DOI: 10.1002/cbic.200700736

Protein Surface Recognition: Structural Characterisation of Cytochrome c–Porphyrin Complexes

Peter B. Crowley,*[a] Prasad Ganji,[b] and Hasim Ibrahim[b]

One of the challenges that faces chemical biology is the design of molecules that inhibit protein–protein interactions.[1] Generally, protein complexes involve hydrophobic surface patches, the size and shape of which contribute to the binding affinity.[2–5] Specificity, on the other hand, is achieved through polar interactions such as hydrogen bonds, salt bridges and cation–π pairs.[2,6,7] The challenge with inhibitor design, therefore, is to encode a small molecule with sufficient recognition information to approximate the surface presented by the protein. Indeed, the broad range of binding affinities (transient complexes have small interfaces whereas permanent protein assemblies employ large rugged interfaces) suggests increasing levels of design difficulty. The investigation of protein surface ligands can provide useful information towards the goal of rational inhibitor design.

Transient protein complexes are exemplified by the high-turnover interactions utilised in biological electron transport.[8] The analysis of crystallographic and NMR spectroscopic data indicate that redox proteins form transient complexes via small flat binding sites ($\leq$500 Å^2 buried surface per protein), the core of which is hydrophobic and encompasses the redox cofactor or a surface-exposed ligand of the active site.[5,9–13] Polar and charged side chains take up the periphery of the binding site, and frequently there is a juxtaposition of complementary charged groups across the interface.[10–13] For instance, the positively charged Lys residues that surround the exposed haem edge of cytochrome c (cyt c) are complemented by clusters of Asp and Glu side chains on cytochrome c peroxidase (Figure 1).[11,12]

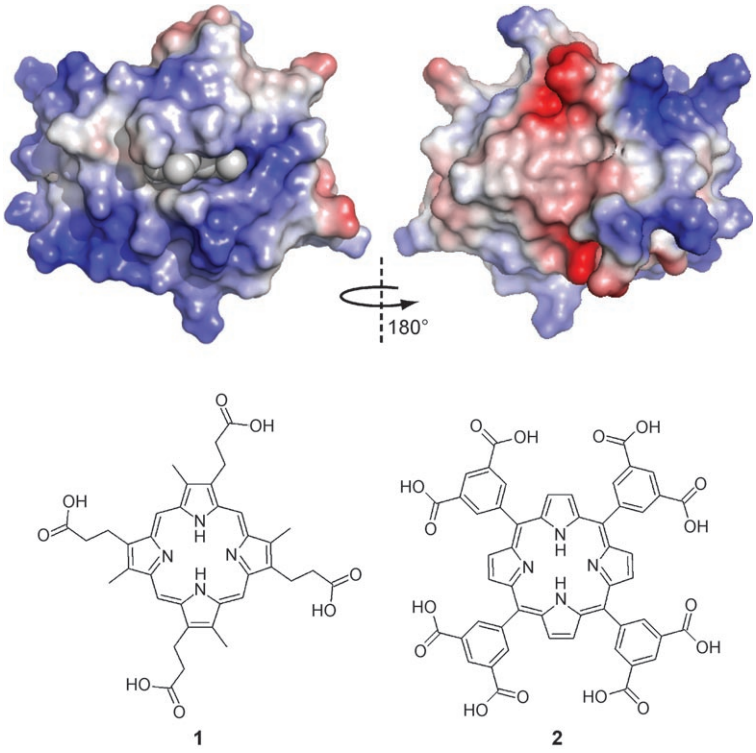

Figure 1. PyMOL surface representation of cyt c[22] illustrating the positive (blue) and negative (red) electrostatic potentials. The exposed haem edge is shown as spheres. The porphyrins used in this study were coproporphyrin I (**1**) and 5,10,15,20-tetrakis-(3,5-dicarboxylatophenyl)-porphyrin (**2**).

The architecture of the redox protein-binding site[10] suggests a surface-ligand design based on a hydrophobic core surrounded by polar/charged groups. Porphyrins have received considerable attention in this regard,[14–20] and the general importance of aromatic rings in protein–ligand interactions has been highlighted.[21] Porphyrins are especially versatile templates as their synthesis allows the introduction of a range of substituents to control the solubility, polarity and charge of the molecule. Thus, in the case of surface ligands for cyt c, porphyrins that bear carboxyphenyl or negatively charged amino acid derivatives have been studied.[15,16,20] While protein–ligand binding affinities in the μM–nM range have been achieved, depending on the relative proportion of acidic and aromatic groups on the porphyrin periphery,[15,16] structural information is lacking to date and the mode of porphyrin binding is unknown.

We report here the characterisation of complex formation between yeast cyt c (12.7 kDa) and two porphyrins (Figure 1) as revealed by 2D ^{1}H,^{15}N correlation spectroscopy. The naturally occurring coproporphyrin I (**1**), which bears four propionic acid groups, and the synthetic octaacid porphyrin (**2**) were chosen in light of previous studies.[16] Figure 2A illustrates the typical changes to the ^{1}H,^{15}N HSQC spectrum of cyt c over the course of titration with **1**. Concentration-dependent chemical-shift perturbations were observed for fifty-three backbone amide resonances. The magnitude of the perturbations increased with the amount of added ligand; this is indicative of fast ex-

[a] *Dr. P. B. Crowley*
UCD School of Biomolecular and Biomedical Science
Conway Institute, University College Dublin
Belfield, Dublin 4 (Ireland)
Fax: (+353) 1-716-6701
E-mail: peter.crowley@ucd.ie

[b] *P. Ganji, Dr. H. Ibrahim*
UCD Centre for Synthesis and Chemical Biology
UCD School of Chemistry and Chemical Biology
University College Dublin
Belfield, Dublin 4 (Ireland)

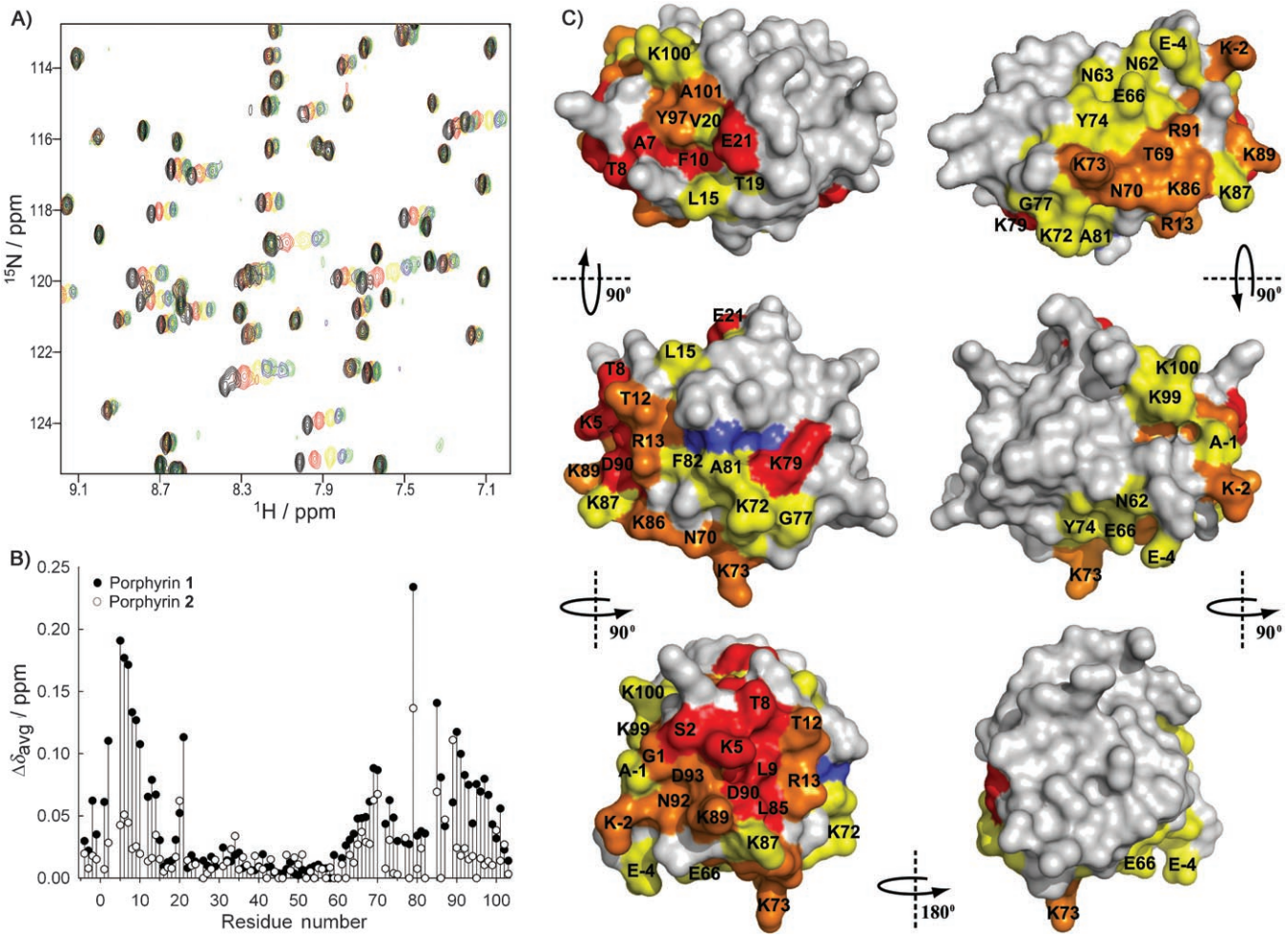

Figure 2. A) A region from the overlaid ^{1}H,^{15}N HSQC spectra of cyt c in the presence of increasing concentrations of **1**. The reference spectrum (0.1 mM pure cyt c) is black while successive titration points are coloured red (0.14), yellow (0.40), blue (0.86) and green (1.62 mM porphyrin **1**). Note the large upfield shifts, which are indicative of ring current effects. B) Plot of the average chemical-shift perturbations ($\Delta\delta_{avg}$) experienced by the backbone amide resonances of cyt c in the presence of excess **1** and **2**. Gaps correspond to unassigned and Pro residues. The residues of cyt c are numbered from −5 to 103.[22] C) Chemical-shift perturbation map of cyt c in the presence of **1**. The molecular surface was generated in PyMOL by using the crystal structure of cyt c.[22] Residues for which the amide resonances experienced small ($\Delta\delta_{avg} > 0.05$ ppm), medium ($\Delta\delta_{avg} \geq 0.10$ ppm) or large ($\Delta\delta_{avg} \geq 0.15$ ppm) shifts are coloured yellow, orange and red, respectively; the haem is shown in blue. The two images in the middle are oriented as in Figure 1, and the left and right panels are related by a 180° rotation. Upper images are related to lower images by 90° rotations as indicated.

change (on the NMR timescale) between the ligand-free and ligand-bound forms of cyt c. Strikingly, the majority of the perturbed resonances experienced an upfield shift in the ^{1}H dimension (Figure 2A).

A significant downfield shift was observed for one resonance (Val20) only. As ring-current effects give rise to upfield shifts, the chemical-shift changes indicate that the interactions of cyt c with **1** are dominated by the porphyrin ring. The upfield shifts might be due to weakening hydrogen bonds[23] since porphyrin binding can destabilise the protein.[19] However, such an explanation is unsatisfactory as changes in the protein structure would give rise to larger and more varied effects in the NMR spectra.

Similar cyt c resonances were affected by **1** and **2** but the perturbations induced by **2** were smaller (Figure 2B) and there were fewer upfield shifts. In fact, one of the largest perturbations was a downfield shift of 0.15 ppm for the $^{1}H^{N}$ of Lys89. While a number of resonances had a similar shift pattern in the presence of either ligand, it appears that the mode of interaction is different. In agreement with previous studies,[16] the binding curves (Figure 3) demonstrate tighter binding for **2** (K_d ~0.07 mM), which saturates at approximately one equivalent. It should be note that the lower affinity interaction with **1** (K_d = 0.4 mM) gave rise to larger chemical-shift perturbations (Figures 2B and 3).

Porphyrin binding had a marginal effect on the line widths of most cyt c resonances. The average line width, which was measured for 70 backbone amide resonances was 17.2 Hz; this increased to 20.0 Hz in the presence of excess porphyrin **1**. Large line width increases were observed for a few resonances including Lys5, Arg13, Thr69 and Tyr97, and the resonance of Glu21 was broadened beyond detection during titrations with

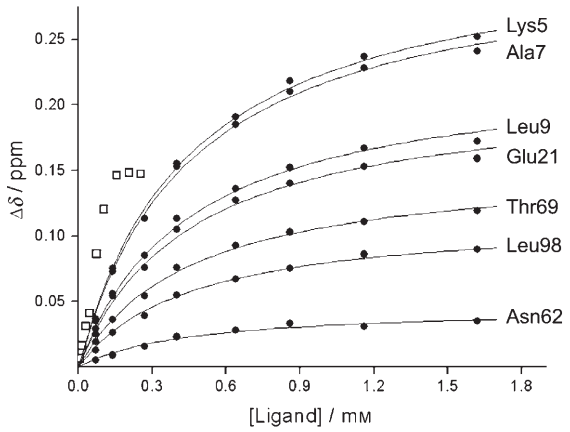

Figure 3. Binding isotherms for cyt c–porphyrin complexes. Labels indicate the $^1H^N$ resonance that was monitored in the presence of porphyrin **1** (●). A global fit of the data to a 1:1 binding model (solid line) yields a $K_d = 0.4$ mM. Also shown is a representative (Lys89) binding curve for porphyrin **2** (□).

2. Such exchange broadening suggests that certain side chains experience conformation changes[24] in order to accommodate ligand binding.

Figure 2B shows a plot of the $^1H^N$ and ^{15}N chemical-shift changes averaged for each backbone amide. Mapping these perturbations onto the crystal structure of cyt c[22] (Figure 2C) reveals a large contiguous patch that covers about half of the protein surface. Compared to transient protein complexes that involve cyt c, the porphyrin binding surface is more extensive and is not limited to the known protein-binding site around the exposed haem edge.[10–12,25,26] The most pronounced effects of porphyrin binding occur in and around the N-terminal helix. Lys5, Leu9 and Asp90 are found at the centre of this patch (Figure 2C). Additional hydrophobicity is provided by Leu85, while the guanidinium of Arg13 is available to form a cation–π interaction with the porphyrin ring. Significant chemical-shift perturbations were detected for the resonances of ten Lys residues, eight of which are located in the vicinity of the haem.

An obvious feature of the chemical-shift map is that it encompasses a surface area several-fold larger than the area of the porphyrin. This raises the question of whether there are several distinct porphyrin-binding sites. However, it is clear from the binding curves (Figure 3) that the data represent a single binding event. In particular, the concentration dependence of $\Delta\delta$ was the same for amides located in different regions of the cyt c structure. The extensive chemical-shift map could also be an indication that ligand binding is experienced indirectly. For instance, complex formation necessitates side-chain conformation changes as well as alterations in protein hydration, which can trigger chemical-shift changes in neighbouring amides. Such indirect or transmitted effects serve to explain why amides in the C-terminal helix experience shifts despite being partly buried by the N-terminal helix. Nevertheless, it is difficult to envisage how indirect effects could perturb the resonances of amides up to 20 Å distance from the binding site.

A more plausible explanation for the extensive chemical-shift map relates to the nature of the protein–porphyrin interaction. If the complex is of low specificity, then the porphyrin might be free to explore different surface patches on the protein. Such a dynamic ensemble of energetically similar interactions has been proposed before for transient protein complexes that involve cyt c. In such cases, despite reasonable association constants ($K_a \sim 10^4$–10^5 M^{-1}), the binding of cyt c does not result in a single orientation.[12,25,26] The chemical-shift changes associated with highly dynamic complexes can be quite small ($\Delta\delta_{avg} \sim 0.05$ ppm),[26] and it has been concluded that $\Delta\delta$ tends to zero as it is averaged over numerous different contributions. Such "dynamic averaging" could be a contributing factor to the smaller perturbations observed for porphyrin **2** compared to **1**.

Docking simulations of the interaction of cyt c and **1** support the hypothesis that the protein–porphyrin complex exists as a dynamic ensemble. Calculations were performed in PatchDock,[27] which yields receptor–ligand configurations based on optimal-shape complementarity. The similarity between the chemical-shift map and the docking results is remarkable. PatchDock predicts three regions on the protein surface, which can accommodate a porphyrin ligand (Figure 4). Each of these

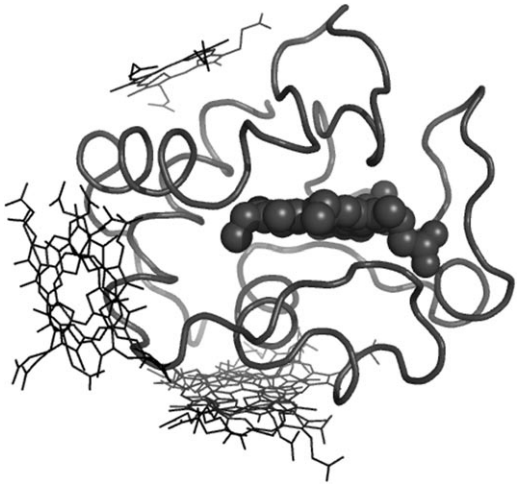

Figure 4. Cyt c–porphyrin docked complexes calculated in Patchdock.[27] Cyt c (oriented as in Figure 1) is illustrated as a ribbon diagram with the haem group shown as spheres. The top ten docking results for porphyrin **1** are shown as sticks. Three of the top ten docking solutions involve interactions at the Lys5-Leu9-Asp90 cluster on cyt c.

surfaces was highlighted by the chemical-shift mapping (Figure 2C). While the docking calculations favour interactions with the "bottom" of cyt c (around Thr69), the NMR spectroscopy data indicate that the predominant interaction is focused around Lys5 and Leu9. Also, PatchDock did not identify the flat surface around the haem edge as a binding site. This discrepancy arises because the scoring function implemented in PatchDock favours configurations that involve concave–convex packing. (Note that the flatness of the haem face of cyt c is a contributing factor to transient protein interactions.)[10] Overall, the docking results suggest that the NMR spectroscopy data represent an ensemble of protein–porphyrin interactions.

How do these results fit in the context of previous studies? The magnitude of the perturbations experienced by cyt c resonances is surprisingly similar to the chemical-shift changes reported for the haem-binding protein p22HBP in complex with an ethylene glycol-modified porphyrin.[28] Protein p22HBP has a broad specificity for porphyrins, and similar to the present study, an extensive surface of the protein was found to be affected by porphyrin binding. From the binding map (Figure 2C) it is evident that porphyrin recognition by cyt c is achieved through surface patches of varying charge and polarity. The variability of protein–porphyrin interactions is highlighted also by protein crystal structures. For instance, in the complex of peanut lectin and tetrasulfonatophenyl porphyrin (PDB ID: 1rir)[29] the protein–porphyrin contacts are mediated by the hydrophobic side chains of Val and Ile. In contrast, in the *N*-methyl-*meso*-porphyrin–antibody complex (PDB ID: 1n7m)[30] the porphyrin is sandwiched between an Asp and Arg on the L chain, and a Tyr on the H chain. Finally, the mode of porphyrin binding contrasts sharply with the complex of cyt c and *p*-aminophenol.[31] The small size of the latter permits it to occupy cavities on the protein surface, resulting in an interaction with Phe36 on the back of cyt c.

In conclusion, the ligand design of a hydrophobic core with a charged periphery results in a low-specificity interaction with the protein surface. As a result, the complex exists as a dynamic ensemble of energetically similar interactions. Current studies are concerned with increasing the specificity of the protein–porphyrin interaction. It is likely that porphyrins that bear amino acid substituents[19] will provide the necessary recognition information to enhance specificity.

Experimental Section

Porphyrin **1** was purchased from Sigma. The synthesis and characterisation of **2** was performed according to known methods.[32] Stock concentrations of ligands were determined by UV/Vis spectroscopy by assuming extinction coefficients of $\varepsilon_{548} = 16.8\ \text{mM}^{-1}\,\text{cm}^{-1}$ (0.1 M HCl) for **1**[18] and $\varepsilon_{516} = 18.6\ \text{mM}^{-1}\,\text{cm}^{-1}$ (0.1 M NaOH) for **2**.[32]

^{15}N-labelled *Saccharomyces cerevisiae* cyt c (that contained the C102T mutation) was prepared and characterised according to published methods.[33] For NMR spectroscopy studies, the typical sample composition was 0.1 mM reduced cyt c, 25 mM potassium phosphate, 50 mM NaCl, 0.1 mM sodium ascorbate (as a reductant) and 10% D_2O, pH 6.0. The sample pH was corrected to pH 6.0 after each addition of ligand.

^{1}H,^{15}N HSQC spectra were acquired at 303 K with spectral widths of 14.1 ppm (^{1}H) and 40.0 ppm (^{15}N) by using a Varian 600 MHz NMR system spectrometer. The analysis of ligand-induced chemical-shift perturbations ($\Delta\delta$ with respect to the spectrum of pure cyt c) was performed in CARA (http://www.nmr.ch/). The perturbations were averaged for each backbone amide resonance by using the formula:

$$\Delta\delta = [(\Delta\delta_N{}^2/25 + \Delta\delta_{HN}{}^2)/2]^{1/2} \quad (1)$$

where $\Delta\delta_N$ and $\Delta\delta_{HN}$ correspond to the change in the ^{15}N and the $^1H^N$ chemical shifts, respectively.

Binding curves were obtained by plotting the magnitude of the chemical-shift change ($\Delta\delta$) as a function of the concentration of added ligand. The data were fit (nonlinear least squares) to a one-site binding model, with $\Delta\delta$ and [ligand] as the dependent and independent variables, respectively, and the maximum chemical-shift change ($\Delta\delta_{MAX}$) and the dissociation constant (K_d) as the fit parameters. A global data analysis was performed in which the curves were fit simultaneously to a single K_d value, while $\Delta\delta_{MAX}$ was varied for each resonance.

Docking calculations were implemented in PatchDock[27] by using the crystal structure coordinates of cyt c (PDB ID: 1ycc[22]). The coordinates for coproporphyrin I were obtained from PDB ID: 1nf4.[34]

Acknowledgements

The authors thank University College Dublin for funding, and Dr. J. Muldoon for assistance with the NMR facilities. The 600 MHz spectrometer was funded by the Higher Education Authority through PRTLI cycle 3. P.G. is the recipient of a UCD Ad Astra scholarship.

Keywords: molecular recognition · NMR spectroscopy · proteins · surface ligands

[1] M. R. Arkin, J. A. Wells, *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317.
[2] S. J. Wodak, J. Janin, *Adv. Protein Chem.* **2002**, *61*, 9–73.
[3] R. P. Bahadur, P. Chakrabarti, F. Rodier, J. Janin, *J. Mol. Biol.* **2004**, *336*, 943–955.
[4] H. Ponstingl, T. Kabir, D. Gorse, J. M. Thornton, *Prog. Biophys. Mol. Biol.* **2005**, *89*, 9–35.
[5] K. Sato, P. B. Crowley, C. Dennison, *J. Biol. Chem.* **2005**, *280*, 19281–19282.
[6] P. B. Crowley, A. Golovin, *Proteins Struct. Funct. Bioinf.* **2005**, *59*, 231–239.
[7] D. Reichmann, O. Rahat, M. Cohen, H. Neuvirth, G. Schreiber, *Curr. Opin. Struct. Biol.* **2007**, *17*, 67–76.
[8] Ed. D. S. Bendall, *Protein Electron Transfer,* 1st ed., BIOS Scientific, Oxford, **1996**.
[9] P. B. Crowley, M. Ubbink, *Acc. Chem. Res.* **2003**, *36*, 723–730.
[10] P. B. Crowley, M. A. Carrondo, *Proteins Struct. Funct. Bioinf.* **2004**, *55*, 603–612.
[11] H. Pelletier, J. Kraut, *Science* **1992**, *258*, 1748–1755.
[12] A. N. Volkov, J. A. R. Worrall, E. Holtzmann, M. Ubbink, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18945–18950.
[13] G. Hagelueken, L. Wiehlmann, T. M. Adams, H. Kolmar, D. W. Heinz, B. Tummler, W.-D. Schubert, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12276–12281.
[14] K. K. Clark-Ferris, J. Fisher, *J. Am. Chem. Soc.* **1985**, *107*, 5007–5008.
[15] R. K. Jain, A. D. Hamilton, *Org. Lett.* **2000**, *2*, 1721–1723.
[16] T. Aya, A. D. Hamilton, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2651–2654.
[17] S. N. Gradl, J. P. Felix, E. Y. Isacoff, M. L. Garcia, D. J. Trauner, *J. Am. Chem. Soc.* **2003**, *125*, 12668–12669.
[18] J. C. Croney, M. K. Helms, D. M. Jameson, R. W. Larsen, *Biophys. J.* **2003**, *84*, 4135–4143.
[19] K. Groves, A. J. Wilson, A. D. Hamilton, *J. Am. Chem. Soc.* **2004**, *126*, 12833–12842.
[20] H. Zhou, L. Baldini, J. Hong, A. J. Wilson, A. D. Hamilton, *J. Am. Chem. Soc.* **2006**, *128*, 2421–2425.
[21] E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem.* **2003**, *115*, 1244–1287; *Angew. Chem. Int. Ed.* **2003**, *42*, 1210–1250.
[22] G. V. Louie, G. D. Brayer, *J. Mol. Biol.* **1990**, *214*, 527–555.
[23] G. Wagner, A. Pardi, K. Wuthrich, *J. Am. Chem. Soc.* **1983**, *105*, 5948–5949.
[24] E. R. Valentine, A. G. Palmer, *Biochemistry* **2005**, *44*, 3410–3417.

[25] P. B. Crowley, K. S. Rabe, J. A. R. Worrall, G. W. Canters, M. Ubbink, *ChemBioChem* **2002**, *3*, 526–533.
[26] J. A. R. Worrall, W. Reinle, R. Bernhardt, M. Ubbink, *Biochemistry* **2003**, *42*, 7068–7076.
[27] D. Schneidman-Duhovny, Y. Inbar, V. Polak, M. Shatsky, I. Halperin, H. Benyamini, A. Barzilai, O. Dror, N. Haspel, R. Nussinov, H. J. Wolfson, *Proteins Struct. Funct. Genet.* **2003**, *52*, 107–112.
[28] D. A. Gell, B. J. Westman, D. Gorman, C. Liew, J. J. Welch, M. J. Weiss, J. P. Mackay, *J. Mol. Biol.* **2006**, *362*, 287–297.
[29] M. Goel, R. S. Damai, D. K. Sethi, K. J. Kaur, B. G. Maiya, M. J. Swamy, D. M. Salunke, *Biochemistry* **2005**, *44*, 5588–5596.
[30] J. Yin, S. E. Andryski, A. E. Beuscher, R. C. Stevens, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 856–861.
[31] M. Assfalg, I. Bertini, R. Del Conte, A. Giachetti, P. Turano, *Biochemistry* **2007**, *46*, 6232–6238.
[32] W. Fudickar, J. Zimmermann, L. Ruhlmann, J. Schneider, B. Roder, U. Siggel, J.-H. Fuhrhop, *J. Am. Chem. Soc.* **1999**, *121*, 9539–9545.
[33] A. S. Morar, D. Kakouras, G. B. Young, J. Boyd, G. J. Pielak, *J. Biol. Inorg. Chem.* **1999**, *4*, 220–222.
[34] S. Macedo, C. V. Romão, E. Mitchell, P. M. Matias, M. Y. Liu, A. V. Xavier, J. LeGall, M. Teixeira, P. Lindley, M. A. Carrondo, *Nat. Struct. Biol.* **2003**, *10*, 285–290.

Received: December 5, 2007
Published online on April 3, 2008

