Binding of small molecules to cavity forming mutants of a *de novo* designed protein

Aditi Das, Yinan Wei, Istvan Pelczer, and Michael H. Hecht*

Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1009

Received 18 October 2010; Revised 18 January 2011; Accepted 20 January 2011
DOI: 10.1002/pro.601
Published online 16 February 2011 proteinscience.org

Abstract: A central goal of protein design is to devise novel proteins for applications in biotechnology and medicine. Many applications, including those focused on sensing and catalysis will require proteins that recognize and bind to small molecules. Here, we show that stably folded α-helical proteins isolated from a binary patterned library of designed sequences can be mutated to produce binding sites capable of binding a range of small aromatic compounds. Specifically, we mutated two phenylalanine side chains to alanine in the known structure of *de novo* protein S-824 to create buried cavities in the core of this four-helix bundle. The parental protein and the Phe→Ala variants were exposed to mixtures of compounds, and selective binding was assessed by saturation transfer difference NMR. The affinities of benzene and a number of its derivatives were determined by pulse field gradient spin echo NMR, and several of the compounds were shown to bind the mutated protein with micromolar dissociation constants. These studies suggest that stably folded *de novo* proteins from binary patterned libraries are well-suited as scaffolds for the design of binding sites.

Keywords: *de novo* design; binary code for protein design; cavity forming mutants; NMR; saturation transfer difference; pulse field gradient spin echo

Introduction

Molecular recognition is a prerequisite for enzymatic catalysis, biosensing, and signal transduction. Therefore, initial steps toward the design of novel proteins with biological functions often focus on developing new binding sites. Such sites can be engineered into pre-existing natural protein scaffolds,[1–5] or into proteins designed *de novo*.[6–14] In general, two approaches are used to produce novel binding proteins: (i) In some cases, researchers use principles of rational design to engineer hydrophobic packing and buried polar interactions in the core of a protein to facilitate binding of a small molecule.[6–8,15] (ii) Alternatively, high throughput methods, such as phage display or mRNA display, can be used to find rare binding proteins amidst large combinatorial libraries of nonbinding sequences.[16–18]

The hydrophobic core of a protein is generally well packed, with few cavities. Previous studies showed that creation of new cavities in a protein core can produce binding sites where small molecules can bind. In particular, mutations in which a bulky residue, such as leucine or phenylalanine, is replaced by a small residue such as alanine can produce buried cavities capable of binding small molecules.[19] For example, mutation of Leu99 to Ala in T4 lysozyme produced a cavity volume of ~150 Å^3, which bound benzene (96 Å^3). This binding also led to increased protein stability.

The sizes and shapes of cavities engineered into the core of a protein depend on several factors such

Aditi Das's current address is Beckman Institute of Advanced Science and Technology, University of Illinois-Urbana Champaign, Urbana 61801.
Yinan Wei's current address is Department of Chemistry, University of Kentucky, Lexington, KY 40506.

Grant sponsor: NIH; Grant number: GM062869; Grant sponsor: NSF; Grant number: MCB-0817651; Grant sponsor: FMC graduate summer fellowship.

*Correspondence to: Michael H. Hecht, Department of Chemistry, Princeton University, Princeton, NJ 08544-1009. E-mail: hecht@princeton.edu

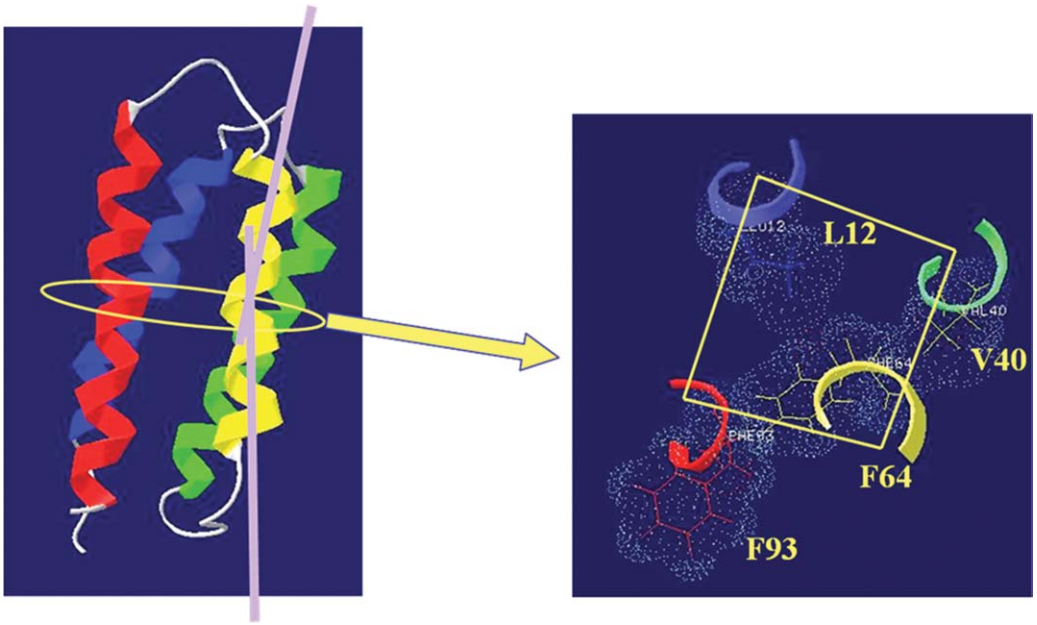

Figure 1. (Left) NMR structure of protein S-824. The yellow circle highlights the region where there is a kink in the helix and a missing hydrogen bond. (Right) Residues involved in the central layer of the protein S-824, which leads to overcrowding in this region.

as the local geometry, whether or not a cavity already exists near the site of substitution, and the extent to which the protein structure relaxes to occupy the space vacated by the mutated residue. The binding of polar molecules typically requires extensive complementarity between the ligand and the site, with specific interactions involving electrostatic interactions and hydrogen bonding.[20] In contrast, it may be relatively easy to devise binding sites for nonpolar molecules in the hydrophobic interior of a protein. Since binding at nonpolar sites is governed by hydrophobicity and steric interactions, a binding site can sometimes be created simply by truncating a buried residue that leaves behind an environment that is sterically complementary to the deleted side-chain.[21]

Four-helix bundles provide a simple and useful motif for the introduction of novel binding sites. We previously reported the design of three libraries of de novo four-helix bundles. These libraries were based on the binary code strategy for protein design, which posits that appropriate patterning of polar and nonpolar residues can suffice to specify the overall fold of a simple globular structure.[22–26] We showed previously that several proteins from our libraries bind heme and display rudimentary enzymatic activities.[27–31] Recently, we showed that de novo proteins from a binary patterned library of sequences function in *E. Coli* and enable the growth of living cells.[32]

The solution structures of two of our de novo proteins, S-824 and S-836, were determined by NMR, and both structures have well packed cores similar to natural proteins.[33] The hydrophobic core of protein S-824, however, is so well packed that it appears overcrowded. This crowding leads to several structural defects, including a kink in the third helix, and the solvent exposure of a phenylalanine side chain (Phe93). Although it is difficult to pinpoint residues responsible for these defects, the close proximity of two phenylalanines (Phe64 and Phe93) at the same layer in the bundle seems to play a role (Fig. 1). To test whether alleviation of this crowding might affect protein stability and/or facilitate small molecule binding, we mutated Phe64 and/or Phe93 to Ala. Interestingly, although these mutations did not increase protein stability, they endowed the protein with an enhanced ability to bind small aromatic molecules.

Results and Discussion

Two large hydrophobic phenylalanine residues in the central layer of the de novo four-helix bundle S-824 were mutated from Phe to Ala. The F64A and F93A mutations were made individually, and as the F64A/F93A double mutant. To assess the effects of these Phe→Ala mutations on protein stability, the proteins were characterized structurally and thermodynamically. Structural perturbation was probed by 1D proton NMR. Stably folded structures typically give rise to well dispersed spectra with relatively sharp peaks. Good dispersion indicates that different parts of the protein occupy unique chemical environments, as observed for well-ordered native structures. The spectra of the four de novo proteins are shown in Figure 2. The parental protein S-824 and the mutant F64A show better dispersion (in the amide region and the methyl region) than either F93A or F64A/F93A. We chose not to characterize the double mutant further because poor dispersion

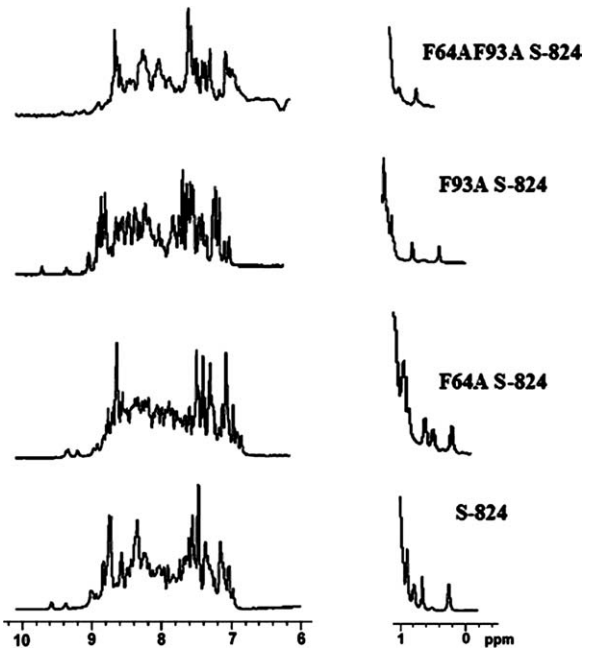

Figure 2. 1D proton spectra of amide (left) and methyl (right) regions of the parental protein S-824, and mutants F64A, F93A, and F64AF93A.

and broad peaks in its spectrum suggests that perturbation of the hydrophobic core caused a dynamic structure that would unlikely to form a binding site capable of distinguishing between similar molecules. Moreover, for practical reasons, poor dispersion and broad peaks render this protein difficult to characterize with the NMR methods described below.

We used circular dichroism (CD) spectroscopy to compare the α-helicity and thermodynamic stability of the mutants to the parental protein, S-824. CD spectra showed the α-helicity of mutants F64A and F93A are about 15% lower than S-824 (data not shown). Thermal denaturation curves of S-824 and both mutants showed the cooperative profiles typical of native proteins. Although the melting temperatures of the mutants are slightly lower than that of S-824, all three proteins are quite stable with denaturation midpoints above 65°C [Fig. 3(A)]. Chemically induced denaturation with guanidine hydrochloride (GdmHCl) also showed that all three proteins are stable and cooperative [Fig. 3(B)]. In agreement with the thermal melts, the chemically induced denaturations showed that both Phe→Ala mutations slightly reduced stability. F93A destabilized the protein by 0.5 kcal/mol, while F64A destabilized the protein by 2.0 kcal/mol. These results are consistent with the fact that residue 93 is exposed, and so its mutation would perturb the structure to a lesser extent than alterations at residue 64, which is buried.

The structures of mutants F64A, F93A, and F64A/F93A were calculated based on the known structure of the parental protein, S-824. We used MODELLER and SWISS PDB Viewer and MODELLER to calculate and view the structures respectively.[34–38] We then used CASTp[39–42] to estimate the size of cavities and pockets in protein S-824 and its mutants. (Predictions by Q-site pocket finder[43] agree well with those by CASTp). Protein S-824 is well packed and has no cavities or pockets large enough to bind aromatic compounds. Likewise, the calculated structure of mutant F93A has no cavities/pockets larger than 45 Å^3. Residue 93 is exposed to the solvent in the structure of S-824; therefore, the Phe93→Ala mutation did not produce a significant cavity/pocket. In contrast, the F64A mutant and the double mutant (F64A/F93A) show one large cavity/pocket (Fig. 4). This cavity is near the mutated residue 64 and has a calculated volume of 133 Å^3 in both F64A and F64A/F93A.

Based on the modeling described above, we expected mutant F64A would bind small aromatic molecules more effectively than the parent protein, S-824. To test whether the proteins actually bind

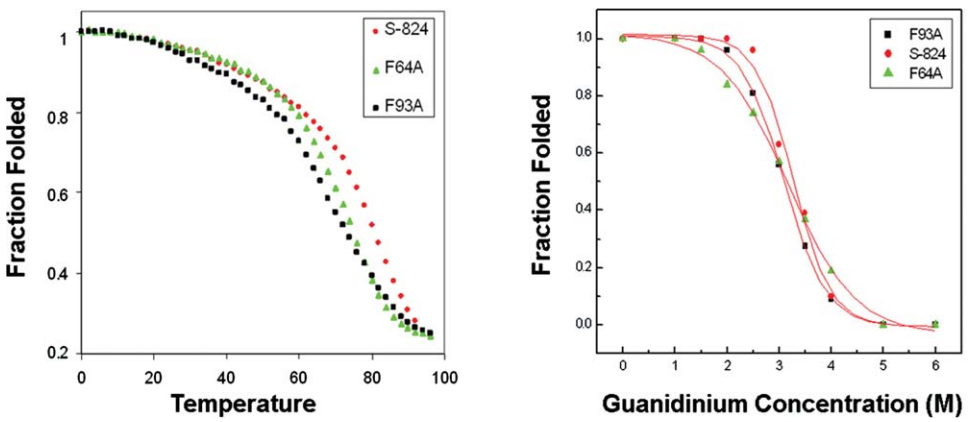

Figure 3. (A) Thermal denaturation of mutant F64A (green triangles), F93A (black squares), and protein S-824 (red diamonds). Data from different proteins were brought to scale by normalizing each individual protein by its own ellipticity at 0°C. (B) GdmHCl denaturation curves of F93A (black squares), F64A (green triangles), and S-824 (red diamonds). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

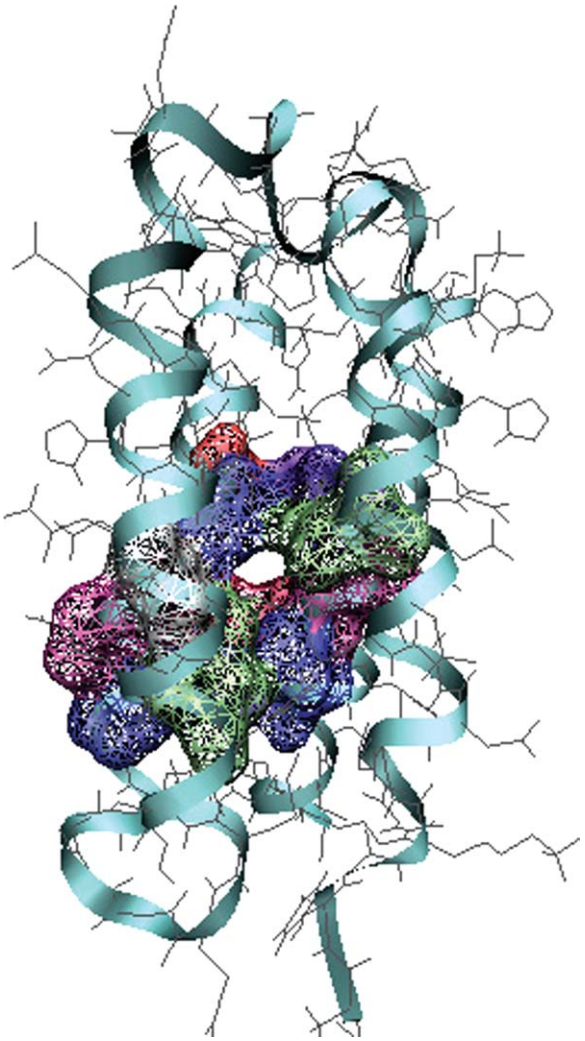

Figure 4. Predicted structure of the cavity in the F64A mutant. The known structure of parental protein S-824 was used as starting point and SWISS PDB viewer and MODELLER were used to predict the effects of the F64A mutation. CASTp was used to determine the cavities and pockets in the modeled structure of the F64A mutant. Residues surrounding the cavity are shown in space filling representation. These residues are L12, V40, I44, K56, L57, M60, M61, A64, Q65, I89, K90, I92, A93, and L96. The predicted volume of this cavity is 133 Å^3.

these compounds, we used saturation transfer difference (STD) NMR. STD is a double resonance method used to probe low affinity interactions ($K_D = 10^{-8}$–$10^{-3}M$) between ligands and proteins.[44–46] In this technique, the hydrogens on a protein are saturated by irradiating a spectral region that contains broad resonances of the macromolecule, but is free of any signals from the small molecule. Spin diffusion causes the saturation to propagate rapidly throughout the protein and to any bound ligands. This causes an attenuation of the signal of the unbound small molecule in proportion to its binding characteristics. The STD NMR technique involves comparing the ^{1}H NMR spectrum of a solution of protein and ligand probed with either on-resonance (I_{on}) or off-resonance (I_{off}) irradiation. A difference spectrum is generated by subtracting the off-resonance from the on-resonance spectrum ($\Delta I = I_{on} - I_{off}$). The magnetization of a bound ligand is partially saturated, which is transferred to the free ligand by exchange. The difference spectrum shows the signals of the compounds that bind the protein.

Because we are interested in eventually using this technique as a high throughput screen to probe libraries of small molecules for binding to our de novo proteins, we assayed binding using mixed samples containing several compounds in "one pot" experiments. The compounds were divided into five groups such that within each group, each compound produces at least one unique chemical shift that can be used as a signature for its detection amidst the mixture.

Two representative STD-NMR spectra are shown in Figure 5. Such spectra were used to identify which small molecules in each of our mixtures bound to proteins S-824 and F64A. The results are summarized in Table I: Toluene, benzene, *p*-cresol, 2,4,6-trimethoxytoluene, and *m*-xylene bind to F64A but not to S-824. Cyclohexane and *p*-xylene bind to both proteins. To confirm binding, we subtracted the difference spectra at 1 s saturation time from difference spectra at 3 s. The resulting spectra show ligands that bind to the protein.

As summarized in Table I, F64A binds several aromatic molecules. Most of these compounds are hydrophobic, including benzene, toluene, *p*-xylene, and *m*-xylene. However, F64A also binds 2,4,6-trimethoxy toluene, which is somewhat polar, and *p*-cresol, with its polar hydroxyl group capable of forming H-bonds. Presumably, these compounds bind in a pocket that is partially exposed to solvent, thereby allowing the polar substituents to interact with water.

In the STD experiments, the solubility of the small molecule in aqueous solution plays an important role in what is visible in the spectrum. Moreover, the strength of the binding (Kd) is not always reflected in the strength of the STD signal. Therefore, we used PFGSE (pulse field gradient spin echo) NMR[47,48] to measure the affinities of several of the small molecules to proteins S-824 and F64A.

Assuming the protein (P) and ligand (L) are in rapid equilibrium, the dissociation constant (K_D) of the protein/ligand complex (PL) is given by the following equation.

$$K_D = [P][L]/[PL] \quad (1)$$

In this equation $[P] = [P]_T - [PL]$ and $[L] = [L]_T - [PL]$ where $[P]_T$ and $[L]_T$ are the initial total concentrations of the protein and ligand, respectively. In analyzing this system, we assumed (i) binding is a

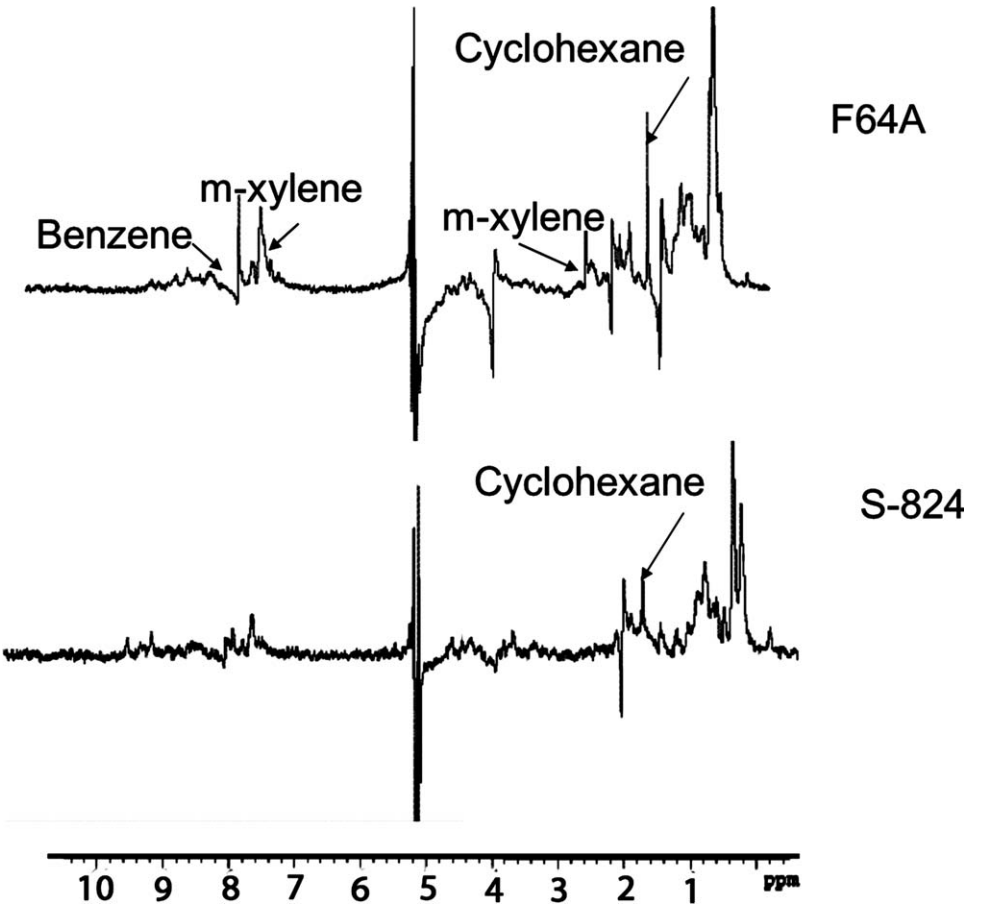

Figure 5. Representative saturation transfer difference (STD) NMR difference spectra (ΔI) of compounds interacting with proteins S-824 and F64A. In this group of organic compounds, F64A binds benzene, *m*-xylene and cyclohexane, while protein S-824 binds cyclohexane.

first order, reversible, and rapid; (ii) all binding sites are independent; and (iii) all binding interactions have the same equilibrium constant K_D.To solve for K_D, we need only to determine the concentration of [*PL*]. This can be determined from diffusion experiments.

The diffusion coefficient of a molecule depends on its size and shape.[49] Small molecules move rapidly, while large molecules move slowly. A small molecule bound to a large molecule diffuses slowly with its host. When experiments are performed at a high ratio of protein to ligand, one can readily determine the amount of PL from its diffusion behavior. The observed diffusion coefficient (D_{OBS}) for a ligand bound to protein is the mole fraction weighted average of the free and bound states of the ligand as shown below. (Note that mole fractions add up to unity: $\chi_{PL} + \chi_L = 1$)

$$D_{OBS} = \chi_L D_L + \chi_{PL} D_{PL} \quad (2)$$

$$\chi_{PL} = (D_{OBS} - D_L)/(D_{PL} - D_L) \quad (3)$$

$$\chi_{PL} = (D_{OBS} - D_L)/(D_P - D_L) \quad (4)$$

$$K_D = \frac{(1 - \chi_{PL})(P_T - \chi_{PL} L_T)}{\chi_{PL}} \quad (5)$$

Here, D_{OBS} is a weighted average of the diffusion coefficient for the free ligand, D_L, and that of the protein-ligand complex, D_{PL}. χ_L is the mole fraction of the ligand and χ_{PL} is the mole fraction of the protein/ligand complex. Equation 5, which is used to determine the binding constant, is modified from Eqs (1) and (2). An advantage of this approach is that D_{PL} is not an unknown quantity. Because the association of a small molecule with a protein does not change the diffusion coefficient of the protein appreciably, it is reasonable to assume that $D_{PL} = D_P$. Hence, titrations are not required, and K_D can be derived from a single experiment.[50]

The diffusion coefficients can be extracted from the PFGSE NMR experiment.[47,50,51] To calculate the diffusion coefficient, a series of 1D proton NMR spectra of the protein-ligand system is recorded while systematically varying the gradient field strength (g) (Fig. 6).

Intensities are recorded for several characteristic peaks, which are chosen because they do not shift in frequency or broaden too much. For the same change in field strength, the intensity of the small molecule peaks diminishes quickly compared to the protein peaks. This is because small molecules

Table I. *Compounds that were screened for binding to F64A and S-824 using the STD method at on-resonance tuned to 0.35 ppm*

	Results of screening for binding to protein	
Ligand	F64A	S-824
Group I		
Toluene	Yes	No
Phenol	No	No
Dioxane	No	No
Pyridine	No	No
1,3,5-Trichlorobenzene	No	No
Group II		
p-Cresol	Yes	No
1,3,5-Trimethylbenzene	No	No
Benzene	Yes	No
2,4,6-Trimethoxytoluene	Yes	No
Group III		
p-Xylene	Yes	Yes
o-Xylene diamine	No	No
1,3,5-Triethyl benzene	No	No
Imidazole	No	No
Group IV		
1,3,5-Triethylbenzene	No	No
p-Xylene	No	No
Iodo benzene	No	No
p-Cresol	Yes	No
Group V		
Cyclohexane	Yes	Yes
m-Xylene	Yes	No
Benzene	Yes	No

diffuse far more rapidly than proteins (or protein ligand complexes).

The variation of peak intensity with the changing field strength g is related by the Stejskal-Tanner equation:

$$\ln I = -\gamma^2 g^2 \delta^2 (\Delta - \delta/3) D \qquad (6)$$

Table II. *K_D of benzene binding was determined for S-824 and its mutants*

Benzene + Protein	K_D (M)
S-824	7×10^{-3}
F93A	2×10^{-3}
F64A	2×10^{-4}
F64AF93A	3×10^{-3}

Here, I is the intensity of the peak, γ is the gyromagnetic ratio (26750 rad g^{-1} s^{-1} for ^1H), δ is the duration of the gradient pulse (1.5 ms), and Δ is the interval between gradient pulses (100 ms and 300 ms for ^1H in two different sets of experiments). The field strength g (gauss/cm) is systematically varied over the experiment. D is the translational diffusion coefficient and is obtained from the slope of the linear plot of $\ln I$ vs. $-\gamma^2 g^2 \delta^2 (\Delta - \delta/3)$.

Values of K_D determined by PFGSE NMR are summarized in Tables II and III. These data show that although binding of benzene is relatively weak for all four proteins, it binds 35-fold more strongly to mutant F64A than to the parent protein, S-824. This is consistent with the prediction of a large cavity in F64A, but not in S-824. (Weak binding to S-824 can be attributed to nonspecific interactions with aromatic residues in S-824, as shown by H,N-HSQC–data not shown). The weak binding to F64A/F93A could be attributed to the presence of large pockets, which might lead to incomplete sequestration of the benzene.

As described above, and summarized in Table I, the STD-NMR experiments showed that F64A also bound several derivatives of benzene. In particular, F64A bound methyl substituted benzenes. We used PFGSE NMR to determine the binding affinity of several of these compounds to both F64A and S-824. The results, summarized in Tables II and III, show

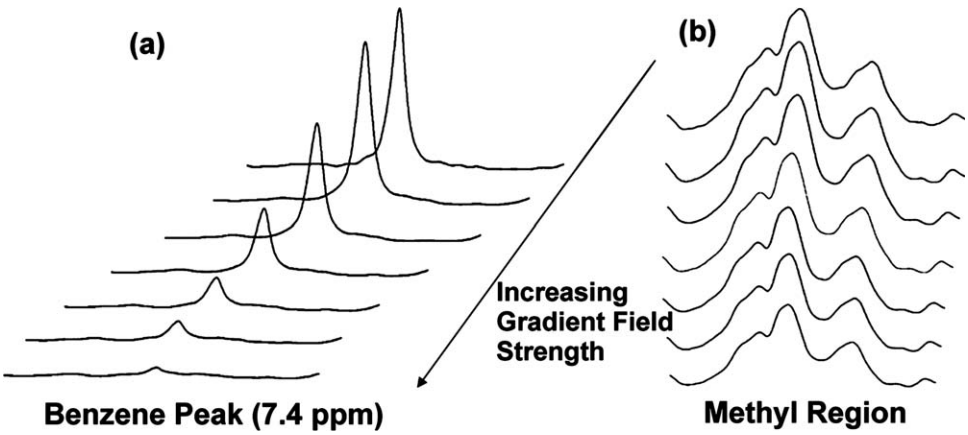

Figure 6. (A) The benzene peak at 7.4 ppm is monitored with changing field strength for a benzene-protein system. (B) The protein peaks monitored at methyl region with changing field strength for a benzene-protein system. The peaks become smaller as the gradient field strength increases from top to bottom. The value of the gradient field strengths from top to bottom are 5.53, 14.39, 23.23, 32.08, 40.93, 49.78, and 58.62 Gauss/cm respectively. For the same change in gradient field strength, in (A) the benzene peaks diminishes quickly while in (B) the protein peaks diminish slowly.

Table III. *The binding constant for F64A and S-824 interacting with several ligands were determined using PFGSE diffusion methods*

Ligand	Binding dissociation constant (M) S-824	Binding dissociation constant (M) F64A
Benzene	7×10^{-3}	2×10^{-4}
Toluene	1×10^{-2}	2×10^{-5}
Cyclohexane	8×10^{-3}	1.5×10^{-4}
p-Cresol	9×10^{-3}	8×10^{-3}
p-Xylene	5×10^{-3}	8.6×10^{-6}
m-Xylene	7×10^{-3}	6.1×10^{-7}

that all aromatic molecules tested bind S-824 with millimolar dissociation constants. Binding to the F64A mutant is considerably stronger, with several compounds having K_D values near or even below 1 μM. Thus, the replacement of a large aromatic side chain (Phe) in the hydrophobic core of this four-helix bundle with a smaller side chain (Ala) has produced a cavity that is well-suited for binding a range of hydrophobic aromatic compounds.

Conclusions

The field of protein design has two main goals: Design for structure and design for function. Our laboratory has approached the first task (structure) in an unusual way: Rather than designing proteins atom-by-atom, we developed a general approach using a binary code strategy to design and construct large collections of novel sequences that successfully fold into stable three-dimensional structures.[26,33] These collections of folded proteins can now be used as a feedstock for the development of new functions, both *in vitro*[27–29,31] and *in vivo*.[32]

Most types of protein function require binding to another molecule. In many cases, the binding partner is a small organic compound that becomes sequestered in a cavity or pocket in the interior of the protein. Our collections of de novo four-helix bundles were not explicitly designed to have such binding sites; however, we reasoned that many of our binary patterned proteins might be sufficiently stable to allow such cavities to be engineered into their structures. In particular, protein S-824, shown previously to have a well-ordered three-dimensional structure and high thermal stability, seemed like an ideal candidate to test this hypothesis. Inspection of the structure of S-824 suggested two phenylalanine side chains that could be truncated to alanines without disrupting the structure (and which might even increase stability). We constructed three mutant proteins. Although the double mutant, F64A/F93A, seems to form a dynamic molten globule, each of the single mutations was tolerated reasonably well.

Computational analyses predicted that F64A would have a cavity large enough to bind benzene or other small aromatic compounds. To test the accuracy of these predictions, we used STD-NMR to search for binding partners among mixtures of aromatic compounds. Binding constants calculated from diffusion coefficients determined by PFGSE-NMR showed that protein F64A binds *p*-xylene and *m*-xylene with micromolar affinity. These studies suggest that stably folded de novo proteins from binary patterned libraries are well-suited as scaffolds for the design of ligand binding sites. The approaches described here can be extended toward the design of sites for transition state analogues, and may ultimately yield de novo proteins with novel catalytic activities.[52]

Materials and Methods

Protein expression and purification

Protein S-824 and its mutants F64A, F93A, and F64A/F93A were expressed and purified using methods similar to those described previously.[23,53,54] DNA encoding protein S-824 was mutated using PCR. The sequence of each mutant was checked by DNA sequencing and the expected mass of each protein was confirmed by mass spectroscopy. Each modified sequence was cloned into plasmid pET11a (Novagen). Proteins were expressed in *E. coli* strain BL21 (DE3) grown in 2xYT medium. Protein was extracted from the cells using the freeze-thaw method[53] and then solubilized in 100 mM $MgCl_2$. Cellular contaminants were removed by acid precipitation in 50 mM sodium acetate buffer (pH 4.2). The resulting supernatant was loaded onto a POROS HS cation exchange column (PerSeptive Biosystems) and eluted using a gradient of NaCl from 0 to 1.5 M. Purified protein was concentrated and exchanged into buffer using Centricon Plus-20 filters (Millipore). The monomeric state of the protein was confirmed using size exclusion chromatography (Superdex 75 HR/10/30–Pharmacia) in 50 mM sodium phosphate buffer pH 7.4, 100 mM NaCl. The concentration of protein was measured by UV spectroscopy at 280 nm using a Hewlett Packard 8452A diode array spectrophotometer. The extinction coefficient used is 6990 M^{-1} cm^{-1}.

CD and denaturation experiments

CD measurements were performed at 20°C in 10 mM phosphate buffer (pH 6.8), 40 mM NaF using an Aviv model 62 DS spectropolarimeter. Mean residual ellipticity is calculated by: MRE = (θobs $\times$ MWR/10 cl) where MWR is the mean molecular weight per residue, c is the concentration of the protein in mg/mL and l is the path length in cm.

Thermal denaturation was monitored by recording the ellipticity at 222.0 nm every 2°C from 0°C to 98°C with 1 min equilibration time for each point.

GdmHCl-induced denaturation

Protein samples containing various concentrations of GdmHCl were incubated for at least 2 h before measurements were taken. Data were recorded in 50

m*M* sodium phosphate buffer (pH 7.0) at 20°C. The free energy of folding, ΔG, was derived as described previously.[55]

Preparation of labeled protein sample for NMR

Uniformly 15N-labeled protein sample was prepared by growing cells in M9 minimal media supplemented with 1 g/L [^{15}N] NH_4Cl and 10 g/L glucose as the sole nitrogen and carbon sources.[56] Protein was purified as described before. The extent of isotope labeling was checked for the ^{15}N-labeled sample by comparing the observed mass (by mass spectroscopy) with that expected from the sequence. The protein was determined to be >98% labeled. Sample concentration was ~2.0 m*M* as determined by absorbance at 280 nm. The purified protein sample was exchanged into a NMR buffer (50 m*M* sodium acetate buffer, pH 4.2 at 25°C, 10% D_2O).

Preparation of NMR samples for STD-NMR experiments

The NMR samples were made with 10% D_2O in 50 m*M* sodium phosphate buffer at pH 7.4. The samples were transferred in low volume NMR (Shigemi) tubes to prevent the evaporation of volatile molecules from the sample during study. Protein concentration was ~0.2 m*M* while total ligand concentration of the mixture was ~100 m*M*.

PFGSE NMR

Multiple samples were prepared in order to determine the diffusion coefficient. First, the ligand was added to the protein, vortexed, and spun down to remove excess ligand. Then the ligand is considered at saturated concentration depending on the solubility of the ligand. Second, the protein-ligand solution is diluted to various concentrations using another protein solution of the same concentration.

The solubilities of the ligands were obtained from literature or calculated from 1D NMR spectra of the ligand and dioxane in water using the following equation.

$$M_{\text{Ligand}} = \frac{M_{\text{Dioxane}} \times \text{INT}_{\text{Ligand}} \times 4}{\text{INT}_{\text{Dioxane}} \times (\text{No. of Protons of Ligand})}$$

Here, M_{Ligand} is the molarity of the ligand, M_{Dioxane} is the molarity of dioxane, and INT are the integrals of the peak of ligand and dioxane. All experiments for PFG studies for ligand binding were conducted at 25°C using a Varian Unity/INOVA 600 MHz spectrometer.

Pulse sequence of PFG-NMR

In the basic sequence of pulse field gradient spin-echo technique, a 90° RF (radio frequency) pulse transfers magnetization to the *xy* plane where the magnetization dephases. A 180° refocusing pulse produces a spin echo after an appropriate interval. Only spins that have undergone no net displacement during the interval δ (duration of the gradient pulse) refocuses and hence the intensity of the echo amplitude is related to diffusion by equation 6. ^{1}H pulsed field gradient (PFG) experiments were run for each sample beginning at 1.11 G/cm and continuing in increments of 4.42 G/cm for 15 steps (up to 63.06 G/cm) of 1.5 min duration each. The characteristic ligand signal was monitored for the ligand alone (D_L) and protein plus ligand (D_{OBS}), and the ^{1}H methyl signals at ~0.9 ppm were monitored for the protein alone (D_P) and protein plus ligand (D_{PL}). The internal standard for measuring the diffusion coefficients was acetate, which did not bind to the protein and showed a constant diffusion value.

Modeling

The mutant structure was calculated using MODELLER and the Swiss PDB viewer mutation tool. The resulting structure was minimized using GROMOS 43B1 force field (built in tool in Swiss PDB viewer). The resulting coordinates were input into the CASTp program to estimate the sizes of pockets and cavities.

Acknowledgments

We thank Dr. Debanu Das (SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA), Dr. Allison Doerr (Nature Methods), Dr. Luke Bradley (University of Kentucky), and Dr. Michael Ackermann (Bristol-Myers Squibb) for helpful discussions.

References

1. Lu Y (2005) Design and engineering of metalloproteins containing unnatural amino acids or non-native metal-containing cofactors. Curr Opin Chem Biol 9: 118–126.
2. Lu Y, Valentine JS (1997) Engineering metal-binding sites in proteins. Curr Opin Struct Biol 7:495–500.
3. Lu Y, Berry SM, Pfister TD (2001) Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. Chem Rev 101:3047–3080.
4. Cordova JM, Noack PL, Hilcove SA, Lear JD, Ghirlanda G (2007) Design of a functional membrane protein by engineering a heme-binding site in glycophorin A. J Am Chem Soc 129:512–518.
5. Park H-S, Nam S-H, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim H-S (2006) Design and evolution of new catalytic activity with an existing protein scaffold. Science 311:535–538.
6. Doerr AJ, Case MA, Pelczer I, McLendon GL (2004) Design of a functional protein for molecular recognition: specificity of ligand binding in a metal-assembled protein cavity probed by 19F NMR. J Am Chem Soc 126:4192–4198.
7. Doerr AJ, McLendon GL (2004) Design, folding, and activities of metal-assembled coiled coil proteins. Inorg Chem 43:7916–7925.

8. Obataya I, Sakamoto S, Ueno A, Mihara H (2001) Design and synthesis of 3alpha-helix peptides forming a cavity for a fluorescent ligand. Biopolymers 59:65–71.
9. Case MA, McLendon GL (2004) Metal-assembled modular proteins: toward functional protein design. Acc Chem Res 37:754–762.
10. Koder RL, Anderson JL, Solomon LA, Reddy KS, Moser CC, Dutton PL (2009) Design and engineering of an O(2) transport protein. Nature 458:305–309.
11. Cui T, Bondarenko V, Ma D, Canlas C, Brandon NR, Johansson JS, Xu Y, Tang P (2008) Four-alpha-helix bundle with designed anesthetic binding pockets. Part II: halothane effects on structure and dynamics. Biophys J 94:4464–4472.
12. Ma D, Brandon NR, Cui T, Bondarenko V, Canlas C, Johansson JS, Tang P, Xu Y (2008) Four-alpha-helix bundle with designed anesthetic binding pockets. Part I: structural and dynamical analyses. Biophys J 94: 4454–4463.
13. Pidikiti R, Shamim M, Mallela KM, Reddy KS, Johansson JS (2005) Expression and characterization of a four-alpha-helix bundle protein that binds the volatile general anesthetic halothane. Biomacromolecules 6:1516–1523.
14. Solt K, Johansson JS, Raines DE (2006) Kinetics of anesthetic-induced conformational transitions in a four-alpha-helix bundle protein. Biochemistry 45:1435–1441.
15. Gonzalez L, Jr., Woolfson DN, Alber T (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat Struct Biol 3:1011–1018.
16. Kim M, Shin DS, Kim J, Lee YS (2010) Substrate screening of protein kinases: detection methods and combinatorial peptide libraries. Biopolymers 94:753–762.
17. Lipovsek D, Pluckthun A (2004) In-vitro protein evolution by ribosome display and mRNA display. J Immunol Methods 290:51–67.
18. Keefe AD, Szostak JW (2001) Functional proteins from a random-sequence library. Nature 410:715–718.
19. Eriksson AE, Baase WA, Wozniak JA, Matthews BW (1992) A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature 355:371–373.
20. Liu L, Baase WA, Michael MM, Matthews BW (2009) Use of stabilizing mutations to engineer a charged group within a ligand-binding hydrophobic cavity in T4 lysozyme. Biochemistry 48:8842–8851.
21. Baldwin E, Baase WA, Zhang X, Feher V, Matthews BW (1998) Generation of ligand binding sites in T4 lysozyme by deficiency-creating substitutions. J Mol Biol 277:467–485.
22. Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) De novo proteins from designed combinatorial libraries. Protein Sci 13:1711–1723.
23. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. Science 262:1680–1685.
24. Bradley LH, Kleiner RE, Wang AF, Hecht MH, Wood DW (2005) An intein-based genetic selection allows the construction of a high-quality library of binary patterned de novo protein sequences. Protein Eng Des Sel 18:201–207.
25. Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003) Solution structure of a de novo protein from a designed combinatorial library. Proc Natl Acad Sci U S A 100: 13270–13273.
26. Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH (2003) Stably folded de novo proteins from a designed combinatorial library. Protein Sci 12:92–102.
27. Wei Y, Hecht MH (2004) Enzyme-like proteins from an unselected library of designed amino acid sequences. Protein Eng Des Sel 17:67–75.
28. Das A, Hecht MH (2007) Peroxidase activity of de novo heme proteins immobilized on electrodes. J Inorg Biochem 101:1820–1826.
29. Moffet DA, Case MA, House JC, Vogel K, Williams RD, Spiro TG, McLendon GL, Hecht MH (2001) Carbon monoxide binding by de novo heme proteins derived from designed combinatorial libraries. J Am Chem Soc 123:2109–2115.
30. Moffet DA, Foley J, Hecht MH (2003) Midpoint reduction potentials and heme binding stoichiometries of de novo proteins from designed combinatorial libraries. Biophys Chem 105:231–239.
31. Das A, Trammell SA, Hecht MH (2006) Electrochemical and ligand binding studies of a de novo heme protein. Biophys Chem 123:102–112.
32. Fisher MA, McKinley KL, Bradley LH, Viola SR, Hecht MH (2011) De novo designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. PLoS ONE 6:e15364.
33. Go A, Kim S, Baum J, Hecht MH (2008) Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. Protein Sci 17:821–832.
34. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325.
35. Sanchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, Madhusudhan MS, Mirkovic N, Sali A (2000) Protein structure modeling for structural genomics. Nat Struct Biol 7:986–990.
36. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9:1753–1773.
37. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815.
38. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2006) Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics. John Wiley & Sons, Supplement 15, 5.6.1–5.6.30.
39. Binkowski TA, Naghibzadeh S, Liang J (2003) CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res 31:3352–3355.
40. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7:1884–1897.
41. Liang J, Edelsbrunner H, Fu P, Sudhakar PV, Subramaniam S (1998) Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. Proteins 33:18–29.
42. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34:W116–W118.
43. Laurie AT, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 21:1908–1916.
44. Meyer B, Klein J, Mayer M, Meinecke R, Moller H, Neffe A, Schuster O, Wulfken J, Ding Y, Knaie O (2004) Saturation transfer difference NMR spectroscopy for identifying ligand epitopes and binding specificities. Ernst Schering Res Found Workshop 44: 149–167.
45. Mayer M, Meyer B (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc 123:6108–6117.

46. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chem Int Ed 38:1784–1788.
47. Altieri AS, Hinton DP, Byrd RA (1995) Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. J Am Chem Soc 117: 7566–7567.
48. Dvinskikh SV, Furo II (2000) Cross-relaxation effects in stimulated-echo-type PGSE NMR experiments by bipolar and monopolar gradient pulses. J Magn Reson 146:283–289.
49. Derrick TS, McCord EF, Larive CK (2002) Analysis of protein/ligand interactions with NMR diffusion measurements: the importance of eliminating the protein background. J Magn Reson 155:217–225.
50. Fielding L (2003) NMR methods for the determination of protein-ligand dissociation constants. Curr Top Med Chem 3:39–53.
51. Stejskal EO, Tanner JE (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J Chem Phys 42:288–292.
52. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. Proc Natl Acad Sci U S A 98:14274–14279.
53. Johnson BH, Hecht MH (1994) Recombinant proteins can be isolated from E. coli cells by repeated cycles of freezing and thawing. Biotechnology 12:1357–1360.
54. Roy S, Hecht MH (2000) Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. Biochemistry 39:4603–4607.
55. Fairman R, Chao HG, Mueller L, Lavoie TB, Shen L, Novotny J, Matsueda GR (1995) Characterization of a new four-chain coiled-coil: influence of chain length on stability. Protein Sci 4:1457–1469.
56. Wei Y, Fela D, Kim S, Hecht M, Baum J (2003) 1H, 13C and 15N resonance assignments of S-824, a de novo four-helix bundle from a designed combinatorial library. J Biomol NMR 27:395–396.