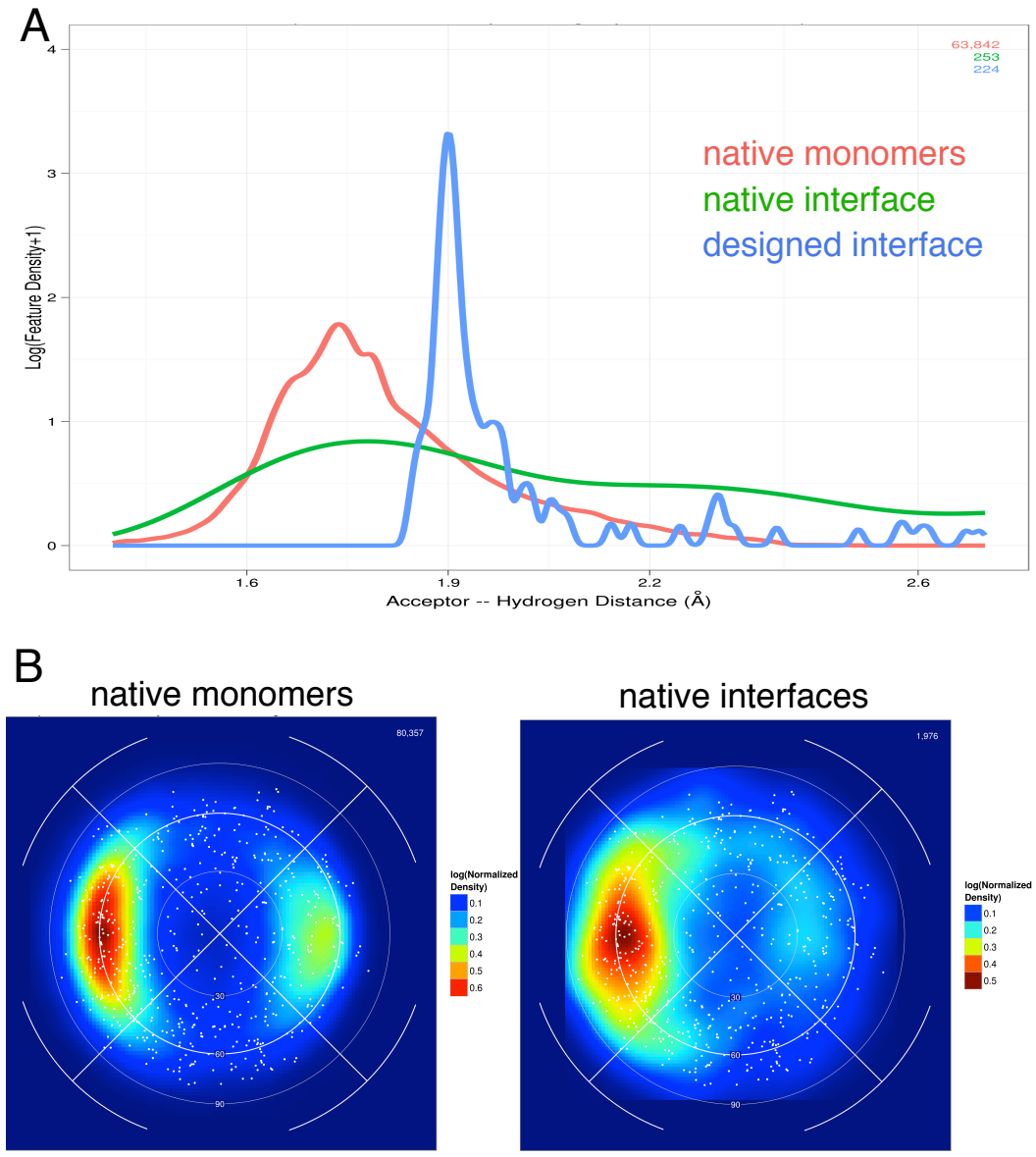

Figure 4.8: Examples of Rosetta incorrectly modeling natural hydrogen bond geometry. Hydrogen bonds from native monomers are intra-chain hydrogen bonds in the top4400 set of structures. Native and designed hydrogen bonds are interchain hydrogen bonds at the interfaces discussed here. Hydrogen bonds are only counted if the sequence separation of the donor and the acceptor is greater than 5 residues. A) Kernel density (log scale) of the distance from the hydrogen on a hydroxyl oxygen to the atom accepting the hydrogen bond. Numbers in upper right represent the number of hydrogen bonds in the plot. B) Lambert azimuthal equal-area projection of the location of a hydrogen donating a hydrogen bond to a side-chain sp^2 acceptor atom. Locations of the hydrogen for native monomer and native interfaces is shown as a density. Positions of the hydrogens in designed interfaces are shown as white dots in both density plots. The number of hydrogen bonds used in the density is shown in the upper right.

is over-penalizing design of polar atoms at the interface. However, designs with many buried polar atoms tend not to form experimentally (Figure 4.5).

4.4 Discussion

These results indicate that although it is possible to computationally design a novel protein-protein interactions there are many more designs that fail to form the predicted complex. In general, the designed interactions tend to be more hydrophobic and smaller than most natural protein-protein interfaces. The successful designed interactions have even fewer polar atoms at their interfaces than most of the designs and have little to no buried hydrogen bonds or unsatisfied polar atoms. It was not possible to completely discriminate successful designs, failed designs, and native interactions using the metrics described here. Previous attempts to discriminate binding from non-binding interface designs based on polarity succeeded in separating designed interfaces from natural heterodimers but failed to identify a design that was experimentally found to bind (6). The lack of large interface designs could be explained by lack of suitable protein scaffolds to form a large contact area. It has been observed that the number of candidate scaffolds decreases rapidly as more design restraints are applied (23).

This work highlights the difficulty of designing polar interactions at protein-protein interfaces. There have been several examples of the successful design of new hydrogen bonds at a natural interface (28; 29; 30), however, the redesigns have lower affinity than the wild type interaction. New hydrogen bonds can increase the affinity of a natural interaction in some cases (31), typically by designing a interface spanning salt bridge (32). However, there are no buried salt bridges in the successful designs outlined here. Another strategy for increasing affinity involves replacing a polar residue with a non-polar one, or a small hydrophobic residue with a larger one (33). None of the examples of successful novel interface design derive a large portion of their interface from polar interactions. Nature is able to sample and make interfaces that computational methods have yet to realize; ones with substantial polar area and hydrogen bond interactions (Figure 4.5A and 4.6B). Computational methods seem adept

at designing good hydrophobic packing interactions (34) but are unable to successfully model favorable polar interactions.

When Rosetta designs hydrogen bonds at an interface there are clear differences between the modeled hydrogen bond and those found in X-ray structures (Figure 4.8). Of the four designs that succeeded only one had a buried hydrogen bond that was recovered in the crystal structure. Mistakes in the hydrogen bond model could explain why no designs with a large amount of polar interface area have formed experimentally. If the designed interactions are tailored to a suboptimal hydrogen bond geometry then the hydrogen bonds formed between the chains may not be sufficiently strong to overcome a desolvation penalty. Current efforts are underway in our laboratory to modify the hydrogen bond parameters to more accurately recapitulate the hydrogen bonds observed in native crystal structures.

There are many more examples of successful computational redesign of natural protein-protein interactions for increased affinity (31; 33; 35; 36; 37; 32) or altered specificity (38; 28; 29; 39; 30) than of the design of a new protein interface. Energy and search functions are able to optimize the local interactions required for binding in the context of a known partner. The design of a novel interface requires searching for alignments of two proteins and the addition of new residue interactions without a native like context to help direct the simulation (2). A search strategy that is able to orient two protein scaffolds into an arrangement similar to a native conformation could turn the difficult problem of novel interface design into the more tractable one of redesign of native interactions.

The successful protein-protein interaction designs outlined here show that it is now possible to design interactions using a variety of strategies as long as the interaction is small and hydrophobic. In addition, all of the successfully designed interactions are dominated by residues in α-helices or β-strands (Figure 4.2C-F). Three important challenges in computational protein-protein interface design remain; 1) The design of an interaction where over 40% of the atoms at the interface are polar and several buried hydrogen bonds are made, 2) The design of an interface larger than 1,600 Å, 3) The design of a loop based interaction. The absence of a successful loop mediated design is surprising given the prevalence of loops in interfaces from phage display (7) and the development of methods to accurately design and

model loops (40; 41). To achieve these goals it is likely that there will need to be improvements in conformational search methods and in energy functions for protein design.

4.5 Supporting Information

4.5.1 Figures

The distance from a low B-factor water ($B < 30$) to protein atoms within 4.5 Å was found for 500 of high resolution structures. The protein atoms near water were then sorted by atom type; carbon, nitrogen and oxygen. The number of atoms at a binned distance from water were then normalized by the volume of a sphere drawn around a water. This calculation was performed as follows:

$$\text{Normalized Counts}_{\text{ atom type}} = \frac{\text{counts}_{\text{ atom type}}}{\frac{4}{3}\pi r^3}$$

Where r represents the distance of an atom from water and 'counts' is the number of atoms at that distance. The overall fraction was calculated by setting the largest number of normalized counts (in this case for carbon) to 1.0 and scaling oxygen and nitrogen appropriately. The result of this calculation is shown in Figure S4.1. The mean distances seen for oxygen and nitrogen closely match previous observations (17).

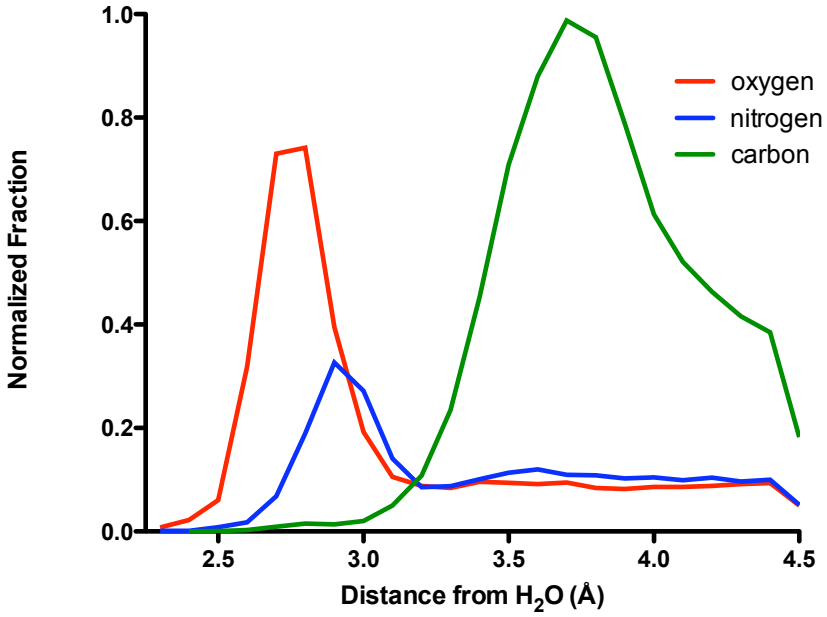

Figure S4.1: Histogram of the distance from low B-factor waters ($B < 30.0$) to oxygen, nitrogen, and carbon in a protein. Counts were normalized by the sphere volume then the maximum value for these counts was set to 1.0.

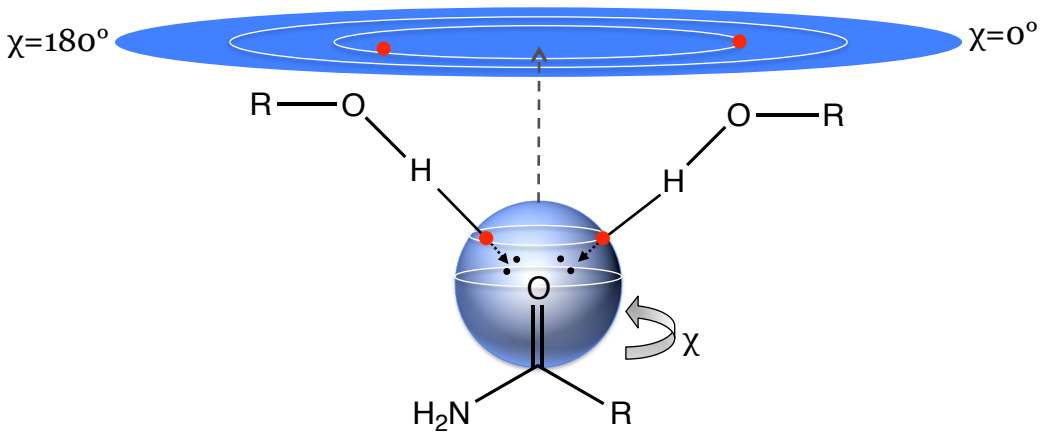

Figure S4.2: Diagram of how a Lambert azimuthal equal-area projection is performed for a hydrogen bond donated to an sp^2 acceptor oxygen on a protein side chain. A vector is the donor hydrogen to the acceptor oxygen. The intersection point of this vector on a unit sphere around the acceptor oxygen is plotted onto a two-dimensional plane. The χ angle is defined as the torsion angle defined by R, the carbonyl carbon, the acceptor oxygen and the hydrogen.

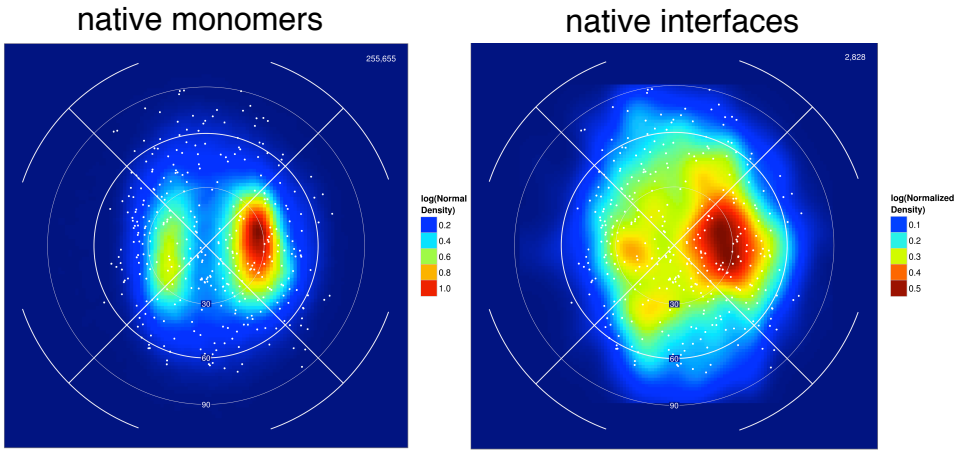

Figure S4.3: Location of polar hydrogens donating hydrogen bond to the protein backbone carbonyl oxygen shown as a Lambert azimuthal equal-area projection. Locations of the hydrogen for native monomer and native interfaces is shown as a density. Positions of the hydrogens in designed interfaces are shown as white dots in both density plots. The number of points used in the density is shown in the upper right. Hydrogen bonds are only counted if the sequence separation of the donor and the acceptor is greater than 5 residues.

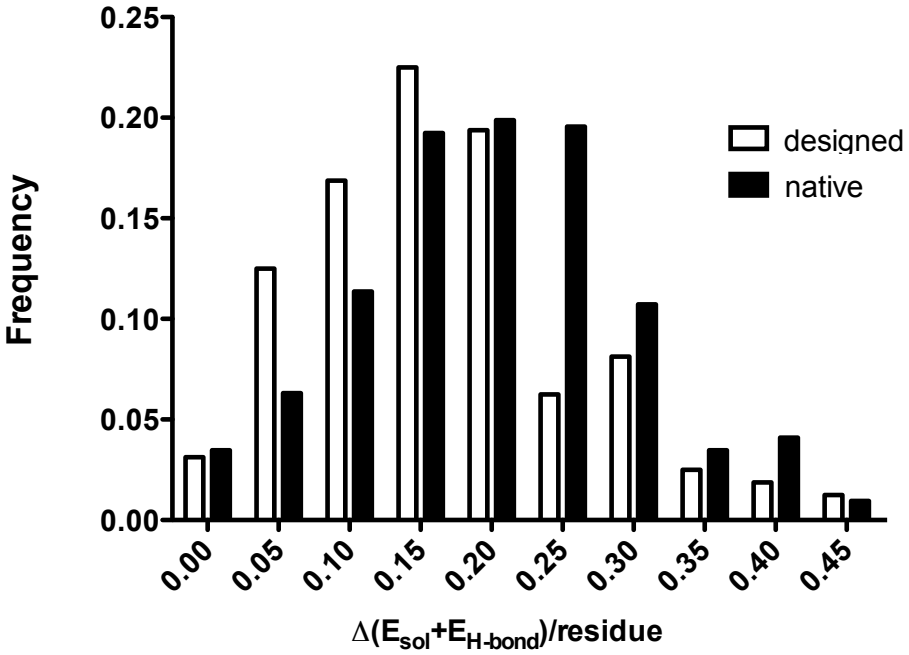

Figure S4.4: Change in solvation energy offset by hydrogen bonding upon binding. The change in solvation energy (ΔE_{sol}) plus hydrogen bond energy (ΔE_{H-bond}) per interface residue for designed and native interfaces.

4.5.2 Protein structures and models used

Table S4.1: PDB IDs of native heterodimers and homodimers used in this analysis. All of the structures listed below can be read and minimized by Rosetta. Chains were selected based on those described by Zhanhua *et al.* (25).

Heterodimer (Chains)	Homodimer (Chains)
1ACB (E,I)	1A4I (A,B)
1AVA (A,C)	1A4U (A,B)
1AVW (A,B)	1AA7 (A,B)
1AXI (B,A)	1AD1 (A,B)
1AY7 (A,B)	1ADE (A,B)
1B27 (A,D)	1AFW (A,B)
1BLX (A,B)	1ALK (A,B)
1BND (A,B)	1AOR (A,B)
1BPL (B,A)	1AQ6 (A,B)
1BRB (E,I)	1AUO (A,B)
1BRL (A,B)	1BBH (A,B)
1BVN (P,T)	1BD0 (A,B)
1C1Y (A,B)	1BH5 (A,B)
1CGI (E,I)	1BJW (A,B)
1CSE (E,I)	1BMD (A,B)
1CT4 (E,I)	1BXG (A,B)
1CXZ (A,B)	1C6X (A,B)
1D4V (B,A)	1CBK (A,B)
1D4X (A,G)	1CDC (A,B)
1D6R (A,I)	1CHM (A,B)
1DFJ (E,I)	1CNZ (A,B)
1DHK (A,B)	1COZ (A,B)
1DS6 (A,B)	1CQS (A,B)
1DTD (A,B)	1D1G (A,B)
1DZB (A,X)	1DOR (A,B)
1E44 (B,A)	1DPG (A,B)
1E96 (B,A)	1DQP (A,B)
1EAI (A,C)	1DQT (A,B)
1EAY (A,C)	1DVJ (A,B)
1EFV (A,B)	1EAJ (A,B)
1EM8 (A,B)	1EBL (A,B)
1EUC (B,A)	1EHI (A,B)
1EUV (A,B)	1EKP (A,B)
1F2T (A,B)	1EN5 (A,B)
1F34 (A,B)	1EN7 (A,B)
1F5Q (A,B)	1EOG (A,B)
1F5R (A,I)	1EV7 (A,B)
1F60 (A,B)	1EWZ (A,C)
1FCD (A,C)	1EXQ (A,B)
1FFG (B,A)	1EYV (A,B)
1FIN (A,B)	1EZ2 (A,B)
1FR2 (B,A)	1F13 (A,B)
1FT1 (B,A)	1F17 (A,B)
1FYH (A,B)	1F4Q (A,B)
1G4U (S,R)	1F6D (A,B)

Continued on next page...

Table S4.1 – continued from previous page

Heterodimer (Chains)	Homodimer (Chains)
1G4Y (R,B)	1F89 (A,B)
1GL4 (A,B)	1FC5 (A,B)
1GPW (A,B)	1FJH (A,B)
1GUA (A,B)	1FL1 (A,B)
1H1S (A,B)	1FP3 (A,B)
1H2A (L,S)	1FUX (A,B)
1H2S (A,B)	1FWL (A,B)
1H31 (A,B)	1FYD (A,B)
1HE1 (C,A)	1G0S (A,B)
1HL6 (A,B)	1G1A (A,B)
1HX1 (A,B)	1G1M (A,B)
1I1R (A,B)	1G64 (A,B)
1IAR (B,A)	1G8T (A,B)
1IBR (B,A)	1GD7 (A,B)
1ITB (B,A)	1GGQ (A,B)
1J7D (B,A)	1H8X (A,B)
1JIW (P,I)	1HDY (A,B)
1JKG (B,A)	1HJ3 (A,B)
1JLT (B,A)	1HJR (A,C)
1JQL (A,B)	1HQO (A,B)
1JTD (A,B)	1HSJ (A,B)
1JTP (A,L)	1HSS (A,B)
1JTT (A,L)	1I0R (A,B)
1JW9 (B,D)	1I2W (A,B)
1K9O (E,I)	1I4S (A,B)
1KA9 (F,H)	1I8T (A,B)
1KI1 (B,A)	1IPI (A,B)
1KSH (A,B)	1IRI (A,B)
1KTZ (B,A)	1J30 (A,B)
1KU6 (A,B)	1JD0 (A,B)
1KXP (D,A)	1JMV (A,B)
1KXV (A,C)	1JOG (A,B)
1L4Z (A,B)	1JP3 (A,B)
1LFD (B,A)	1JR8 (A,B)
1LUJ (A,B)	1JV3 (A,B)
1LW6 (E,I)	1JYS (A,B)
1M4U (A,L)	1K3S (A,B)
1M9E (A,D)	1K6Z (A,B)
1MA9 (A,B)	1K75 (A,B)
1MEE (A,I)	1KGN (A,B)
1MG9 (B,A)	1KIY (A,B)
1N0L (A,B)	1KSO (A,B)
1NF3 (A,C)	1L5X (A,B)
1NLV (A,G)	1LBQ (A,B)
1NPE (A,B)	1LHP (A,B)
1NQI (B,A)	1LHZ (A,B)
1NRJ (B,A)	1LNW (A,B)
1NW9 (B,A)	1LQ9 (A,B)
1O6S (A,B)	1M0W (A,B)
1OHZ (A,B)	1M3E (A,B)
1OKK (D,A)	1M4I (A,B)
1ONQ (A,B)	1M6P (A,B)

Continued on next page...

Table S4.1 – continued from previous page

Heterodimer (Chains)	Homodimer (Chains)
1OO0 (A,B)	1M7H (A,B)
1OP9 (B,A)	1M98 (A,B)
1OPH (B,A)	1M9K (A,B)
1OR0 (B,A)	1MI3 (A,B)
1OXB (A,B)	1MJH (A,B)
1P2J (A,I)	1MKB (A,B)
1P2M (A,B)	1MNA (A,B)
1P5V (A,B)	1N1B (A,B)
1PDK (A,B)	1N2A (A,B)
1PPF (E,I)	1N2O (A,B)
1PQZ (A,B)	1N80 (A,B)
1PVH (A,B)	1NA8 (A,B)
1Q40 (B,A)	1NFZ (A,B)
1QAV (B,A)	1NU6 (A,B)
1R0R (E,I)	1NW1 (A,B)
1R8S (E,A)	1NWW (A,B)
1RE0 (B,A)	1NY5 (A,B)
1RJ9 (A,B)	1O4U (A,B)
1RKE (A,B)	1OAC (A,B)
1S0W (A,C)	1ON2 (A,B)
1S1Q (A,B)	1OR4 (A,B)
1SCJ (A,B)	1ORO (A,B)
1SGD (E,I)	1OTV (A,B)
1SHW (B,A)	1OX8 (A,B)
1SLU (B,A)	1P3W (A,B)
1SMP (A,I)	1P43 (A,B)
1SPB (S,P)	1P6O (A,B)
1STF (E,I)	1PE0 (A,B)
1SVX (B,A)	1PJQ (A,B)
1TA3 (B,A)	1PN0 (A,C)
1TE1 (B,A)	1PN2 (A,B)
1TMQ (A,B)	1PP2 (R,L)
1TX4 (A,B)	1PT5 (A,B)
1UBK (L,S)	1Q8R (A,B)
1UGH (E,I)	1QFH (A,B)
1UJZ (B,A)	1QHI (A,B)
1US7 (B,A)	1QMJ (A,B)
1USU (A,B)	1QR2 (A,B)
1UUZ (D,A)	1QXR (A,B)
1UZX (A,B)	1QYA (A,B)
1V74 (A,B)	1R5P (A,B)
1VG0 (A,B)	1R7A (A,B)
1WQ1 (G,R)	1R8J (A,B)
1YCS (B,A)	1R9C (A,B)
1YVN (A,G)	1REG (X,Y)
2HBE (B,A)	1RN5 (A,B)
2KIN (A,B)	1RQL (A,B)
2NGR (B,A)	1RVE (A,B)
2SIC (E,I)	1RYA (A,B)
2SNI (E,I)	1S2Q (A,B)
2TEC (E,I)	1S44 (A,B)
3FAP (A,B)	1SCF (A,B)

Continued on next page...

Table S4.1 – continued from previous page

Heterodimer (Chains)	Homodimer (Chains)
3YGS (P,C)	1SMT (A,B)
4SGB (E,I)	1SO2 (A,B)
	1SOX (A,B)
	1TLU (A,B)
	1TRK (A,B)
	1UC8 (A,B)
	1V26 (A,B)
	2DAB (A,B)
	2GSA (A,B)
	2HHM (A,B)
	2NAC (A,B)
	2SQC (A,B)
	3LYN (A,B)
	3SDH (A,B)
	7AAT (A,B)
	8PRK (A,B)
	9WGA (A,B)

Table S4.2: List of computational design models used here and the experimental results for each one. All design models from our laboratory are listed here along with the successful designs from other labs. HA design models are available from Fleishman *et al* (20), Pdar/Prb models were obtained from Karanicolas *et al* (24). If the designed complex was crystalized a PDB ID is listed.

Project	Model	Experimental result	Success definition	PDB
β-strand homodimer	1cc8_3742	oligomer	failure	
	1cc8_0966	oligomer	failure	
	1cc8_1894	oligomer	failure	
	1cc8_4097	oligomer	failure	
	1cc8_LeuZipper	monomer	failure	
	1cc8_4579	monomer	failure	
	βdimer1	dimer $K_D = 1$ μM	strong	3zy7
	βdimer2	dimer low expression	weak	
	βdimer3	monomer	failure	
	βdimer4	monomer	failure	
RalA binder design	1JL1_anti_des1	no binding	failure	
	1JL1_parl_des1	no binding	failure	
	1sc0_des1	dimer	failure	
Rac1 binder design	1c1y_tf	no binding	failure	
	1c1y_v	no binding	failure	
	11y_wy	no binding	failure	
Ubiquitin binder design	1JJV.ubq.11_0005	binds 30 μM / nonspecific	failure	
	1Y8C.ubq.123_0002	$K_D > 50$ μM	failure	
	2ODV.ubq.9_0003	$K_D > 50$ μM	failure	
Zn^{+2}-mediated heterodimer	2d4x_0028	binds with Zn^{+2} 20 uM	weak	
	2d4x_0320	40 uM	failure	
	2fz4_design	$K_D > 100$ μM	failure	
	2onu_design	aggregates	failure	
Zn^{+2}-mediated homodimer	1g2r_334	oligomer	failure	
	1rz4_436	no expression	failure	
	1yzm_329	binds with Zn^{+2} $K_D = 30$ nM	strong	3v1c 3v1e
	2il5_335	oligomer	failure	
	1he9_180	oligomer	failure	
	2a90_308	monomer/dimer - poor solubility	failure	
	2d4x_161	always monomer	failure	
	2qov_414	monomer w/o zinc, tet with zinc	failure	
UbcH7 and Ankyrin designs	ubch7_10266	no binding	failure	
	helical_peptide	no binding	failure	
Ubc12 Zn Ankyrin designs	0032_I141A	aggregates with zinc	failure	
	0097_D108A_D99G	aggregates with zinc	failure	
PAK1 binder design	1i2t_233	no binding	failure	
	spider_roll	binds $K_D = 100$ μM	weak	
	1i2t_1212	$K_D = 330$ μM	failure	
	1i2t_3533	not soluble	failure	
	s032	not soluble	failure	
	s037	$K_D = 160$ μM	failure	
	mbp_17	no binding	failure	
	mbp_42	no binding	failure	
	1mn8_17567	not soluble	failure	
	1mn8_4957	not soluble	failure	

Continued on next page...

Table S4.2 – continued from previous page

Project	Model	Experimental result	Success definition	PDB
β-pix anchor design	bpix1	100 uM	failure	
	bpix2	155 uM	failure	
	bpix3	148 uM	failure	
G-tail C-term helix	gtail1	non-specific binding	failure	
	gtail2	non-specific binding	failure	
	gtail3	non-specific binding	failure	
	gtail4	non-specific binding	failure	
GoLoco extension	1255	no binding	failure	
	1680	no binding	failure	
	4091	no binding	failure	
	aa_0273	no binding	failure	
	aa_0951	binds $K_D = 800$ nM	strong	2xns
	aa_0971	no binding	failure	
	aa_0976	no binding	failure	
Pdar/Prb	Design_11	$K_D < 10$ nM wrong orrientation	weak	3q9n
HA binder	HB36	Evolved binding to$K_D = 4$ nM	strong	3r2x
	HB80	Evolved binding to$K_D = 3$ nM	weak	
Helical bundle		tetramer as designed	strong	1rh4

4.5.3 Analysis protocol

The following XML script was used as input for interface analysis:

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <s12_prime weights="score12prime"/>
  </SCOREFXNS>
  <TASKOPERATIONS>
    <RestrictToInterfaceVector name=vectorTask chain1_num=1 chain2_num=2
        CB_dist_cutoff=10.0 nearby_atom_cutoff=5.5 vector_angle_cutoff=65.0
        vector_dist_cutoff=8.0/>
    <InitializeFromCommandline name=cmdTask/>
    <IncludeCurrent name=currentTask/>
  </TASKOPERATIONS>
  <FILTERS>
  </FILTERS>
  <MOVERS>
    <InterfaceAnalyzerMover name=fullanalyze scorefxn=s12_prime packstat=1
        pack_input=0 pack_separated=1 jump=1 tracer=0 use_jobname=1 resfile=0
        task_operations=cmdTask,currentTask,vectorTask/>
  </MOVERS>
  <APPLY_TO_POSE>
  </APPLY_TO_POSE>
  <PROTOCOLS>
    <Add mover_name=fullanalyze/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

4.5.4 Hydrogen bond features

The following XML script was used as input to extract hydrogen bond geometry for the input proteins.

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <s weights=score12prime/>
  </SCOREFXNS>
  <MOVERS>
    <ReportToDB name=features_reporter db="features_hbonds.db3"
        sample_source="Interfaces">
      <feature name=ResidueFeatures/>
      <feature name=ResidueSecondaryStructureFeatures/>
      <feature name=PdbDataFeatures />
      <feature name=HBondFeatures scorefxn=s />
      <feature name=HBondParameterFeatures scorefxn=s/>
    </ReportToDB>
  </MOVERS>
  <PROTOCOLS>
    <Add mover_name=features_reporter/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

This will result in the output of a SQL database named `features_hbonds.db3`.

References

1. Mandell, D. J. and Kortemme, T. (2009) Computer-aided design of functional protein interactions. Nature chemical biology **5**, 797–807

2. Karanicolas, J. and Kuhlman, B. (2009) Computational design of affinity and specificity at protein-protein interfaces. Current opinion in structural biology **19**, 458–63

3. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. Science (New York, N.Y.) **282**, 1462–7

4. Havranek, J. J. and Harbury, P. B. (2003) Automated design of specificity in molecular recognition. Nature structural biology **10**, 45–52

5. Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature **458**, 859–864

6. Fleishman, S. J., Whitehead, T. A., Strauch, E. M., Corn, J. E., Qin, S., Zhou, H. X., Mitchell, J. C., Demerdash, O. N., Takeda-Shitaka, M., Terashi, G., Moal, I. H., Li, X., Bates, P. A., Zacharias, M., Park, H., Ko, J. S., Lee, H., Seok, C., Bourquard, T., Bernauer, J., Poupon, A., Aze, J., Soner, S., Ovali, S. K., Ozbek, P., Tal, N. B., Haliloglu, T., Hwang, H., Vreven, T., Pierce, B. G., Weng, Z., Perez-Cano, L., Pons, C., Fernandez-Recio, J., Jiang, F., Yang, F., Gong, X., Cao, L., Xu, X., Liu, B., Wang, P., Li, C., Wang, C., Robert, C. H., Guharoy, M., Liu, S., Huang, Y., Li, L., Guo, D., Chen, Y., Xiao, Y., London, N., Itzhaki, Z., Schueler-Furman, O., Inbar, Y., Potapov, V., Cohen, M., Schreiber, G., Tsuchiya, Y., Kanamori, E., Standley, D. M., Nakamura, H., Kinoshita, K., Driggers, C. M., Hall, R. G., Morgan, J. L., Hsu, V. L., Zhan, J., Yang, Y., Zhou, Y., Kastritis, P. L., Bonvin, A. M., Zhang, W., Camacho, C. J., Kilambi, K. P., Sircar, A., Gray, J. J., Ohue, M., Uchikoga, N., Matsuzaki, Y., Ishida, T., Akiyama, Y., Khashan, R., Bush, S., Fouches, D., Tropsha, A., Esquivel-Rodriguez, J., Kihara, D., Stranges, P. B., Jacak, R., Kuhlman, B., Huang, S. Y., Zou, X., Wodak, S. J., Janin, J., and Baker, D. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol **414**, 289–302

7. Sidhu, S. S. and Koide, S. (2007) Phage display for engineering and analyzing protein interaction interfaces. Current opinion in structural biology **17**, 481–7

8. Colwill, K. and Gräslund, S. (2011) A roadmap to generate renewable protein binders to the human proteome. Nature Methods **8**, 551–8

9. Ofran, Y. and Rost, B. (2003) Analysing six types of protein-protein interfaces. Journal of molecular biology **325**, 377–87

10. Schreiber, G. (2002) Kinetic studies of proteinprotein interactions. Current Opinion in Structural Biology **12**, 41–47

11. Reichmann, D., Phillip, Y., Carmi, A., and Schreiber, G. (2008) On the contribution of water-mediated interactions to protein-complex stability. Biochemistry **47**, 1051–60

12. Sheinerman, F. B. and Honig, B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces. Journal Of Molecular Biology **318**, 161–177

13. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic acids research **35**, W375–83

14. Lewis, S. M. and Kuhlman, B. A. (2011) Anchored design of protein-protein interfaces. PloS one **6**, e20872

15. Le Grand, S. M. and Merz, K. M. (1993) Rapid Approximation to Molecular Surface Area via the Use of Boolean Logic and Look-Up Tables. Journal Of Computational Chemistry **14**, 349–352

16. Leaver-Fay, A., Butterfoss, G., Snoeyink, J., and Kuhlman, B. (2007) Maintaining solvent accessible surface area under rotamer substitution for protein design. Journal Of Computational Chemistry **28**, 1336–1341

17. Matsuoka, D. and Nakasako, M. (2009) Probability distributions of hydration water molecules around polar protein atoms obtained by a database analysis. The journal of physical chemistry. B **113**, 11274–92

18. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-chain Amide Orientation. Journal Of Molecular Biology **285**, 1735–1747

19. Jha, R. K., Leaver-Fay, A., Yin, S., Wu, Y., Butterfoss, G. L., Szyperski, T., Dokholyan, N. V., and Kuhlman, B. (2010) Computational design of a PAK1 binding protein. Journal Of Molecular Biology **400**, 257–270

20. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science **332**, 816–821

21. Sammond, D. W., Bosch, D. E., Butterfoss, G. L., Purbeck, C., Machius, M., Siderovski, D. P., and Kuhlman, B. (2011) Computational design of the sequence and structure of a protein-binding peptide. Journal of the American Chemical Society **133**, 4190–2

22. Der, B. S., Machius, M., Miley, M. J., Mills, J. L., Szyperski, T., and Kuhlman, B. (2012) Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. Journal of the American Chemical Society **134**, 375–385

23. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A., and Kuhlman, B. (2011) Computational design of a symmetric homodimer using β-strand assembly. Proceedings of the National Academy of Sciences of the United States of America **108**, 20562–7

24. Karanicolas, J., Corn, J., Chen, I., Joachimiak, L., Dym, O., Peck, S., Albeck, S., Unger, T., Hu, W., Liu, G., Delbecq, S., T.áMontelione, G., P.áSpiegel, C., Liu, D., and Baker, D. (2011) A de novo protein binding pair by computational design and directed evolution. Molecular Cell **42**, 250–260

25. Zhanhua, C., Gan, J. G.-k., Lei, L., Sakharkar, M. K., and Kangueane, P. (2005) Protein subunit interfaces: heterodimers versus homodimers. Bioinformation **1**, 28

26. Xu, D., Tsai, C. J., and Nussinov, R. (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. Protein engineering **10**, 999–1012

27. Kortemme, T., Morozov, A. V., and Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. Journal of molecular biology **326**, 1239–59

28. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., and Baker, D. (2004) Computational redesign of protein-protein interaction specificity. Nature Structural & Molecular Biology **11**, 371–379

29. Joachimiak, L. A., Kortemme, T., Stoddard, B. L., and Baker, D. (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. Journal of molecular biology **361**, 195–208

30. Kapp, G. T., Liu, S., Stein, A., Wong, D. T., Reményi, A., Yeh, B. J., Fraser, J. S., Taunton, J., Lim, W. A., and Kortemme, T. (2012) Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair. Proceedings of the National Academy of Sciences of the United States of America

31. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G., and Herrmann, C. (2004) Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proceedings of the National Academy of Sciences of the United States of America **101**, 9223–8

32. Filchtinski, D., Sharabi, O., Rüppel, A., Vetter, I. R., Herrmann, C., and Shifman, J. M. (2010) What makes Ras an efficient molecular switch: a computational, biophysical, and structural study of Ras-GDP interactions with mutants of Raf. Journal of molecular biology **399**, 422–35

33. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P., and Kuhlman, B. (2007) Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. Journal of molecular biology **371**, 1392–404

34. Borgo, B. and Havranek, J. J. (2012) Automated selection of stabilizing mutations in designed and natural proteins. Proceedings of the National Academy of Sciences of the United States of America **109**, 1494–9

35. Lippow, S. M., Wittrup, K. D., and Tidor, B. (2007) Computational design of antibody-affinity improvement beyond in vivo maturation. Nature biotechnology **25**, 1171–6

36. Potapov, V., Reichmann, D., Abramovich, R., Filchtinski, D., Zohar, N., Ben Halevy, D., Edelman, M., Sobolev, V., and Schreiber, G. (2008) Computational Redesign of a Protein-Protein Interface for High Affinity and Binding Specificity Using Modular Architecture and Naturally Occurring Template Fragments. Journal Of Molecular Biology **384**, 109–119

37. Reynolds, K. A., Hanes, M. S., Thomson, J. M., Antczak, A. J., Berger, J. M., Bonomo, R. A., Kirsch, J. F., and Handel, T. M. (2008) Computational redesign of the SHV-1 β-Lactamase/β-Lactamase inhibitor protein interface. Journal Of Molecular Biology **382**, 1265–1275

38. Shifman, J. M. and Mayo, S. L. (2002) Modulating calmodulin binding specificity through computational protein design. Journal Of Molecular Biology **323**, 417–423

39. Yosef, E., Politi, R., Choi, M. H., and Shifman, J. M. (2009) Computational Design of Calmodulin Mutants with up to 900-Fold Increase in Binding Specificity. Journal Of Molecular Biology **385**, 1470

40. Hu, X., Wang, H., Ke, H., and Kuhlman, B. (2007) High-resolution design of a protein loop. Proceedings of the National Academy of Sciences of the United States of America **104**, 17668–73

41. Mandell, D. J., Coutsias, E. A., and Kortemme, T. (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nature methods **6**, 551–2

Chapter 5

Conclusion

This dissertation illustrates progress in the computational design of protein-protein interactions and presents a new way of overcoming some of the challenges in interface design. β-strand pairing at an interface can ensure orientation specificity and satisfy hydrogen bonding potential. The accurate design of a symmetric homodimer is a result of this approach (Chapter 2). It is also possible to find proteins with exposed strands that are complementary to two target proteins, RalA and PCSK9 (Chapter 3). An analysis of successes and failures in computational interface design shows that there has been progress devising new strategies to create novel interactions, however most designs fail to interact with their target (Chapter 4). All of these stories point to the conclusion that it is difficult to design hydrogen bonds involving side-chains at protein-protein interfaces. β-strand interactions at an interface make it possible to minimize the need for designed hydrogen bonds. This chapter introduces additional uses of β-strand pairing to engineer new protein-protein interactions and some of the problems encountered using proteins with exposed strands. It concludes by addressing ways to improve interface design methodology and goals for computational protein-protein interface design.

5.1 Additional applications of β-strand interface design

Chapter 2 establishes the first computationally designed interaction involving β-strand pairing between two globular domains. Chapter 3 explores how this idea could be extended to design a heterodimeric interaction. The immediate goal is to design β-strand mediated

binding partners to a small Ras superfamily GTPase and PCSK9. These designed proteins could be used to alter RalA signaling in cells or lower blood LDL levels. Yeast display of the designed scaffold proteins will allow rapid screening and affinity maturation of the designs for their target. As highlighted in Figure 1.2, β-strand pairing is used to form many other natural interfaces. All of these could be targeted using the methods presented in Chapter 3.

One possible extension is to redesign a β-strand mediated homodimer to form a heterodimer. βdimer1 or another strand mediated homodimer could be used as the scaffold for design. The design method would be similar to that employed by Bolon *et al* to redesign α-helix pairing to favor formation of a heterodimer over the native homodimer (1). The design protocol will need to simultaneously optimize the energy of the desired heterodimeric complex while disfavoring the stability of the homodimer. A fitness function that mirrors this goal is described in the development of a `MultistateDesign` program in Rosetta (2). Initial attempts to redesign βdimer1 to form a heterodimer resulted in aggregation of the protein. A scaffold homodimer that is more malleable to design or a rapid way to experimentally test the heterodimer designs is required for further progress.

An ambitious project could be the design of a novel protein that interacts with a partner by β-strand pairing. This could be accomplished by creating an idealized partner strand to a protein of interest using the method described in Section 3.2.1. The remainder of the novel protein would then be built to stabilize the idealized strand fragment while presenting additional contacts to the target protein. Charged residues should be favored for design on the idealized strand to serve as protecting groups to prevent aggregation of the designed protein. This *de novo* protein would most likely have mixed α/β secondary structure because the design of novel all β protein has proved elusive (3).

5.2 Difficulties in β-strand interface design

One of the major impediments to designing additional β-strand mediated protein-protein interactions is the availability of scaffold proteins with complementary exposed strands. It is unsurprising that most proteins do not fit the desired criteria given previous observations

that proteins with exposed β-strands are negatively designed to avoid edge strand pairing(4). Although a search protocol found 1,100 exposed strands in a starting set of 5,500 protein structures (see figure 2.1 on page 31), only 50 of these were finally considered to be designable as homodimers. Of those 50, only two (PDB IDs: 1CC8 and 2A7B) were chosen as scaffolds for further design and experimental confirmation. This illustrates the difficulty of finding proteins with exposed β-strands suitable for design.

Many of the proteins identified as having exposed strands have specific problems, other than those described by Richardson and Richardson (4), that make them impractical as potential design scaffolds. An exposed strand can be occluded by atoms that do not appear because of missing density or the formation of a domain swap in the crystal structure (Figure 5.1A). Several identified strands in the asymmetric unit of a crystal structure actually form an intermolecular β-sheet in a homo-oligomeric complex (Figure 5.1A). Exposed strands are also identified at the N or C terminus of a potential scaffold. These regions can be highly flexible (Figure 5.1C), making it unreasonable to expect the strand to maintain conformation during design. While screening of scaffolds to find potential binding partners to RalA we identified several exposed β-strands forming crystal contacts to other exposed strands in the crystal lattice. This fact informed our decision to identify for and avoid while these scaffolds while picking candidate monomers for homodimer design. Better annotation of structures in the PDB would help alleviate many of these problems.

Proteins with exposed β-strands proved difficult to work with experimentally. Four of six designs based on the scaffold 1CC8 formed large oligomers in solution (Table S4.2). Of the four designs of 2A7B (βdimer1-4), two expressed poorly and could not be obtained in quantities sufficient for additional experimental characterization. Redesigns of βdimer1 to form a heterodimer led to the protein aggregating uncontrollably in solution. Redesigns of a protein to form a heterodimer with a GTPase were less problematic, most likely because the targeted interfaces were polar (Tables 3.1 and S4.2).

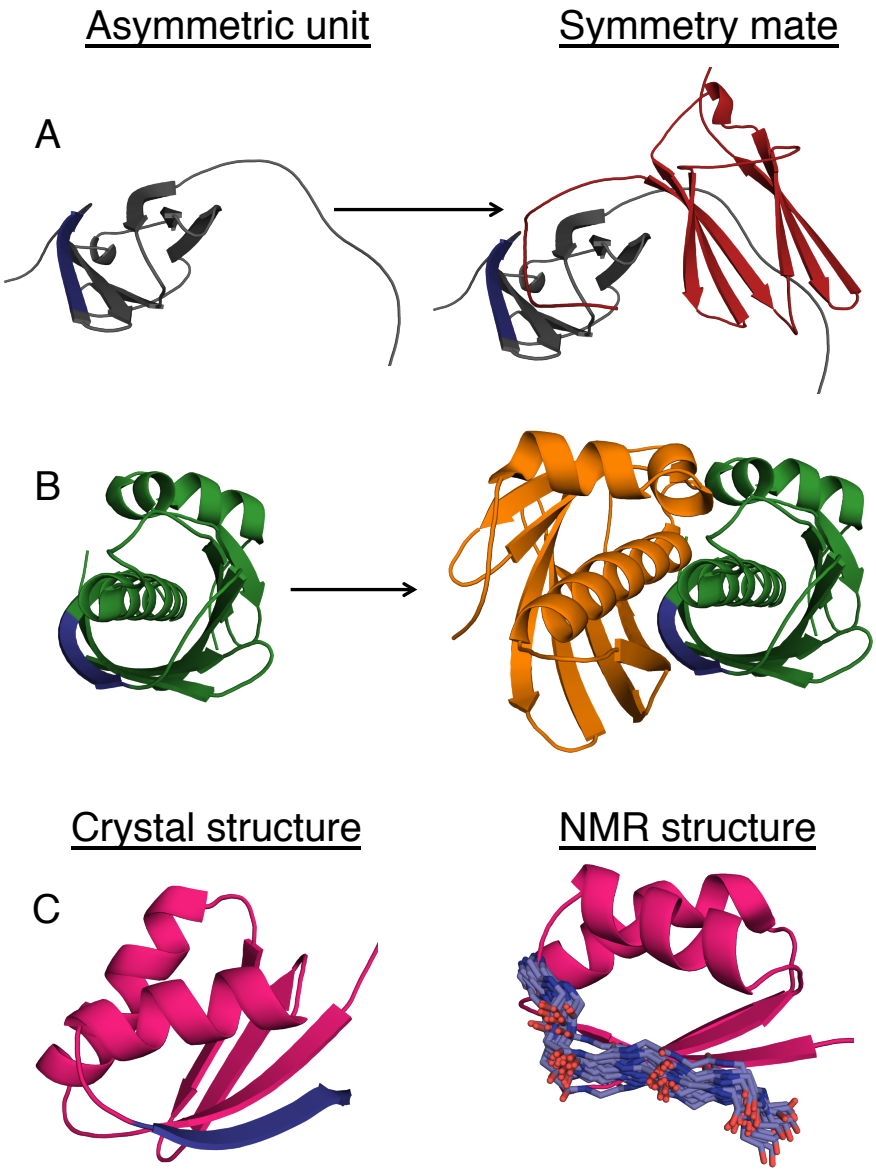

Figure 5.1: Problems frequently encountered when identifying exposed β-strands from protein crystal structures. β-strands identified as solvent exposed are shown in blue. A) A strand appears to be exposed but the overall conformation of the extended loops in the asymmetric unit does not seem realistic (PDB ID: 1WWC). Building in symmetry related protein molecules reveals a strand swapped dimer that now partially occludes the exposed strand. B) An exposed strand in the asymmetric unit of a protein (PDB ID: 1SC0) is clearly involved in forming a β-strand mediated homodimer when the appropriate symmetry mate is modeled. C) Exposed strands are frequently identified at the flexible N-terminus or C-terminus of a protein (PDB ID: 1CC8). These regions can be highly flexible as observed in an NMR structure of the same protein (PDB ID: 2GGP).

5.3 Improvements in interface design methodology

The high failure rate of computational interface design (Table 4.1) indicates that improvements are needed in both design methodology and metrics for discriminating design simulation outputs. Successful interface design has been made possible by devising new strategies within existing energy functions and search methods. Among these are ways to ensure proper orientation of an interface using either β-strands pairing (5) or metal coordination (6; 7). Other methods involve grafting known side-chain hot-spots onto new scaffolds (8; 9; 10; 11), transferring know binding epitopes to a different protein (12; 13; 14), or fitting a helix into a hydrophobic groove (10; 15; 12). The need for specialized strategies to create novel interfaces underscores the difficulty of the problem. None of the methods highlighted above could be described as a general method for designing any interaction.

Energy functions for computational design need to be improved to make interface design a more tractable goal. Rosetta has difficulty designing native-like hydrogen bonds (Chapter 4) and modeling electrostatic complementation at protein interfaces (16; 10). Attempts to reweight the existing scoring terms in the Rosetta energy function failed to improve sequence recovery or prediction of $\Delta\Delta G_{binding}$, suggesting that the energy terms for hydrogen bonding, Lennard-Jones interactions and electrostatic interactions need to be altered to improve interface design. The Rosetta community is currently parameterizing new potential functions for hydrogen bonding and electrostatics. The use of these new potentials could allow for successful design of polar interactions that closely resemble native heterodimers.

Alternative, more computationally expensive, simulations have been used to discriminate and rank Rosetta designed models. A molecular dynamics protocol successfully ranked the activities of several Rosetta designed enzymes performing a Kemp elimination reaction (17). Rusio *et al* found that protein dynamics could distort the active site conformation of a computationally designed retrol-aldolase and lead to lower activity than expected (18). A set of protein-protein interface metrics including knowledge based interaction networks, molecular mechanics force fields, and electrostatic encounter complex modeling can help determine which interface designs that are unlikely to form a complex (16; 19). Additional computational met-

rics or robust molecular mechanics simulations could help eliminate designs that are likely to fail. However, given the low cost of ordering genes and use of yeast display to screen designed interactions, time intensive discrimination methods might not be worthwhile.

The successes in *de novo* computational protein interaction design suggest there are a few general guidelines that appear work across design goals. Rosetta is most successful when designing small hydrophobic interfaces. The successes thus far involve interactions of α-helices with other α-helices or β-strand pairing, thus the scaffold secondary structure chosen for design should match that of the targeted interface. With the exception of β-strands pairing, buried polar atoms are detrimental to a designed interface. Even buried polar atoms involved in hydrogen bonds are not favorable due to Rosetta's difficulty in designing hydrogen bonds with native-like geometry. Finally, it is imperative to have experimental methods to rapidly evaluate and improve a successful design. Fleishman *et al* used yeast display to test as many designs in two years as our laboratory had in seven 4.1.

5.4 Future of computational interface design

As discussed in Chapter 4 there are several remaining challenges for computational design to successfully recapitulate the diversity seen in natural protein complexes. Despite these fundamental difficulties, there are foreseeable contributions that interface design could make to biological engineering. The rise of synthetic biology has illustrated the need for new, predictable, protein-protein interactions (20). Computational protein-protein can contribute new tools to enable design of novel biological functions.

The predominant focus of computational protein interface design has been on creating and modifying dimeric protein-protein interactions (21). Oligomers with three or more subunits represent over 40% of proteins in *E. coli* (22). Computational interface design has, with one notable exception (23), not been able to recapitulate the abundance of oligomeric complexes seen in nature. Accurate design of an oligomeric scaffold could allow for for precise control of orientation and function a pathway in cells (24). Dueber *et al* demonstrated that modular protein scaffolds can be used to control metabolic flux (25). Computational interface design

could also be applied to the creation of a large symmetric homooligomer. Rosetta's ability to model a user defined symmetry (26) and clear rules for the formation of symmetric complexes (27; 22) should allow for the design of a multi-subunit structure similar to hemoglobin or bacterial microcompartments (28). Most previous attempts to make β-strand mediated assemblies have resulted in uncontrolled fibrillization (29). The ability to accurately design β-strand pairing could make it possible to create new assemblies that are well behaved.

Engineered protein-protein interactions can be as a tool to predictably manipulate cell functions (21). One of the challenges is creating orthogonal interactions that avoid cross talk with natural cellular pathways (20). Several groups have used computational to create second-site suppressor mutations that make an interaction specific for the redesigned pair and avoid interacting with a native target (30; 31; 32; 33). An important next step is the design a signaling hub that interacts with various partners (21). The second-site supressor strategy would not be useful for proteins that serve as signaling hubs and share a common interface with several downstream effectors, such as the Ras superfamily. A recently developed `MultistateDesign` protocol in Rosetta is able to optimize an arbitrary fitness function for an unlimited number of different desired and undesired interactions (2). `MultistateDesign` is an ideal tool to design orthogonal protein interactions. As a proof of concept, `MultistateDesign` was able to recover known mutations that ablate one interaction of but maintain two others at a common location on a signaling hub (2). It could be applied to other known interfaces to create a specificity switch or tune the affinity to alter a signaling output.

Protein-protein interface design could create new scaffolds for computational enzyme design. Several novel enzymes have been designed with Rosetta in recent years (34; 35; 36). However, the activity of the computationally enzymes are well below most natural metabolic enzymes. Alternative backbone conformations can provide a boost in activity of a computationally designed enzyme by increasing hydrophobic contacts with the substrate (37). The active site of many natural enzymes is formed between protein domains or at the interface of an oligomer (38; 39). A novel homodimer or other homo-oligomer could provide a better active site for enzyme design than is currently available. Small molecules from the crystallization solution were found at the interfaces of the computationally designed homodimers MID1 (7)

and βdimer1 (5) (Figure 5.2. These interfaces could be re-engineered to increase affinity or specificity for alternative small molecules. In fact, MID1 has hydrolase activity towards some substrates (40).

Fin.

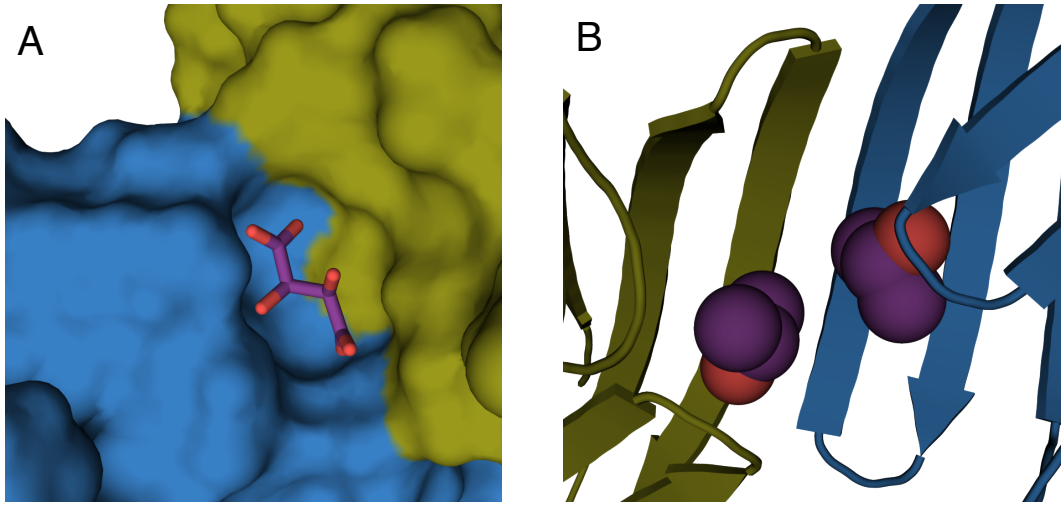

Figure 5.2: Ligands from the crystallization solutions are present at the interfaces of two computationally designed homodimers. A) A tartrate molecule (purple) is observed at the interface of the Zn^{+2}mediated homodimer MID1 (chain A purple, chain B olive) (PDB ID: 3V1C). B) Two symmetrically related isopropanol molecules (purple) are buried at the interface of βdimer1 (chain A purple, chain B olive) (PDB ID: 3ZY7).

References

1. Bolon, D. N., Grant, R. A., Baker, T. A., and Sauer, R. T. (2005) Specificity versus stability in computational protein design. Proceedings of the National Academy of Sciences of the United States of America **102**, 12724–9

2. Leaver-Fay, A., Jacak, R., Stranges, P. B., and Kuhlman, B. (2011) A generic program for multistate protein design. PloS one **6**, e20937

3. Shifman, J. M. (2008) Intricacies of Beta sheet protein design. Structure (London, England : 1993) **16**, 1751–2

4. Richardson, J. S. and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proceedings of the National Academy of Sciences of the United States of America **99**, 2754–9

5. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A., and Kuhlman, B. (2011) Computational design of a symmetric homodimer using β-strand assembly. Proceedings of the National Academy of Sciences of the United States of America **108**, 20562–7

6. Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B., and Tezcan, F. A. (2010) Metal templated design of protein interfaces. Proceedings of the National Academy of Sciences of the United States of America **107**, 1827–32

7. Der, B. S., Machius, M., Miley, M. J., Mills, J. L., Szyperski, T., and Kuhlman, B. (2012) Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed Protein Homodimer. Journal of the American Chemical Society **134**, 375–385

8. Liu, S., Liu, S. Y., Zhu, X. L., Liang, H. H., Cao, A. N., Chang, Z. J., and Lai, L. H. (2007) Nonnatural protein-protein interaction-pair design by key residues grafting. Proceedings of The National Academy of Sciences **104**, 5330–5335

9. Fleishman, S. J., Corn, J. E., Strauch, E. M., Whitehead, T. A., Karanicolas, J., and Baker, D. (2011) Hotspot-centric de novo design of protein binders. J Mol Biol **413**, 1047–1062

10. Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E.-M. M., Wilson, I. A., and Baker, D. (2011) Computational design of proteins targeting

the conserved stem region of influenza hemagglutinin. Science **332**, 816–821

11. Lee, S.-C., Park, K., Han, J., Lee, J.-j., Kim, H. J., Hong, S., Heu, W., Kim, Y. J., Ha, J.-S., Lee, S.-G., Cheong, H.-K., Jeon, Y. H., Kim, D., and Kim, H.-S. (2012) Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. Proceedings of the National Academy of Sciences **109**

12. Azoitei, M. L., Correia, B. E., Ban, Y.-E. A., Carrico, C., Kalyuzhniy, O., Chen, L., Schroeter, A., Huang, P.-S., McLellan, J. S., Kwong, P. D., Baker, D., Strong, R. K., and Schief, W. R. (2011) Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. Science (New York, N.Y.) **334**, 373–6

13. Correia, B. E., Holmes, M. A., Huang, P.-S., Strong, R. K., and Schief, W. R. (2011) High-resolution structure prediction of a circular permutation loop. Protein science : a publication of the Protein Society **20**, 1929–34

14. Azoitei, M. L., Ban, Y.-E. A., Julien, J.-P., Bryson, S., Schroeter, A., Kalyuzhniy, O., Porter, J. R., Adachi, Y., Baker, D., Pai, E. F., and Schief, W. R. (2012) Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. Journal of molecular biology **415**, 175–92

15. Sammond, D. W., Bosch, D. E., Butterfoss, G. L., Purbeck, C., Machius, M., Siderovski, D. P., and Kuhlman, B. (2011) Computational design of the sequence and structure of a protein-binding peptide. Journal of the American Chemical Society **133**, 4190–2

16. Fleishman, S. J., Whitehead, T. A., Strauch, E. M., Corn, J. E., Qin, S., Zhou, H. X., Mitchell, J. C., Demerdash, O. N., Takeda-Shitaka, M., Terashi, G., Moal, I. H., Li, X., Bates, P. A., Zacharias, M., Park, H., Ko, J. S., Lee, H., Seok, C., Bourquard, T., Bernauer, J., Poupon, A., Aze, J., Soner, S., Ovali, S. K., Ozbek, P., Tal, N. B., Haliloglu, T., Hwang, H., Vreven, T., Pierce, B. G., Weng, Z., Perez-Cano, L., Pons, C., Fernandez-Recio, J., Jiang, F., Yang, F., Gong, X., Cao, L., Xu, X., Liu, B., Wang, P., Li, C., Wang, C., Robert, C. H., Guharoy, M., Liu, S., Huang, Y., Li, L., Guo, D., Chen, Y., Xiao, Y., London, N., Itzhaki, Z., Schueler-Furman, O., Inbar, Y., Potapov, V., Cohen, M., Schreiber, G., Tsuchiya, Y., Kanamori, E., Standley, D. M., Nakamura, H., Kinoshita, K., Driggers, C. M., Hall, R. G., Morgan, J. L., Hsu, V. L., Zhan, J., Yang, Y., Zhou, Y., Kastritis, P. L., Bonvin, A. M., Zhang, W., Camacho, C. J., Kilambi, K. P., Sircar, A., Gray, J. J., Ohue, M., Uchikoga, N., Matsuzaki, Y., Ishida, T., Akiyama, Y., Khashan, R., Bush, S., Fouches, D., Tropsha, A., Esquivel-Rodriguez, J., Kihara, D., Stranges, P. B., Jacak, R., Kuhlman, B., Huang, S. Y., Zou, X., Wodak, S. J., Janin, J., and Baker, D. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol **414**, 289–302

17. Kiss, G., Röthlisberger, D., Baker, D., and Houk, K. N. (2010) Evaluation and ranking of enzyme designs. Protein science : a publication of the Protein Society **19**, 1760–73

18. Ruscio, J. Z., Kohn, J. E., Ball, K. A., and Head-Gordon, T. (2009) The influence of protein dynamics on the success of computational enzyme design. Journal of the American Chemical Society **131**, 14111–5

19. Moal, I. H., Agius, R., and Bates, P. A. (2011) Protein-protein binding affinity prediction on a diverse set of structures. Bioinformatics (Oxford, England) **27**, 3002–9

20. Clancy, K. and Voigt, C. A. (2010) Programming cells: towards an automated 'Genetic Compiler'. Current opinion in biotechnology **21**, 572–81

21. Mandell, D. J. and Kortemme, T. (2009) Computer-aided design of functional protein interactions. Nature chemical biology **5**, 797–807

22. Goodsell, D. S. and Olson, A. J. (2000) Structural symmetry and protein function. Annual review of biophysics and biomolecular structure **29**, 105–53

23. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom. Science (New York, N.Y.) **282**, 1462–7

24. Good, M. C., Zalatan, J. G., and Lim, W. A. (2011) Scaffold proteins: hubs for controlling the flow of cellular information. Science (New York, N.Y.) **332**, 680–6

25. Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., Prather, K. L. J., and Keasling, J. D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nature biotechnology **27**, 753–9

26. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and Andre, I. (2011) Modeling Symmetric Macromolecular Structures in Rosetta3. PLoS One **6**, e20450

27. Caspar, D. L. and Klug, A. (1962) Physical principles in the construction of regular viruses. Cold Spring Harbor symposia on quantitative biology **27**, 1–24

28. Tanaka, S., Sawaya, M. R., and Yeates, T. O. (2010) Structure and mechanisms of a protein-based organelle in Escherichia coli. Science (New York, N.Y.) **327**, 81–4

29. Yeates, T. O. and Padilla, J. E. (2002) Designing supramolecular protein assemblies. Curr Opin Struct Biol **12**, 464–470

30. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., and Baker, D. (2004) Computational redesign of protein-protein interaction specificity. Nature Structural & Molecular Biology **11**, 371–379

31. Joachimiak, L. A., Kortemme, T., Stoddard, B. L., and Baker, D. (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. Journal of molecular biology **361**, 195–208

32. Sammond, D. W., Eletr, Z. M., Purbeck, C., and Kuhlman, B. (2010) Computational design of second-site suppressor mutations at protein-protein interfaces. Proteins **78**, 1055–65

33. Kapp, G. T., Liu, S., Stein, A., Wong, D. T., Reményi, A., Yeh, B. J., Fraser, J. S., Taunton, J., Lim, W. A., and Kortemme, T. (2012) Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair. Proceedings of the National Academy of Sciences of the United States of America **109**, 5277–82

34. Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. Nature **453**, 190–5

35. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) De novo computational design of retro-aldol enzymes. Science (New York, N.Y.) **319**, 1387–91

36. Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science (New York, N.Y.) **329**, 309–13

37. Eiben, C. B., Siegel, J. B., Bale, J. B., Cooper, S., Khatib, F., Shen, B. W., Players, F., Stoddard, B. L., Popovic, Z., and Baker, D. (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. Nature biotechnology **30**, 190–2

38. Grishin, N. V. and Phillips, M. A. (1994) The subunit interfaces of oligomeric enzymes are conserved to a similar extent to the overall protein sequences. Protein science : a publication of the Protein Society **3**, 2455–8

39. Ali, M. H. and Imperiali, B. (2005) Protein oligomerization: how and why. Bioorganic & medicinal chemistry **13**, 5013–20

40. Der, B. S., Edwards, D. R., and Kuhlman, B. (2012) Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. Biochemistry **51**, 3933–40

Appendix A

Fluorescence polarization titrations and fitting protocol

A.1 Titration calculations

How to calculate volumes for your titrations.

We will start with:

$$M_T V_T = M_n V_n$$

Where:

$$\begin{aligned} M_T &= \text{titrant concentration} \\ V_T &= \text{volume of titrant to be added} \\ M_n &= \text{concentration after } n \text{ titrations} \\ V_n &= \text{total volume after } n \text{ titrations} \end{aligned}$$

Since we want to keep a running tally we can rewrite M_n and V_n as:

$$\begin{aligned} V_n &= V_T + V_{n-1} \\ M_n &= \frac{M_F(V_T + V_{n-1}) - M_{n-1}V_{n-1}}{V_T + V_{n-1}} \end{aligned}$$

Where M_F is the desired final concentration of titrant. Note that $M_F = M_n$ for all intents and purposes. Thus we can rewrite:

$$M_T V_T = \frac{M_F(V_T + V_{n-1}) - M_{n-1}V_{n-1}}{V_T + V_{n-1}}(V_T + V_{n-1})$$

Solving for V_T:

$$V_T = \frac{(M_{n-1} - M_F)V_{n-1}}{M_F - M_T}$$

or

$$V_{T_n} = \frac{(M_{n-1} - M_{F_n})V_{n-1}}{M_{F_n} - M_{T_n}}$$

Now that you know this, you can put together a table (one of the few times that Excel is acceptable to use) that calculates what your titrations (V_{add}) need to be assuming you know the initial volume in your cuvette (V_i), the concentration of titrant (M_T) and the final concentration (M_{F_n}) you want at a given n titration. Use the equation for V_{T_n} given above.

Titration #	M_F	V_{add}	V_{total}
0	0	0	V_i
1	M_{F_1}	V_{T_1}	$V_i + V_{T_1}$
2	M_{F_2}	V_{T_2}	$V_1 + V_{T_2}$
3	M_{F_3}	V_{T_3}	$V_2 + V_{T_3}$
n	M_{F_n}	V_{T_n}	$V_{n-1} + V_{T_n}$

A.2 Homodimer fitting protocol for Prism

The following is the code to fit to fluorescence polarization data (1) using Prism4 (Graphpad Software). In this case X is the concentration of unlabeled protein titrated into the cuvette and Y is the polarization measurement.

```
;write concentration of labeled protein here
labled=0.0034
Ptotal=labled+X
; Atotal is labled, X is total unlabled []

deltaPol=Polmax-Polmin
; polarization change

; now stuff for monomer concentration
;which you need the quadratic formula for
a=2
b=Kd
c=-Ptotal*Kd
MonomerConc=(-b+sqrt(b*b-4*a*c))/(2*a)

;now the dimer concentration can be written as
DimerConc=(Ptotal-MonomerConc)/2

;write heterodimer conc as fraction of total protein
HeteroDimer=2*(labled / Ptotal)*(X/Ptotal)*DimerConc
fractionHet=HeteroDimer / labled

;Now fit to this equation:
Y=deltaPol*fractionHet+Polmin
```

References

1. Lakowicz, J. R. (2006) Principles of Fluorescence Spectroscopy. Springer, 3rd edition. ISBN 978-0-387-31278-1