



Computational Design of a Protein-Based Enzyme Inhibitor

Erik Procko¹, Rickard Hedman², Keith Hamilton³, Jayaraman Seetharaman⁴, Sarel J. Fleishman⁵, Min Su⁴, James Aramini³, Gregory Kornhaber³, John F. Hunt⁴, Liang Tong⁴, Gaetano T. Montelione^{3,6} and David Baker¹

1 - Department of Biochemistry and Howard Hughes Medical Institute (HHMI), University of Washington, Seattle, WA 98195, USA

2 - Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

3 - Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

4 - Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, USA

5 - Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

6 - Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, NJ 08854, USA

Correspondence to David Baker: dabaker@u.washington.edu

<http://dx.doi.org/10.1016/j.jmb.2013.06.035>

Edited by A. Keating

Abstract

While there has been considerable progress in designing protein–protein interactions, the design of proteins that bind polar surfaces is an unmet challenge. We describe the computational design of a protein that binds the acidic active site of hen egg lysozyme and inhibits the enzyme. The design process starts with two polar amino acids that fit deep into the enzyme active site, identifies a protein scaffold that supports these residues and is complementary in shape to the lysozyme active-site region, and finally optimizes the surrounding contact surface for high-affinity binding. Following affinity maturation, a protein designed using this method bound lysozyme with low nanomolar affinity, and a combination of NMR studies, crystallography, and knockout mutagenesis confirmed the designed binding surface and orientation. Saturation mutagenesis with selection and deep sequencing demonstrated that specific designed interactions extending well beyond the centrally grafted polar residues are critical for high-affinity binding.

Published by Elsevier Ltd.

Introduction

Computational design of protein–protein interactions has the potential to rapidly generate new binding proteins for any specified site of interest on a target protein, bypassing many of the steps associated with current technologies such as antibody development. This may lead to new biochemical reagents, diagnostics, and therapeutics. There has been significant progress designing protein–protein interactions, often by taking advantage of special structural properties, including generating β -sheet extensions, coiled coils, and metal ion-dependent interfaces [1–3]. More general “two-sided design” approaches have generated protein pairs that have been shown to interact experimentally,

though the structures of the complexes either have not been solved [4] or have been found to be different from the original design conception [5], highlighting the challenges to designing specific and accurate interactions.

Two approaches have recently been described for “one-sided design”—the design of proteins that bind to a specific site on a target protein. The first involves computationally docking a non-interactive scaffold protein to the target surface and designing the sequence to optimize the interaction energy. This dock-and-design approach has generated a weak affinity binder for p21-activated kinase 1 (PAK1) [6]. A second, hot-spot-centric approach has utilized a common feature of native interfaces: the presence of a few key conserved interaction residues often

centrally located, called hot spots, that disproportionately contribute the bulk of the binding energy within a much larger interface [7–9]. By grafting three hot spot residues from erythropoietin onto the surface of an unrelated protein of correct shape, a new binder to the erythropoietin receptor with modest affinity was designed [10]. Most recently, novel binders to a conserved epitope on influenza hemagglutinin (HA) were created by a two-step process, in which first a hot spot region was computed and then many scaffold proteins were scanned to identify those that could be designed to incorporate the hot spot residues [11]. The two designed HA-binding proteins were small and helical, shared no sequence or structural features with other proteins that bind HA, and presented hydrophobic surfaces for binding.

Recent analyses of success and failure in design of function [12,13] have emphasized that successful design has involved hydrophobic interfaces, and it is unclear whether computational design methodology can target the polar surfaces common in naturally occurring protein–protein interactions. Here, we investigate the design of a protein inhibitor that binds a deeply recessed and highly charged enzyme active site using the dock-and-design and hot-spot-based approaches. We show that binding can be achieved using appropriately placed polar hot spot residues and that additional designed interactions across the interface are critical to provide sufficient binding energy.

Results

Hen egg lysozyme (HEL) is a glycoside hydrolase that breaks down sugar linkages in the bacterial cell wall. It was the first enzyme to have its crystal structure solved [14], and many antibodies, antigen receptors, and bacterial defensive proteins that bind HEL with high affinity have been identified and their bound atomic structures have been determined [15–17]. Therefore, not surprisingly, HEL has become a favorite model system for investigating protein–protein interactions. Due to this extensive prior literature, we chose HEL as a target for the challenge of computational design of a protein inhibitor that binds a polar enzyme active site. As described in the following sections, we tested both a dock-and-design and a hot-spot-based strategy for designing proteins that bind to this polar site.

Dock-and-design approach

The dock-and-design strategy (outlined in Fig. 1a) begins by coarsely docking structures from a library of hundreds of potential scaffolds (using the set described in Ref. [11]) to the active-site region of HEL, followed by several rounds of refined docking

and sequence design of scaffold residues proximate to the interface using the ROSETTA design methodology [18–20]. The principles of this dock-and-design approach are very similar to the strategy used to target PAK1 [6], except that here we used not one but hundreds of scaffolds, increasing the chances of sampling a docked conformation conducive for designing high-affinity interactions. Designed HEL-binding proteins were filtered by calculated interface binding energy, shape complementarity, packing, and size, aiming for interface quality metrics similar to those of native HEL complexes (Table S1).

Twenty-four designs based on different scaffolds were selected (Table S1), and yeast codon-optimized genes encoding the designs were synthesized and expressed in the EBY100 yeast strain as a fusion between N-terminal Aga2p for surface display and a C-terminal Myc epitope tag for detection with a fluorescent antibody [21]. Binding was assessed by flow cytometry after incubating yeast with an oligomeric HEL-streptavidin ligand to enhance apparent affinity by avidity [21,22]. Among the 24 designs, DnvLB16 bound HEL but not other biotinylated proteins tested for non-specific interactions (Fig. 1b). DnvLB16 differed from its starting scaffold [Protein Data Bank (PDB) ID: 3DT5; a protein of unknown function] by 15 substitutions.

DnvLB16 interface residues were reverted to their wild-type scaffold identities by site-directed mutagenesis to assess whether the designed interactions bound lysozyme as intended (Fig. 1c). Some of these mutants still bound HEL despite substituting hydrophobic residues buried in the design model deep within the interface to polar or charged amino acids (Fig. 1c and e). Further, we affinity matured a variant of DnvLB16 (mutant M40R;W41H;H42S, with significantly higher surface expression on yeast to aid selection, see Supplementary Fig. S1) by diversifying the gene with error-prone PCR and selecting a transformed yeast library with three rounds of dual-color fluorescence-activated cell sorting (FACS). Three mutations that enhanced affinity were found, and all three targeted positions outside the designed interface and cannot be explained by the computational model (Fig. 1d and e). Since mutations at the designed interface did not affect binding, whereas mutations outside the designed binding surface did, we concluded that DnvLB16 does not bind HEL as designed. The unanticipated binding mode may arise because the dock-and-design procedure gives broadly hydrophobic interfaces with few specific polar contacts, in contrast with native interfaces (cf. 29% and 39% of interface contacts involve polar atoms, respectively; Table S1). The authors were also uncertain about the binding model of the previously designed PAK1 binder, again due to the excessive hydrophobic character of the designed surface [6]. The bias towards hydrophobic surfaces is primarily due to

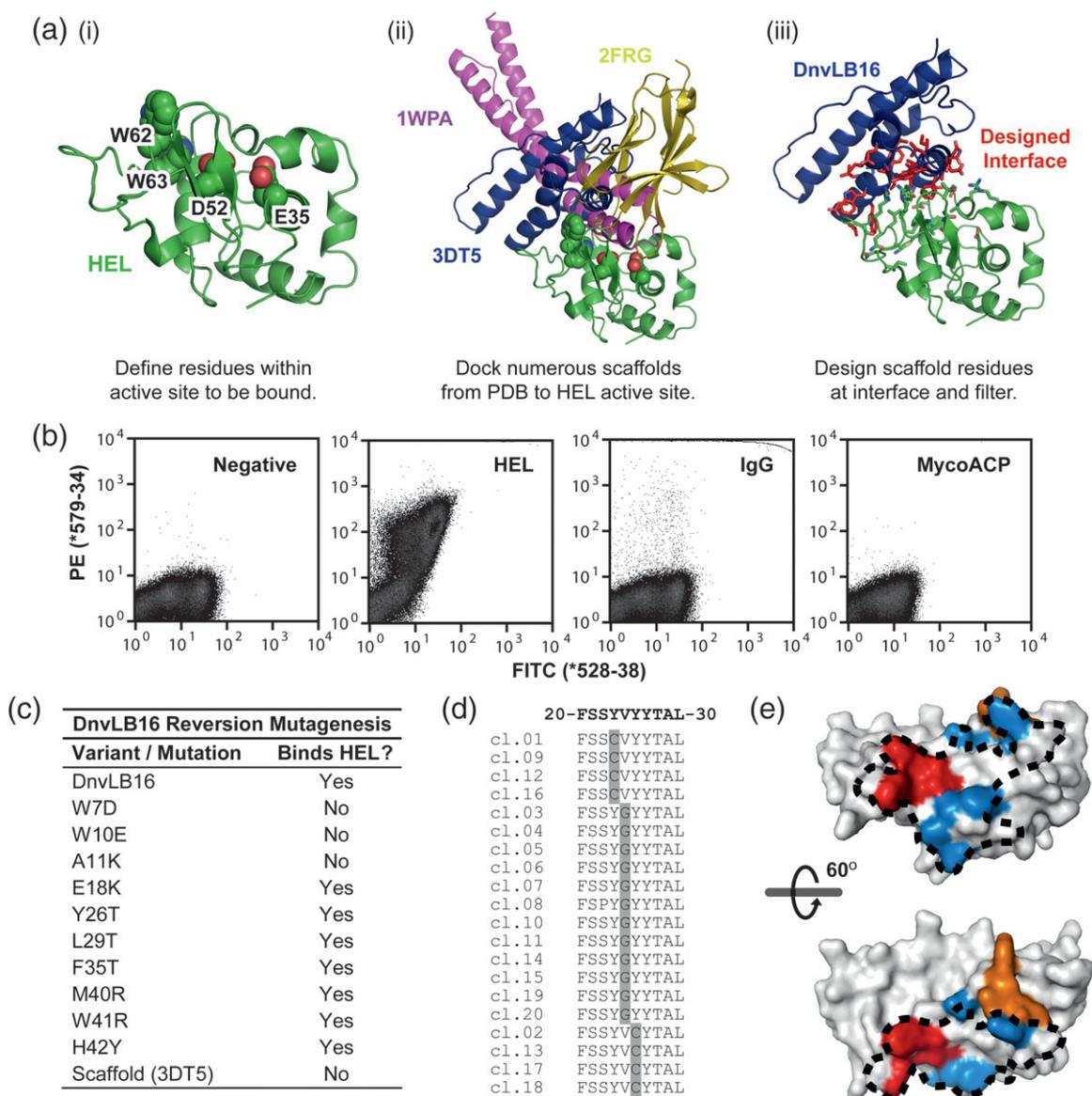


Fig. 1. A HEL-binding protein built using a dock-and-design strategy makes interactions inconsistent with the computational model. (a) Outline of the computational design strategy. Residues within the HEL active site were selected to be part of the target interaction surface (i). A curated set of 865 protein structures (referred to as scaffolds) was individually docked to the target site using PatchDock (ii). Docked configurations were refined using ROSETTA, followed by sequence design of the scaffold protein within 8 Å of the interface to minimize the assembly's energy (iii). Designed protein–HEL complexes were filtered by interface metrics. (b) Designed proteins were expressed on the yeast surface with a C-terminal myc epitope tag detected with a fluorescein isothiocyanate-conjugated antibody (x-axis). Biotinylated HEL (1 μM) was premixed with phycoerythrin (PE)-conjugated streptavidin (0.5 μM) to form an oligomeric/avidin complex, which was incubated with the yeast cells to detect surface interactions (y-axis) by flow cytometry. Shown are yeast display data for HEL-binding design DnvLB16. An identical analysis with negative control proteins (biotinylated IgG and biotinylated *Mycobacterium tuberculosis* acyl-carrier protein, MycoACP) failed to show interactions with DnvLB16. (c) Designed residues on DnvLB16 were mutated back to their original amino acid identities in the starting scaffold, and binding to avidin HEL was tested as above. (d) DnvLB16-M40R;W41H;H42S with improved expression and solubility (Supplementary Fig. S1) was evolved. The yeast display library consisted of 1×10^6 transformants with 0 to ~5 amino acid substitutions per DnvLB16 clone and was sorted for three rounds. For each round, cells were stained with 50 nM monomeric HEL followed by PE-streptavidin after washing off unbound ligand. Plasmids from improved HEL-binding mutants were isolated and the DnvLB16 gene was sequenced. Shown is a region of the sequence alignment for isolated clones (designated cl.X). (e) The designed HEL-interaction site is shown with a black broken line on a surface representation of DnvLB16. Positions of reversion mutations that maintained binding (blue) or lost binding (red) with HEL are colored. Positions identified as important for high-affinity binding by directed evolution are colored orange.

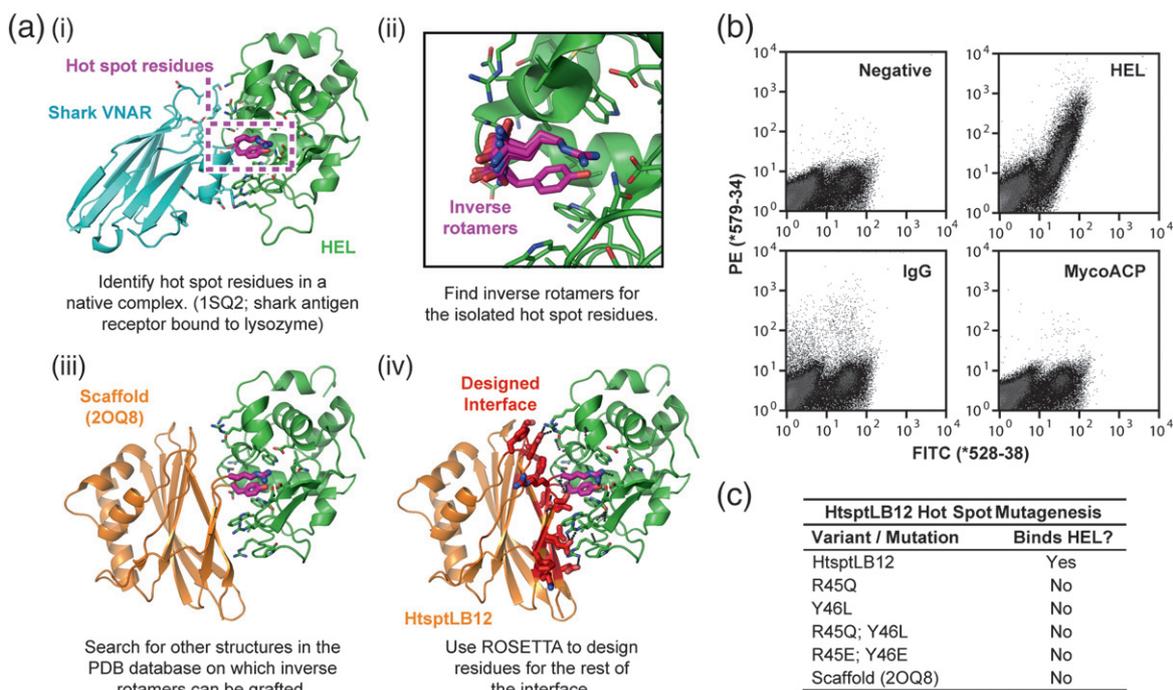


Fig. 2. Construction of a HEL-binding protein by transplanting hot spot residues and computationally designing the surrounding interface. (a) Schematic of the design process. Hot spot residues R100 and Y101 (magenta sticks) were taken from shark VNAR (cyan cartoon) bound to HEL (green cartoon) (i). Rotamers for the disembodied hot spot residues compatible with the binding geometry are enumerated (ii). Then, protein scaffolds are docked against the target surface using PatchDock and ROSETTA with a modified energy function that biases towards backbone overlap between scaffold and hot spots. Inverse hot spot rotamers are placed sequentially on the scaffold backbone (iii), and the surrounding surface of the scaffold in contact with HEL is redesigned with ROSETTA to minimize the total energy (iv). Designs are filtered by multiple criteria to assess interface quality. (b) Binding of HtsptLB12 to HEL assessed by yeast display and flow cytometry using protocol described in Fig. 1b legend. HtsptLB12 binds HEL but not IgG or MycoACP. (c) The grafted hot spot residues, R45 and Y46 of HtsptLB12, were mutated back to their original identities in the scaffold protein, or to charged glutamates. Binding of the mutants to HEL was assessed as in (b).

sampling limitations; satisfying multiple buried hydrogen bond donors and acceptors is more difficult than packing nonpolar side chains.

A hot-spot-centric approach to designing a HEL inhibitor

Given the failure of the dock-and-design protocol to control binding geometry, we applied a hot-spot-based design strategy to the deep and acidic active-site cleft of HEL. Incorporation of a cluster of interacting residues has been proposed to lead to preorganized surfaces that are less compatible with alternative interactions [23], and hydrogen-bonding hot spot residues can help overcome the difficulty in satisfying buried hydrogen-bonding groups noted above.

To identify possible polar hot spot interactions, we inspected structures of proteins bound to HEL. R100 and Y101 in a shark variable new antigen receptor (VNAR) domain (PDB ID: 1SQ2) form van der Waals, hydrogen-bonding, and salt-bridge interactions deep within the HEL active site [15]. Computational alanine

scanning confirmed the disproportionate contribution of these two residues to the binding energy (Supplementary Fig. S2). Side-chain rotamer conformations were generated for each of the two hot spot residues in which the functional group (guanidinium for Arg, the aromatic ring for Tyr) was kept fixed on the HEL surface in the location observed in the VNAR-bound structure (these functional group locations were the only information taken from this structure) (Fig. 2a). These inverse rotamers provide favorable HEL interactions but allow a diversity of backbone positions to facilitate placement on designed scaffolds. Scaffold proteins (from the same set used above for dock design [11]) were then docked in random orientations to the HEL active-site region using the PatchDock surface features matching algorithm [24]. The docked poses were then refined with RosettaDock using a biased force field that favors configurations positioning the scaffold backbone near the backbone of the disembodied hot spot rotamers [23]. Inverse rotamers within 3 Å of the scaffold backbone were chosen and sequentially transplanted on the scaffold, interspersed with

rigid-body minimization. Following hot spot placement, the surrounding scaffold residues that contribute the majority of the interface were optimized with three rounds of sequence design with structure minimization using RosettaDesign (Fig. 2a).

After filtering and selection (see Materials and Methods), 21 hot-spot-based designs (Table S1) were expressed on the yeast surface as Aga2 fusions and their binding to avid HEL-streptavidin oligomers was assessed. One design, HtsptLB12, bound HEL but not two control proteins (Fig. 2b). Not only did the original scaffold protein (nuclease AFV1-157; PDB ID: 2OQ8 [25]) show no affinity for HEL, but targeted mutations of the HtsptLB12 hot spot residues (R45 and Y46) to their native identities or to glutamates abolished binding (Fig. 2c). HtsptLB12 has a mixed α/β -topology very different from the VNAR immunoglobulin fold, the two proteins have overlapping but different interaction surfaces on HEL, and the compositions of the interface residues, apart from a general electrostatic complementarity, are quite different (Supplementary Fig. S3).

Improved binding following evolution of the interface periphery

In the absence of avidity, HtsptLB12 binds monomeric HEL weakly with micromolar affinity. HtsptLB12 was diversified by error-prone PCR and higher-affinity variants were enriched using yeast display and FACS to investigate how the design

methods could be improved [26]. Two HtsptLB12 mutations improving binding signal, S50A (which could allow HEL-R73 to make more ideal hydrogen bonds to neighboring acidic residues) and K52M (which could relieve electrostatic repulsion with HEL-R73), were identified at the edge of the designed interface (Fig. 3a and Supplementary Fig. S4a). The two substitutions together (variant HtsptLB12.v1) increase the apparent affinity by yeast display from micromolar to 69 ± 6 nM.

HtsptLB12.v1 was then subjected to a second round of affinity maturation, and affinity-enhancing mutations were again found at the edge of the designed interface, all localized on or impacting the structure of HtsptLB12 interface loop 18–22 (Fig. 3b and Supplementary Fig. S4b). Combining these individually identified mutations in different arrangements led to variant HtsptLB12.v2 with two substitutions (S19Y and K67T) with an apparent $K_d = 8 \pm 2$ nM.

Structural characterization of the designed protein–protein interaction

The X-ray crystal structure of unbound HtsptLB12.v1 was determined to 2.9 Å resolution (PDB ID: 3VB8). The asymmetric unit contained two copies, referred to as crystal chains A and B. The backbones of the HtsptLB12 computational model and the crystal structures are very similar (root-mean-square deviations to crystal chains A and B were 0.930 and

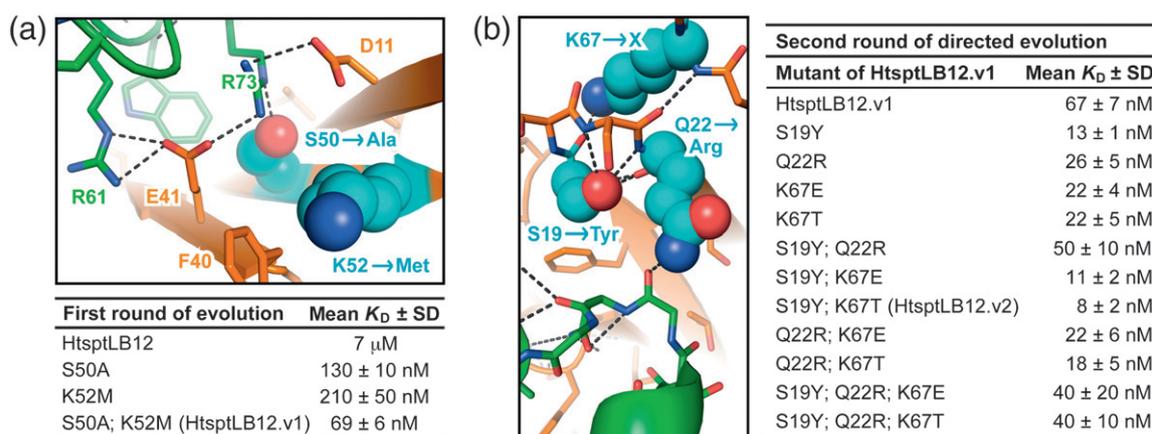


Fig. 3. Directed evolution of HtsptLB12 reveals affinity-enhancing mutations at the interface periphery. (a) HtsptLB12 was diversified by error-prone PCR, a yeast display library was sorted, and affinity-enhancing mutations S50A and K52M were identified (Supplementary Fig. S4). Shown at the top is the mutated region in the computational model, with HEL in green and HtsptLB12 in orange. S50 and K52 are highlighted with cyan spheres. Below are the apparent dissociation constants of targeted HtsptLB12 mutants, determined by titrating monomeric HEL and measuring binding signals to HtsptLB12-expressing yeast ($n = 3-4$, excluding the K_d for parental HtsptLB12, which was determined from a single titration series due to limited reagent). (b) A second round of directed evolution was applied to a library of HtsptLB12.v1 mutants. Three mutations improved affinity (S19Y, Q22R, and K67X, where X is any of several amino acids), shown on the left as cyan spheres on the modeled structure of HtsptLB12–HEL (orange and green, respectively). To the right are the apparent dissociation constants of HtsptLB12.v1 mutants determined by yeast display ($n = 4-5$).

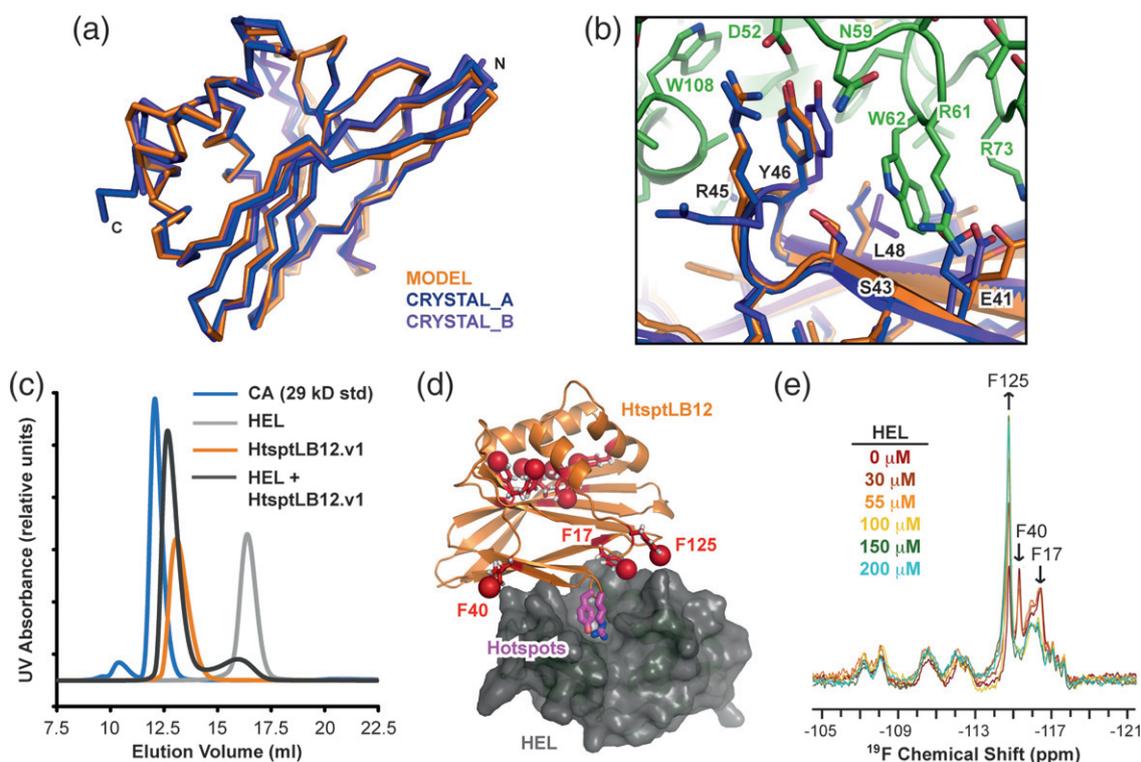


Fig. 4. HtsptLB12 structure is nearly identical to the designed model and interacts with HEL at the designed surface. (a) The crystal structure of HtsptLB12.v1 (two copies A, blue, and B, purple, in the asymmetric unit) is superimposed with the computational model of unbound HtsptLB12 (orange). (b) Zoomed-in region of the modeled HtsptLB12 (orange) and HEL (green) complex encompassing the hot spot residues. The crystal structure of unbound HtsptLB12.v1 (crystal chains A and B are blue and purple) is overlaid with HtsptLB12 in the computational model. The hot spot residue conformations are very close to the design model; this is particularly notable for the conformationally flexible arginine of chain A. (c) HEL (14 kDa, 20 nmol, gray line) elutes as a higher-MW complex (black line) from a SEC column when mixed with purified HtsptLB12.v1 (20 kDa, 20 nmol, orange line). An MW standard, bovine carbonic anhydrase (CA, 29 kDa, 20 nmol), is shown as a blue line for comparison. HEL elutes anomalously from dextran-based gel-filtration resins with an apparent MW of 7 kDa [27]. Results are representative of two repeats. (d) 4-Fluorophenylalanine was incorporated into HtsptLB12.v2 for ^{19}F NMR studies. Shown is the computational model of HtsptLB12 (orange cartoon) bound to HEL (gray surface), with the nine phenylalanines of HtsptLB12 as red sticks and the fluorine-substituted *para*-positions as red spheres. The three phenylalanines of HtsptLB12 in or near the interface are labeled. (e) ^{19}F NMR spectra of fluorophenylalanine-substituted HtsptLB12.v2 (100 μM), titrated with HEL (0 to 200 μM). The titration spectra are overlaid to highlight the appreciable differences in the intensities and frequencies of ^{19}F resonances, assigned to F17 and F40, both of which are in or near the HEL binding site. The increase in the ^{19}F resonance intensity of F125 is attributed to degeneracy of this resonance with one or both of F17 and F40 in the complex.

0.843 Å, respectively), indicating that the 14 surface mutations distinguishing HtsptLB12.v1 from the starting scaffold did not cause unanticipated structural perturbations (Fig. 4a). In native complexes, interface residues are often pre-ordered in their bound conformations even in the unbound proteins [28]. This reduces the entropic cost of binding and possibly prevents binding residues adopting alternative configurations that can interact with off-targets. Of the 15 side-chain rotamers at the interface in the design model, 8 had similar conformations in the crystal structure chains A and B (Fig. 4b). A total of 6 out of 15 interface residue rotamers match between both crystal chains A and B and the bound HtsptLB12–HEL model. The designed interface

residues form asymmetric packing interactions between chains A and B in the crystal lattice, and some rotamers differ between the two crystal chains because of the local packing environment (Supplementary Fig. S5). Overall, the structure suggests a partial pre-ordering of interface residues as designed. Attempts to obtain high-resolution diffracting crystals of the bound HtsptLB12–HEL complex were unsuccessful.

To test whether the designed protein binds in solution and not just via interactions on the yeast cell surface, we analyzed an equimolar mixture of HtsptLB12.v1 with HEL by size-exclusion chromatography (SEC). HEL shifted from a low-molecular-weight (MW) peak to a higher-MW complex (Fig. 4c)

upon addition of HtsptLB12.v1. Chromatography of mixtures of the two proteins in different proportions supported an equimolar ratio within the complex as designed (Supplementary Fig. S6), and comparison of the elution volumes of the bound and unbound proteins to an MW standard is consistent with 1:1 stoichiometric binding (Fig. 4c).

To provide more direct information on the interaction surface, we incorporated 4-fluorophenylalanine into HtsptLB12.v2 and measured the ^{19}F NMR spectrum. HtsptLB12.v2 has nine phenylalanine residues, but only three are anticipated to be within or proximal to the designed interface (Fig. 4d). Upon addition of increasing amounts of HEL, three of the HtsptLB12.v2 ^{19}F -Phe resonances changed resonance frequency and/or intensity due to HEL interactions (Fig. 4e). ^{19}F NMR spectra of phenylalanine-to-alanine mutants unambiguously assigned these resonance peaks to F17, F40, and F125, the three phenylalanine residues located at the designed interface

(Supplementary Fig. S7). The increased intensity of the ^{19}F resonance at -114.8 ppm at higher HEL concentrations is attributed to degeneracy of the ^{19}F resonances of F125, F40, and/or F17 in the complex. This provides further evidence that the HtsptLB12–HEL interaction in the complex involves the *in silico* designed interface.

A binding fitness landscape reveals the importance of designed residues surrounding the grafted hot spots

While the limited mutagenesis, affinity maturation, unbound crystal structure, and SEC and NMR data are all consistent with the designed HtsptLB12–HEL model, to more thoroughly characterize the binding interface, we used site-saturation mutagenesis combined with selection and deep sequencing to generate a sequence fitness landscape. The landscape is mapped by determining the frequency of

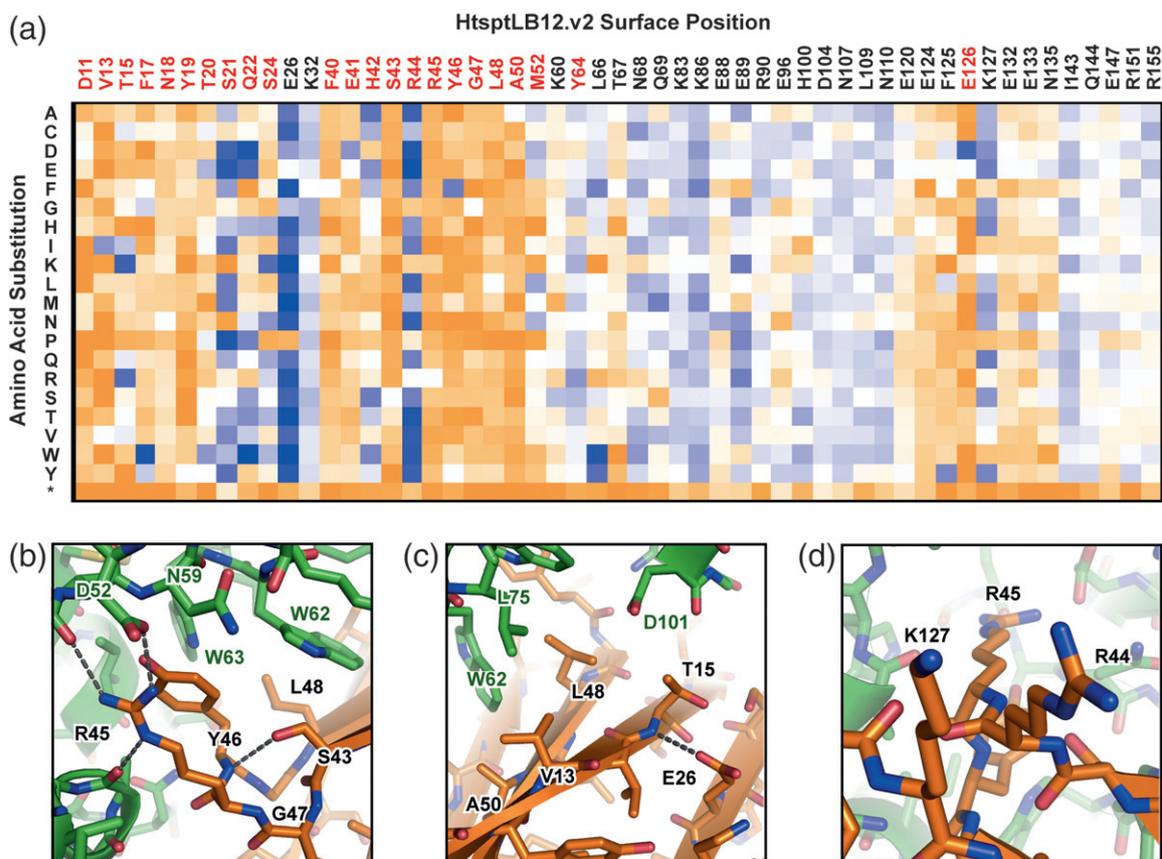


Fig. 5. A binding fitness landscape indicates that interface residues surrounding the grafted hot spots are critical for activity. (a) Fifty-three surface positions of HtsptLB12.v2 were chosen for single site-saturation mutagenesis. Mutants were selected by yeast display and one round of FACS after incubation with 4 nM HEL. The number of transformants in the yeast library was 1.5×10^6 , sufficient for the 1696 unique DNA sequences. Cells falling within the top 1.5% of events measured by binding signal relative to protein expression were collected, and plasmid DNA was harvested and sequenced. The \log_2 enrichment ratio for each amino acid substitution is plotted from -3.5 (i.e., depleted, orange) to $+3.5$ (i.e., enriched, blue). Residues within 10 Å of the designed interface are in red text. *, stop codon. (b–d) Regions of the computationally designed interface discussed in the main text. HEL is green and HtsptLB12 is orange.

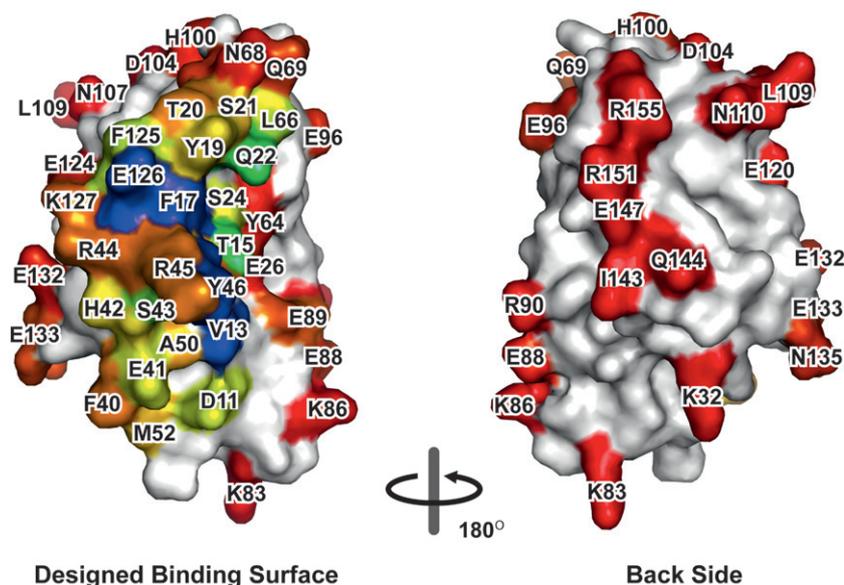


Fig. 6. The designed binding surface is conserved during *in vitro* evolution. Sequence conservation measured by Shannon entropy is mapped onto the surface of HtsptLB12. Entropy color scale is from ≤ 3.2 (blue) to 4.3 (red).

recovery of a clone in a selected population relative to an unselected population [29–31].

Fifty-three surface positions of HtsptLB12.v2 were independently diversified by introduction of an NNK codon (N is any base, K is G or T) using PCR fragment assembly (Supplementary Fig. S8). The combined library containing 53 sites \times (20 amino acids + 1 stop codon) = 1113 protein variants was transformed into yeast and cells were sorted with a single round of FACS after incubation with 4 nM HEL. The frequencies of each variant were then compared between the naive/pre-sorted and enriched/post-sorted yeast populations using Illumina deep sequencing. A total of 1,765,230 sequences passing a quality filter were obtained for the naive population; the numbers of reads for all single amino acid substitution variants ranged from 21 to 7306, with a median of 511 reads per mutant. Sequences encoding the parental protein accounted for $\sim 25\%$ of reads. A total of 1,650,263 sequences passing the quality filter were obtained for the sorted population.

Enrichment ratios are plotted in Fig. 5a for the 53 mutated surface positions of HtsptLB12.v2 after selection for high-affinity HEL binding. The data are closely consistent with the computational model. The hot spot residues R45 and Y46 are highly conserved, with all substitutions (except for the conservative Y46F change) depleted. Other designed residues that make direct interfacial contacts, such as V13, L48, and A50, are similarly highly conserved, as are S43 and G47 that structurally support the hot-spot-bearing loop conformation (Fig. 5b and c). Substitutions HtsptLB12-T15K/R are enriched, likely reflecting increases in affinity resulting from interactions with HEL-D101 (Fig. 5c). Substitutions to HtsptLB12-E126, which is designed to form a

salt-bridge interaction with HEL-R112, are depleted save a conservative E126D mutation. HtsptLB12-R44 and K127 are adjacent on the back side of the hot-spot-bearing loop (Fig. 5d) and are likely to interact unfavorably both with each other (destabilizing the loop) and with strongly positively charged HEL; not surprisingly, mutations of either to non-basic amino acids are enriched. Mutations of HtsptLB12-E26, which hydrogen bonds to the backbone of the N-terminal edge β -strand (Fig. 5c), are also favored; these may allow small backbone adjustments for enhanced HEL contacts.

The designed binding surface is strikingly more conserved than predicted non-interacting regions. Figure 6 shows the Shannon entropy for each mutated position mapped on the HtsptLB12 surface. The grafted hot spots R45 and Y46 form only a fraction of the conserved interaction surface. Indeed, E126 has the lowest sequence entropy, and V13, F17, and hot spot Y46 follow with nearly identical sequence entropies. Specific interactions across the entire interface were critical to the success of the design; simply grafting a couple of centrally located hot spot residues alone cannot provide all the interactions necessary to stabilize a noncovalent protein–protein complex with polar character.

The designed protein inhibits lysozyme catalytic activity

Mutations to five HtsptLB12.v2 positions (T15K/R, R44D/E/N/T, and E126D described above, and S21D/E/P and Q22D/E targeting the same loop as the second round of affinity maturation) were combined in a yeast display library, which was sorted for higher affinity to HEL. These mutations were chosen because they were localized at the

interface and had \log_2 enrichment ratios of at least 2 from the deep sequencing data. The sorted library was well converged, with 65% of clones being a single sequence (HtsptLB12.v3, Fig. 7a) with nine mutations from the starting design. Apparent dissociation constants from yeast display show improved binding after each round of directed evolution (Fig. 7b) to a final apparent K_d for HtsptLB12.v3 of 1.4 ± 0.3 nM ($n = 5$). The protein was expressed and purified from *Escherichia coli*, and the K_d obtained using bio-layer interferometry was ~ 3 nM (Supplementary Fig. S9).

Purified HtsptLB12 variants inhibited HEL-catalyzed hydrolysis of the bacterial cell wall, measured by following the decrease in optical density of a HEL-treated *Micrococcus lysodeikticus* culture (Fig. 7c). The inhibitory efficacies of the HtsptLB12 variants matched their ranked apparent affinities by yeast display, with the concentration at which 500 nM HEL acting on 0.1 mg/ml cell substrate is 50% inhibited (IC_{50}) ranging from >20 μ M for the original design to 300 nM for the most evolved variant. Due to poor assay sensitivity with low substrate/enzyme concentrations, we have been unable to measure the K_i accurately; the IC_{50} for HtsptLB12.v3 indicates the K_d in solution must be less than 50 nM. Inhibition of HEL catalysis provides further evidence that the designed inhibitor interacts with the lysozyme active site.

Discussion

We have used hot-spot-based protein interface design to generate a protein inhibitor of a polar enzyme active site. Compared to a simple dock-and-design strategy that failed to preclude alternative binding interactions, a combination of grafting hot spot residues and appropriately designing the surrounding surface created an enzyme inhibitor with interactions that matched the computational model. Thus, computational algorithms can successfully design interfaces with native-like polar surfaces using hot spot residues to make the necessary buried polar interactions.

Several important lessons can be drawn from the design and its characterization. First, specific interactions across the entire interface were critical. Many substitutions of surrounding interface residues were as disfavored as deleterious hot spot mutations or introduction of nonsense stop codons. This contrasts with computational epitope grafting methods (exemplified by grafting antigenic epitopes to protein scaffolds for vaccines [32]), where the grafted epitope provides all the necessary binding energy. It also contrasts to the designed erythropoietin receptor binder, in which grafting just three hot spot residues to a scaffold was sufficient for moderate affinity binding [10]. Here, hot spot

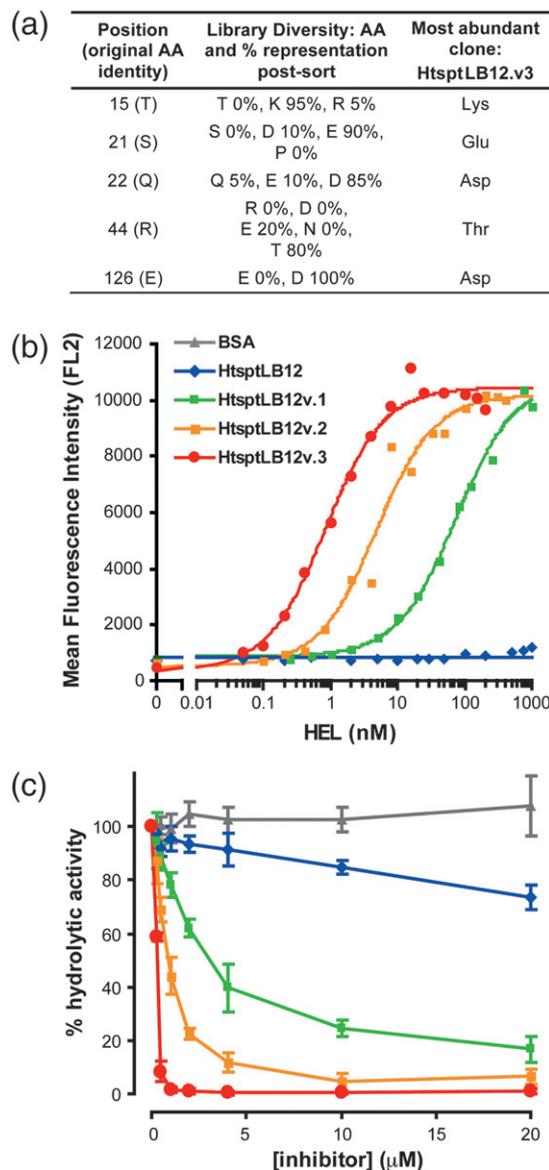


Fig. 7. HtsptLB12 inhibits lysozyme activity. (a) An HtsptLB12.v2 variant library was constructed combining multiple affinity-enhancing mutations. The yeast display library (containing 2×10^6 transformants) was sorted for four rounds with increasing stringency: 2 nM HEL incubation in round 1, 0.5 nM HEL in round 2, and 0.2 nM HEL for rounds 3 and 4. Twenty clones from the final enriched population were sequenced, and the proportion of sequences with a particular amino acid at each of the diversified positions is tabulated (middle column). The most abundant clone sequence, representing 65% of the final enriched population, is shown on the right. (b) Yeast display titration curves. HtsptLB12 variants were expressed on the yeast surface, and cells were incubated with monomeric HEL at the indicated concentrations. HEL binding signal is detected by flow cytometry in the FL2 fluorescence channel. (c) Cell wall hydrolysis of an *M. lysodeikticus* suspension by 500 nM HEL is inhibited by HtsptLB12 variants.

placement is the initial seed for forming a specific interface, around which many new essential interactions must still be designed.

Second, while residues that form close-packing interactions with the target site were conserved during the *in vitro* optimization and in the sequence-fitness landscape, residues at the periphery, where long-range electrostatics and solvation have prominent roles in binding energetics, were not optimal in the original design and suggest areas for energy function improvement. Poor hydrogen-bonding geometry and electrostatic repulsion between nearby basic residues were corrected in the first round of directed evolution, while residues impacting a single and perhaps flexible loop 18–22 were found in the second round of evolution (Fig. 3). Affinity-enhancing substitutions identified from the deep sequencing experiment likely improve peripheral electrostatic contacts and peripheral loop 18–22 further (Figs. 5 and 7a). Further illustrating the need for energy function improvements, the optimized design is computed to have slightly higher interaction energy [~ 2 Rosetta energy units (REU)] than the original. The computational method is therefore essentially neutral to the changes that occur for reasons stated above.

Third, accurate design of new protein–protein interactions works best with rigid contact surfaces, when the mutated scaffold is unlikely to change backbone structure in unforeseen ways. Most successful interface designs have placed critical side chains on rigid secondary structural elements [6,11,33], and a connectivity metric for assessing how well the binding residues are “connected” to the protein core fared favorably in a community-wide challenge to discriminate verified designed binding proteins from decoys [12]. While HtsptLB12 does not stand out from other nonfunctional designs on any single measurable attribute at the designed interface (Table S1), HtsptLB12 does have a high average degree of connectivity (9.5 for HtsptLB12, measured within 8 Å of the interface, with the other hot-spot-based designs experimentally tested here having an average connectivity of interface residues from 6.9 to 10.4). HtsptLB12 primarily interacts with HEL via a rigid concave β -sheet and a four-residue loop on which the hot spots are placed, also almost certainly rigid due to its short length, intricate backbone hydrogen bonding, and moderate *B*-factors in the crystal structure.

Protein design is a rigorous test of understanding of macromolecular energetics. Our results show that there are no fundamental difficulties in designing binders to polar sites on protein surfaces, which has not been accomplished previously. The hot spot residues in this case were critical since they compensate for the interactions of polar residues in the active site with water molecules in the unbound

state, but many additional interactions were required for a favorable free energy of binding. A key next challenge is to develop methods for identifying such polar hot spot residues *de novo* to enable design of binders to polar surfaces more generally.

Materials and Methods

Computational design

Scaffold proteins were docked to the HEL active site using PatchDock [24]. A PatchDock constraints file that defines the receptor active site is shown in Supplementary Fig. S10. For each scaffold, docked configurations were chosen at random from the top scoring hundred and were refined, hot spot residues were grafted, and interface residues were designed using the RosettaScripts interface from the ROSETTA software package [34]. RosettaScripts protocols for dock-and-design without and with loop remodeling and for hot spot placement followed by design are shown in Supplementary Figs. S11, S12, and S13, respectively. For dock-and-design, scaffold loops within 8 Å of the interface were remodeled [35] in some cases and their sequences were designed to stabilize new conformations that might make improved atomic contacts. Backbone changes were slight to modest due to very short loops on most scaffolds (Supplementary Table S1). Stubs files containing the inverse rotamers libraries for hot spot placement were generated as previously described [23] using RosettaScripts (Supplementary Fig. S14). Designs were filtered based on calculated interface binding energy < -18 REU, shape complementarity > 0.6 , interface packing score < 1.0 , and buried solvent accessible surface area > 800 Å². Manual adjustments were made in FoldIt [36]. This included substituting some free cysteines to A or S depending on local structure if there was concern of spurious disulfide formation, and mutated residues were reverted back to the scaffold's native amino acid if calculated binding affinity remained similar. This increases folding probability by making the fewest mutations necessary.

Plasmid and library construction

Genes for designed proteins were synthesized (Genescript) with codon usage optimized for yeast expression and cloned between the NdeI and XhoI sites of pETCON [11] for yeast surface display as an Aga2p-fusion protein. Sequences for DnvLB16 and HtsptLB12 are shown in Supplementary Fig. S15 and are deposited with Addgene as plasmids 45120–45125. Site-directed mutations were made by overlapping PCR and confirmed by sequencing (Genewiz). Plasmids were transformed into chemically competent EBY100 yeast cells. Error-prone PCR libraries were generated using GeneMorph II Random Mutagenesis (Agilent Technologies). The single site-saturation mutagenesis library was generated using overlapping PCR (Supplementary Fig. S8 and Table S2). Libraries were transformed as linear PCR product together with linear cut pETCON (digested with NheI, XhoI, and BamHI) into EBY100 yeast cells using electroporation.

Yeast surface display

Transformed yeast cells were grown overnight in SDCAA media (2% w/v glucose, 0.67% w/v yeast nitrogen base, 0.5% w/v casamino acids, and 0.1 M sodium phosphate, pH 6.6) at 30 °C and induced at an OD₆₀₀ (optical density at 600 nm) of 0.5 in SGCAA media (glucose is replaced with galactose) at 22 °C for 2 days. Cells were washed in phosphate-buffered saline (PBS) (140 mM NaCl, 1 mM KCl, 12 mM Na₂HPO₄, and 1.2 mM KH₂PO₄, pH 7.4) supplemented with 0.1% bovine serum albumin (BSA) and resuspended in PBS-BSA containing biotinylated ligand. A concentrated stock solution of biotinylated lysozyme (Sigma-Aldrich) was prepared in 20 mM HEPES (pH 5) and 50 mM NaCl for adding at the indicated concentrations. Cells were incubated at 22 °C for 2–4 h with agitation, washed with cold PBS-BSA, incubated for 10 min on ice with PBS-BSA containing 10 µg/ml streptavidin–phycoerythrin (Invitrogen) and 5 µg/ml fluorescein isothiocyanate-conjugated chicken anti-c-myc (Immunology Consultants Laboratory), and washed and resuspended in PBS-BSA for analysis. Cells were analyzed on a C6 flow cytometer operated by CFlow Plus software (Accuri) and sorted with a BD Influx cytometer operated by Spigot (BD Biosciences). Sorted cells were recovered in SDCAA at 30 °C for another round of sorting or for plasmid DNA preparation.

Deep sequencing analysis

Naive and sorted yeast cultures were lysed by incubation with 125 U/ml Zymolase at 37 °C for 5 h and freeze-thawing, followed by DNA purification as per manufacturer's directions (ZymoPrep kit from Zymo Research). Contaminating genomic DNA was removed with treatment for 90 min at 30 °C with 2 U/µl Exonuclease I (New England Biolabs) and 0.25 U/µl Lambda exonuclease (New England Biolabs) and plasmid DNA was purified by QIAquick kit (Qiagen). The HtsptLB12 gene was amplified in two fragments to provide full sequencing coverage. A first round of PCR used primers that annealed to the plasmid with overhangs adding sequences for the recommended Illumina MiSeq sequencing primers to anneal later. A second round of PCR added additional end sequences for annealing to the Illumina flow cell oligonucleotides and included a short 6-bp barcode for unique sample identification. Each round of PCR was limited to 12 cycles with high-fidelity Phusion polymerase (New England Biolabs) to minimize errors. Gel-extracted and purified DNA was sequenced with a MiSeq sequencer using a 300-cycle paired-end reads reagent kit as per the manufacturer's directions (Illumina), and sequences passing the chastity quality filter were analyzed with scripts adapted from Enrich [37]. To calculate conservation score, we applied Shannon's entropy equation to a hypothetical selected population that began with an even distribution of all possible point substitutions to which the experimentally determined enrichment ratios were applied.

Protein sample production

HtsptLB12 constructs were cloned between the NdeI and XhoI sites of pET29b (Novagen), placing a 6His-tag on the protein's C-terminus. *E. coli* BL21(DE3) cells trans-

formed with plasmid were grown in Terrific Broth at 37 °C to OD₆₀₀ ~0.5 and induced with 0.1 mM IPTG overnight at 20 °C. Cells were lysed in lysis buffer (0.3 M NaCl, 20 mM Tris-Cl, pH 8.0, 0.5 mM phenylmethylsulfonyl fluoride, and 20 mM imidazole) containing 0.05 mg/ml DNase I and 0.1% Triton X-100 by sonication. Cleared lysate was loaded on NiNTA resin (Qiagen) and washed with 30 column volumes of lysis buffer. Protein was eluted with a 25–250 mM imidazole step gradient in lysis buffer and ethylenediaminetetraacetic acid was added to 1 mM to inhibit proteases. Protein was dialyzed overnight at 4 °C against buffer A (20 mM Tris-Cl, pH 8.0, 50 mM NaCl, and 0.25 mM DTT) containing 1 mM ethylenediaminetetraacetic acid. Protein was further purified on a HiTrap Q ion-exchange column (GE Healthcare) using gradient elution from buffer A to buffer B (containing 0.5 M NaCl). Concentrated protein was separated on a Sephacryl-200/16/60 column (GE Healthcare) with running buffer (20 mM Tris, pH 7.5, and 100 mM NaCl). Protein was concentrated by centrifugal ultrafiltration and concentration was determined by absorbance at 280 nm using calculated extinction coefficients.

SEC of complexes

Proteins were separated on a Superdex-75/10/300 column (GE Healthcare) using a 200-µl injection loop. Individual proteins have excellent solubilities, but the HEL-HtsptLB12 complex has reduced solubility in low salt, necessitating a moderately high salt running buffer (20 mM Tris-Cl, pH 7.5, and 400 mM NaCl).

Preparation of ¹⁹F-Phe-labeled samples

Proteins were expressed as previously described [38] with some modifications. Briefly, proteins were expressed in BL21-CodonPlus (DE3)-RILP cells (Agilent) in MJ9 media, chilled on ice water for 15 min once OD₆₀₀ reached ~0.5, and 50 mg DL-5-F-Trp was added as powder. Cultures were then grown at 17 °C for 1 h and induced with 1 mM IPTG for 24 h.

¹⁹F NMR spectroscopy

The experiments were performed on a Varian INOVA 500-MHz spectrometer equipped with a 5-mm ¹H/¹⁹F probe with a sample temperature of 25 °C. The transmitter frequency was set to 470.182 MHz with an offset of –6390.4 Hz. The spectra were referenced to an external sample 0.05% α,α,α-trifluorotoluene in C₆D₆ at –62.7 ppm. Data were acquired and processed with VNMRJ2.1 software, with the first four points of the FID replaced using linear prediction to eliminate a baseline roll. The spectra were acquired using 20,000 scans, with a 0.4-s acquisition time and a pulse width of 5 µs.

HtsptLB12.v1 structure determination

HtsptLB12.v1 was crystallized by microbatch under oil method at 4 °C from drops containing a 1:1 mixture of protein solution (7.0 mg/ml) and well precipitant solution

(1.59 M MgSO₄ and 0.1 M Tris–Cl, pH 8.0). Crystals were cryo-protected with 20% glycerol and flash frozen in liquid nitrogen. Diffraction data were collected at beamline X4A, and the initial phases were solved by molecular replacement with PDB 3II3 as the search model using MOLREP [39]. A few regions of the model were built manually and refined with CNS [40] to a final $R_{\text{cryst}}/R_{\text{free}}$ of 0.24/0.29. Diffraction and refinement statistics are in Supplementary Table S3.

HEL activity assay

HEL (Sigma-Aldrich) activity was measured at 500 nM enzyme in PBS at 24 °C against 0.1 mg/ml *M. lysodeikticus* (Sigma-Aldrich) by following the decrease in OD at 450 nm. Recordings were taken with a Spectramax M5e plate reader (Molecular Devices) for 10 min with mixing, and the initial linear OD₄₅₀ decrease was used for determining the hydrolysis rate.

Accession numbers

The structure of HtsptLB12.v1 is deposited in the PDB with accession code 3VB8.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jmb.2013.06.035>

Acknowledgements

This work was supported by the National Institute of General Medical Studies of the National Institutes of Health under award number P41 GM103533, Defense Threat Reduction Agency grant HDTRA1-10-0040 and by a grant from the Protein Structure Initiative of the National Institutes of Health (U54-GM094597) to the Northeast Structural Genomics Consortium. S.J.F. was supported by the Human Frontier Science Program.

Received 23 April 2013;

Received in revised form 15 June 2013;

Accepted 25 June 2013

Available online 1 July 2013

Keywords:

protein–protein interactions;

hot spot;

Rosetta molecular modeling program;

protein engineering and design

Abbreviations used:

FACS, fluorescence-activated cell sorting; HA, hemagglutinin; HEL, hen egg lysozyme; PAK1, p21-activated kinase 1; PDB, Protein Data Bank; VNAR, variable new antigen receptor; SEC, size-exclusion chromatography; MW, molecular weight; PBS, phosphate-buffered saline; BSA, bovine serum albumin.

References

- [1] Stranges PB, Machius M, Miley MJ, Tripathy A, Kuhlman B. Computational design of a symmetric homodimer using beta-strand assembly. *Proc Natl Acad Sci USA* 2011;108:20562–7.
- [2] Lanci CJ, MacDermaid CM, Kang SG, Acharya R, North B, Yang X, et al. Computational design of a protein crystal. *Proc Natl Acad Sci USA* 2012;109:7304–9.
- [3] Salgado EN, Ambroggio XI, Brodin JD, Lewis RA, Kuhlman B, Tezcan FA. Metal templated design of protein interfaces. *Proc Natl Acad Sci USA* 2010;107:1827–32.
- [4] Huang PS, Love JJ, Mayo SL. A de novo designed protein–protein interface. *Protein Sci* 2007;16:2770–4.
- [5] Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, et al. A de novo protein binding pair by computational design and directed evolution. *Mol Cell* 2011;42:250–60.
- [6] Jha RK, Leaver-Fay A, Yin S, Wu Y, Butterfoss GL, Szyperski T, et al. Computational design of a PAK1 binding protein. *J Mol Biol* 2010;400:257–70.
- [7] Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol* 1998;280:1–9.
- [8] Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science* 1995;267:383–6.
- [9] Moreira IS, Fernandes PA, Ramos MJ. Hot spots—a review of the protein–protein interface determinant amino-acid residues. *Proteins* 2007;68:803–12.
- [10] Liu S, Zhu X, Liang H, Cao A, Chang Z, Lai L. Nonnatural protein–protein interaction-pair design by key residues grafting. *Proc Natl Acad Sci USA* 2007;104:5330–5.
- [11] Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 2011;332:816–21.
- [12] Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, et al. Community-wide assessment of protein–interface modeling suggests improvements to design methodology. *J Mol Biol* 2011;414:289–302.
- [13] Stranges PB, Kuhlman B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci* 2013;22:74–82.
- [14] Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 1965;206:757–61.
- [15] Stanfield RL, Dooley H, Flajnik MF, Wilson IA. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* 2004;305:1770–3.
- [16] Davies DR, Cohen GH. Interactions of protein antigens with antibodies. *Proc Natl Acad Sci USA* 1996;93:7–12.
- [17] Aberger C, Monchois V, Byrne D, Chenivresse S, Lembo F, Lazzaroni JC, et al. Structure and evolution of the Ivy protein family, unexpected lysozyme inhibitors in Gram-negative bacteria. *Proc Natl Acad Sci USA* 2007;104:6394–9.
- [18] Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 2011;487:545–74.
- [19] Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, et al. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 2003;331:281–99.
- [20] Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–8.

- [21] Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* 2006;1:755–68.
- [22] Hackel BJ, Kapila A, Wittrup KD. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *J Mol Biol* 2008;381:1238–52.
- [23] Fleishman SJ, Corn JE, Strauch EM, Whitehead TA, Karanicolas J, Baker D. Hotspot-centric de novo design of protein binders. *J Mol Biol* 2011;413:1047–62.
- [24] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 2005;33:W363–7.
- [25] Goulet A, Pina M, Redder P, Prangishvili D, Vera L, Lichiere J, et al. ORF157 from the archaeal virus Acidianus filamentous virus 1 defines a new class of nuclease. *J Virol* 2010;84:5025–31.
- [26] Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 1997;15:553–7.
- [27] Whitaker JR. Determination of molecular weights of proteins by gel filtration on Sephadex. *Anal Chem* 1963;35:1950–3.
- [28] Fleishman SJ, Khare SD, Koga N, Baker D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci* 2011;20:753–7.
- [29] Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence–function relationships. *Nat Methods* 2010;7:741–6.
- [30] McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. *Nature* 2012;491:138–42.
- [31] Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 2012;30:543–8.
- [32] Ofek G, Guenaga FJ, Schief WR, Skinner J, Baker D, Wyatt R, et al. Elicitation of structure-specific antibodies by epitope scaffolds. *Proc Natl Acad Sci USA* 2010;107:17880–7.
- [33] King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 2012;336:1171–4.
- [34] Fleishman SJ, Leaver-Fay A, Corn JE, Strauch EM, Khare SD, Koga N, et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* 2011;6:e20161.
- [35] Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods* 2009;6:551–2.
- [36] Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. *Nature* 2010;466:756–60.
- [37] Fowler DM, Araya CL, Gerard W, Fields S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* 2011;27:3430–1.
- [38] Acton TB, Xiao R, Anderson S, Aramini J, Buchwald WA, Ciccocanti C, et al. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol* 2011;493:21–60.
- [39] Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* 1997;30:1022–5.
- [40] Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905–21.