J|A|C|S
ARTICLES
Published on Web 03/10/2004

Design of a Functional Protein for Molecular Recognition: Specificity of Ligand Binding in a Metal-Assembled Protein Cavity Probed by ^{19}F NMR

Allison J. Doerr, Martin A. Case,† István Pelczer, and George L. McLendon*

Contribution from the Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received April 25, 2003; E-mail: glm@princeton.edu

Abstract: A metal-assembled homotrimeric coiled coil based on the GCN4-p1 sequence has been designed that noncovalently binds hexafluorobenzene and other similar ligands in a hydrophobic cavity, created by making the core substitution Asn16Ala ($[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$). The K_D of binding of hexafluorobenzene with $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ was observed to be $1.1(9) \times 10^{-4}$ M by diffusion NMR experiments. A control coiled coil with the core substitution Asn16Val ($[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$) exhibited a significantly weaker association with hexafluorobenzene, providing evidence that even in the absence of structural data, benzene-like ligands bind in the cavity created by the Asn16Ala substitution. ^{19}F NMR was employed to observe hexafluorobenzene binding and to monitor titrations with competing hydrophobic and polar ligands similar in size and shape to hexafluorobenzene. All hydrophobic ligands bound with greater affinity than the polar ligands in the hydrophobic core, although the cavity seems to be somewhat flexible in terms of the sizes of molecules it can accommodate. Thus ^{19}F NMR has proved to be a useful spectral tool to probe molecular recognition in a hydrophobic cavity of a metal-assembled coiled coil.

Introduction

The design of functional proteins is of considerable interest.[1] The construction of protein receptors for molecular recognition of ligands is the first step toward tailor-made sensors and enzymes.[2] The engineering of existing protein scaffolds to induce or improve enzymatic function has been intensely studied;[3] however, the design of novel functional proteins from first principles is a more challenging task. Nevertheless, the packing interactions between side chains within the core of de novo proteins have been extensively explored.[4] Thus the stage has been set for the investigation of the noncovalent binding of organic ligands within designed protein cores.[5] Here we report evidence of specificity of molecular recognition within the hydrophobic core of a metal-assembled homotrimeric coiled coil protein.

The coiled coil (or helix bundle) is well known from fibrous proteins, yet also plays an important role in globular proteins,[6] and thus provides a tractable architecture for the design of a small molecule receptor. Protein−subunit and protein−protein interactions are often controlled by the oligomerization of two or more α-helices, implicating molecular recognition in the formation of a coiled coil. Coiled coil helical sequences tend to adopt a canonical heptad repeat $(abcdefg)_n$.[7] Nonpolar side chains in the a and d positions make up the well-packed hydrophobic core. However, polar substitutions in the hydrophobic core are often key for specifying the oligomerization state of the coiled coil.[8]

The wild-type sequence of the 33-residue GCN4-p1 leucine zipper peptide forms parallel helix dimers.[9] The hydrophobic interior is packed in a V_aL_d sequence, with valine residues in

† Current address: Department of Chemistry, University of Vermont, Burlington, VT 05405.

(1) (a) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.* **1999**, *68*, 779−819. (b) Baltzer, L. *Curr. Opin. Struct. Biol.* **1998**, *8*, 466−470. (c) Kohn, W. D.; Hodges, R. S. *Tibtech* **1998**, *16*, 379−389. (d) Hill, R. B.; Raleigh, D. P.; Lombardi, A.; DeGrado, W. F. *Acc. Chem. Res.* **2000**, *33*, 745−754. (e) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935−941.

(2) (a) Baltzer, L.; Nilsson, J. *Curr. Opin. Biotech.* **2001**, *12*, 355−360. (b) Kennan, A. J.; Haridas, V.; Severin, K.; Lee, D. H.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2001**, *123*, 1797−1803. (c) Lombardi, A.; Bryson, J. W.; Ghirlanda, G.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 12378−12379. (d) Allert, M.; Baltzer, L. *ChemBioChem* **2003**, *4*, 306−318. (e) Broo, K.; Brive; L. Lundh, A.-C.; Ahlberg, P.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 8172−8173. (f) Di Costanzo, L.; Wade, H.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F.; Lombardi, A. *J. Am. Chem. Soc.* **2001**, *123*, 12749−12757.

(3) (a) Penning, T. M.; Jez, J. M. *Chem. Rev.* **2001**, *101*, 3027−3046. (b) Harris, J. L.; Craik, C. S. *Curr. Opin. Chem. Biol.* **1998**, *2*, 127−132.

(4) (a) Case, M. A.; McLendon, G. L. *J. Am. Chem. Soc.* **2000**, *122*, 8089−8090. (b) Schnarr, N. A.; Kennan, A. J. *J. Am. Chem. Soc.* **2001**, *123*, 11081−11082. (c) Schnarr, N. A.; Kennan, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9779−9783. (d) Kashiwada, A.; Hiroaki, H.; Kohda, D.; Nango, M.; Tanaka, T. *J. Am. Chem. Soc.* **2000**, *122*, 212−215. (e) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401−1407. (f) Tripet, B.; Wagschal, K.; Lavigne, P.; Mant, C. T.; Hodges R. S. *J. Mol. Biol.* **2000**, *300*, 377−402. (g) Suzuki, K.; Hiroaki, H.; Kohda, D.; Tanaka, T. *Protein Eng.* **1998**, *11*, 1051−1055.

(5) (a) Obataya, I.; Sakamoto, S.; Ueno, A.; Mihara, H. *Biopolymers* **2001**, *59*, 65−71. (b) Johansson, J. S.; Scharf, D.; Davies, L. A.; Reddy, K. S.; Eckenhoff, R. G. *Biophys. J.* **2000**, *78*, 982−993.

(6) (a) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162−1164. (b) Burkhard, P.; Stetefeld, J.; Strelkov, S. V. *Trends Cell Biol.* **2001**, *11*, 82−88.

(7) McLachlan, A. D.; Stewart, M. *J. Mol. Biol.* **1975**, *98*, 293−304.

(8) (a) Gonzalez, L.; Woolfson, D. N.; Alber, T. *Nat. Struct. Biol.* **1996**, *3*, 1011−1018. (b) Akey, D. L.; Malashkevich, V. N.; Kim, P. S. *Biochemistry* **2001**, *40*, 6352−6360. (c) Campbell, K. M.; Sholders, A. J.; Lumb, K. J. *Biochemistry* **2002**, *41*, 4866−4871. (d) Beck, K.; Gambee, J. E.; Kamawal, A.; Bachinger, H. P. *EMBO J.* **1997**, *16*, 3767−3777.

10.1021/ja035798b CCC: $27.50

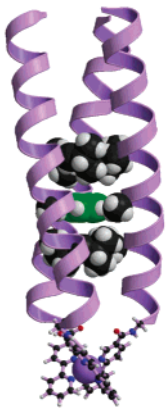

Figure 1. Model[12] of the benzene-binding metal-assembled $[Fe(bpy)GCN4\text{-}N16A)_3]^{2+}$; benzene shown in green, flanking leucines and alanines shown in black, N-terminal bipyridine moieties at base of the figure coordinating an iron atom. Shown as Λ-*fac*, the predicted major isomer.

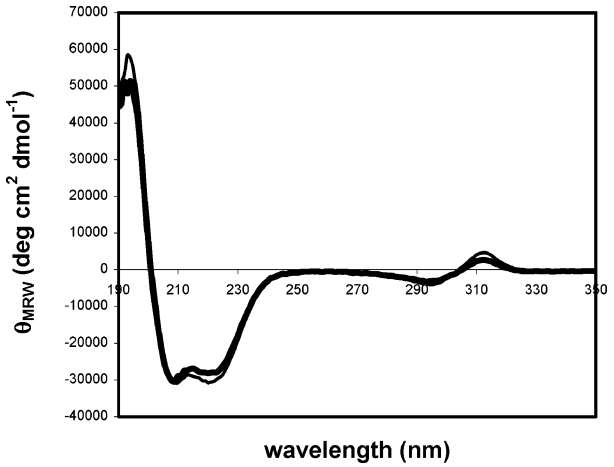

Figure 2. Circular dichroism spectra of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ (bold line) and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ (thin line) at 25 °C.

the a positions and leucine residues in the d positions, with the exception of a pair of asparagines at position 16 (an a position) responsible for dimer formation. Alber and co-workers[10] substituted these asparagines with alanine residues to create a small hydrophobic cavity. With the addition of benzene or cyclohexane, a trimeric complex formed with the ligand occupying the void created by the Asn16Ala substitution. This remarkable result provides attractive precedent for creating ligand-binding cavities in a metal-assembled parallel homotrimeric coiled coil.

Metal-assisted assembly of parallel 3-helix bundles via covalent attachment of 5-carboxy-2,2′-bipyridine (bpy) to the N-termini of helical peptides has been shown to provide a rigid template for protein construction.[11] Moreover, with the addition of iron(II) as the coordinating metal, the oligomeric state is unambiguously trimeric, circumventing the complication of accounting for multiple equilibria (and hence multiple binding site geometries) when measuring binding constants. Accordingly, a more accurate binding assessment may be made.

Two 25-residue peptides were synthesized, beginning with residue 6 of GCN4-p1 and ending with residue 30. The key residues changed are underlined.

Wild-type GCN4-p1: Ac-RMKQLEDKVEELLSK**N**YHLENEVARLKKLVGER

bpyGCN4-N16A: bpy-GDKVEELLSK**A**YHLENEVARLKKLV-NH_2

bpyGCN4-N16V: bpy-GDKVEELLSK**V**YHLENEVARLKKLV-NH_2

Homotrimers were assembled upon addition of iron(II) to form $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$; the resulting solutions are an intense magenta color.[11] A representation of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ with benzene bound in the central cavity is shown in Figure 1.[12] The predicted major isomer, Λ-*fac*, is shown. There are four potential isomers of the octahedral complex of $[Fe(bpy)_3]^{2+}$: Λ-*fac*, Δ-*fac*, Λ-*mer*, and Δ-*mer*. However, it has been shown in similar systems that the facial isomer is present in excess of 95% over the meridional isomer due to a propensity for in-register helical packing.[13] In addition, it has been shown that at room temperature the more stable Λ-*fac* isomer is present in a 40% excess over the Δ-*fac* isomer.[13a]

How can the binding of benzene and other small organic ligands within this structure be observed? How can binding be quantified? How selective is binding? The molecular approach chosen to address these questions requires the use of ^{19}F NMR and hexafluorobenzene as a reference molecule. ^{19}F NMR is an excellent spectral probe to monitor binding interactions, presuming either the host or guest molecule contains fluorine atoms perturbed by a binding event.[14] Like ^{1}H, ^{19}F is an abundant and highly NMR-sensitive nucleus. Unlike ^{1}H NMR, the problem of spectral overlap is eliminated in ^{19}F NMR. ^{19}F NMR is very sensitive to local environment; significantly larger chemical shift changes may result upon a binding interaction than with ^{1}H NMR. Moreover, competition titrations with a variety of nonfluorinated ligands may be performed simply by monitoring the chemical shift of the hexafluorobenzene signal without deconvolution of complicated spectra. Even ligands with very limited solubility in water may be tested.

$[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ serves as a control: the valine residues at position 16 are expected to pack the core without creating a cavity. Even in the absence of structural data, the evidence of diminished ligand binding with the control protein should provide proof that a positive binding interaction is occurring in the cavity of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$.

Results and Discussion

Circular Dichroism Spectroscopy. Secondary structure and stability of the designed proteins were examined by circular dichroism (CD) spectroscopy. The CD spectra of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ are shown in Figure 2. Both proteins exhibit spectra characteristic of α-helices, with

(9) (a) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539−544. (b) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989** *243*, 538−542.
(10) Gonzalez, L., Jr.; Plecs, J. J.; Alber, T. *Nat. Struct. Biol.* **1996**, *3*, 510−515.
(11) (a) Gochin, M.; Khorosheva, V.; Case, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 11018−11028. (b) Ghadiri, M. R.; Soares, C.; Choi, C. *J. Am. Chem. Soc.* **1992**, *114*, 825−831. (c) Lieberman, M.; Sasaki, T. *J. Am. Chem. Soc.* **1991**, *113*, 1470−1471.
(12) (a) Crystal coordinates from PDB code 1SWI, www.rcsb.org/pdb. (b) Image produced with MolScript: Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946−950. (c) Image produced with Raster3d: Bacon, D. J.; Anderson, W. F. *J. Mol. Graphics* **1988**, *6*, 219−220.
(13) (a) Case, M. A.; Ghadiri, M. R.; Mutz, M. W.; McLendon, G. L. *Chirality* **1998**, *10*, 35−40. (b) Ghadiri, M. R.; Case, M. A. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1594−1597.
(14) (a) Dalvit, C.; Flocco, M.; Veronesi, M.; Stockman, B. J. *Comb. Chem. High Throughput Screening* **2002**, *5*, 605−611. (b) Dalvit, C.; Fagerness, P. E.; Hadden, D. T. A.; Sarver, R. W.; Stockman, B. J. *J. Am. Chem. Soc.* **2003**, *125*, 7696−7703.

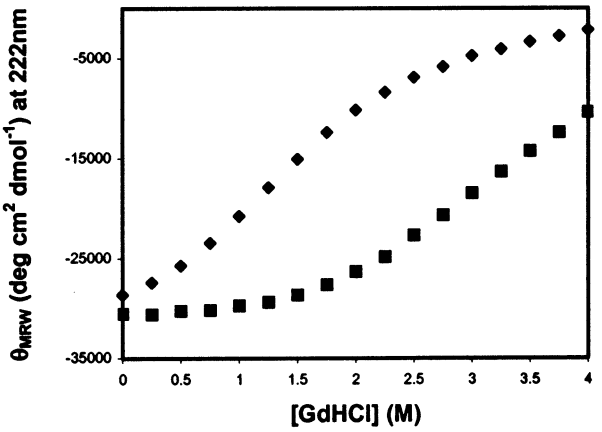

Figure 3. Guanidine hydrochloride denaturation monitored by CD at 222 nm of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ (◆) and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ (■) from 0 to 4 M GdHCl. The transition midpoint of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ occurs at approximately 1.5 M GdHCl; the transition midpoint of $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ occurs at approximately 3.5 M GdHCl.

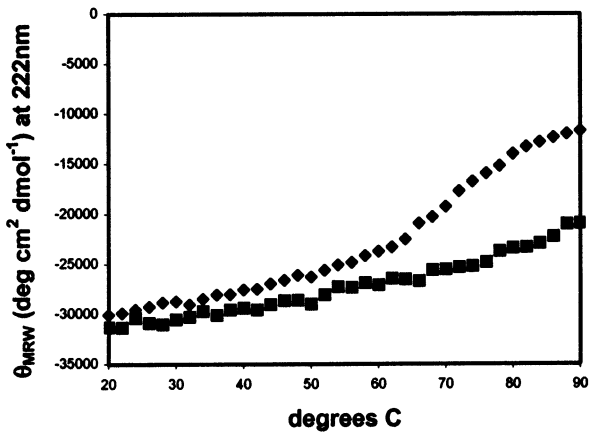

Figure 4. Thermal melts monitored by CD at 222 nm of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ (◆) and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ (■) from 20 to 90 °C.

ellipticity minima at 222 and 208 nm and a maximum at 192 nm. In addition, the minimum at 295 nm and maximum at 310 nm indicate predominance of the Λ-*fac* isomer.[15]

Percent helicity can be derived by comparison of the mean residue weight ellipticity at 222 nm with −33 000 deg cm^2 dmol^{-1} as determined for a fully folded helical protein.[16] $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ was found to be 93% helical, and $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ was about 85% helical prior to ligand addition. Addition of cyclohexane (a hydrophobic yet only slightly soluble ligand) to $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ induced a slight increase in helicity to 88%, and addition of tetrahydropyran (a soluble yet polar ligand) increased the helicity to 92%.

Chemical melts using the denaturant guanidine hydrochloride were performed to assess protein stability. $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ was more stable to GdHCl denaturation than $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$, as illustrated in Figure 3.

Thermal stability was measured by CD for $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ up to 90 °C. $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ was more stable to thermal denaturation than $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$, as illustrated in Figure 4. Fe^{II} leaves the complex at about 60 °C and is rapidly oxidized to Fe^{III}, so that the complex does not reassemble upon cooling, as was revealed by a lack of pink color post-melt.

Pulsed Field Gradient Spin−Echo NMR: Determination of K_D of the Hexafluorobenzene−$[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ System. Pulsed field gradient spin−echo (PFGSE) NMR experiments using the water-suppressed longitudinal encode-decode (sLED) pulse sequence[17] with bipolar gradient pulses[18] were performed to determine the dissociation constant K_D of the hexafluorobenzene−$[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ system. This method is based on size discrimination of the small ligand versus the large protein. Small molecules diffuse much more rapidly in solution than large molecules such as proteins. Moreover, when a small molecule associates with a larger host molecule, the rate of diffusion is greatly slowed for that small molecule, whereas the rate of diffusion for the larger host does not appreciably change. When a protein−ligand system is in fast exchange (K_D is on the order of 10^{-3} M), the ligand diffusion coefficient is a weighted average of the free and bound states.

The diffusion coefficient, D, can be extracted from a PFGSE experiment using the Stejskal−Tanner equation:[19]

$$\ln I = -\gamma^2 g^2 \delta^2 (\Delta - \delta/3) D \qquad (1)$$

where I is the signal intensity, γ is the nuclear gyromagnetic ratio (25 180 rad g^{-1} s^{-1} for ^{19}F, 26 750 rad γ^{-1} s^{-1} for ^{1}H), δ is the duration of the gradient pulse (1.5 ms), and Δ is the interval between gradient pulses (100 ms for ^{19}F and 300 ms for ^{1}H). The field gradient strength, g (gauss cm^{-1}), is systematically varied over the experiment. D is therefore the slope of a linear plot of ln I vs $-\gamma^2 g^2 \delta^2 (\Delta - \delta/3)$. The following equation is used to solve for the mole fractions χ_L and χ_{PL}:[20]

$$D_{obs} = D_L \chi_L + D_{PL} \chi_{PL} \qquad (2)$$

where D_{obs} is the diffusion coefficient for the ligand (hexafluorobenzene) in the presence of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ (^{19}F), D_L is for the ligand alone (^{19}F), and D_{PL} is for the fully bound protein plus ligand (^{1}H). However, the assumption $D_{PL} = D_P$ can be made if the diffusion coefficient of the larger molecule is unaffected by the binding of the smaller molecule. This was indeed found to be the case for this system. So, D_{PL} is found by obtaining D_P of the protein alone (^{1}H). K_D can then be solved using eq 3 and its modification, eq 4:

$$K_D = [P][L]/[PL] \qquad (3)$$

$$K_D = \chi_L (P_T - L_T \chi_{PL})/\chi_{PL} \qquad (4)$$

where P_T and L_T are the total (nonequilibrium) protein and ligand concentrations, respectively.

Although theoretically one can solve for K_D with just one experiment, multiple samples were prepared to determine each diffusion coefficient (including several different [protein]:[ligand] ratios). The mean diffusion coefficients obtained were

(15) (a) Lieberman, M.; Tabet, M.; Sasaki, T. *J. Am. Chem. Soc.* **1994**, *116*, 5035−5044. (b) It was not possible to separate the isomers using reversed-phase HPLC because the Fe^{II} complex dissociates under HPLC conditions. Furthermore, the complexes would quickly racemize once purified due to the lability of Fe^{II}.

(16) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34*, 11645−11651.

(17) (a) Alteri, A. S.; Hinton, D. P.; Byrd, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 7566−7567. (b) Gibbs, S. J.; Johnson, C. S., Jr. *J. Magn. Reson.* **1991**, *93*, 395−402.

(18) Dvinskikh, S. V.; Furó, I. *J. Magn. Reson.* **2000**, *146*, 283−289.

(19) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, *42*, 288−292.

(20) (a) Cameron, K. S.; Fielding, L. *J. Org. Chem.* **2001**, *66*, 6891−6895. (b) Fielding, L. *Tetrahedron* **2000**, *56*, 6151−6170.

$D_L = 6(1) \times 10^{-6}$ cm^2 s^{-1} for hexafluorobenzene in buffer and $D_{PL} = 8(1) \times 10^{-7}$ cm^2 s^{-1} for [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ in buffer (neither of which are concentration-dependent). D_{obs} of hexafluorobenzene ranged from 3×10^{-6} cm^2 s^{-1} (high [protein]:[ligand] ratio) to 5×10^{-6} cm^2 s^{-1} (low [protein]:[ligand] ratio). The mean K_D for the hexafluorobenzene−[Fe(bpyGCN4-N16A)$_3$]$^{2+}$ system found by PFGSE NMR was $1.1(9) \times 10^{-4}$ M.

1D ^{19}F NMR Titration: Determination of K_D of the Hexafluorobenzene−[Fe(bpyGCN4-N16A)$_3$]$^{2+}$ System. Hexafluorobenzene in the presence of [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ results in a 1D ^{19}F NMR spectrum quite distinct from that of hexafluorobenzene in an aqueous environment. The single peak is shifted downfield with the chemical shift dependent on [protein]:[ligand] ratio, and the line width is significantly broadened. This suggests that the system is in the "fast exchange" regime (K_D is on the order of 10^{-3} M). Thus the observed chemical shift represents an average of the free and bound states of the ligand. This information from a simple 1D ^{19}F NMR spectrum may be used to calculate the binding constant K_D. For comparison with the diffusion methods described above, traditional direct titration experiments were performed. Equations 5 and its modification, eq 6, can be employed in order to calculate K_D (eq 3):[20]

$$\delta_{obs} = \delta_L \chi_L + \delta_{PL} \chi_{PL} \quad (5)$$

$$[PL] = (\delta_L L_T - \delta_{obs} L_T)/(\delta_L - \delta_{PL}) \quad (6)$$

where δ_{obs} is the observed NMR chemical shift for each titration point, δ_L is the chemical shift of the free ligand in buffer (−163.5 ppm), and δ_{PL} is the chemical shift of the fully bound ligand (estimated at −162.2 ppm). Notice the form of eq 5 is exactly like that of eq 2. However, unlike the diffusion experiments, where D_{PL} is a measurable quantity and only χ_L and χ_{PL} are unknowns, δ_{PL} is a titration unknown and must be estimated by measuring the chemical shift of the guest molecule in the presence of excess protein host at the lowest possible ligand concentration still detectable by NMR. Titration of [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ with hexafluorobenzene yielded a mean K_D of $4(3) \times 10^{-4}$ M. Although this value is close to the diffusion-measured K_D of $1.1(9) \times 10^{-4}$ M, the error associated with the direct titration method is larger than the error associated with the diffusion method. Furthermore the diffusion calculations do not require the estimation of key parameters. Therefore, the K_D measured by diffusion experiments is accepted as more accurate and is used in all further calculations.

Competitive Binding: Specificity of the Designed Cavity for Small Molecules. ^{19}F NMR is a useful tool to monitor competition for a protein binding site, given that the reference ligand is fluorinated and the competing ligand is not. The resulting δ_{obs} of this system is used to calculate the apparent binding constant of the fluorinated ligand in the presence of the competing ligand I (the "inhibitor"). This in turn can be used to calculate the inhibition constant K_I of the competing ligand. In this manner, even molecules of limited solubility (which cannot be tested using traditional ^{1}H NMR titration methods) may be evaluated in order to determine binding constants. The problem of ^{1}H spectral overlap is also eliminated.

Visually, the effect of a strongly competing ligand on the ^{19}F NMR spectra of hexafluorobenzene is quite dramatic. When hexafluorobenzene is displaced from the [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ cavity, the signal shifts upfield and the line width narrows, appearing more like the free hexafluorobenzene signal, reflecting transfer from the bound protein state to the free solution state. For example, Figure 5 shows the competition titration of toluene into the hexafluorobenzene−[Fe(bpyGCN4-N16A)$_3$]$^{2+}$ system.

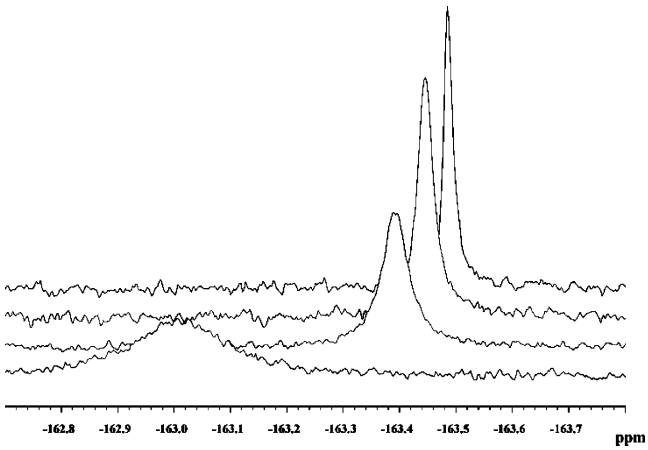

Figure 5. ^{19}F NMR spectra of competition titration of hexafluorobenzene in the presence of [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ with toluene, shown from bottom (no toluene addition) to top with increasing toluene concentration (top spectrum, highest toluene concentration).

The concentration of the complex [PL] (hexafluorobenzene bound to [Fe(bpyGCN4-N16A)$_3$]$^{2+}$) at each titration point was calculated using eq 6. Competing ligands were treated as inhibitors (I). The following equations were employed to solve for the inhibition constant K_I at each titration point:[21]

$$K_D^{app} = (P_T L_T - P_T[PL] + [PL]^2 - L_T[PL])/[PL] \quad (7)$$

$$K_I = [I]K_D/(K_D^{app} - K_D) \quad (8)$$

where K_D^{app} is the apparent K_D for the hexafluorobenzene−[Fe(bpyGCN4-N16A)$_3$]$^{2+}$ system in the presence of the competing ligand I. Mean K_I values were calculated for the competing ligands based on the K_D of $1.1(9) \times 10^{-4}$ M for the binding of hexafluorobenzene with [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ (from diffusion experiments).

Table 1 reports the mean K_I values of [Fe(bpyGCN4-N16A)$_3$]$^{2+}$ for each of the ligands. Toluene and cyclohexane occupy the designed cavity with the highest affinity. The somewhat smaller benzene and hexafluorobenzene and the somewhat larger 1,3,5-trimethylbenzene and *m*-xylene all bind more weakly, although the values are not significantly different enough to draw a major conclusion about the size of the cavity. However, it is interesting to note that cyclohexane has higher binding affinity than benzene, and there is some precedent for this result. Schnarr and Kennan[4b,c] created a similar GCN4 system with an unnatural cyclohexylalanine side chain (Chx) built into the peptide sequence. Heterotrimerization occurred by steric matching of a 2:1 complex of Ala16:Chx16.

Not surprisingly, hydrophobic molecules bind more strongly in the hydrophobic cavity than any of the polar molecules evaluated. Polar molecules bind in the [Fe(bpyGCN4-N16A)$_3$]$^{2+}$

(21) Dalvit, C.; Flocco, M.; Knapp, S.; Mostardini, M.; Perego, R.; Stockman, B. J.; Veronesi, M.; Varasi, M. *J. Am. Chem. Soc.* **2002**, *124*, 7702−7709.

Table 1. Inhibition Constants (K_i) for Binding of Small Molecules to $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$, Determined by ^{19}F NMR Competition Titrations with Hexafluorobenzene, Ranked from Strongest to Weakest Binding Affinity

small molecule ligand	K_i, M
toluene	$3(1) \times 10^{-6}$
cyclohexane	$8(5) \times 10^{-6}$
benzene	$4(3) \times 10^{-5}$
1,3,5-trimethylbenzene	$5(2) \times 10^{-5}$
m-xylene	$1.1(5) \times 10^{-4}$
hexafluorobenzene	$1.1(9) \times 10^{-4}$ (K_D)
1,3,5-trimethoxybenzene	$1.1(6) \times 10^{-4}$
phenol	$1.25(7) \times 10^{-4}$
3,5-dimethylphenol	$3.9(6) \times 10^{-4}$
tetrahydropyran	$1.41(4) \times 10^{-3}$
pyridine	$4(2) \times 10^{-3}$
1,4-dioxane	$3(2) \times 10^{-2}$

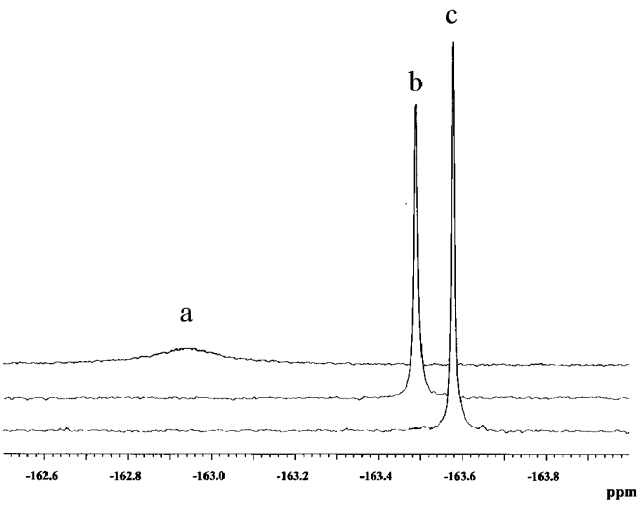

Figure 6. ^{19}F NMR spectra of saturated hexafluorobenzene solutions in the presence of (a) $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$, (b) $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$, and (c) deuterioacetate buffer.

cavity, but with much lower affinity, since they are preferentially solvated in buffer rather than the hydrophobic protein core. However, it is difficult to draw a conclusion about a size trend; for example, 1,3,5-trimethoxybenzene, phenol, and 3,5-dimethylphenol bind almost equally well. All of the smaller molecules (tetrahydropyran, pyridine, and 1,4-dioxane) bind the poorest; however, these are all also the most polar. It is likely that the $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ cavity is quite flexible, especially when accommodating unfavorable polar molecules.[22]

Control Experiments: Diminished Ligand Binding to $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$. In the absence of structural data, a control is necessary to provide evidence that the observed "binding" is not an artifact. $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ is in every way alike to $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$, except in lacking an alanine cavity. Therefore, any difference in the NMR spectrum of the guest molecule in the presence of the proteins can be attributed to a binding event taking place in the cavity. Both 1D NMR and PFGSE NMR experiments provide evidence for diminished binding of hexafluorobenzene to $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$.

1D ^{19}F NMR evidence is illustrated in Figure 6. The hexafluorobenzene signal shifts slightly downfield in the presence of $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ to a maximum of -163.4 ppm, no matter the [protein]:[ligand] ratio. The line width remains quite sharp, similar to the hexafluorobenzene signal when in buffer. Both pieces of information indicate that the overwhelming proportion of hexafluorobenzene remains in the buffer environment. However, the signal shifts significantly downfield in the presence of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ to a maximum of -162.2 ppm, and the line width broadens considerably, signifying a change of environment for the hexafluorobenzene.

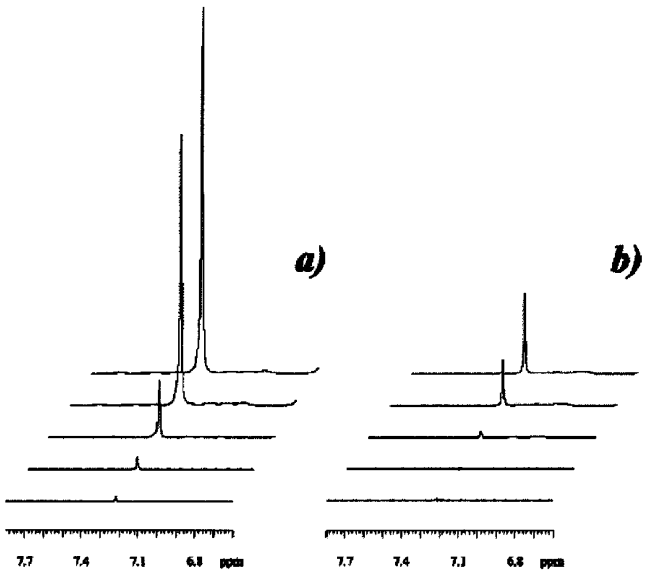

Figure 7. Transferred saturation enhancement of the benzene signal with 0.05, 0.1, 0.4, 2, and 10 s selective irradiation time (from bottom to top), respectively, on the methyl groups of the proteins (a) $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ and (b) $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$. The spectra were normalized to the saturated methyl signals at 1 ppm and are plotted on a comparable scale.

Diffusion experiments also point toward negligible binding of hexafluorobenzene to $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$. D_{obs} of hexafluorobenzene did not appear to be dependent on $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ concentration, and as a result the mean D_{obs} for the system was $5.9(8) \times 10^{-6}$ cm^2 s^{-1}. This is essentially equivalent to the diffusion coefficient of a solution of hexafluorobenzene in buffer (no protein present), $D_L = 6(1) \times 10^{-6}$ cm^2 s^{-1}. This indicates there is no significant interaction between hexafluorobenzene and $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$, at least not an interaction quantifiable by PFGSE NMR.

In addition, saturation transfer difference (STD) NMR was used to study the association of benzene with each designed protein. A saturation pulse is applied on the methyl groups of the proteins at 1 ppm. Magnetization is transferred to the ligand (benzene) only if there is a significant association with the protein. The intensity of the benzene signal of the difference spectrum shows the extent of saturation transfer, hence the strength of binding.[23] The more intense the signal, the stronger the interaction; the more dispersed the signal, the weaker the interaction. The STD NMR data indicate that the binding interaction of benzene with $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ is significantly stronger than that of benzene with $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$. Figure 7 illustrates the differential binding affinities of the two proteins for benzene. Accordingly, this provides another control in the absence of structural data.

(22) Feher, V. A.; Baldwin, E. P.; Dahlquist, F. W. *Nat. Struct. Biol.* **1996**, *3*, 516−521.

(23) (a) Mayer, M.; Meyer, B. *Angew. Chem., Int. Ed.* **1999**, *38*, 1784−1788. (b) Meyer, B.; Peters, T. *Angew. Chem., Int. Ed.* **2003**, *42*, 864−890.

These three independent methods' results confirm that small benzene-like ligands interact negligibly with $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$. However, these structures are likely to be quite flexible, and so occasionally, $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ may accommodate a small molecule, resulting in evidence of a very weak interaction. Further structural evaluation of site(s) for weak binding is currently in progress.

Conclusion

In summary, a hydrophobic cavity has been constructed within a metal-assembled trimeric coiled coil protein that binds hexafluorobenzene and other small molecules. This binding site distinguishes between otherwise similar molecules on the basis of polarity and, to some extent, size. A method for accurately assessing binding constants as quantified by ^{19}F NMR competition titrations using a fluorinated reference molecule has been introduced. The binding constants for a variety of small molecules have been quantified. We anticipate that additional structural refinement and sculpting of the binding site will provide further sensitivity in molecular recognition. Such studies are in progress.

Experimental Section

Materials. All chemicals were purchased from Aldrich Chemical Co., Milwaukee, WI, and Fisher Scientific, Pittsburgh, PA, unless otherwise stated. Protected amino acids and solid phase peptide synthesis (SPPS) resins were purchased from NovaBiochem, San Diego, CA, and Advanced ChemTech, Louisville, KY. Deuterium oxide for NMR was purchased from Cambridge Isotope Laboratories, Inc., Andover, MA. Commercially obtained chemicals were used without further purification, with the exceptions of DMF (dried over 4 Å molecular sieves) and DIEA.[24]

Peptide Synthesis. Peptides were synthesized on a model 396MPS multiple peptide synthesizer (Advanced ChemTech) by standard Fmoc SPPS protocol on MBHA resin. 5-Carboxy-2,2′-bipyridine was synthesized[25] and coupled according to Fmoc protocol. Side chain protecting groups were removed by TFA deprotection. Peptides were cleaved from the resin with anhydrous HF containing 10% v/v anisole as a scavenger at 0 °C for 45 min. Following an ether rinse, free peptide was washed from the resin with 1 mL neat TFA followed by 100 mL of 5% acetic acid. After lyophilization, crude peptides were purified by reverse-phase HPLC on a Vydac C18 preparative column with gradient elution by solvents A (99% H_2O, 1% CH_3CN, 0.1% TFA) and B (90% CH_3CN, 10% H_2O, 0.1% TFA) and lyophilized to dryness. The peptide masses were confirmed by electrospray mass spectroscopy on a Hewlett-Packard 5989B mass spectrometer. Calculated mass for bpyGCN4-N16A 3064 g/mol, observed 3064 g/mol. Calculated mass for bpyGCN4-N16V 3092 g/mol, observed 3092 g/mol.

Protein Homotrimer Assembly. Peptides were dissolved in 0.1% TFA solution, and excess solid $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ was added and dissolved. After complex formation (solution turns magenta), excess Fe was removed using a PD-10 desalting column (Amersham Pharmacia Biotech, Piscataway, NJ) and the resulting solution of homotrimers was lyophilized. TFA (0.1%) was exchanged for 20 mM deuterioacetate buffer, 150 mM NaCl, pH 6.0, using a PD-10 column (with the exception of CD experiments, where the buffer was 5 mM NaOAc, pH 6.0). Samples were lyophilized and reconstituted in D_2O for NMR purposes. Protein concentrations were determined by UV spectroscopy on a model Lambda 6 Perkin-Elmer spectrophotometer ($\epsilon_{545\ nm} = 7000$); protein concentrations ranged from 0.1 to 1.0 mM for all NMR experiments. Protein concentration was independently determined by ^1H NMR spectroscopy of the iron-assembled protein with 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt as an integration reference; from this the UV extinction coefficient at 545 nm was extrapolated.

CD Spectroscopy. Solutions (10−50 mM, as quantified by UV spectroscopy) of protein in 5mM NaOAc buffer, pH 6, were prepared for all CD experiments. CD experiments were performed on a model 62DS Aviv spectrophotometer. Wavelength scanning spectra and thermal melts were recorded in 1.0 mm quartz cells. Four repeats for each of the full spectra were obtained and averaged. All spectra were obtained with 1.0 nm bandwidth, averaging time of 2.0 s, and a step size of 1.0 nm from 350 to 190 nm (or as low as possible). With the exception of thermal melts, all spectra were recorded at 25 °C. Thermal melts were recorded at 222 nm from 20 to 90 °C with a thermal equilibration time of 2min, per 2 °C step. GdHCl titrations were recorded in 1.00 cm quartz cells equipped with cap and small stir bar. Saturated (8 M) GdHCl solution was injected with an external KDS200 syringe pump (KD Scientific Inc.; Boston, MA) in 0.25 M additions, stirring constantly, with an equilibration time of 20 min per injection. GdHCl melts were recorded at 222 and 260 nm for reference after each 0.25 M injection.

^{19}F NMR Spectroscopy. All ^{19}F NMR spectra were recorded on a Varian Unity/INOVA 400 MHz spectrometer equipped with a 4-nucleus PFG AutoSwitchable probe unless otherwise stated. Hexafluorobenzene was added in excess (approximately 5 μL) to 0.5−1.0 mL solutions, mixed well, and centrifuged 5 min at 10 000 rpm. The heavier excess hexafluorobenzene sinks to the bottom ($d = 1.61$ g/mL). In this manner, the maximum amount was dissolved, since solubility is limited. Hexafluorobenzene solubility in buffer (2.8(4) mM) was determined by saturating hexafluorobenzene in buffer with a known amount of trifluoroethanol for integration reference. Average hexafluorobenzene solubility in the presence of protein (1.4(3) mM) was determined in the same manner, but was determined individually for each experiment whenever possible. All NMR samples were made up in low-volume NMR tubes (Shigemi, Inc., Allison Park, PA). Trifluoroacetic acid (TFA) or trifluoroethanol (TFE) were used as internal standards for chemical shift reference and hexafluorobenzene concentration reference for all ^{19}F NMR experiments. Protein concentrations were determined by UV spectroscopy (see above).

PFGSE NMR. Protein solutions were prepared in D_2O deuterioacetate buffer and saturated with hexafluorobenzene as described above. Multiple samples (at least 5) were prepared for the assessment of each diffusion coefficient. Several $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ solutions were prepared without hexafluorobenzene, several $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ solutions at different protein concentrations were prepared saturated with hexafluorobenzene, and several saturated hexafluorobenzene solutions in buffer were prepared. For each hexafluorobenzene-containing sample, a known amount of trifluoroethanol (TFE) was added for a concentration reference, since the diffusion calculations of K_D are sensitive to ligand concentration. Control experiments with $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ were prepared in the same manner.

^1H and ^{19}F pulsed field gradient spin−echo (PFGSE) NMR experiments using the sLED pulse sequence with bipolar gradient pulses[17,18] were run for each sample from 500 to 28 500 DAC in 15 increments of 2000 (32 000 DAC = 74.5 g cm^{-1}). The ^{19}F hexafluorobenzene signal was monitored in all hexafluorobenzene-containing samples, and the ^1H Val and Leu methyl signal at ∼1 ppm was monitored for samples containing solely protein. Signal intensities were measured with a ruler. All samples for PFGSE studies were run at 25 °C on a Varian Unity/INOVA 600 MHz spectrometer.

1D ^{19}F NMR Titration. To determine δ_L, hexafluorobenzene was saturated into D_2O deuterioacetate buffer as described above. Indirectly referenced to fluorotrichloromethane and directly referenced to TFA (−76.050 ppm), the observed chemical shift δ_L of the free hexafluorobenzene is −163.5 ppm.

(24) Armagero, W. L. F.; Perrin, D. D. *Purification of Laboratory Chemicals*, 4th ed.; Butterworth-Heinemann: Oxford, 1996.
(25) (a) Krôhnke, F.; Gross, K. F. *Chem. Ber.-Recl.* **1959**, *92*, 22−36. (b) Huang, T. L. J.; Brewer, D. G. *Can. J. Chem.* **1981**, *59*, 1689−1700. (c) Black, G.; Depp, E.; Corson, B. B. *J. Org. Chem.* **1949**, *14*, 14−21.

To determine δ_{PL}, hexafluorobenzene was added in a 1:3.5 ratio of hexafluorobenzene (0.28 mM) to $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ (0.98 mM). This is the lowest possible hexafluorobenzene concentration that is still observable by NMR. The observed chemical shift δ_{PL} of the "completely" bound hexafluorobenzene is −162.2 ppm. This value was also independently confirmed by back calculation from the K_D as determined from PFGSE NMR experiments.

To solve for K_D via the titration method, solutions of $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ in D_2O deuterioacetate buffer were prepared and divided; one portion was saturated with hexafluorobenzene. Solutions for ^{19}F NMR were mixed to contain various proportions of hexafluorobenzene. Protein concentration was determined by UV, and hexafluorobenzene concentration was determined by adding a TFE standard. Multiple samples were prepared and the δ_{obs} recorded for each. Control experiments with $[Fe(bpyGCN4\text{-}N16V)_3]^{2+}$ were performed in the same manner.

Competitive Binding Assessment. Competition titrations were performed in a manner similar to the hexafluorobenzene titrations. Holding $[Fe(bpyGCN4\text{-}N16A)_3]^{2+}$ at a constant concentration, the hexafluorobenzene−protein complex was titrated with a secondary ligand−protein solution for several data points, and δ_{obs} of hexafluorobenzene was recorded for each sample. Experiments showing complete replacement by the inhibitor (as indicated by an upfield return of the hexafluorobenzene signal to −163.5 ppm) or resulting in negative K_I values were excluded in the calculation of the mean K_I.

Ligand concentration could not be readily calculated due to the limited solubility of many of those evaluated for competitive binding affinity versus hexafluorobenzene. Therefore, the solubilities of saturated solutions of ligands in buffer were determined in order to know ligand concentration. Hydrophobic ligands were saturated into aqueous solution by adding a small aliquot (approximately 5 μL) to buffer (approximately 0.5 mL), mixing well, and centrifuging 5 min at 10 000 rpm. The molarity of each saturated solution (aqueous layer) of ligand in buffer was determined by ^1H NMR. Deuterioacetate (20 mM), 150 mM NaCl, pH 6.0 buffer was prepared with 3-(trimethylsilyl)propionic acid or 1,4-dioxane standard in D_2O for integration reference. The mean concentration was determined from multiple samples (at least 6). For the freely miscible ligands, the amount dissolved in a given sample was noted. Calculated solubilities are given in Table 2.

Table 2. Molar Solubilities of Ligands Saturated in 20 mM Deuterioacetate, 150 mM NaCl, pH 6.0 D_2O Buffer

small molecule ligand	solubility in D_2O buffer, M
toluene	$3.5(3) \times 10^{-3}$
cyclohexane	$3.4(3) \times 10^{-4}$
benzene	$1.01(5) \times 10^{-2}$
1,3,5-trimethylbenzene	$2.9(2) \times 10^{-4}$
m-xylene	$1.4(6) \times 10^{-3}$
1,3,5-trimethoxybenzene	$1.7(2) \times 10^{-3}$
phenol	1.0(1)
3,5-dimethylphenol	$2.2(3) \times 10^{-2}$
tetrahydropyran	freely miscible
pyridine	freely miscible
1,4-dioxane	freely miscible

Saturation Transfer Difference NMR. Saturation transfer difference experiments[23] were acquired at 25 °C on a Varian Unity/INOVA 600 MHz spectrometer. Protein solutions were saturated with benzene as described above. The protein was irradiated at 1 ppm (methyl signal) at various mixing times (0.05, 0.1, 0.4, 2, and 10 s). The benzene signal was monitored at 7.2 ppm. Spectra were plotted as the difference of on- and off-resonance irradiations.

Acknowledgment. This work was supported by the National Science Foundation (Grant CHE-0106342).

JA035798B