A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds

P. Benjamin Stranges[1] and Brian Kuhlman[1,2]*

[1]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Abstract: The accurate design of new protein–protein interactions is a longstanding goal of computational protein design. However, most computationally designed interfaces fail to form experimentally. This investigation compares five previously described successful *de novo* interface designs with 158 failures. Both sets of proteins were designed with the molecular modeling program Rosetta. Designs were considered a success if a high-resolution crystal structure of the complex closely matched the design model and the equilibrium dissociation constant for binding was less than 10 μ*M*. The successes and failures represent a wide variety of interface types and design goals including heterodimers, homodimers, peptide-protein interactions, one-sided designs (i.e., where only one of the proteins was mutated) and two-sided designs. The most striking feature of the successful designs is that they have fewer polar atoms at their interfaces than many of the failed designs. Designs that attempted to create extensive sets of interface-spanning hydrogen bonds resulted in no detectable binding. In contrast, polar atoms make up more than 40% of the interface area of many natural dimers, and native interfaces often contain extensive hydrogen bonding networks. These results suggest that Rosetta may not be accurately balancing hydrogen bonding and electrostatic energies against desolvation penalties and that design processes may not include sufficient sampling to identify side chains in preordered conformations that can fully satisfy the hydrogen bonding potential of the interface.

Keywords: protein interface design; computational protein design; Rosetta; hydrogen bond modeling

Introduction

The computational design of new protein–protein interactions has proven to be a difficult challenge.[1,2] Experimental measurements have shown that the majority of designed interactions do not form tight complexes ($K_D < 10$ μM)[3] or bind in an alternate conformation to the design model.[4] However, there has been exciting progress as a handful of *de novo* designed interfaces have been shown to bind with submicromolar binding affinities and adopt the intended binding orientation.[5–9] Here, we compare models that interact as predicted to those that fail with the goal of identifying common themes between the two sets. For instance, are current design methods more or less likely to succeed when designing interfaces enriched in polar or nonpolar amino acids?

One recent study sought to improve design selection methodology by asking the computational
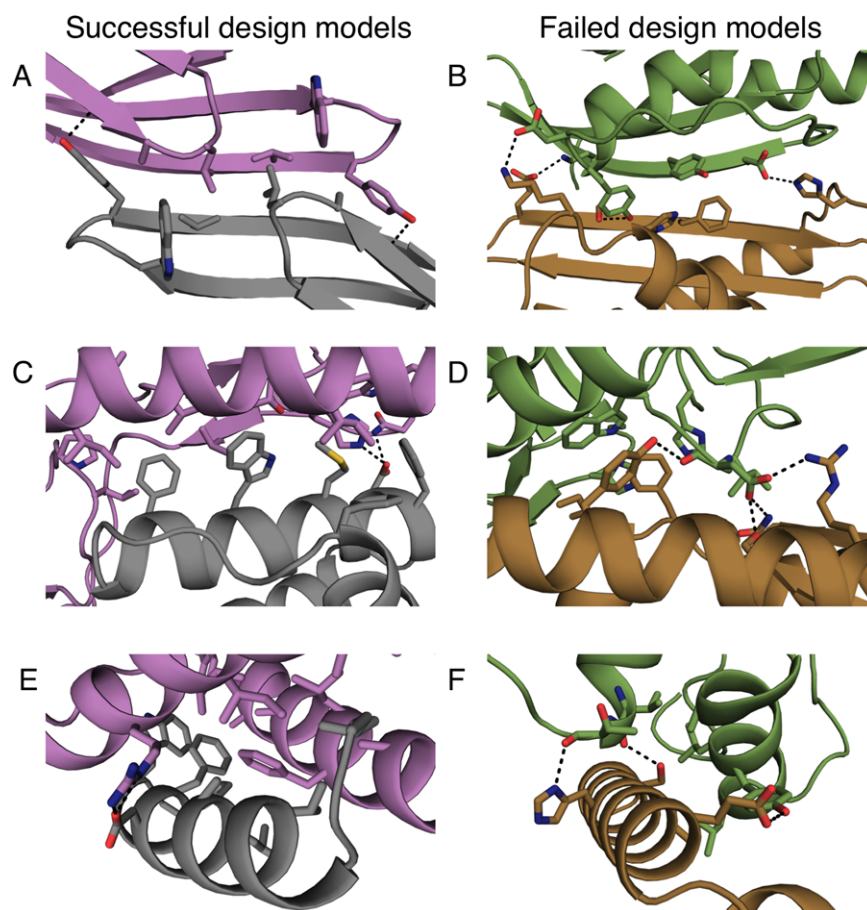

Figure 1. Examples of successful (left) and unsuccessful (right) protein interface design models. Separate chains of successful designs are shown in purple and gray; the different chains of failed models are colored green and brown; dashed black lines represent interface spanning side-chain involved hydrogen bonds. The structures shown represent examples design models of β-strand mediated interface (A, B), two models targeting flu HA (C, D), and design of helix secondary structure to bind a target protein. The successful models shown are βdimer1 (A), HB36 (C), and GL$_{helix}$-4 (D). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

protein docking community to establish metrics that discriminate designed proteins that were known not to bind from natural interfaces.[3] 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 metrics failed to distinguish natural small hydrophobic interfaces from designed small hydrophobic interfaces.

Here, we focus on the differences between failed and successful designs in addition to comparing design models with native interfaces. We choose to investigate designs made with the molecular modeling program Rosetta because we have access to a large number of design models and many of the recent successful designs were made using Rosetta. Our data set contains five successful interface designs and 153 failures. The successes and failures represent a wide variety of design goals including the creation of both heterodimers and homodimers. In all cases, the design models were created using Rosetta's rotamer optimization algorithms and full atom energy

function to optimize contacts at the target interface. The Rosetta energy function emphasizes short range forces including steric repulsion, London dispersion forces, hydrogen bonding, and bond torsion strain.[10] Solvent is modeled implicitly with the pair-wise additive desolvation model from the EEF1 force field.[11]

In general we find that the designs are smaller and more hydrophobic than native protein interactions. Though most designs fail to form experimentally, the ones that successfully interact are dominated by hydrophobic packing interactions. All attempts to design polar, hydrogen bond rich, interfaces have failed to produce proteins that bind. We address possible causes and solutions to the discrepancies between designed and native protein–protein interfaces.

Results

Definition of a successful design

For the purpose of this study, the computational interface designs were divided into three categories,

Table 1. *The Numbers of Experimentally Tested Computational Protein–Protein Interfaces Examined in This Work*

Design goal	No. tested	Express/ soluble	Strong success	Weak success
PAK1 binders[14]	10	6	0	1
GTPase binders	6	6	0	0
$G_{\alpha i}$ binding peptides[5]	11	11	1	0
Ubiquitin or UbcH7 binders	5	5	0	0
Metal mediated homodimers[8]	8	6	1	0
Metal-mediated heterodimers	6	5	0	1
β-strand-mediated homodimers[7]	10	10	1	1
FNIII to SH3-domain	3	3	0	0
Flu-hemagglutinin binders[6,9]	88	73	2	0
Prb/Pdar[4]	11	11	0	1
Total	158	136	5	4

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 ($K_D \leq 10\ \mu M$) where an X-ray crystal structure matches the design model. Weak successes ($K_D \leq 100\ \mu M$) have moderate to high affinity and other experimental evidence that the interface forms as designed. Citations are given when available.

strong success, weak success, and failure. A strong success is defined as a high affinity interaction ($K_D \leq 10\ \mu M$) where the X-ray crystal structure of the complex closely matches the computational prediction. A weak success has at least a moderate affinity ($K_D \leq 100\ \mu M$) and either mutational or NMR chemical shift data suggest the interface forms as designed. A failed design does not meet the previous criteria. Figure 1 shows several successful and failed *de novo* protein interface design models. Table I shows a summary of how many designs satisfy either definition of success. A complete list of structures used is given in Supporting Information Table SI.

There are five examples in our list of models 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).[5] Two others are proteins that were redesigned to bind influenza hemagglutinin (HB36 and HB80).[6,9] The final two are redesigns of natural monomeric proteins to form symmetric homodimers via Zn^{+2} binding (MID1)[8] or β-strand pairing (βdimer1).[7] 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 model. Some successful interface designs that are not included in Table I area novel helical tetramer from Harbury *et al.*[12] and two large assemblies designed by King *et al.*[13] We choose to not include these designs in our analysis because the other design goals were the construction of dimers.

Four designed interfaces are classified as weak successes because there is not a crystal structure of the modeled complex. These include a low affinity binder to PAK1,[14] a Zn^{+2} mediated heterodimer and a β-strand mediated homodimer (βdimer2).[7] The designed interaction of Prb and Pdar is classified as a weak success because a crystal structure of the complex is a 180° rotation from the computational model.[4]

The designed interfaces are small

Natural dimeric interfaces broadly sample different contact areas ranging from a change in solvent accessible surface area ($\triangle SASA$) of 850–10,000 $Å^2$ (up to 7000 $Å^2$ for heterodimers) (Supporting Information Fig. 1).[15] The designs sample much smaller interfaces ranging from 850 to 2400 $Å^2$, with the majority of designs having an interface area between 1000 and 1600 $Å^2$ (Supporting Information Fig. S1 inset). Successful designs are represented over the range of designed interface sizes and the crystal structures of the successful designs show a similar $\triangle SASA$ to the design models. There have been no successful dimer designs where the interface area is over 1600 $Å^2$ suggesting that better sampling and additional effort is required to recapitulate the sizes of native complexes.

Native proteins have a similar interface energy density as calculated with Rosetta ($\triangle G_{bind}/\triangle SASA$) across all sizes of interfaces [Fig. 2(A,B)] while the designed interfaces vary in energy density depending on the size of the interface [Fig. 2(C)]. Larger designed interfaces tend to have a less favorable $\triangle G_{bind}/\triangle SASA$ than smaller ones. This observation suggests that, unlike native complexes, current sampling and design strategies are unable to create high-quality contacts across large interfaces. It should be noted that most of the protocols used to produce the computational designs allow rigid body motion between the protein chains but not substantial backbone rearrangement. Crystal structures of the successful designs maintain a similar $\triangle G_{bind}/\triangle SASA$ to the computational models (Supporting Information Fig. S2). However, $\triangle G_{bind}/\triangle SASA$ cannot be used to clearly separate failed from successful designs. Both the failed and successful designs have $\triangle G_{bind}/\triangle SASA$ values that are similar to native interfaces, which reflects the fact that $\triangle G_{bind}$ is optimized during the design process.

Success is not determined by packing quality

Proteins designed with Rosetta can exhibit lower packing quality than natural proteins.[16] Atomic packing defects at modeled protein interfaces can indicate that a complex is unlikely to form experimentally.[17] We analyzed the design models to determine if poor packing quality was responsible for the failure of designed interfaces to form. Two measures of packing contacts at a protein interface were used to interrogate the designs and native structures, a
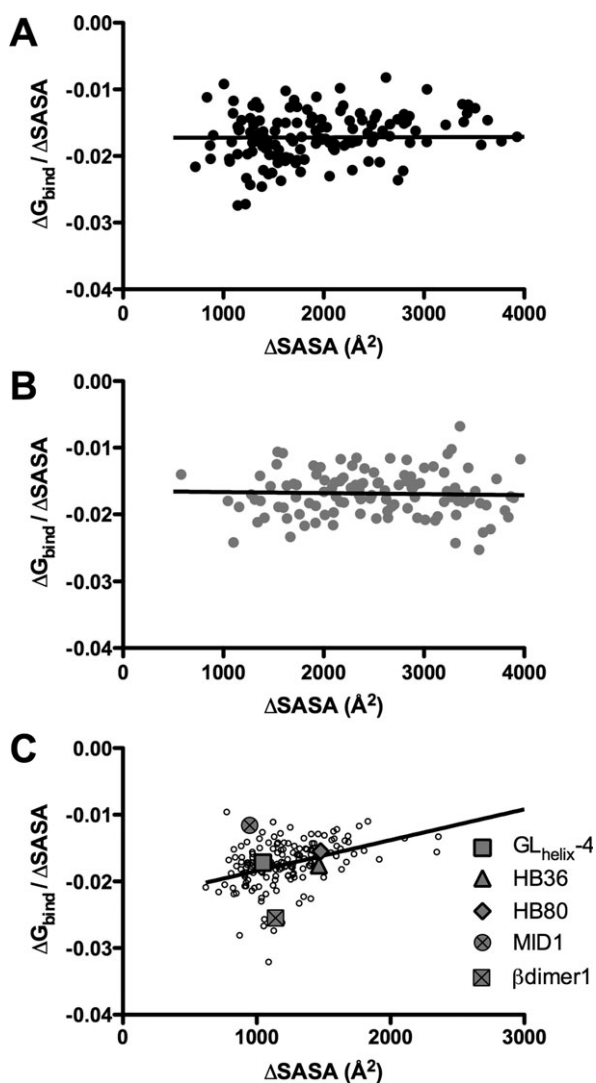
Figure 2. Interface energy density as computed by Rosetta for ($\triangle G_{bind}/\triangle SASA$) designed and natural interfaces. The change in SASA upon binding versus $\triangle G_{bind}/\triangle SASA$ is shown for native heterodimers (A), native homodimers (B), and all designed interfaces(C). Large points with gray interiors represent the successful design models. Least-squares lines were fit to each set of interfaces. The correlation coefficient for the design models is $r = 0.33$.

shape complementarity score[18] and the RosettaHoles score.[19] Residues at the interface of designed and native interfaces do not show a difference in packing quality (Fig. 3). Furthermore, the successful designs do not cluster towards better shape complementarity [Fig. 3(A)] or a better RosettaHoles score [Fig. 3(B)]. Rotamer optimization and minimization of crystal structures of the natural interfaces does not significantly alter the packing quality and the crystal structures of the successful designs span a similar range of packing quality to the design models (Supporting Information Fig. S3). Thus, these two metrics are not sufficient to distinguish which designs were likely to form from those that failed.

Natural protein–protein interfaces contain hot-spot residues that contribute a large amount to the binding energy of the complex.[20] We employed an alanine-scanning method in Rosetta to determine if designed interfaces exhibited a similar trend. The designed heterodimeric interfaces have a similar number of hot-spot residues to natural proteins (defined as $\Delta\Delta G_{bind} > +2.0$ REU), further demonstrating that Rosetta can form packing interactions similar to natural proteins (Supporting Information Fig. S4).

Successful designs have few polar interactions

Successful Rosetta designed interfaces have a low amount of polar area at their interface compared with many of the other computational models and native interfaces. The amount of $\triangle SASA$ that polar atoms contribute to an interface normalized by the total $\triangle SASA$ of the interface shows that the successful designs all have below 40% of their interface made up of polar atoms (Fig. 4). The successful design that has the largest fraction of polar interface area, βdimer1, has six main-chain to main-chain hydrogen bonds across the interface which account for a large amount of buried polar area. The
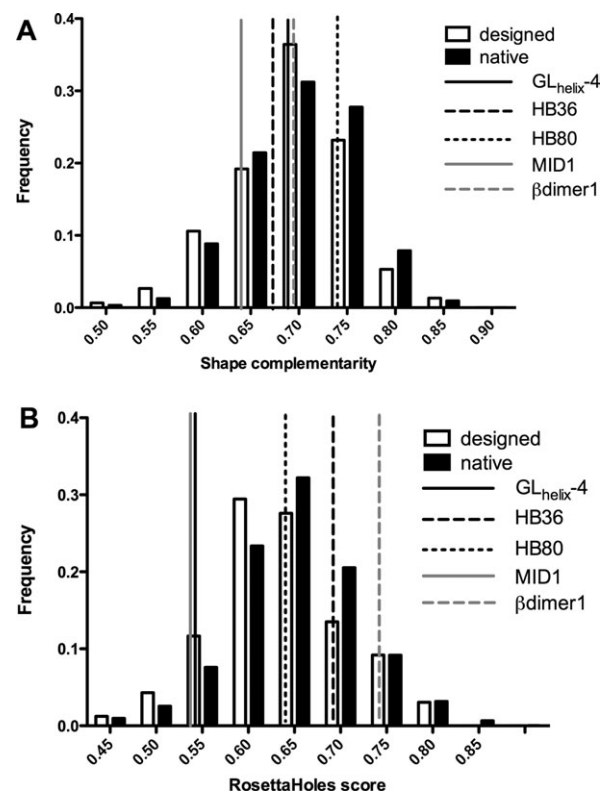

Figure 3. Packing quality measure of the design models and Rosetta minimized natural interfaces. Two independent measures of packing quality are shown; (A) the shape complementarity score[18] for the interface and (B) the RosettaHoles score for residues at the interface. For each metric a value of 1.0 represents perfect packing, while lower values represent packing defects. Lines represent the successful design models.
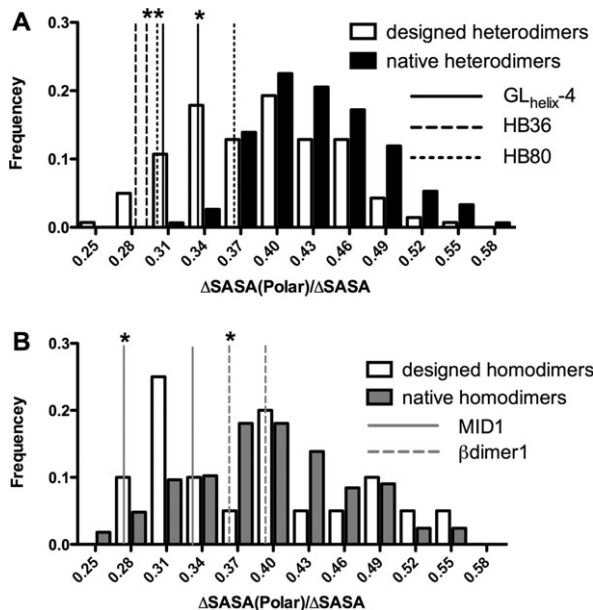
Figure 4. Polar content of designed and natural interfaces. The polar fraction of interface area is shown for designs versus heterodimers (A) and homodimers (B). Successful designs highlighted by lines. An asterisk above the line denotes the value for the crystal structure while no asterisk is present above the successful design models.

remainder of the βdimer1 interface is predominantly hydrophobic.

Overall, Rosetta designed interfaces have less contribution from polar interactions at the interface than natural dimers. This was noticed previously when comparing proteins designed to bind HA to natural heterodimers.[3] Even after including the design models from our lab, Rosetta designed heterodimers still have lower polar content at the interface compared with natural heterodimers [Fig. 4(A)]. The amount of polar interface area for designed and natural homodimers is similar [Fig. 4(B)]. The proteins designed to bind HA targeted a hydrophobic region on HA, thus raising the possibility that those designs skew the data set to disfavor successful polar interactions. However, both the designed HA binders and the proteins designed in our lab include designs that span a range of polar content ranging from ∼30–50% of the interface area (Supporting Information Fig. S5).

The design of polar residues at an interface can result in the burial of a polar atom without a hydrogen-bonding partner. Native interfaces tend to have no more than two buried, unsatisfied, polar atoms per 1000 Å^2 of interface [Fig. 5(A)]. The design models have a similar number of buried unsatisfied polar atoms as native interfaces. The crystal structures of three of the strong successes (GL$_{helix}$-4, MID1, and βdimer1) have no buried unsatisfied polar atoms at the interface [Fig. 5(A)]. The design models for these three interfaces also have no buried unsatis-

fied polar atoms (Supporting Information Fig. S6A). The two successful HA binders (HB36 and HB80) have buried unsatisfied polar atoms at their interface (Supporting Information Fig. S6A). Following affinity maturation, the crystal structure of these interfaces shows a drop in the number of buried unsatisfied polar atoms compared with the design models [Fig. 5(A)]. This reduction could indicate one way in which directed evolution was able to raise the affinity of the interaction. All design models that had more than two buried polar groups without a hydrogen bond partner in the design model failed to form high affinity complexes.

One of the most striking differences between successful designs, unsuccessful designs, and native complexes is the amount of binding energy, as calculated by Rosetta, contributed by buried hydrogen bonds involving side chains. The successful designs all have few or zero buried hydrogen bonds [Fig. 5(B)]. The design models of GL$_{helix}$-4, HB36, and HB80 each have one buried hydrogen bond across the interface (Supporting Information Fig. 6B). A buried hydrogen bond in GL$_{helix-4}$ was not observed
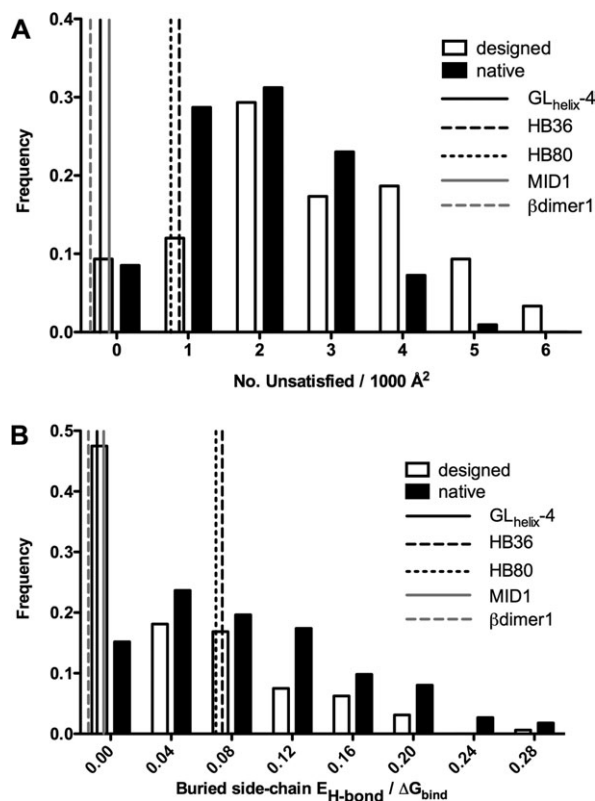

Figure 5. Buried polar atoms and buried hydrogen bonds at interfaces. Values for Rosetta minimized crystal structures are shown by lines. A: The number of buried polar atoms without a hydrogen-bonding partner per 1,000 Å^2 of interface area. B: The total energy of a buried, side-chain involved hydrogen bond at the interface as a fraction of total binding energy (△G$_{bind}$). HB36 and HB80 acquired an additional buried hydrogen bond due to mutations introduced by affinity maturation.

in the crystal structure. MID1 and βdimer1 have no buried side chain hydrogen bonds across the interface. One new buried hydrogen bond was introduced to HB36 and HB80 during affinity maturation of the computational design. In cases where multiple buried hydrogen bonds were present in the design model, the designed complex failed to form. This is not because buried hydrogen bonds are forbidden by the rules of physical chemistry, many native interfaces have multiple buried hydrogen bonds and a significant portion of the binding energy as calculated by Rosetta is derived from hydrogen bonds.

Successful interface designs made with computational programs other than Rosetta have also had interfaces that are predominately hydrophobic. For example, the *de novo* designed helical bundle RH4[12] has a polar △SASA fraction of 0.23, which is lower than the successful Rosetta designs described here as well as many native interfaces. It also has no buried unsatisfied polar atoms and no buried hydrogen bonds.

Discussion

These results indicate that successful Rosetta designed protein–protein interactions differ from unsuccessful designs and native interactions in the polar makeup of the interface. Designed interactions tend to be more hydrophobic and smaller than most natural protein–protein interfaces. The successful designs have less polar area at the interface when compared to most design models and few buried hydrogen bonds or unsatisfied polar atoms. Burying polar atoms, even those modeled to form hydrogen bonds appears detrimental to the success of a computational interface design.

The scarcity of polar interactions in the successful designs 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,[21–23] however, these redesigns have lower affinity than the wild type interaction. New hydrogen bonds can increase the affinity of a natural interaction in some cases, typically by designing a interface spanning salt bridge at the edge of an interface.[24,25] However, there are no buried salt bridges in the successful designs investigated here. Another strategy for increasing affinity involves replacing a polar residue with a nonpolar one, or a small hydrophobic residue with a larger one.[26] None of the examples of successful novel interface design derive a large portion of their interface from polar interactions. Unlike computational methods, nature is able to make protein interfaces with substantial polar area and hydrogen bond interactions [Fig. 4(A,B)].

There are more examples of successful computational redesign of natural protein–protein interactions for increased affinity[24–29] or altered specificity[21–23,30,31] 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.

Inaccuracies in the Rosetta energy function could account for the failure of polar designs. Rosetta does not explicitly model water during design, in part because a previous effort to model with explicitly solvated rotamers did not yield improvements in computational benchmarks.[32] Natural homodimers and heterodimers contain about 10 water molecules per 1000 Å^2 of interface area. On average 30% of these waters are buried from bulk solvent.[33] The inability to account for waters at polar interfaces could prevent computational methods from finding a sequence that allows for binding. We are unable to draw conclusions about the solvent content of the designed interfaces because the crystal structures for HB36, HB80, and $GL_{\text{helix-4}}$ were not determined at a resolution high enough to allow for accurate water placement near the designed interface. Another reason that polar designs fail could be the design of poor hydrogen bonding geometries. The Rosetta hydrogen bond energy function used to make the design models does not include a term to ensure that a hydrogen bond donor is in the plane of lone-pair electrons on acceptor carbonyl oxygens.[34] Some of the failed designs have interface spanning hydrogen bonds that are more than 60° away from optimal sp^2 acceptor geometry. In addition, none of the design protocols that produced the models investigated here, made use of long-range electrostatic interactions. Rosetta employs a course grain energy term that favors the proximity of residues with opposite charges. Including a more complex electrostatic potential can improve prediction of $\Delta\Delta G_{\text{bind}}$ in Rosetta.[9] An alternative target function, for instance optimizing for ΔG_{bind} instead of total energy, could also be a way to approach future design goals.[35]

Another reason that polar designs could be failing is insufficient sampling and stabilization of a binding competent conformation. Fleishman *et al.* previously noted that interface residues in natural complexes tend to favor a similar rotamer in both the bound and unbound form. Designed interfaces did not favor the bound rotamer in the unbound state.[36] In addition residues with three or four dihedral angles tend to undergo rotameric shifts upon binding,[37] suggesting that sampling large rotamer libraries for these residues might be necessary at

protein–protein interfaces. Failed designs are rarely investigated further to determine other possible reasons the design did not interact with the target. Mutating a large number of residues in the design process could destabilize the designed protein or alter the intrachain contacts such that the designed protein's conformation does not match the model. In fact, some MID1 structures show noticeable backbone rearrangements from the starting structure.[8] Experimental determination of structures of failed designs could help inform new design methodology.

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, residues in either α-helices or β-strands dominate all successful designs [Fig. 1(A,C,D)]. Three important challenges in *de novo* 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 a single interface larger than 1,600 Å^2; (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[38] and the development of methods to accurately design and model loops.[39,40] 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.

Materials and Methods

Set of designed interfaces

The computational models used in this analysis represent a wide array of interface design goals (Table I). The design models fall in two main categories: (i) design of one protein chain to bind to a natural target and (ii) design of both chains involved in an interaction to create a novel heterodimer or homodimer. The majority of the designs, 140 out of 158, 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, PAK1,[14] proteins involved in ubiquitin transfer, and influenza hemagglutinin from Fleishman *et al*.[6] Another 11 models represent the design of both the structure and sequence of a peptide to bind $G_{\alpha i1}$.[5] The second category is comprised of 18 redesigns of natural proteins to form homodimers mediated by metal binding[8] or β-strands,[7] and 11 models from Karanicolas *et al*. where both interface forming chains are designed to form a new heterodimer.[4]

Of the 59 designs from our laboratory 52 of them successfully expressed in *E. coli*. All designs made by our group are available in Supporting Information. Seventy-three of the 88 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.[15] 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 (Supporting Information Table SII).

Computational evaluation of protein interfaces

The natural and designed interactions were all minimized with Rosetta to make energy evaluations between them comparable. The minimized and X-ray crystal structures were then evaluated for several metrics including computed binding energy ($\Delta G_{\text{bind}} = E_{\text{AB}} - E_A - E_B$) buried solvent accessible surface area upon binding (ΔSASA) and buried unsatisfied polar atoms at the interface (discussed below). A full description of the computational protocols used and the command lines is given in Supporting Information.

Polar burial definition

Rosetta calculates SASA using the Le Grand and Merz method.[41] 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 a hydrogen-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[42] we chose to use atomic radii from Reduce[43] and a water probe radius of 1.2 Å to find buried polar atoms and hydrogen bonds. Buried and unsatisfied hydrogen bonds for the natural interfaces were calculated based on the conformation in the crystal structure because it has been observed that repacking a structure with Rosetta can increase the number of buried unsatisfied polars.[16]

Acknowledgements

The authors would like to think those who shared designed models and{ unpublished experimental results with us including Bryan Der, Ramesh Jha, Steven Lewis, and Deanne Sammond.

References

1. Mandell DJ, Kortemme T (2009) Computer-aided design of functional protein interactions. Nat Chem Biol 5:797–807.
2. Karanicolas J, Kuhlman B (2009) Computational design of affinity and specificity at protein-protein interfaces. Curr Opin Struct Biol 19:458–463.

Successful Interface Designs Avoid Polar Interactions

3. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, Mitchell JC, Demerdash ON, Takeda-Shitaka M, Terashi G, et al. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol 414: 289–302.

4. Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, et al. (2011) A de novo protein binding pair by computational design and directed evolution. Mol Cell 42:250–260.

5. Sammond DW, Bosch DE, Butterfoss GL, Purbeck C, Machius M, Siderovski DP, Kuhlman B (2011) Computational design of the sequence and structure of a protein-binding peptide. J Am Chem Soc 133:4190–4192.

6. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E-MM, Wilson IA, Baker D (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 332:816–821.

7. Stranges PB, Machius M, Miley MJ, Tripathy A, Kuhlman B (2011) Computational design of a symmetric homodimer using β-strand assembly. Proc Natl Acad Sci USA 108:20562–20567.

8. Der BS, Machius M, Miley MJ, Mills JL, Szyperski T, Kuhlman B (2012) Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. J Am Chem Soc 134:375–385.

9. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, et al. (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotech 30:543–548.

10. Rohl CA, Strauss CEM, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383:66–93.

11. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35:133–152.

12. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS (1998) High-resolution protein design with backbone freedom. Science 282:1462–7.

13. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science 336: 1171–1174.

14. Jha RK, Leaver-Fay A, Yin S, Wu Y, Butterfoss GL, Szyperski T, Dokholyan NV, Kuhlman B (2010) Computational design of a PAK1 binding protein. J Mol Biol 400:257–270.

15. Zhanhua C, Gan JG, Lei L, Sakharkar MK, Kangueane P (2005) Protein subunit interfaces: heterodimers versus homodimers. Bioinformation 1:28.

16. Borgo B, Havranek JJ (2012) Automated selection of stabilizing mutations in designed and natural proteins. Proc Natl Acad Sci USA 109:1494–1499.

17. Bahadur RP, Zacharias M (2008) The interface of protein-protein complexes: analysis of contacts and prediction of interactions. Cell Mol Life Sci 65:1059–1072.

18. Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. J Mol Biol 234: 946–950.

19. Sheffler W, Baker D (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein Sci 18: 229–239.

20. Irina S. Moreira PAFMJR (2007) Hot spots—a review of the protein-protein interface determinant amino-acid residues. Proteins 68:803–812.

21. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D (2004) Computational redesign of protein-protein interaction specificity. Nature Struct Mol Biol 11:371–379.

22. Joachimiak LA, Kortemme T, Stoddard BL, Baker D (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. J Mol Biol 361:195–208.

23. Kapp GT, Liu S, Stein A, Wong DT, Reményi A, Yeh BJ, Fraser JS, Taunton J, Lim WA, Kortemme T (2012) Control of protein signaling using a computationally designed GTPase/GEF orthogonal pair. Proc Natl Acad Sci USA 109:5277–5282.

24. Kiel C, Selzer T, Shaul Y, Schreiber G, Herrmann C (2004) Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. Proc Natl Acad Sci USA 101:9223–9228.

25. Filchtinski D, Sharabi O, Rüppel A, Vetter IR, Herrmann C, Shifman JM (2010) What makes Ras an efficient molecular switch: a computational, biophysical, and structural study of Ras-GDP interactions with mutants of Raf. J Mol Biol 399:422–435.

26. Sammond DW, Eletr ZM, Purbeck C, Kimple RJ, Siderovski DP, Kuhlman B (2007) Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. J Mol Biol 371:1392–1404.

27. Lippow SM, Wittrup KD, Tidor B (2007) Computational design of antibody-affinity improvement beyond in vivo maturation. Nature Biotech 25:1171–1176.

28. Potapov V, Reichmann D, Abramovich R, Filchtinski D, Zohar N, Ben Halevy D, Edelman M, Sobolev V, Schreiber G (2008) Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. J Mol Biol 384:109–119.

29. Reynolds KA, Hanes MS, Thomson JM, Antczak AJ, Berger JM, Bonomo RA, Kirsch JF, Handel TM (2008) Computational redesign of the SHV-1 β-Lactamase/β-Lactamase inhibitor protein interface. J Mol Biol 382: 1265–1275.

30. Shifman JM, Mayo SL (2002) Modulating calmodulin binding specificity through computational protein design. J Mol Biol 323:417–423.

31. Yosef E, Politi R, Choi MH, Shifman JM (2009) Computational design of calmodulin mutants with up to 900-fold increase in binding specificity. J Mol Biol 385:1470.

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

33. Rodier F, Bahadur RP, Chakrabarti P, Janin J (2005) Hydration of protein-protein interfaces. Proteins 60: 36–45.

34. Kortemme T, Morozov AV, Baker D (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J Mol Biol 326:1239–1259.

35. Leaver-Fay A, Jacak R, Stranges PB, Kuhlman B (2011) A generic program for multistate protein design. PloS One 6:e20937.

36. Fleishman SJ, Khare SD, Koga N, Baker D (2011) Restricted sidechain plasticity in the structures of native proteins and complexes. Protein Sci 20:753–757.

37. Kirys T, Ruvinsky AM, Tuzikov AV, Vakser IA (2012) Rotamer libraries and probabilities of transition between rotamers for the side chains in protein-protein binding. Proteins 80:2089–2098.

38. Sidhu SS, Koide S (2007) Phage display for engineering and analyzing protein interaction interfaces. Curr Opin Struct Biol 17:481–487.

39. Hu X, Wang H, Ke H, 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 16:1799–1805.

40. Mandell DJ, Coutsias EA, Kortemme T (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nature Meth 6:551–552.

41. Le Grand SM, Merz KM (1993) Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. J Comp Chem 14:349–352.

42. Matsuoka D, Nakasako M (2009) Probability distributions of hydration water molecules around polar protein atoms obtained by a database analysis. J Phys Chem B 113:11274–11292.

43. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 285:1735–1747.

Successful Interface Designs Avoid Polar Interactions