Journal of Structural Biology 185 (2014) 193–202

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Computational design of protein-small molecule interfaces

Brittany Allison [a], Steven Combs [a], Sam DeLuca [b], Gordon Lemmon [b], Laura Mizoue [c,f], Jens Meiler [a,b,d,e,f,g,*]

[a] Department of Chemistry, 7330 Stevenson Center, Station B 351822, Nashville, TN 37235, USA
[b] Chemical and Physical Biology Program, 340 Light Hall, Nashville, TN 37232, USA
[c] Department of Biochemistry, 607 Light Hall, Nashville, TN 37232, USA
[d] Department of Pharmacology, 476 Robinson Research Building, 2220 Pierce Avenue, Nashville, TN 37232, USA
[e] Department of Biomedical Informatics, 400 Eskind Biomedical Library, 2209 Garland Ave, Nashville, TN 37232, USA
[f] Center for Structural Biology, 465 21st Avenue South, Nashville, TN 37232, USA
[g] Institute for Chemical Biology, 896 Preston Research Building, Nashville, TN 37232, USA

ARTICLE INFO

Article history:
Available online 17 August 2013

Keywords:
Rosetta
RosettaLigand
Computational interface design
Protein-small molecule interaction
Ligand docking
Sequence optimization

ABSTRACT

The computational design of proteins that bind small molecule ligands is one of the unsolved challenges in protein engineering. It is complicated by the relatively small size of the ligand which limits the number of intermolecular interactions. Furthermore, near-perfect geometries between interacting partners are required to achieve high binding affinities. For apolar, rigid small molecules the interactions are dominated by short-range van der Waals forces. As the number of polar groups in the ligand increases, hydrogen bonds, salt bridges, cation–π, and π–π interactions gain importance. These partial covalent interactions are longer ranged, and additionally, their strength depends on the environment (e.g. solvent exposure). To assess the current state of protein-small molecule interface design, we benchmark the popular computer algorithm Rosetta on a diverse set of 43 protein–ligand complexes. On average, we achieve sequence recoveries in the binding site of 59% when the ligand is allowed limited reorientation, and 48% when the ligand is allowed full reorientation. When simulating the redesign of a protein binding site, sequence recovery among residues that contribute most to binding was 52% when slight ligand reorientation was allowed, and 27% when full ligand reorientation was allowed. As expected, sequence recovery correlates with ligand displacement.

1. Introduction

Engineering protein-small molecule interactions is key for advancement of several grand challenges in computational biology. Protein-small molecule interactions are the basis for enzymatic catalysis, receptor–small molecule signaling, and transporter selectivity and are thus essential for carrying out biological processes and maintaining overall homeostasis in the body. Designed proteins that bind small molecule targets can act as therapeutics by sequestering ligands, stimulating or extinguishing signaling pathways, delivering other molecules to sites of action, and serving as *in vivo* diagnostics (Golan et al., 2008). For example, small molecule depletion has been suggested as a strategy for treatment of prostate cancer (Knudsen and Scher, 2009), cocaine abuse (Zhu et al., 2006), and bacterial infection (Clifton et al., 2009). Proteins that bind small molecules also have applications in environmental chemistry and food chemistry as biosensors (Baeumner, 2003). Thus, the ability to engineer highly precise and specific interactions at protein interfaces can serve in many capacities.

Computational design of protein-small molecule interfaces continues to present challenges. Although the creation of new enzymes is a landmark achievement in protein design (Baker et al., 2008a,b, 2010; Zanghellini et al., 2006), the success rate is low and the designed proteins are poor catalysts compared to naturally-occurring enzymes. To help pinpoint the causes, a systematic study was conducted introducing mutations into the active site of three designed retro-aldolases (RA34, RA45, and RA95) derived from the TIM-barrel scaffold IGPS. In RA34 and RA95, mutations that increase substrate binding affinity and thereby enzymatic activity involve increases in side chain volume and hydrophobicity, including G233F/I/V/Y in RA34 (Wang et al., 2012) and T51Y, T83K, S110H, M180F and R182M in RA95 (Althoff et al., 2012). In contrast, many improvements to the RA45 design arose from large-to-small mutations including W8A/T/V, F133L, V159C, and R182V/I (Althoff et al., 2012). In all cases, key functional groups that engage the ligand are introduced or removed. These observations indicate that neither the hydrophobic packing nor the

* Corresponding author. Office: MRBIII, Room 5144 B, 465 21st Ave South, Nashville, TN 37232, USA. Mail: 7330 Stevenson Center, Station B 351822, Nashville, TN 37235, USA. Tel.: +1 (615) 936-5662. Fax: +1 (615) 936-2211.
E-mail address: jens@meilerlab.org (J. Meiler).
URL: http://www.meilerlab.org (J. Meiler).

1047-8477/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.08.003

positioning of substrate within the binding pocket were optimal in the initial designs. Similarly, a previously reported successful computational design of a protein-small molecule interface (Allert et al., 2004) did not withstand close examination (Hayden, 2009; Schreier et al., 2009).

Rosetta, a protein modeling software suite for protein structure prediction and design (Schueler-Furman et al., 2005), has been successfully used to tackle a number of interface design problems. Some of these successes include creating novel enzymes (Baker et al., 2008a,b, 2010), altering the specificity of protein-peptide (Sood and Baker, 2006), protein–DNA (Ashworth et al., 2010) and protein–protein interfaces (Kortemme et al., 2003), and designing proteins that bind a selected surface of a virus (Fleishman et al., 2011). Rosetta seeks to find the lowest energy conformation for a design by combining discrete side chain conformation (rotamer) optimization with Monte Carlo minimization (Schueler-Furman et al., 2005). This includes sampling random perturbations of the backbone torsion angles, rigid body degrees of freedom, and rotamer conformations, followed by an all-over local minimization to resolve clashes (Schueler-Furman et al., 2005). These methods enable much faster and larger exploration of sequence and conformational space compared to experimental methods such as phage display (Weng and DeLisi, 2002).

The energy function that Rosetta uses to discriminate between native-like and non-native-like atom arrangements includes a van der Waals-like attractive and repulsive potential, solvation term, hydrogen bonding potential, electrostatics potential, rotamer probability, and (φ, ψ) angle probabilities in the protein backbone (Meiler and Baker, 2006). The total energy of the system is computed as a weighted sum of all interactions with weights optimized through a series of benchmarks. All energy functions are pairwise decomposable (i.e. they depend on no more than two interacting partners). This design of the energy function maximizes algorithm speed since interaction energies can be pre-computed and stored. However, it also limits the accuracy of the energy function, particularly electrostatic and partial covalent interactions which vary greatly in strength depending on the environment of the interacting partners. Experimental characterization of some of the best scoring designs is used to validate and improve the computational protocols. In this way, both design successes and failures help test and expand our understanding of the fundamental forces involved in molecular recognition.

RosettaLigand is an application within Rosetta that was originally developed to dock small molecules into a protein with full protein and ligand flexibility (Baker and Davis, 2009; Meiler and Baker, 2006). In these studies, we expand RosettaLigand to include amino acid optimization (design) at the protein-small molecule interface. Using the full-atom energy function and Monte Carlo minimization procedure, RosettaLigand optimizes the small molecule position and protein side chain rotamers simultaneously (Meiler and Baker, 2006). RosettaLigand allows for protein backbone flexibility, side chain rotamer searching, and full ligand flexibility, all of which are necessary for accurately modeling the interface (Baker and Davis, 2009; Meiler and Baker, 2006). Fig.1 details each step of the ligand docking protocol. For each model, RosettaLigand calculates an 'interface energy' as the total score of the protein–ligand complex minus the total score of the apo-protein (Meiler et al., 2009). The accuracy of models in terms of ligand placement is determined by computing the root-mean-square distance (RMSD) over all ligand atoms between model and co-crystal structure. RosettaLigand is the foundation (Zanghellini et al., 2006) of a number of the successfully design enzymes (Jiang et al., 2008; Rothlisberger et al., 2008; Siegel et al., 2010) with the before-mentioned caveat that the computationally predicted residues are often sub-optimal even in the first shell surrounding the ligand. In order to understand its capabilities and limitations, the present work systematically

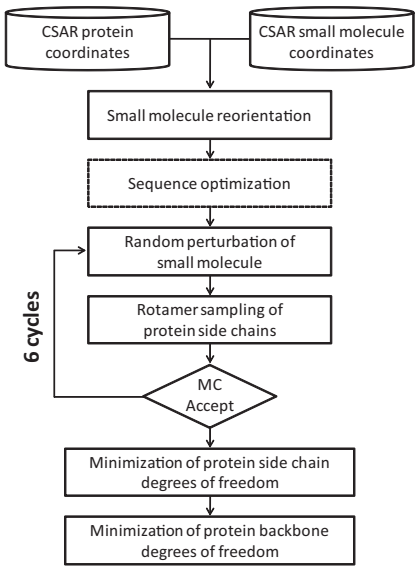

Fig.1. Flowchart of small molecule docking with design. The RosettaLigand protocol was modified to include interface design (dotted line box). From the input coordinates, the small molecule is allowed to rotate and translate before sequence optimization of nearby residues. After 6 cycles of small molecule perturbation, side chain rotamer sampling, and Monte Carlo (MC) minimization, a final gradient-based minimization of the protein is performed to resolve any clashes.

assesses RosettaLigand's ability to design protein-small molecule interfaces. This analysis is an important, and so far omitted, benchmark to identify design challenges that can currently be solved and to work towards improvements needed to achieve consistent success.

Recovering native protein-small molecule interfaces in sequence and conformation is a benchmark for designing novel interfaces. Creating new interfaces or even modifying existing ones requires computational tools that sample and select native-like interactions. In this study, we examine how RosettaLigand performs in sequence recovery within protein-small molecule interfaces while allowing for small molecule reorientation and side chain conformational changes. The benchmark consists of two parts. Part one tests overall sequence recovery when all residues within the protein-small molecule interface are allowed to change identity. Part two simulates a protein-smallmolecule design more closely by mutating up to five residues that contribute most to the interaction with the small molecule to alanine. This effectively removes the binding site's memory of the native ligand. In the design experiment a scoring bonus is given to the starting sequence. These experiments test RosettaLigand's ability to distinguish between native and non-native binding interaction and whether RosettaLigand can identify key mutations needed to bind the small molecule while limiting the total number of mutations. The results illustrate the types of ligands that Rosetta handles best and provide insights into weaknesses where continued method development is required.

2. Results and discussion

The setup of the experiments allows us to determine overall protein–ligand interface sequence recovery as well as an optimal strategy for re-designing proteins to recognize different small molecules using a minimal set of mutations. For this purpose separate measures for sequence recovery among the residues critical for ligand binding are determined. We investigate how sequence recovery varies with ligand size, binding affinity, and RosettaLigand interface energy. We appreciate that sequence recovery measures

have one critical limitation: they assume that the native protein-small molecule interface is optimal for tight small molecule binding, which is certainly incorrect. In result, if a position fails to recover to the native amino acid it can be because of an actual failure of the design algorithm or because the alternate amino acid is tolerated or even favorable in an actual protein-small molecule interface – a distinction that only the experiment can make. Therefore, we do not expect a 100% success rate for sequence recovery. This poses a dilemma for the development of protein design algorithms: at what point can we stop optimizing for increased sequence recovery? Ideally, one wants to capture native-like designs and interactions that would be seen in nature, but not to a point where the algorithm over-fits the designs. To circumvent part of this problem, we developed a Position-Specific Scoring Matrix (PSSM) recovery measure (DeLuca et al., 2011) which computes the fraction of residues that revert to an amino acid observed in evolution. PSSM recovery is a more robust measure of design success as it tolerates mutations that have been observed in evolution.

2.1. Experimental setup

A set of 43 high resolution protein-small molecule crystal structures were selected from the Community Structure-Activity Resource (CSAR) database and used directly in testing. In practice, however, the task is often to redesign a binding site to recognize a (different) specific small molecule. In this setting, one is interested in identifying the minimal number of mutations needed to achieve the desired functionality and avoiding additional mutations that provide little or no benefit. This can be achieved by including a 'favor native' residue bonus (FNRB) energy that must be overcome before a mutation is accepted.

The starting sequences in the CSAR benchmark set are already (close to) optimal for binding the target small molecule. Therefore, we created mutant proteins that are expected to have reduced or no binding to the target small molecule. First, the five residues that contribute most to small molecule binding according to the RosettaLigand energy function were determined and then sequentially mutated to alanine. Next, these artificial mutants were employed to test RosettaLigand's ability to recognize sub-optimal interactions and replace them with those that are optimal for binding.

A total of four experiments were conducted: (1) re-designing the protein-small molecule interface in the native protein without reorienting the ligand (design native), (2) re-designing the protein-small molecule interface in the native protein with ligand reorientation (dock/design native), (3) re-designing the protein-small molecule interface in the alanine mutants using a FNRB without reorienting the ligand (design alanine mutants), and (4) re-designing the protein-small molecule interface in the alanine mutants using a FNRB and ligand reorientation (dock/design alanine mutants). The latter experiment tests if RosettaLigand can identify critical mutations and distinguish them from arbitrary sequence changes. For each experiment, 1000 models were generated, filtered by RosettaLigand interface energy, and the top 50 were selected for analysis.

2.2. Trends for sequence recovery across all four experiments

The ligand RMSD, sequence recovery, number of mutations, and interface energy from each experiment were averaged to identify trends across experiments (Table 1). As expected, comparison of the Design (1, 3) vs. Dock/Design (2, 4) experiments shows that lower RMSDs and better interface scores are observed when an optimal ligand pose is inputted and allowed to move only slightly (0.1 Å translation, 2° rotation) versus when the ligand pose needs to be identified (2 Å translation, 360° rotation). The conformational space increases if the ligand is allowed to reorient and other binding poses with different sequences achieve favorable interface scores. Large changes in ligand position alter the interactions with the protein and encourage mutations. Unfortunately, other than extensive experimental studies, there is no way to test the plausibility of these alternative protein-small molecule interfaces. Applying a FNRB that favors retention of the native amino acid (3, 4) yields fewer mutations and higher sequence recovery only when the ligand is allowed full reorientation. The highest sequence recovery, lowest number of mutations, and best interface energy is observed in experiment (1) where the ligand is held in its approximate initial pose and no alanine mutations are introduced into the sequence. However, in the design of novel protein-small molecule interfaces, full ligand reorientation must be allowed in order for the ligand to search the entire binding pocket for optimal placement (4).

2.3. Trends in recovery of wild type (WT) residues across alanine mutants experiments

Similar results occur when introducing alanine mutations in the binding site. To maximize sequence recovery, these had to be converted to the native amino acid overcoming the FNRB. For each experiment, the ligand RMSD, sequence recovery, recovery of WT residue, retention of alanine, and mutation to another residue were determined for each model and averaged (Table 2). Since both of these experiments include a FNRB, the only variable is whether or not the ligand was allowed full reorientation. The penalty for allowing full ligand movement is larger in recovering the specific WT residue (51.7% vs. 26.8%) than the penalty in overall sequence recovery (58.9% vs. 47.6%). This result was not surprising because in cases where the ligand position is not recovered there is a minimal chance that the correct residue will be selected. Considering that the chance of randomly selecting the correct residue is 5% (1 out of 20), RosettaLigand's ability to recover 26.8% of WT residues under the most stringent conditions represents a significant improvement. Some designs have been selected to show the diversity of ligands that have high sequence and wild type recovery versus ones that have low recoveries (Fig.2).

2.4. Detailed analysis of the dock/design alanine mutants experiment (4)

Since the ultimate goal is to use RosettaLigand to design novel protein-small molecule interfaces, we took an in-depth look at

Table 2
Recovery of wild type (WT) residue from the alanine mutants experiments.

	Design alanine mutants (3)	Dock/design alanine mutants (4)
Ligand RMSD (Å)	0.7 ± 0.6[a]	2.5 ± 1.0[a]
Sequence recovery (%)	58.9 ± 14.4[a]	47.6 ± 11.8[a]
Alanine mutations recovered to WT (%)	51.7 ± 32.8	26.8 ± 20.5
Alanine mutations designed to other amino acids (%)	26.5 ± 25.1	36.8 ± 17.1
Alanine mutations remaining alanine (%)	21.8 ± 25.2	36.3 ± 22.4

[a] The slight differences in values compared to Table 1 are due to exclusion of data from the native complexes.

Table 1
Varying parameters of experiments and sequence recovery results.

	Design native (1)	Dock/design native (2)	Design alanine mutants (3)	Dock/design alanine mutants (4)
Ligand translation (Å)[a]	0.1	2.0	0.1	2.0
Ligand rotation (deg)[a]	2	360	2	360
Ligand RMSD (Å)	0.5 ± 0.4	2.2 ± 1.0	0.7 ± 0.6	2.4 ± 1.0
Sequence recovery (%)	64.0 ± 12.4	36.6 ± 11.9	59.8 ± 14.2	48.7 ± 11.9
Number of mutations	6.5 ± 2.5	11.5 ± 3.4	7.2 ± 2.6	9.3 ± 2.7
Interface energy (REU)[b]	−19.0 ± 5.4	−15.1 ± 3.5	−17.5 ± 5.2	−14.1 ± 3.5

[a] Reorientation allowed from initial pose during docking.
[b] REU – Rosetta energy units.

the results from the experiment that most closely resembles this scenario. Sequence recovery was plotted against a number of variables to see if RosettaLigand performs better with different types of ligand and/or protein properties (Fig.3, Fig.S1). As seen in previous results, sequence recovery decreases with increasing RMSD (Fig.3A). Binding pocket crowdedness measures how tightly packed the ligand is in binding pocket, calculated as the number protein/ligand atom pairs within 3 Å of each other, divided by the total number of ligand atoms. A high crowdedness indicates a ligand surrounded by protein contacts, whereas low crowdedness indicates that only a portion of the ligand is in contact with the protein. Sequence recovery was best when the ligand had 2–3 protein contacts per atom (Fig.3B). As crowdedness deviated from this range, recovery decreased, implying that RosettaLigand has difficulties in tightly packed as well as under-packed protein/small molecule interfaces. For the number of ligand hydrogen bond donors and acceptors, WT recovery remained consistent until there is a drop at 13+ donors/acceptors (Fig.3C). A complex hydrogen bonding network would be more difficult to recover than a simple network, so this was expected. For number of ligand rotatable bonds, ligands with 1, 2, or 3 rotatable bonds have the best recoveries (Fig.3D). A decrease in sequence recovery is observed with increasing ligand flexibility. Accordingly, also ligands with rigid ring systems have increased sequence recovery values (Fig.3E). This was expected, since rings provide the ligand a more defined shape, making it easier for RosettaLigand to identify the correct binding pose. LogP, a measure of lipophilicity comparing the concentration of ligand in octanol vs. water, shows the best recoveries for ligands with a logP around zero (between ∼−2.5 and 2.5, Fig.3F). As the ligand becomes more hydrophilic, sequence recovery decreases. Aside from a few outliers, as the number of ligand atoms increases, recovery decreases (Fig.S1A). Our interpretation is that larger ligands have fewer well-defined contacts that are more difficult to recover. Surprisingly, the number of residues considered for design has little impact on sequence recovery; one may expect that more residues in the binding pocket would decrease recovery, but this was not the case (Fig.S1B). Binding affinity showed little effect on recovering the interface (Fig.S1C). There appears to be a drop in recovery for very tight binders, however there are few of these complexes to begin with. Binding affinity normalized by ligand molecular weight does not influence recovery (Fig.S1D). Topological polar surface area (Fig.S1E) and van der Waals surface area (Fig.S1F) both show the same trend; as surface area increases, maximum recovery decreases. This was not surprising, considering that surface area and number of ligand atoms correlate with each other. Ligand interface energy correlates little with sequence recovery (Fig.S1G) even if the interface energies were normalized by small molecule molecular weight (Fig.S1H).

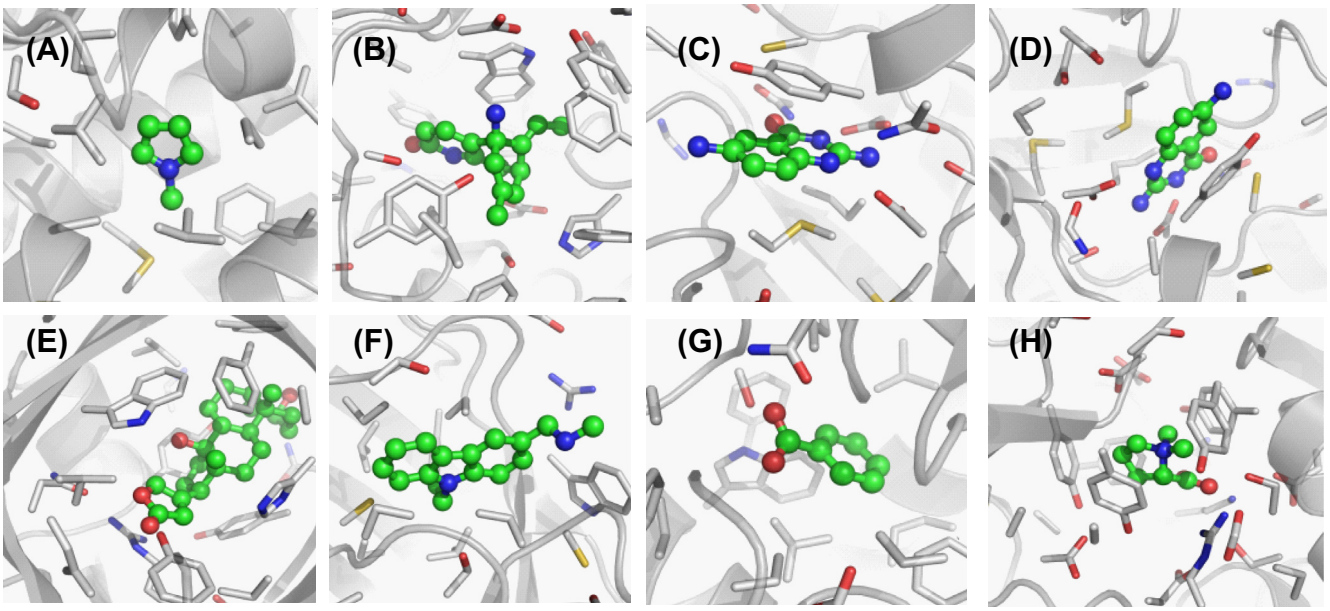

Fig.2. Examples of the best and worst designs from each experiment. Best designs from each experiment (A, B, C, D) are shown in contrast to the worst designs from each experiment (Panels E, F, G, H). For experiment 1 design native, a best design model had a sequence recovery of 90% (A), while a worst design model had a sequence recovery of 58% (E). For experiment 2 dock/design native, a best design model had a sequence recovery of 80% (B), while a worst design model had a sequence recovery of 51% (F). For experiment 3 design alanine mutants, a best design model had a wild type recovery of 94% (C), while a worst design model had a wild type recovery 8% (G). For experiment 4 dock/design alanine mutants, a best design model had a wild type recovery of 55% (D), while a worst design model had a wild type recovery 10% (H).

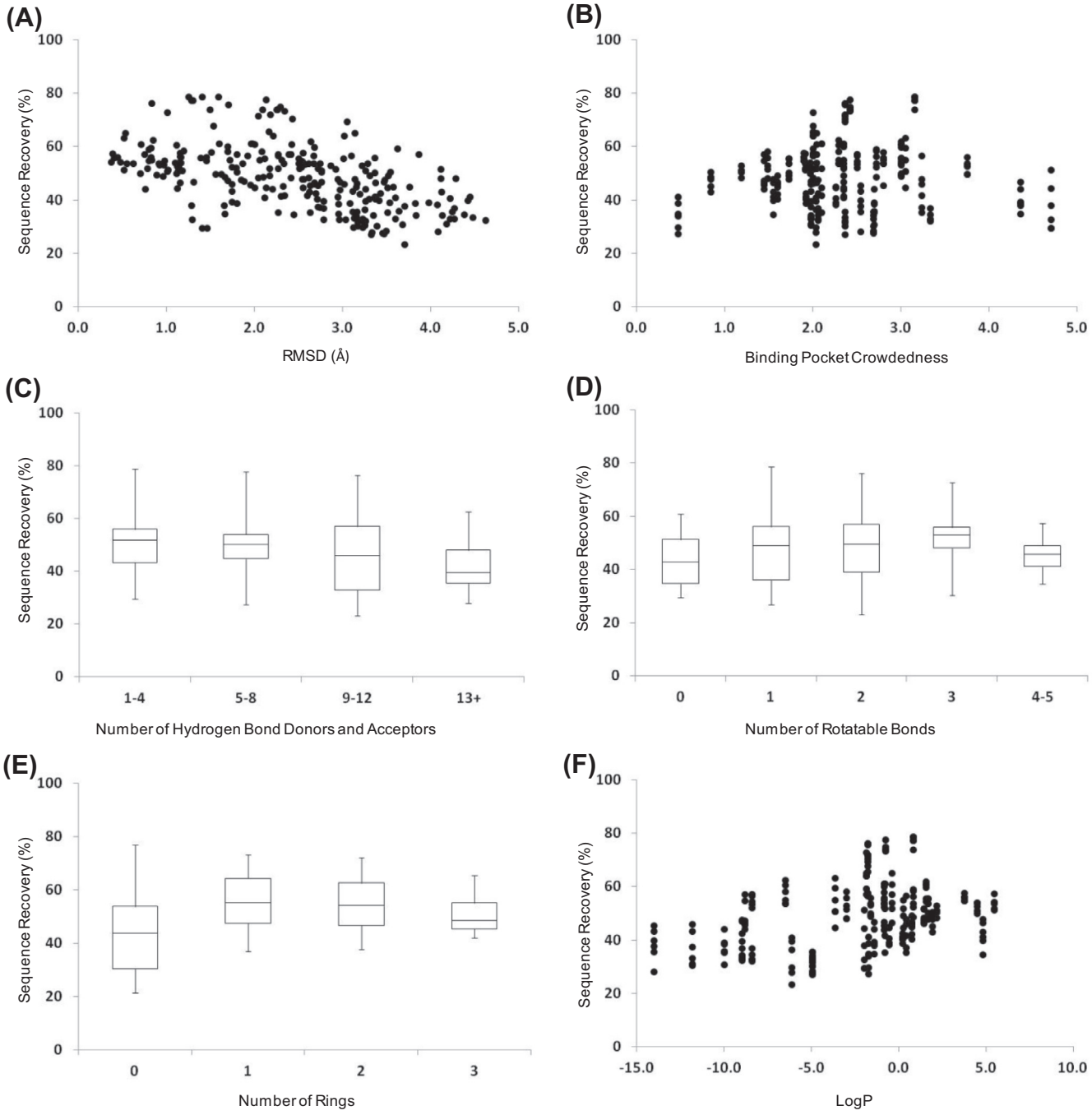

Fig.3. Sequence recovery from dock/design alanine mutants experiment. Increasing RMSD decreases the sequence recovery (A). Ligand atoms in contact with ~2–3 protein atoms show the best recoveries (B). Ligands having very many hydrogen bond donors and acceptors show a decrease in recovery (C). Ligands containing 1–3 rotatable bonds achieve the best recoveries (D). The number of rings in a ligand has slight correlation with recovery; having at least one ring decreases the range of recoveries (E). Amphipathic ligands have better recoveries than more hydrophilic and hydrophobic ligands (F). These results imply that RosettaLigand best recovers the protein–ligand interface when the ligand has a moderate number of hydrogen bond donors and acceptors, contains fewer rotatable bonds, is amphipathic, and when the binding pocket is not too loose or too crowded.

Taken together, these results suggest that RosettaLigand is biased towards non-polar, rigid ligands for achieving maximum recovery. For moderately sized ligands, the interface was recovered better than larger ligands. Interfaces with many hydrogen bond donors and acceptors were difficult to recover. Overall, there seems to be a preference for ligands that are moderately sized and not too polar, because interfaces containing ligands with these properties are recovered the best. However, correlations were generally weak with many outliers confirming that there is no single parameter that identifies an interface that is easier to design.

The same parameters were analyzed for recovery of alanine mutants to the wild type residue (Fig.4 and Fig.S2). Trends are more difficult to discern because in sequence recovery varies in a larger range. Not surprisingly, most of the complexes with high RMSDs had low recovery rates, while those with low RMSDs displayed a wider range of recovery rates, from very high to very low (Fig.4A). Similar to sequence recovery, ligands with 2–3 protein contacts per atom showed the best WT recoveries (Fig.4B). The number of ligand hydrogen bond donors and acceptors shows a trend, where recovery drops with 9+ donors/acceptors (Fig.4C). It

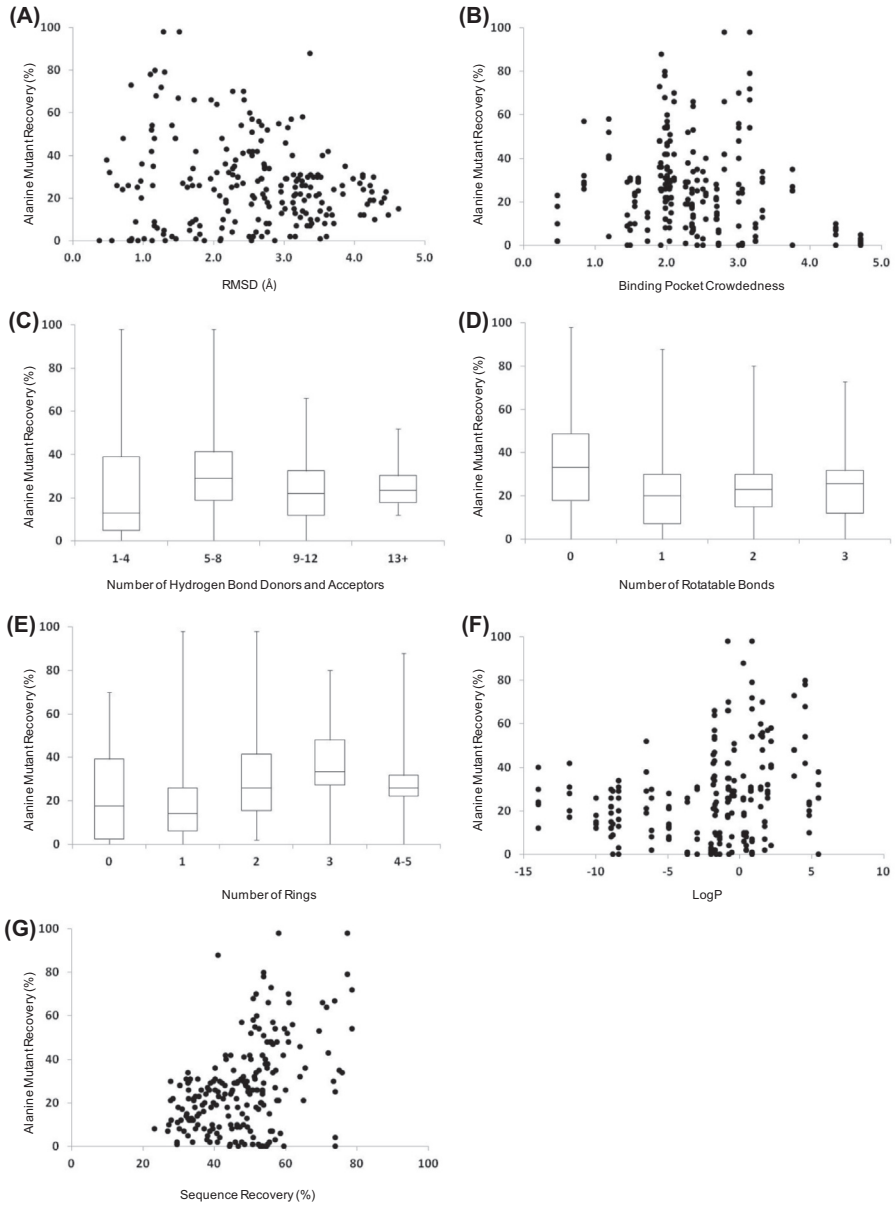

Fig.4. Alanine mutant to wild type recovery from dock/design alanine mutants experiment. Most complexes with high RMSDs had low alanine to WT recovery, whereas complexes with low RMSDs had a range of WT recovery (A). Ligand atoms in contact with ~2–3 protein atoms show the best recoveries (B). WT recovery drops when the ligand has 9 or more hydrogen bond donors and acceptors (C). As the number of ligand rotatable bonds increases, the maximum WT recovery decreases (D). Ligands with zero rings have the lowest recoveries; ligands with 3 rings, although they do not reach maximum recovery, the average is the highest (E). Amphipathic ligands have better recoveries than more hydrophilic ligands (F). Positive correlation seen between sequence recovery and the recovery of the alanine mutants to WT (G). These results imply that RosettaLigand best recovers the WT residues when the ligand has less than 8 hydrogen bond donors and acceptors, contains fewer rotatable bonds, contains more rings, is amphipathic, and when the binding pocket is not too loose or too crowded.

is expected that under the more stringent condition of only measuring alanine to WT recovery, a simpler hydrogen bonding network is easier to recover, also compared with overall sequence recovery which dropped after 13+ donors/acceptors. As the number of ligand rotatable bond increases, the maximum recovery decreases (Fig.4D). Ligands containing at least one ring have better recovery than ligands without rings (Fig.4E). Ligands with 1 or 2 rings reach the highest maximum, while ligands with 3 rings have the best average. This confirms that protein/ligand interfaces for rigid ligands are easier to design for RosettaLigand. As seen with sequence recovery, amphipathic ligands with a logP between −2.5 and 2.5 have the highest WT recoveries (Fig.4F). Hydrophobic ligands perform well, and hydrophilic ligands worst. Comparing overall sequence recovery to WT sequence recovery shows a positive correlation (Fig.4G), which demonstrates as expected that the complexes that recovered the most of the interface had the best chance of recovering the specific WT residues; instances where interface sequence recovery was very low resulted in very low WT recovery as well. This implies that Rosetta is able to discern which protein residues are most critical for ligand binding when most of the non-critical residues are correct as well. Other than a few outliers, increasing the number of ligand atoms decreases the maximum recovery (Fig.S2A). The number of residues considered for design (Fig.S2B) and the binding affinity (Fig.S2C) had little impact on WT recovery. Lower normalized binding affinities had a better chance of recovering the WT residues (Fig.S2D). As seen with sequence recovery, ligands with a high topological polar surface area have decreased recoveries (Fig.S2E). Van der Waals surface area (Fig.S2F), Rosetta interface energy (Fig.S2G), and normalized Rosetta interface energy (Fig.S2H) show no effect on WT recovery.

Overall, the trends for WT recovery correlate with sequence recovery, which is expected. Rosetta performed best when the ligand contains a small/moderate number of hydrogen bond donors and acceptors, at least one ring, low number of rotatable bonds, moderately sized, and amphipathic.

2.5. PSSM recovery results

The Position-Specific Scoring Matrix (PSSM) identifies amino acid mutations that are tolerated in homologous proteins. Thus, evaluating protein designs based on PSSM score provides a more robust assessment of favorable mutations than sequence recovery alone (DeLuca et al., 2011). By tolerating amino acids seen in evolution a more robust judgment of RosettaLigand's ability to capture biological sequences is made. Evolutionarily advantageous mutations may contribute to the interface in ways other than stability or low energy, and it is important to consider these mutations as well. In addition to designing in residues that contribute to binding and promote strong interactions, we also want to design an interface that is native-like. Is the most optimal protein–ligand interface, one that could be seen in nature, the lowest in energy? This is a fair question to consider, and one that can be assessed with PSSM recovery. PSSM recovery is expected to be higher than sequence recovery. However, it also has its limitations: (1) not all amino acids tolerated or beneficial will have been sampled in evolution, (2) the space of known related protein sequences might be incomplete, and (3) some mutation seen in other proteins might alter specificity and are not tolerated for the particular small molecule in the benchmark. The average sequence recovery and average PSSM recovery of every structure in each dataset are plotted in Fig.5. As noted earlier, applying a FNRB improves the sequence recovery, which is helpful because recovery is more difficult in the experiments that include mutated alanine residues. It was expected that allowing full ligand movement would decrease the sequence recovery compared to allowing slight ligand movement. Because PSSM views a mutation favorably if the new amino acid is frequently seen at that position, PSSM recovery will always be higher than sequence recovery, where all mutations are counted as incorrect. The percentage of PSSM recovery was also computed on a per residue basis (Fig.6A). Glycine, Alanine, Leucine, Valine, and Threonine were frequently recovered, while Tryptophan and Glutamine were often mutated. Cysteine was omitted as it was not included during design. All four data sets exhibit similar biases in terms of PSSM recovery.

The change in sequence composition provides information about overall biases in sequence design. Sequence composition change is calculated as (design_count–native_count)/(design_count). The sequence composition difference per residue is plotted in Fig.6B and Table S1. Sequence composition remains consistent for Glutamine, Isoleucine, Methionine, Proline, Threonine, and Valine. There are large negative biases to design out Phenylalanine and Tryptophan. While some degree of unfavorable mutation can be tolerated or even desired at certain positions in a ligand binding pocket, one would not expect to see such a significant loss of aromatic residues. Poor recovery of aromatic amino acids may reflect the absence of π–π and cation–π interaction scoring terms in the Rosetta energy function.

Sequence recovery, alanine to wild-type recovery, and PSSM recovery provide feedback to evaluate RosettaLigand's performance in designing protein-small molecule interfaces. The results demonstrate that by recovering native-like interactions, RosettaLigand showspromise as a tool for designing novel protein-small molecule interfaces. *In silico*, the best assessment of accuracy is to compare designs to the sequences of the protein–ligand complexes in the benchmark set. Other algorithms that seek to computationally design protein-small molecule interfaces include OSPREY (Gainza and Donald, 2012) and PocketOptimizer (Malisi et al., 2012). In a number of studies computationally designed

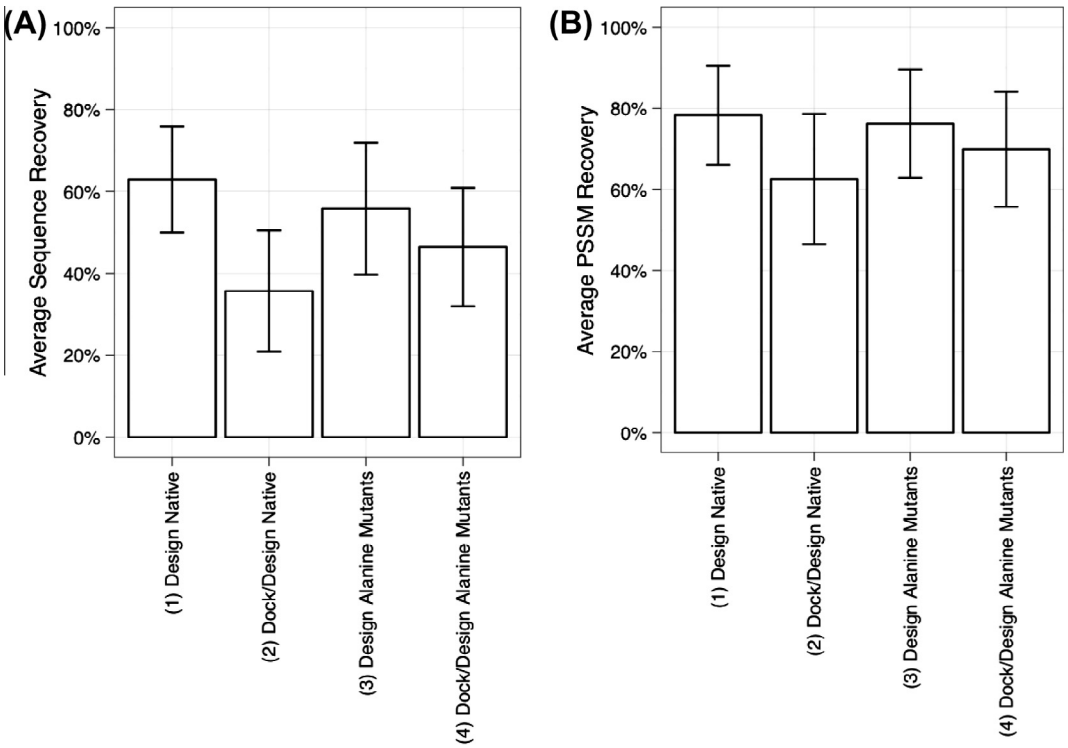

Fig.5. Sequence and PSSM recovery of the experiments. The sequence recovery for each experiment was calculated (A). The PSSM recovery for experiment was calculated (B). For both plots, error bars are 1 standard deviation from the mean. Sequence recovery, although reported earlier, is included in this form for a side-by-side comparison to PSSM recovery. Applying a bonus to the native sequence improves the sequence when the ligand is allowed full reorientation. Allowing the ligand full reorientation decreases the sequence recovery when compared to its similar experiment that only allows slight ligand reorientation.

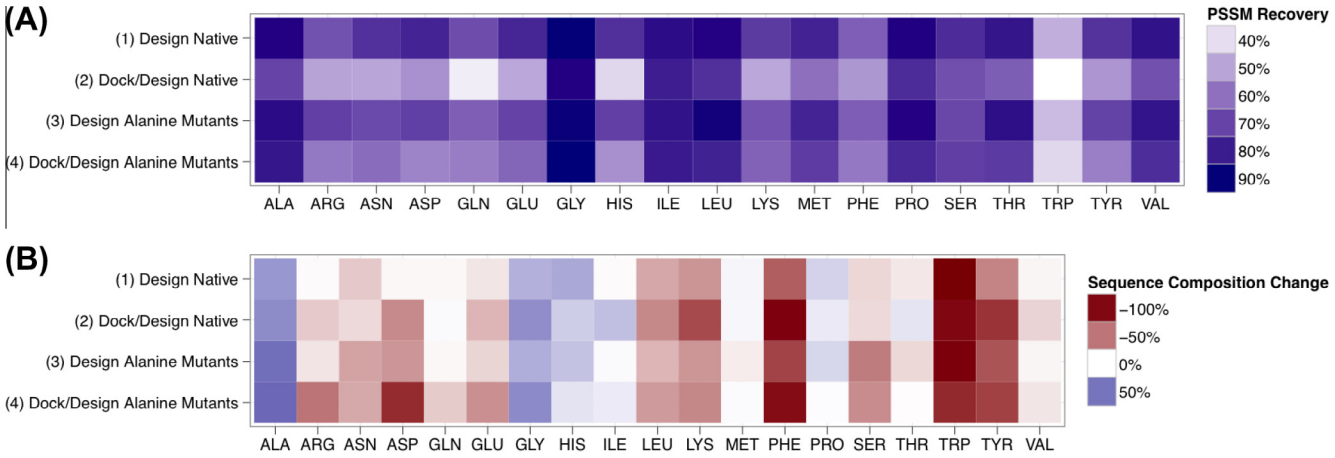

Fig.6. Heatmap of PSSM recovery per residue for each experiment (A). Dark blue indicates that these residues are mutated to residues with good PSSM scores, while light blue indicates a mutation to a residue seldom seen at that particular position. Heatmap of change in sequence composition for each experiment (B). Sequence composition change is calculated as (design_count-native_count)/(design_count). White indicates a residue that remains consistent in composition, red indicates a residue that is designed out of the sequences, blue indicates a residue that is designed into the sequences. Cys was omitted as it was not involved in design.

mutations in protein-small molecule interfaces were experimentally verified. For example when redesigning an enzyme for target substrates (Chen et al., 2009), design of a peptide inhibitor which rescues regulatory activity (Roberts et al., 2012), and to predict mutations that arise from drug resistance (Frey et al., 2010). Some methodological improvements that can be considered to improve RosettaLigand performance further include: continuous flexibility of rotamers (Gainza and Donald, 2012; Gainza et al., 2013), continuous backbone flexibility (Hallen et al., 2013), local backrub motions (Keedy et al., 2012), and computing partition functions over molecular ensembles (Georgiev et al., 2008). However, the first critical step is to perform an experimental verification of RosettaLigand designed protein-small molecule interfaces.

3. Conclusions

RosettaLigand has been used previously to dock small molecules into proteins, allowing full ligand and protein flexibility and recovering small molecule position and most interface side chain conformations within 2 Å of the experimental structure. The results described here have expanded the methods further to include sequence optimization and performed stringent tests on them following the protocol typically used when designing novel protein-small molecule interfaces. We designed experiments to test RosettaLigand's ability to recover the sequence and ligand position while reorienting the small molecule and applying a native sequence bonus. In addition to sequence recovery, we tested RosettaLigand's ability to recover WT residues from those that were intentionally mutated. Most of the trends we saw were expected, such as lower sequence recovery with higher ligand RMSD and higher sequence recovery with fewer rotatable bonds. As overall sequence recovery increased, recovery of WT residues increased as well. This implies that RosettaLigand can recognize residues necessary for binding and not over-design the interface. Recognizing ligand properties that maximize the recovery of native-like interactions and also recognizing the ligand properties that Rosetta struggled with is twofold: (1) it gives us crucial feedback for improving the algorithm, and (2) it gives us an advantage in designing novel interfaces, by starting out with designs for ligands that have shown good results.

Many factors contribute to the difficulty in computationally designing protein-small molecule interfaces. The design algorithm must sample the correct ligand and side chain identity and conformation and also have a comprehensive energy function that can distinguish between interactions that promote binding and those that abolish it. One may naively assume that because the binding pocket is significantly smaller than the entire protein, interface design is less challenging than complete protein design. However, there are several arguments why this is not the case: (1) as the protein-small molecule interface is small compared to the core of a protein, there is less tolerance for error, (2) varying 10 positions with all 20 amino acids yields $20^{10} = 10^{13}$ sequences which is near the limit of the sequence space that can be screened experimentally (Sidhu and Koide, 2007; Sidhu et al., 2000), and (3) designing protein-small molecule interfaces requires often precise positioning of interacting functional groups which is more challenging than optimizing apolar van der Waals interactions. RosettaLigand can more successfully design sites for apolar small molecules whose binding is dominated by van der Waals interactions. This was seen in many of the sequence recovery plots, where recoveries were the worst for ligands that were very hydrophilic, contained many hydrogen bond donors and acceptors, and had high topological polar surface areas. An in-depth analysis of recovery by polar vs. apolar amino acids reveals that for all experiments, sequence recovery and PSSM recovery for apolar resides was higher than for polar residues (Table S2). For example, in the Dock/design alanine mutants experiment, apolar residues at the protein–ligand interface had sequence recovery of 62.4% while, polar residues were recovered 32.8%. Also, PSSM recovery shows that apolar residues were recovered 78.7%, while polar residues recovered 63.1%. Additional energy terms will likely be needed for accurate design of interfaces that rely primarily on partial covalent interactions.

4. Materials and methods

4.1. Compilation of a benchmark of 43 protein–ligand complexes

The Community Structure-Activity Resource (CSAR) database (Dunbar et al., 2011) contains a diverse set of protein-small molecule crystal structures and includes information on binding affinities. The full dataset of 343 complexes was filtered to obtain a suitable subset for the present study. First, complexes where metal ions or water molecules were deemed critical for the interaction were excluded since design of interfaces that contain more than

two interaction partners requires further modification of the design algorithms. From the remaining set, proteins containing more than 800 amino acids were excluded to limit the time needed in the protein minimization step, leaving 102 complexes. Half of these contained ligands with more than 3 rotatable bonds, and these were excluded to limit the degrees of freedom of the ligand. Lastly, complexes where the ligand was at the interface of two protein chains were excluded. The final benchmark set contained 43 protein–ligand crystal structures, with resolutions better than 2.50 Å and ligand molecular weights varying between 70 and 400 g/mol.

4.2. Preparing the benchmark set

Files from the Community Structure-Activity Resource (CSAR) dataset (Dunbar et al., 2011) were prepared as described previously (Lemmon and Meiler unpublished results). The ligand atom coordinates were extracted from the input files, and the script 'mol_file_to_params.py' was used to create .params files that describe chemical properties of each ligand and assign each ligand a Rosetta atom type. BioPython was used to align residue names, and convert non-canonical residues to their canonical base residues. Neutralizing caps were removed from the N- and C-termini of the protein chain. Protein chains were relabeled alphabetically, and the ligand was given the chain identifier 'X' and residue code 'INH' in each file. This dataset was filtered to exclude protein chains longer than 800 amino acids, ligands with more than 3 rotatable bonds, metal ions and water molecules tightly bound at the interface (within 3.0 Å of protein or ligand), leaving 43 complexes that were used for the native complexes datasets (Table S3).

4.3. Determining critical residues in the protein–ligand interface

For each of the 43 protein-small molecule complexes, RosettaLigand was used to generate 100 'relaxed' models, employing a Rosetta protocol that relies on gradient minimization and side chain repacking. The contribution of each residue to the interface energy was determined as the difference in per-residue energy in the free and bound forms of the protein, and averaged among the 100 models generated. The residues with the highest contributions to ligand binding were mutated sequentially to alanine to create five new complexes (e.g. in the first complex only the residue contributing most to the stability of the interface was mutated, in the second complex the two highest contributors were mutated, etc.). The final alanine-modified benchmark set contained $5 \times 43 = 215$ complexes.

4.4. Determining residues in the design sphere

Interface residues were selected for design and repacking based on four distances measured between the Cα of a protein residue and the closest non-hydrogen atom of the ligand. Residues within 6 Å were designed (i.e. side chains can change to anyother amino acid, excluding cysteine). Residues within 6–8 Å were considered for design only if the residue was pointing towards the ligand (i.e. the distance between Cβ and any non-hydrogen ligand atom was less than the distance between Cα and the same ligand atom). Residues within 8–10 Å were repacked (i.e. side chain rotamers were sampled but residues were not mutated). Residues within 10–12 Å were repacked only if the residue was pointing towards the ligand. The cutoff values were chosen to ensure that the design sphere was small enough to allow for mutations close to the ligand, yet large enough to include the longest residue, Arginine, in the sphere as well.

4.5. Determining the optimal favor native residue bonus (FNRB)

In interface design, only the protein residues within the known or putative ligand binding site are allowed to mutate. Designs with the lowest number of mutations are preferred to minimize perturbation of the protein fold. In order to achieve this, a small energy bonus is added to keep the original residue unless introducing an alternate amino acid results in a significant energetic gain. The 'favor native' residue bonus (FNRB) is typically chosen in the range of the per residue standard deviation of the RosettaLigand score. The optimal FNRB cannot be determined by simply redesigning protein–ligand complexes with their native sequence since an increased bonus will always result in increased sequence recovery. Instead, we first determine the five most critical binding residues and mutate these amino acids to alanine. These mutants are then redesigned using RosettaLigand to test whether the alanine reverts to the correct residue.

Of the 43 complexes in the benchmark, a subset was randomly selected to establish the optimal FNRB. This subset contained no duplicate proteins or ligands. For each experiment, the ligand was allowed full reorientation (2 Å translation, 360° rotation) with FNRB values between 0.5 and 1.5. One thousand designs were generated for each alanine mutant with each bonus and then the models with the top 50 interface scores were selected for analysis (Table S4). As expected, when the FNRB is increased, the sequence recovery increases (number of mutations decreases). However, the percent reversion of alanine to WT residue increases until FNRB = 1.0 and then decreases as FNRB is increased further. Based on these results, FNRB = 1.0 was applied for the subsequent experiments.

4.6. Description of each experiment

Four sequence recovery experiments (percentage of designed residues that are identical to native residues) were conducted on the full dataset of 43 complexes. Design native (1) probed sequence recovery of the native complexes when there was no FNRB and the ligand was allowed limited reorientation (0.1 Å translation, 2° rotation). Dock/design native (2) probed sequence recovery of the native complexes when there was no FNRB and the ligand was allowed full reorientation (2 Å translation, 360° rotation). Design with alanine mutants (3) probed sequence recovery and alanine-to-WT recovery of the native and alanine mutant complexes when there was a FNRB and the ligand was allowed slight reorientation. Lastly, dock/design with alanine mutants (4) probed sequence recovery and alanine-to-WT recovery of the native and alanine mutant complexes when there was a FNRB and the ligand was allowed full reorientation.

4.7. Determining the number of residues considered for design to compute sequence recovery

In the experiments that allowed full ligand movement, the total number of residues that interact with the ligand is a moving target. To report overall sequence recovery as a percentage, the number of residues at the interface was chosen as the average number of residues considered for design. This is still somewhat problematic as not all residues were necessarily allowed to change in each of the docking/design trajectories. We counterbalance this limitation by also reporting the absolute number of mutations (Table S5). However, these numbers are not comparable from complex to complex because the number of residues at the interface varies depending on the ligand size and the shape of the binding pocket.

4.8. Individual ligand parameters determined by the BCL

The BioChemistryLibrary (BCL) is a software suite tailored for small molecule modeling, and contains a variety of small molecule descriptors (Butkiewicz et al., 2013). The ligand parameters calculated by the BCL include: number of hydrogen bond donors and acceptors, number of ligand rings, number of atoms, topological polar surface area (Ertl et al., 2000), van der Waals surface area (computed using the BCL's algorithm, which considers the overlapping spheres of neighboring atoms), and logP (Xing and Glen, 2002).

4.9. Evaluating RosettaLigand design performance using position specific scoring matrices (PSSMs)

PSSM data provide a more quantitative insight into specific residues that are successes/failures when subjected to design. PSSMs were generated from the native protein sequences using BLAST (Altschul et al., 1990). The designed residues were then scored using the PSSM. Thus, a residue of a type frequently seen at that position would have a positive PSSM score, while a residue seldom seen at that position would have a negative PSSM score. The percent PSSM recovery is a measure of the percentage of residues with favorable mutations.

Acknowledgment

Work in the Meiler laboratory is supported through NIH (R01 GM099842). B.A. is supported through the National Science Foundation Graduate Research Fellowship Program, under grant number DGE-0909667.

References

Allert, M.; Rizk, S. S.; Looger, L. L.; Hellinga, H. W., 2004. Computational design of receptors for an organophosphate surrogate of the nerve agent soman. Proceedings of the National Academy of Sciences of the United States of America 101, 7907–12. Epub 2004 May 17.

Althoff, E.A., Wang, L., Jiang, L., Giger, L., Lassila, J.K., et al., 2012. Robust design and optimization of retroaldol enzymes. Protein Sci. 21, 717–726.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Ashworth, J., Taylor, G.K., Havranek, J.J., Quadri, S.A., Stoddard, B.L., et al., 2010. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. Nucleic Acids Res. 38, 5601–5608.

Baeumner, A.J., 2003. Biosensors for environmental pollutants and food contaminants. Anal. Bioanal. Chem. 377, 434–445.

Baker, D., Davis, I.W., 2009. RosettaLigand docking with full ligand and receptor flexibility. J. Mol. Biol. 385, 381–392.

Baker, D., Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., et al., 2008a. De novo computational design of retro-aldol enzymes. Science 319, 1387–1391.

Baker, D., Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., et al., 2008b. Kemp elimination catalysts by computational enzyme design. Nature 453, 190–195.

Baker, D., Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., et al., 2010. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 329, 309–313.

Butkiewicz, M., Lowe Jr., E.W., Mueller, R., Mendenhall, J.L., Teixeira, P.L., et al., 2013. Benchmarking ligand-based virtual high-throughput screening with the pubchem database. Molecules 18, 735–756.

Chen, C.Y., Georgiev, I., Anderson, A.C., Donald, B.R., 2009. Computational structure-based redesign of enzyme activity. Proc. Natl. Acad. Sci. USA 106, 3764–3769.

Clifton, M.C., Corrent, C., Strong, R.K., 2009. Siderocalins: siderophore-binding proteins of the innate immune system. Biometals 22, 557–564.

DeLuca, S., Dorr, B., Meiler, J., 2011. Design of native-like proteins through an exposure-dependent environment potential. Biochemistry-US 50, 8521–8528.

Dunbar, J.B., Smith, R.D., Yang, C.Y., Ung, P.M.U., Lexa, K.W., et al., 2011. CSAR benchmark exercise of 2010: selection of the protein-ligand complexes. J. Chem. Inf. Model. 51, 2036–2046.

Ertl, P., Rohde, B., Selzer, P., 2000. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J. Med. Chem. 43, 3714–3717.

Fleishman, S.J., Whitehead, T.A., Ekiert, D.C., Dreyfus, C., Corn, J.E., et al., 2011. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 332, 816–821.

Frey, K.M., Georgiev, I., Donald, B.R., Anderson, A.C., 2010. Predicting resistance mutations using protein design algorithms. Proc. Natl. Acad. Sci. USA 107, 13707–13712.

Gainza, P., Roberts, K.E., Donald, B.R., 2012. In: Protein design using continuous rotamers. Plos Comput. Biol. 8.

Gainza, P., Roberts, K.E., Georgiev, I., Lilien, R.H., Keedy, D.A., et al., 2013. OSPREY: protein design with ensembles, flexibility, and provable algorithms. Methods Enzymol. 523, 87–107.

Georgiev, I., Lilien, R.H., Donald, B.R., 2008. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. J. Comput. Chem. 29, 1527–1542.

Golan, D.E., Leader, B., Baca, Q.J., 2008. Protein therapeutics: a summary and pharmacological classification. Nat. Rev. Drug Discovery 7, 21–39.

Hallen, M.A., Keedy, D.A., Donald, B.R., 2013. Dead-end elimination with perturbations (DEEPer): a provable protein design algorithm with continuous sidechain and backbone flexibility. Proteins 81, 18–39.

Hayden, E.C., 2009. In: Key protein–design papers challenged. Nature 461, 859.

Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Rothlisberger, D., et al., 2008. De novo computational design of retro-aldol enzymes. Science 319, 1387–1391.

Keedy, D.A., Georgiev, I., Triplett, E.B., Donald, B.R., Richardson, D.C., et al., 2012. The role of local backrub motions in evolved and designed mutations. PLoS Comput. Biol. 8, e1002629.

Knudsen, K.E., Scher, H.I., 2009. Starving the addiction: new opportunities for durable suppression of AR signaling in prostate cancer. Clin. Cancer Res. 15, 4792–4798.

Kortemme, T., Morozov, A.V., 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.

Malisi, C., Schumann, M., Toussaint, N.C., Kageyama, J., Kohlbacher, O., et al., 2012. Binding pocket optimization by computational protein design. PLoS ONE 7, e52505.

Meiler, J., Baker, D., 2006. RosettaLigand: protein-small molecule docking with full side-chain flexibility. Proteins 65, 538–548.

Meiler, J., Kaufmann, K.W., Dawson, E.S., Henry, L.K., Field, J.R., et al., 2009. Structural determinants of species-selective substrate recognition in human and Drosaphila serotonin transporters revealed through computational docking studies. Proteins 74, 630–642.

Roberts, K.E., Cushing, P.R., Boisguerin, P., Madden, D.R., Donald, B.R., 2012. Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity. Plos Comput. Biol. 8.

Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., et al., 2008. Kemp elimination catalysts by computational enzyme design. Nature 453, 190–195.

Schreier, B., Stumpp, C., Wiesner, S., Hocker, B., 2009. Computational design of ligand binding is not a solved problem. Proc. Nat. Acad. Sci. USA 106, 18491–18496.

Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., Baker, D., 2005. Progress on modeling of protein structures and interactions. Science 310, 638–642.

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

Sidhu, S.S., Lowman, H.B., Cunningham, B.C., Wells, J.A., 2000. Phage display for selection of novel binding peptides. Methods Enzymol. 328, 333–363.

Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., et al., 2010. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 329, 309–313.

Sood, V.D., Baker, D., 2006. Recapitulation and design of protein binding peptide structures and sequences. J. Mol. Biol. 357, 917–927.

Wang, L., Althoff, E.A., Bolduc, J., Jiang, L., Moody, J., et al., 2012. Structural analyses of covalent enzyme-substrate analog complexes reveal strengths and limitations of de novo enzyme design. J. Mol. Biol. 415, 615–625.

Weng, Z.P., DeLisi, C., 2002. Protein therapeutics: promises and challenges for the 21st century. Trends Biotechnol. 20, 29–35.

Xing, L., Glen, R.C., 2002. Novel methods for the prediction of log P, pK(a), and log D. J. Chem. Inf. Comput. Sci. 42, 796–805.

Zanghellini, A., Jiang, L., Wollacott, A.M., Cheng, G., Meiler, J., et al., 2006. New algorithms and an in silico benchmark for computational enzyme design. Protein Sci.: Publ. Protein Soc. 15, 2785–2794.

Zhu, X., Dickerson, T.J., Rogers, C.J., Kaufmann, G.F., Mee, J.M., et al., 2006. Complete reaction cycle of a cocaine catalytic antibody at atomic resolution. Structure 14, 205–216.