

# Specificity versus stability in computational protein design

Daniel N. Bolon\*, Robert A. Grant\*, Tania A. Baker\*<sup>†</sup>, and Robert T. Sauer\*<sup>‡</sup>

\*Department of Biology and <sup>†</sup>Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Robert T. Sauer, July 19, 2005

Protein–protein interactions can be designed computationally by using positive strategies that maximize the stability of the desired structure and/or by negative strategies that seek to destabilize competing states. Here, we compare the efficacy of these methods in reengineering a protein homodimer into a heterodimer. The stability-design protein (positive design only) was experimentally more stable than the specificity-design heterodimer (positive and negative design). By contrast, only the specificity-design protein assembled as a homogenous heterodimer in solution, whereas the stability-design protein formed a mixture of homodimer and heterodimer species. The experimental stabilities of the engineered proteins correlated roughly with their calculated stabilities, and the crystal structure of the specificity-design heterodimer showed most of the predicted side-chain packing interactions and a main-chain conformation indistinguishable from the wild-type structure. These results indicate that the design simulations capture important features of both stability and structure and demonstrate that negative design can be critical for attaining specificity when competing states are close in structure space.

adaptor protein | protein engineering | SspB

Highly specific recognition lies at the core of most cellular processes. The interaction of two partner proteins in a crowded intracellular environment depends on the equilibrium stability of the complex, which is determined by affinity and concentration, but is also controlled by the competing binding of either partner to other cellular macromolecules. Efficient protein recognition requires both stability and specificity. In natural systems, both parameters are subject to evolutionary optimization. Likewise in engineered systems, stability can be targeted for maximization, and known competing states or interactions can be minimized through the use of negative design. The biophysical tradeoff between stability and specificity in molecular recognition is a fascinating problem that has important implications in the development of practical tools for synthesizing and dissecting biological systems.

Focusing on stability without explicit consideration of specificity has resulted in impressive protein-engineering feats, including full-sequence design of a zinc-finger protein that folds in the absence of metal (1), introduction of catalytic activity into previously inert protein scaffolds (2, 3), and the creation of a novel protein fold (4). These successes relied on positive design only. Positive design maximizes favorable interactions in the target conformation. Negative design, by contrast, maximizes unfavorable interactions in competing states and requires modeling of each unwanted conformation (5, 6).

Protein–protein interactions have been successfully reengineered to alter binding specificity both by using and ignoring negative design (7–15). How is success possible in the absence of negative design? One possibility is that most changes that optimize the stability of a target structure or complex have random energetic effects on nontarget conformations, making explicit negative design unnecessary. However, if undesired states are similar in structure to the target state, then energetic effects are more likely to be correlated, and mutations that stabilize the target state may also stabilize competing states. In

such cases, negative design should be critical for achieving specificity.

How important is negative design for achieving specificity in protein–protein interactions? To address this question, we compared a strategy that seeks to maximize specificity through positive and negative design with one that optimizes only the stability of the target conformation (positive design only). Our experimental system is the SspB adaptor protein, which forms a wild-type homodimer (16). We previously reported designed sequence changes that allow two SspB subunits to assemble as a heterodimer, and used this protein to study adaptor-mediated delivery of protein substrates to the ClpXP protease (9). Here, we compare the positive and negative optimization strategies used in this design to a strategy that ignores explicit negative design. We show that SspB mutants designed solely for heterodimer stability were successful in meeting this objective but also formed equally stable homodimers. By contrast, SspB mutants that were computationally optimized for specificity assembled almost exclusively as a heterodimer, but this protein was less stable than the molecule designed for stability. To compare the actual and “designed” conformations, we determined the crystal structure of the specificity-design heterodimer. Collectively, our results emphasize a key role for negative design in engineering protein assembly reactions that are highly specific.

## Methods

**Computational Design.** Interface positions 12, 15, 16, and 101 from each monomer of the 1OU9 crystal structure of *Haemophilus influenzae* SspB (16) were computationally randomized, allowing Gly, Ala, Ser, Val, Thr, Leu, Ile, Phe, Tyr, or Trp. Side-chain geometry was varied by using a rotamer library based on a survey of protein structures (1, 17). Remaining portions of the structure were held fixed. To enable rapid computational searches, a pairwise energy matrix was constructed to describe all side-chain combinations. Energies were calculated based on the Dreiding force field (18) and included a Lennard–Jones potential for van der Waals (vdW) interactions after scaling atomic radii by 0.9 (19), a surface-area based hydrophobic solvation potential including an exposure penalty with a sequence-independent unfolded reference state ( $\sigma_{np} = 0.048 \text{ kcal/mol/\AA}^2$ ,  $\kappa = 1.6$ ) (20), a geometric hydrogen-bond potential with a well depth of 8 kcal/mol, a coulombic electrostatic potential with a distance-dependent dielectric constant of 40R, and a polar-hydrogen burial penalty of 2 kcal/mol (21). Calculations were performed on a Silicon Graphics Indigo2 workstation with a 175-MHz R10000 processor. For the stability design, optimization was performed by using the dead-end elimination (DEE) search algorithm (22, 23), which converged from  $10^{16}$  possible rotamer combinations to the final solution in approximately 1 min.

For the heterodimer specificity design, several modifications

Abbreviations: DEE, dead-end elimination; vdW, van der Waals.

Data deposition: The designed heterodimer structure has been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID code 1Z5Z).

<sup>†</sup>To whom correspondence should be addressed. E-mail: [bobsauer@mit.edu](mailto:bobsauer@mit.edu).

© 2005 by The National Academy of Sciences of the USA

were made to the protein design algorithm. To approximate conformational relaxation in competing states, we placed an energy ceiling of 5 kcal/mol on unfavorable vdW interactions in the homodimer states only. Without this energy ceiling, we noted that homodimer sequences with vdW overlaps led to high homodimer energies, driving the selection of sequences with poor predicted heterodimer stabilities. We sought sequences in which the  $2AB \rightleftharpoons AA + BB$  equilibrium lay as far as possible to the left by searching for low optimization energies [ $E_{\text{opt}} = 2E_{AB} - E_{AA} - E_{BB}$ ]. We did not include the denatured states of the folded subunits in these studies. For every possible homodimer sequence, the energy ( $E_{A2}$  and  $E_{B2}$ ) of the eight designed positions with optimal rotamer geometry was determined in separate DEE calculations. These optimizations started with a pairwise matrix representing all side chains; for each possible homodimer sequence, nonsequence rotamers were eliminated in a first step before optimizing side-chain geometry with DEE. An energy matrix was calculated with a standard Lennard-Jones vdW potential to describe the heterodimer state. Combined with the precalculated homodimer sequence energies, this method allowed the optimization energy to be rapidly determined for any heterodimer conformation. For the final optimization, we used a Monte Carlo search (24) with  $10^9$  total steps. The best solution was found after  $7.7 \times 10^4$  steps. To test the robustness of the search, it was repeated ten times with different random seeds. Each search returned the same solution, suggesting a global optimum.

**Protein Expression and Purification.** The structured portion of *Haemophilus influenzae* SspB (residues 1–106) was included in all protein constructs. Full-length constructs contained 58-residue tails from the C terminus of *Escherichia coli* SspB; this HE chimera has been shown to be fully active in stimulating ClpXP degradation of GFP-SsrA degradation (9). Untagged SspB subunits were coexpressed with His-6-SspB in recombination-deficient BLR (DE3) cells (Novagen) grown at 37°C in 2XYT broth. Cells were lysed by sonication in 50 mM potassium phosphate (pH 8)/20 mM imidazole, and debris was removed by centrifugation at  $20,000 \times g$  for 30 min. The supernatant was bound to Qiagen  $\text{Ni}^{++}$ -NTA resin (4 ml of resin per liter of culture). After washing, untagged SspB was eluted with 6 M GuHCl/50 mM potassium phosphate (pH 8). Residual His-tagged protein in this eluate was removed by passage through another  $\text{Ni}^{++}$ -NTA column (1 ml of resin per liter of culture). For heterodimers, equimolar untagged SspB subunits were mixed in 6 M GuHCl, exchanged into 50 mM potassium phosphate (pH 7) by using a PD10 column (Amersham Pharmacia), and further purified by ion-exchange chromatography on a MonoQ column.

**Biophysical Characterization.** Urea denaturation assays were performed in 50 mM potassium phosphate (pH 6.8) at 30°C. Unfolding was monitored by changes in tryptophan fluorescence using a QM-2000-4SE spectrofluorometer (Photon Technology International). Samples of SspB or variants (1.5  $\mu\text{M}$  dimer) were incubated with different concentrations of urea for at least 1 h, and the fraction unfolded was determined from the center-of-mass fluorescence after correction for the fluorescence of the folded and unfolded states. Based on studies of wild-type SspB (25), denaturation was fit as a transition from folded dimer to unfolded monomers by using a nonlinear least-squares algorithm implemented in KALEIDAGRAPH (Synergy Software). Analytical ultracentrifugation experiments were performed with 30  $\mu\text{M}$  SspB at 20°C in 20 mM potassium phosphate (pH 7) at 16,000 rpm in a Ti-60 rotor on a Beckman XLA instrument with equilibration for 36 h.

To monitor populations of heterodimers and homodimers, equal concentrations of GuHCl-denatured subunits (one with a

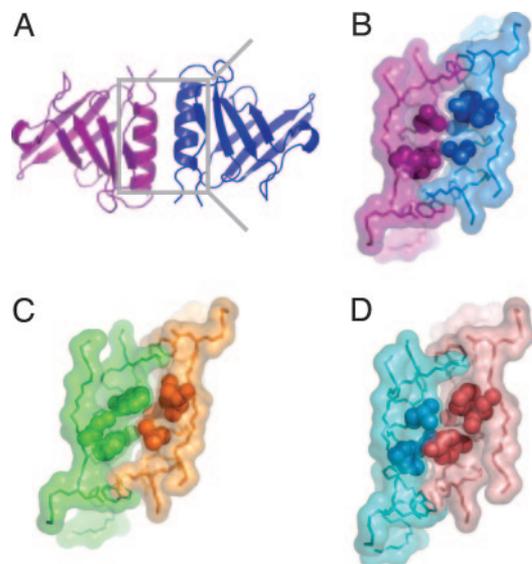
C-terminal tail and one without a C-terminal tail) were mixed at room temperature and allowed to refold after desalting on a PD10 column into R buffer (25 mM Hepes, pH 7.6/1 mM EDTA) plus 100 mM KCl. Heterodimers containing one truncated subunit and one full-length subunit were purified on a MonoQ column by using a linear gradient in R buffer from 180 mM to 315 mM KCl. Fractions containing the heterodimer were pooled and diluted with an equal volume of water, and half of the sample was immediately rechromatographed on the MonoQ column. The remaining sample was incubated for 24 h at 30°C and then rechromatographed.

**X-Ray Crystallography.** For crystallography, one subunit of the heterodimer contained *H. influenzae* SspB residues 1–129 with Tyr-12, Gly-15, Phe-15, Met-101, and an additional A73Q mutation, and the other subunit contain residues 1–111 with Leu-12, Ser-15, Leu-16, and Ala-101. These subunits were expressed, purified, combined, and crystallized in hanging drops containing 2  $\mu\text{l}$  of protein (5.7 mg/ml) and 2  $\mu\text{l}$  of well solution [0.1 M sodium cacodylate, pH 5.7/200 mM  $\text{CaCl}_2$ /150 mM KCl/12% (wt/vol) PEG 6000/9% (vol/vol) glycerol]. Crystals (space group C2) had unit cell dimensions  $a = 120.0 \text{ \AA}$ ,  $b = 61.0 \text{ \AA}$ , and  $c = 62.4 \text{ \AA}$  with  $\beta = 110.9^\circ$ . Crystals were flash-frozen, and diffraction data were collected on the Advanced Photon Source 8BM beamline and processed with the HKL suite of programs (26). A wild-type SspB subunit (16) with the side chains of residues 12, 16, and 101 truncated to Ala was used for molecular replacement with AMORE (27) in the CCP4 program suite (28). The asymmetric unit contained three subunits; two formed a dimer within the asymmetric unit, and one formed a dimer across a crystallographic symmetry axis. After simulated annealing using CNS (29), clear electron density was observed for the side chains of residues 12, 15, 16, and 101 in both subunits of the noncrystallographic dimer and for Gln-73 in one subunit, allowing model building of these positions. The crystallographic dimer contained averaged density at positions 12, 15, 16, 73, and 101, and these positions were modeled as Ala or Gly. The structure was refined with alternate cycles of manual model building using O (30), positional and individual B-factor refinement with CNS, and addition of water molecules. The final model had a working R factor of 0.23 and a free R of 0.26 for data to a resolution of 2.0  $\text{ \AA}$ .

## Results and Discussion

**Computational Design.** The subunit–subunit interface of the wild-type SspB dimer is symmetric (Fig. 1A), consisting of side chains from an  $\alpha$ -helix and  $\beta$ -strand (16, 31). For design studies, we selected four positions in each subunit (Leu-12, Ala-15, and Tyr-16 from the helix, and Val-101 from the strand), which pack together in a complementary fashion to form a shielded hydrophobic cluster (Fig. 1B). These eight residues were chosen because of their central position in the dimer interface, because the energetics of hydrophobic packing are well understood (32), and because computational methods are highly effective at optimizing packing (33).

In one set of calculations, the subunit–subunit interface of SspB was optimized for stability without explicit consideration of specificity. Using the DEE search algorithm (22, 23) implemented in the ORBIT protein-design code (1), the identities and side-chain geometries of eight interface residues were computationally optimized based on energies calculated for vdW interactions, burial of polar hydrogens and hydrophobic surface, hydrogen bonding, and electrostatic interactions (18, 34). The lowest-energy structure was asymmetric (Fig. 1C), containing *Phe-12/Ala-15/Phe-16/Ile-101* (FAFI) in one subunit and *Leu-12/Ala-15/Leu-16/Ile-101* (LALI) in the other subunit (mutant residues italicized). The FAFI/LALI heterodimer was calcu-



**Fig. 1.** SspB dimer interface. (A) Ribbon diagram of SspB from the wild-type crystal structure (16) with one subunit colored purple and the other subunit colored blue. (B–D) Molecular images of the SspB dimer interface showing a transparent surface of strand  $\beta 7$  and helix  $\alpha 1$ . Side chains that were allowed to vary in design calculations are shown in space-fill representation. Nearby side chains whose geometries did not vary during the calculations are shown in stick representation. (B) Wild-type LAYV/LAYV interface. (C) Stability-design FAFI/LALI interface. (D) Specificity-design LSLA/YGFM interface.

lated to be more stable than the parental SspB homodimer (Table 1), largely because of predicted improvements in packing geometry and hydrophobic burial. The prediction of an asymmetric sequence with greater stability than the symmetric wild-type homodimer is not unexpected because asymmetric sequence space is larger than symmetric sequence space. Interestingly, although the mutant homodimer states were not explicitly considered in this design, the energy calculated for the FAFI/LALI heterodimer was significantly lower than the average of those calculated after optimization of the side-chain geometries of the FAFI/FAFI or LALI/LALI homodimers (Table 1). Thus, these calculations suggested that specificity might be obtained simply as a byproduct of optimizing stability.

To design explicitly for specificity, we modified the ORBIT code to optimize the difference in energy between a heterodimer and both homodimers assuming a fixed protein backbone (see *Methods*). The latter assumption was necessary for efficient searching but created a problem for negative design because unfavorable interactions could not be ameliorated by conformational relaxation of the backbone. The unfavorable term in the vdW potential, which has a very steep distance dependence, was particularly prone to producing unrealistically bad energies for competing states. To approximate the process of conformational relaxation, we capped unfavorable vdW contacts at 5 kcal/mol

**Table 1. Calculated energies**

Protein	Seq. A*	Seq. B*	A/B energy <sup>†</sup>	A/B preference <sup>‡</sup>
Wild type	LAYV	LAYV	−76.3	0.0
Stability	FAFI	LALI	−84.1	−33.4
Specificity	LSLA	YGFM	−66.7	−62.9

\*Amino acids at dimer interface positions 12, 15, 16, and 101.

<sup>†</sup>Calculated energy (kcal/mol) of A/B dimer following optimization of side-chain geometry for designed proteins.

<sup>‡</sup>Calculated energy (kcal/mol) favoring heterodimer over homodimers ( $2E_{AB} - E_{AA} - E_{BB}$ ).

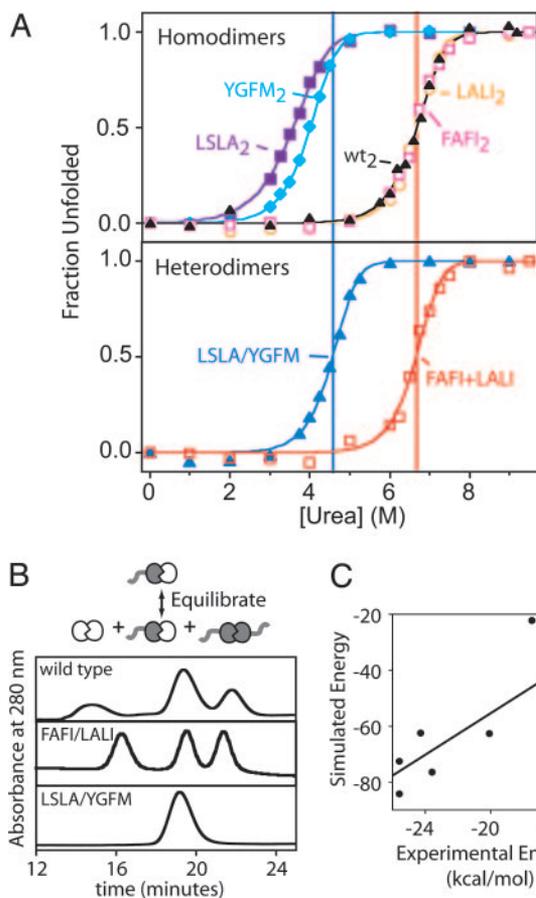
per interaction. From a practical perspective, this approach optimizes the number but not the precise magnitude of unfavorable vdW contacts in competing states.

Our specificity calculations resulted in a heterodimer with Leu-12/Ser-15/Leu-16/Ala-101 (LSLA) and Tyr-12/Gly-15/Phe-16/Met-101 (YGFM) sequences in the two subunits (Fig. 1D). The LSLA/YGFM molecule was calculated to be substantially more stable than the average of the YGFM/YGFM and LSLA/LSLA homodimers. Calculations also suggested that LSLA/YGFM would have greater heterodimer specificity than the stability design but would be less stable than the parental SspB homodimer or the stability-design FAFI/LALI heterodimer (Table 1). In the LSLA/YGFM design, the large side chains of Tyr-12, Phe-16, and Met-101 from one subunit packed efficiently with the smaller Leu-12, Leu-16, and Ala-101 side chains from the partner subunit (Fig. 1). Moreover, the small size of the Gly-15 side chain on one side of the interface permitted accommodation of the larger and opposed Ser-15 side chain. Serine is more frequently observed on the surface of proteins, where its polar hydroxyl group can interact with solvent than in the hydrophobic core, and is typically not included in the design of solvent inaccessible positions (35). However, it provides a steric profile that is not represented in the naturally occurring aliphatic amino acids and, in this design, pairs with glycine to replace the steric bulk of two symmetric alanines in the wild-type homodimer. In theory, replacing Ser-15 with Cys, which is less hydrophilic but of similar size, could result in a more stable heterodimer. However, Cys was not included in the design to avoid experimental complications from disulfide formation. Compared with the side-chain packing in LSLA/YGFM, packing efficiency was compromised in the modeled LSLA/LSLA and YGFM/YGFM homodimers.

**Dimerization Preferences and Thermodynamic Stability.** To test the predictions of the designs described above, we expressed and purified subunits containing the FAFI, LALI, YGFM, and LSLA sequences.<sup>§</sup> Each individual protein as well as mixtures of FAFI+LALI and LSLA+YGFM sedimented as dimers in analytical ultracentrifugation experiments performed with initial protein concentrations of 30  $\mu$ M (data not shown). Thus, each of the four mutant proteins can assemble as a dimer. The stabilities of the mutant homodimers and wild-type SspB were determined in urea-denaturation studies (Fig. 2A Upper), revealing that the FAFI/FAFI, LALI/LALI, and parental SspB homodimers were all of similar stability (Table 2). Moreover, each of these dimers was significantly more stable than YGFM/YGFM, which was more stable than LSLA/LSLA. Denaturation studies were also performed on mixtures of appropriate subunits to probe the stability of heterodimers (Fig. 2A Lower). Importantly, the LSLA+YGFM mixture showed higher stability than either possible homodimer. Calculations based on these results indicated that in the absence of denaturant and at equimolar concentrations of YGFM and LSLA, 99% of all molecules should exist as LSLA/YGFM heterodimers. Thus, designing for heterodimer stability and against homodimer stability achieved the desired goal. By contrast, denaturation of the FAFI+LALI mixture was indistinguishable from that of either homodimer. Therefore, designing for stability alone failed to achieve specificity but resulted in molecules that were more stable than the specificity design.

We previously described an assay for heterodimer preference that depends on having SspB variants with different charges resulting from the presence or absence of an unstructured

<sup>§</sup>In previous work (9), we purified truncated variants of the YGFM and LSLA proteins and found them to be very poorly soluble. In the studies reported here, both proteins contained C-terminal tails that improved solubility.



**Fig. 2.** Heterodimer and homodimer stabilities. (A) Urea-induced unfolding of SspB dimers was monitored by changes in tryptophan fluorescence at 30°C in 50 mM potassium phosphate (pH 6.8). (Upper) Unfolding transitions for the LSLA, YGFM, wild-type, FAFI, and LALI homodimers. (Lower) Unfolding transitions for the LSLA/YGFM and FAFI+LALI proteins. For comparison between homodimers and heterodimers, vertical blue and red lines mark the  $C_m$  for LSLA/YGFM and FAFI+LALI, respectively. Fitted  $\Delta G$  and  $m$  values are listed in Table 2. Based on these values, an equimolar mixture of LSLA and YGFM subunits results in 99% of the LSLA/YGFM heterodimer at equilibrium. Denaturation experiments were fit to a two-state model in which native dimers are in equilibrium with unfolded monomers (25). This model predicts that denaturation should be concentration-dependent and that denaturation monitored by circular dichroism should give the same transition. Both predictions were experimentally confirmed for the LSLA homodimer, the protein of lowest stability (data not shown). (B) Exchange reactions. Heterodimers containing one full-length subunit and one truncated subunit were purified by ion-exchange chromatography, incubated for 24 h, and then rechromatographed. The wild-type heterodimer and FAFI/LALI heterodimer equilibrated to form a mixture of both homodimers and the heterodimer. The LSLA/YGFM heterodimer did not form appreciable quantities of either homodimer. (C) Energies calculated from the design simulations (Table 1) are plotted against the experimental  $\Delta G$  values determined by urea denaturation (Table 2). The simulated energies were calculated for interactions of the optimized positions in the folded state and were intended to capture the relative stability of the sequence variants. However, these calculations do not include terms for main-chain to main-chain interactions or folding entropy and therefore are not intended to represent absolute stability. The line is a linear fit ( $R = 0.78$ ).

C-terminal tail (9). In this assay, two variants are mixed, denatured, and renatured, and then subjected to ion-exchange chromatography, allowing separation of both homodimer species and the heterodimer. The purified heterodimer is then left for 24 h, and a second chromatography step is used to assess the populations of heterodimer and homodimer species. In this experiment, the purified LSLA/YGFM heterodimer was the

**Table 2. Experimental stability**

Protein	Sequence*	$C_m^\dagger$	$m$ value <sup>‡</sup>	$\Delta G^{\S}$
Wild type	LAYV	6.7	2.4	23.6
Stability				
A <sub>2</sub>	FAFI	6.7	2.5	24.3
B <sub>2</sub>	LALI	6.7	2.7	25.6
A/B	FAFI + LALI <sup>¶</sup>	6.7	2.7	25.6
Specificity				
A <sub>2</sub>	LSLA	3.6	1.9	14.5
B <sub>2</sub>	YGFM	4.0	2.5	17.5
A/B	LSLA/YGFM	4.6	2.7	20.1

\*Amino acids at dimer interface positions 12, 15, 16, and 101.

<sup>†</sup>Urea concentration (M) at midpoint of denaturation.

<sup>‡</sup>Slope of  $\Delta G$  versus [urea] ( $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ ).

<sup>§</sup>Free energy ( $\text{kcal}\cdot\text{mol}^{-1}$ ) of dissociation/unfolding (30°C, no urea) calculated assuming a transition from a folded dimer to unfolded monomers.

<sup>¶</sup>This sample formed a mixture of dimer species of similar stability and was fit using a single dissociation/unfolding transition.

only detectable species after the 24-h incubation (Fig. 2B). This result was not caused by slow dissociation of the LSLA/YGFM heterodimer, because heating or addition of denaturant to the purified heterodimer gave the same result (data not shown). By contrast, after 24 h, the second chromatography step revealed that the FAFI/LALI heterodimer had re-equilibrated to form a mixture of homodimers and the heterodimer (Fig. 2B). These experiments confirmed the conclusions from the denaturation experiments. The LSLA/YGFM heterodimer was significantly more stable than the average stabilities of the YGFM/YGFM or LSLA/LSLA homodimers, whereas the FAFI/LALI heterodimer had stability comparable to the FAFI/FAFI and LALI/LALI homodimers.

**Predicted vs. Experimental Stabilities.** To assess the accuracy of our calculations and the underlying physical model, we compared  $\Delta G$  values determined for the different heterodimers and homodimers by urea denaturation with the energies predicted from our designs (Fig. 2C). There was a positive correlation ( $R = 0.78$ ) between the experimental and predicted energies, indicating that the physical model captures general stability features but does not perform sufficiently well to predict experimental stabilities in detail. This weak correlation could be caused by structural relaxation (i.e., calculating the energies of structures that are not the dominant solution conformation) and/or by use of an approximate energy function. In the case of the stability design, the inability to predict dimer stabilities with a high degree of accuracy led to the incorrect prediction that the FAFI/LALI heterodimer would be more stable than the corresponding homodimers.

**Crystal Structure of the Heterodimer.** To evaluate the structural accuracy of the designed LSLA/YGFM heterodimer, we crystallized this protein (P3<sub>1</sub>21 space group) and solved the structure by molecular replacement. The asymmetric unit of this crystal form contained a single subunit that formed a dimer across the crystallographic twofold axis, resulting in averaged densities for the mutant side chains (data not shown). This result indicated that the LSLA and YGFM mutations, which are buried in the dimer interface, do not alter the overall fold and exterior of the protein and thus are not differentiated in crystal packing.

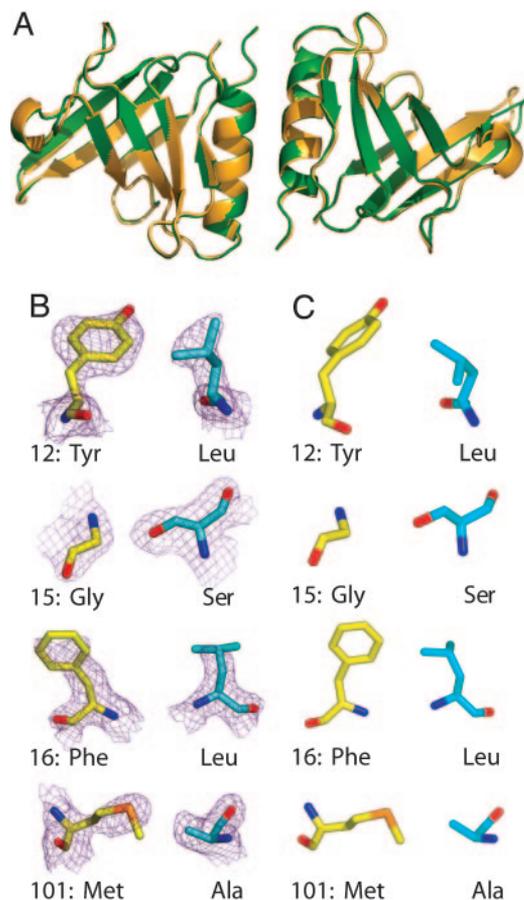
To bias against equivalent crystal packing, the YGFM subunit was engineered to contain an 18-residue C-terminal extension relative to the LSLA subunit. This heterodimer crystallized in the C2 space group, and the structure, which was determined to a resolution of 2.0 Å (Table 3), revealed an asymmetric unit containing one complete LSLA/YGFM heterodimer and one

**Table 3. Data collection and refinement statistics**

Space group	C2
$a = 120.0 \text{ \AA}$ , $b = 61.0 \text{ \AA}$ , $c = 62.4 \text{ \AA}$ , $\beta = 110.9^\circ$	
Resolution, $\text{\AA}$	30.0–2.0
Reflections, measured/unique	206,796/28,612
$R_{\text{merge}}$	0.086 (0.371)
Completeness, %	99.9 (99.8)
Number of atoms, prot/water	2552/93
$R_{\text{cryst}}/R_{\text{free}}$	0.231/0.257
RMSD from target values	
Bonds, $\text{\AA}$	0.006
Angles, $^\circ$	1.3

$R_{\text{merge}} = \sum |I_{\text{obs}} - \langle I \rangle| / \sum I$  summed over all observations and reflections.  $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$ .  $R_{\text{free}} = R_{\text{cryst}}$  calculated for 10% of reflections omitted from the refinement.

subunit that paired with a crystallographically symmetric partner as in the P3<sub>1</sub>21 structure. The side chains of the mutant LSLA and YGFM substitutions were clearly defined in the electron-density map of the heterodimer that was completely contained within the asymmetric unit. Alignment of the LSLA/YGFM structure with the wild-type SspB structure (Fig. 3A) showed that



**Fig. 3.** Crystal structure of LSLA/YGFM heterodimer. (A) Ribbon representation of an alignment of the LSLA/YGFM heterodimer structure (Protein Data Bank entry 1ZSZ; green) with the wild-type SspB homodimer structure (Protein Data Bank entry 1OU9; orange). The main-chain RMSD between these structures is 0.4  $\text{\AA}$ , indicating that the redesigned dimer interface does not perturb the overall structure. (B)  $2F_o - F_c$  simulated annealing omit maps of electron density for designed mutations at the dimer interface contoured to 1.0  $\sigma$ . (C) Side-chain geometries of optimized amino acids in the designed heterodimer model are generally similar to those observed in the crystal structure.

the designed mutations did not perturb the overall SspB fold. Indeed, the RMSD for main-chain atoms (0.4  $\text{\AA}$ ) in this alignment was the same as that for aligning two wild-type SspB structures from different crystal forms (16). Moreover, the same RMSD values were obtained by aligning main-chain atoms from the mutated portions of the LSLA/YGFM structure ( $\alpha 1$  and  $\beta 7$ ) or the unmutated portions of the structure with the corresponding regions of wild-type SspB.

Comparison of the conformations of the side chains of residues 12, 15, 16, and 101 in the experimental (Fig. 3B) and predicted (Fig. 3C) structures of the LSLA/YGFM heterodimer revealed that the majority of designed side chains were accurately predicted. Because of similar main-chain conformations, Gly-15 of the YGFM subunit and Ala-101 of the LSLA subunit overlay well between the crystal structure and the design. Similarly, Tyr-12 and Met-101 in the YGFM subunit and Ser-15 in the LSLA subunit were in the same rotamer bins (all dihedral angles within the same energy well) in both structures. Leu-12 and Leu-16 in the LSLA subunit had the same  $\chi 1$  angle in both structures but differed in  $\chi 2$ . The greatest discrepancy between the designed and experimental structures involved the conformation of Phe-16 in the YGFM subunit, which displayed a 63° difference in  $\chi 1$  angles between the two structures. We note, however, that the electron density for the Phe-16 side chain was weak (Fig. 3) and its  $B$  factors were high relative to the rest of the structure. Thus, this side chain may sample multiple conformations in the crystal.

These structural results demonstrate that the protein-design calculations capture many but not all of the features of the real protein. When we calculated the stability of the experimental structure with the energy function used for design, it was 1 kcal/mol more stable than the predicted design. This result suggests that if the crystallographically determined rotamers were present in the design calculation, then they would have been chosen by the optimization algorithm. In protein design, representing all possible side-chain geometries results in intractably large searches, and side chains are commonly represented by a discrete set of rotamer geometries based on dihedral angles observed in known protein structures (17). Increasing the detail of the rotamer library by including a greater number of discrete side chain geometries can improve the structural and energetic accuracy of prediction but comes at an enormous cost in search time (36). In the design of the LSLA/YGFM heterodimer, the energy calculated for the rotamer-based design was close enough to that calculated from the crystal structure so that imprecision in the prediction of some side-chain geometries was not a serious impediment.

In our design process, only the conformations of eight side chains in the dimer interface were allowed to vary. Consistently, nearby side chains in the core and dimer interface assumed “wild-type conformations” in the LSLA/YGFM structure. It seems likely that favorable subunit–subunit interactions mediated by these nearby symmetric side chains (Fig. 1) also help stabilize the LSLA/LSLA and YGFM/YGFM homodimers. To design a molecule with even greater specificity in terms of heterodimer preference, one could begin with the LSLA/YGFM structure and vary other side chains in the dimer interface to introduce greater asymmetry. It will be interesting to see whether stepwise design processes of this type are more efficient than designs in which the initial number of mutated side chains is simply increased.

**Discussion.** Our work builds on previous protein design concepts, particularly the work of Havranek and Harbury (7), who described a theoretical foundation for specificity optimization. To our knowledge, however, the experiments reported here provide the first head-to-head experimental comparison of stability and specificity strategies for computational protein design. Design-

ing for specificity and stability clearly resulted in different outcomes. To obtain a specific SspB heterodimer, we found that it was necessary to include an energetic penalty for competing homodimer states explicitly in the design process. Optimizing for specificity did not optimize stability, and specificity was achieved at the cost of stability.

Although all negative-design strategies optimize the difference in computed energy between the target and competing states, different groups have used a wide range of force fields, descriptions of the unfolded state, and optimization algorithms to achieve this end (7, 8, 10). For our studies, we used a force field that had been empirically optimized for protein design, implicitly considered the energy of the unfolded state to be constant, and used a combination of DEE and Monte Carlo search algorithms for global optimization. Perhaps the greatest challenge in negative design is to model accurately the energetic effects of destabilizing mutations in competing states that likely involve conformational relaxation. Our approach was to cap unfavorable vdW energies when modeling competing states as an approximation for conformational relaxation that would alleviate atomic overlaps (see *Methods*).

1. Dahiyat, B. I. & Mayo, S. L. (1997) *Science* **278**, 82–87.
2. Bolon, D. N. & Mayo, S. L. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 14274–14279.
3. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003) *Nature* **423**, 185–190.
4. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003) *Science* **302**, 1364–1368.
5. Wilson, C., Mace, J. E. & Agard, D. A. (1991) *J. Mol. Biol.* **220**, 495–506.
6. Yue, K. & Dill, K. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4163–4167.
7. Havranek, J. J. & Harbury, P. B. (2003) *Nat. Struct. Biol.* **10**, 45–52.
8. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004) *Nat. Struct. Mol. Biol.* **11**, 371–379.
9. Bolon, D. N., Wah, D. A., Hersch, G. L., Baker, T. A. & Sauer, R. T. (2004) *Mol. Cell* **13**, 443–449.
10. Ali, M. H., Taylor, C. M., Grigoryan, G., Allen, K. N., Imperiali, B. & Keating, A. E. (2005) *Structure (Cambridge, Mass.)* **13**, 225–234.
11. Shifman, J. M. & Mayo, S. L. (2002) *J. Mol. Biol.* **323**, 417–423.
12. Shimaoka, M., Shifman, J. M., Jing, H., Takagi, J., Mayo, S. L. & Springer, T. A. (2000) *Nat. Struct. Biol.* **7**, 674–678.
13. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L. & Gonzalez, C. (2002) *Nat. Struct. Biol.* **9**, 621–627.
14. Hendsch, Z. S., Nohaile, M. J., Sauer, R. T. & Tidor, B. (2001) *J. Am. Chem. Soc.* **123**, 1264–1265.
15. Nohaile, M. J., Hendsch, Z. S., Tidor, B. & Sauer, R. T. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 3109–3114.
16. Levchenko, I., Grant, R. A., Wah, D. A., Sauer, R. T. & Baker, T. A. (2003) *Mol. Cell* **12**, 365–372.
17. Dunbrack, R. L., Jr., & Karplus, M. (1993) *J. Mol. Biol.* **230**, 543–574.
18. Mayo, S. L., Olafson, B. D. & Goddard, W. A., III (1990) *J. Phys. Chem.* **94**, 8897–8909.
19. Dahiyat, B. I. & Mayo, S. L. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10172–10177.
20. Street, A. G. & Mayo, S. L. (1998) *Fold Des.* **3**, 253–258.
21. Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. (1997) *Protein Sci.* **6**, 1333–1337.
22. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. (2000) *J. Comput. Chem.* **21**, 999–1009.
23. Desmet, J., Maeyer, M. D., Hazes, B. & Lasters, I. (1992) *Nature* **356**, 539–542.
24. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, A. H. (1953) *J. Chem. Phys.* **21**, 1087–1092.
25. Wah, D. A., Levchenko, I., Baker, T. A. & Sauer, R. T. (2002) *Chem. Biol.* **9**, 1237–1245.
26. Otwinowski, Z. & Minor, W. (1997) in *Macromolecular Crystallography*, eds. Carter, C. W. & Sweet, R. M. (Academic, San Diego), Part A, pp. 307–326.
27. Navaza, J. & Saludjian, P. (1997) in *Macromolecular Crystallography*, eds. Carter, C. W. & Sweet, R. M. (Academic, San Diego), pp. 581–594.
28. Bailey, S. (1994) *Acta Crystallogr. D* **50**, 760–763.
29. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) *Acta Crystallogr. D* **54**, 905–921.
30. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991) *Acta Crystallogr. A* **47**, 110–119.
31. Song, H. K. & Eck, M. J. (2003) *Mol. Cell* **12**, 75–86.
32. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992) *Science* **255**, 178–183.
33. Street, A. G. & Mayo, S. L. (1999) *Struct. Fold. Des.* **7**, R105–R109.
34. Gordon, D. B., Marshall, S. A. & Mayo, S. L. (1999) *Curr. Opin. Struct. Biol.* **9**, 509–513.
35. Bolon, D. N., Marcus, J. S., Ross, S. A. & Mayo, S. L. (2003) *J. Mol. Biol.* **329**, 611–622.
36. Gordon, D. B., Hom, G. K., Mayo, S. L. & Pierce, N. A. (2003) *J. Comput. Chem.* **24**, 232–243.
37. Mei, G., Di Venere, A., Rosato, N. & Finazzi-Agro, A. (2005) *FEBS J.* **272**, 16–27.

From a practical point of view, our approach succeeded in designing a heterodimer with a high degree of molecular specificity and should be applicable to other design targets. For example, dimeric proteins, which represent  $\approx 8\%$  of all protein structures (37), play essential roles in a number of biological systems. In addition, the same methods should be useful in designing protein–protein pairs with novel specificity that could be used in the synthesis and study of specific protein signaling events. Although improvements in the design process are clearly needed, our results represent a step toward the engineering of biochemical systems including specific interaction networks.

We thank Steve Mayo for the ORBIT protein design code and A. Keating, J. Kenniston, A. Martin, K. McGinness, S. Moore, P. Strop, and B. Tidor for helpful discussions and comments. Studies at the NE-CAT beamlines of the Advanced Photon Source were supported by National Institutes of Health (NIH) National Center for Research Resources Award RR-15301 and by Department of Energy Office of Basic Energy Sciences Contract W-31-109-Eng-38. This work was supported by NIH Grants AI-16892 and AI-15706 and by an NIH postdoctoral fellowship (to D.N.B.). T.A.B. is an employee of the Howard Hughes Medical Institute.