Peptide Structures

DOI: 10.1002/anie.201001487

Structural Characterization of a Microperoxidase Inside a Metal-Directed Protein Cage**

*Thomas W. Ni and F. Akif Tezcan**

Atomic-level structural characterization of reactive or conformationally fluxional species is hampered by their resistance to crystallization or their short lifetimes, which prevents their interrogation by diffraction methods or NMR spectroscopy. While time-resolved X-ray crystallography has provided detailed structural snapshots of reaction intermediates with lifetimes as short as picoseconds,[1] it is technically challenging and typically restricted to the study of photo-triggerable chemical events and a preformed crystal lattice. In an alternative strategy, unstable species have been isolated and stabilized through encapsulation in molecular hosts, which has subsequently allowed the determination of their structures by NMR methods or crystallography.[2,3] These molecular hosts are typically assemblies of organic building blocks held together by noncovalent or metal-mediated interactions, and have sizes that are in the order of a few nanometers. Owing to their small sizes, which allow extensive interactions with the encapsulated molecular guest(s), these synthetic hosts have enabled the trapping of reactive species for prolonged periods that otherwise would not persist in bulk solution. In this way, several structurally unwieldy or unstable molecules have been characterized, including cyclobutadiene and 1,2,4,6-cycloheptatetraene inside hemicarcerands by the groups of Cram[4] and Warmuth,[5] extended and compressed alkanes inside dimeric cavitand capsules by Ajami and Rebek,[6] reactive organometallic species [Cp(Ru(*cis*-1,3,7-cyclooctatriene)]$^+$ and [Cp′Mn(CO)$_2$] inside gallium- and palladium-directed cages by the groups of Raymond[7] and Fujita,[8] and short peptides with various secondary structures inside palladium-directed bowls and prisms by Fujita et al.[9–11]

We have now extended these studies to a larger substrate captured inside a supramolecular protein assembly. Herein we describe the construction of a novel tetrahedral superprotein architecture (Zn_{30}:CFMC-1_{12}), which is assembled in the crystal lattice through an extensive set of metal-mediated interactions between protein building blocks. We have exploited this well-ordered, crystalline cage to encapsulate and structurally characterize a conformationally fluxional, uncrystallizable heme peptide fragment, namely a microperoxidase, for the first time. The high-resolution crystal structure of this 9-amino-acid-long microperoxidase ($MP9_{cb562}$) provides detailed information on the isolated influence of *c*-type heme protein linkages on the heme structure and the attached polypeptide chain. Significantly, this study highlights a new potential application for designed or natural protein cages to be used as hosts for studying the structure and chemistry of large, biologically significant guests.

The structural evolution of a series of metal-directed protein assemblies that led to our serendipitous discovery of the protein lattice cage Zn_{30}:CFMC-1_{12} is outlined in Figure 1. Because the rationale behind each engineering step is somewhat peripheral to the focus of the current work, we will only give a brief description here (for a more detailed account, see the Supporting Information). The parent species, Zn_4:MBPC-1_4 (Figure 1a), is a D_2-symmetric tetramer of cytochrome cb_{562} building blocks held together by four interfacial zinc(II) ions coordinated to three histidines (63/73/77) and one aspartate (74) side chain.[12] With the original goal of engineering a stable architecture from Zn_4:MBPC-1_4 that can house coordinatively unsaturated and potentially reactive zinc centers, we performed the following modifications: Step 1: Introduction of six hydrophilic-to-hydrophobic mutations (R34A/L38A/Q41W/K42S/D66W/V69I) into one of the interfaces (*i1*) in Zn_4:MBPC-1_4 to stabilize this interface.[13] This step produced the tetrameric complex, Zn_4:RIDC-1_4 (Figure 1b). Step 2: Elimination of one of the zinc ligands (D74) in Zn_4:RIDC-1_4 through the D74A mutation to open up the zinc coordination sphere and split the tetrameric architecture into two, whereby each half contains the stabilized interface *i1*. This modification led to the unexpected formation of a trimeric architecture, Zn_2:^{D74A}RIDC-1_3 (PDB ID: 3M15; Figure 1c) with two coordinatively saturated zinc ions. Step 3: Introduction of unfavorable, sterically demanding interactions into the protein surfaces in Zn_2:^{D74A}RIDC-1_3 to prevent the third protein monomer from saturating the zinc coordination sphere, which was carried out through the incorporation of a surface cysteine residue through the I67C mutation and its covalent modification with a charged group (iodoacetic acid; IAA) to yield CFMC-1. As desired, sedimentation velocity (SV) measurements indicated that CFMC-1 exclusively produces a dimeric species in the presence of zinc (Supporting Information, Figure S2b), which is most likely

[*] T. W. Ni, Prof. Dr. F. A. Tezcan
Department of Chemistry and Biochemistry
University of California, San Diego
9500 Gilman Drive, MC 0356, La Jolla, CA 92093 (USA)
Fax: (+1) 858-534-6157
E-mail: tezcan@ucsd.edu

[**] This work was supported by the National Science Foundation (CHE-0908115), the Hellman Foundation, and the Arnold and Mabel Beckman Foundation. F.A.T. was in part supported by the DOE (DE-FG02-10ER46677, crystallographic analyses). Portions of this research were carried out at SSRL, operated by Stanford University on behalf of DOE.

Supporting information for this article, including experimental details and procedures and a more detailed account of Figure 1, is available on the WWW under http://dx.doi.org/10.1002/anie.201001487.

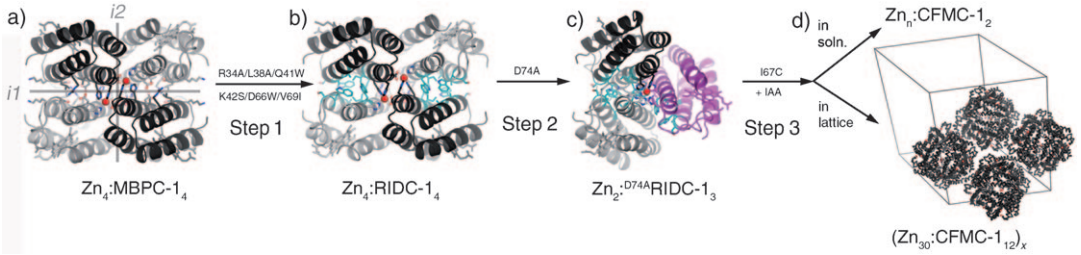

Figure 1. Progression from the parent Zn_4:MBPC-1_4 complex to the Zn_{30}:CFMC-1_{12} tetrahedral units formed in the rhombohedral crystal lattice. Interfaces 1 (*i1*) and 2 (*i2*), present in the Zn_4:MBPC-1_4 and Zn_4:RIDC-1_4 tetramers, are indicated on the former structure. Hydrophobic residues incorporated into *i1* are shown as cyan sticks. Zinc ions are shown as red spheres (see Supporting Information, Figure S1, for a larger view of Zn_4:RIDC-1_4). Only four Zn_{30}:CFMC-1_{12} units are shown for clarity. IAA refers to iodoacetic acid used for the covalent functionalization of C67 to yield CFMC-1.

due to repulsive interactions between the surface acetic acid groups. Surprisingly, upon determining the crystal structure of the Zn-CFMC-1 adduct at 2.5 Å resolution (PDB ID: 3M4B), we encountered a cage-like arrangement of CFMC-1 molecules in the lattice (Figure 1 d), which we subsequently decided to exploit as a crystallographic host.

The rhombohedral (*H*32) lattice of the Zn-CFMC-1 complex contains four protein molecules per asymmetric unit, where chains A, B, and C are related by threefold non-crystallographic symmetry (NCS), and chains B and D are related by a twofold NCS. The crystallographic threefold symmetry operation produces a dodecamer (Zn_{30}:CFMC-1_{12}; Figure 2 a) that has the shape of a truncated tetrahedron

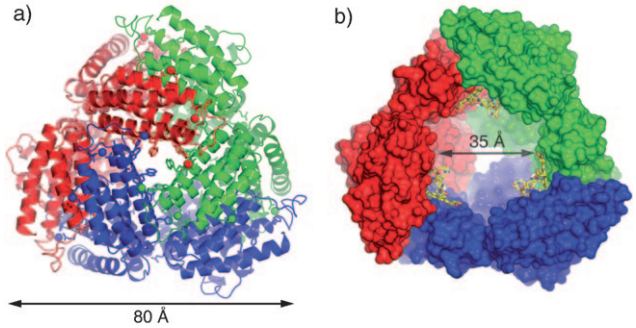

Figure 2. a) Tetrahedral architecture of each Zn_{30}:CFMC-1_{12} unit as viewed down the crystallographic threefold symmetry axis. The contents of each of the three asymmetric units (four protomers and ten Zn ions) that make up the dodecamer are shown in blue, green, and red. b) Surface representation for the inner cavity of Zn_{30}:CFMC-1_{12}, showing the three immobilized $MP9_{cb562}$ molecules (yellow sticks).

measuring approximately 80 Å at the edges. Each dodecameric unit in the lattice is assembled through an extensive network of interfacial zinc ions that form three different sets of coordination environments (Supporting Information, Figure S5). Type 1 zinc atoms (12 in total) are ligated by the amine and carbonyl groups of the N-terminal alanine (A1), D39 from the same monomer, H77′ from a second monomer, and a water molecule. Type 2 zinc atoms (12 in total) are coordinated to H63 and the C67 acetic acid group from one monomer, H59′ from a second monomer, and a water ligand. Type 1 and Type 2 zinc atoms are responsible for mediating the threefold symmetric interprotein interactions and found around the large and the small threefold pores, respectively. Type 3 zinc atoms (6 in total) are ligated to pairs of E8/D12 side chains from two monomers and mediate the twofold symmetric interactions.[14] Significantly, no zinc atoms are found between individual dodecamers, suggesting the possibility that dodecamers could also be formed in solution under the right conditions. Coincidentally, this assembly is reminiscient of tetrahedral aminopeptidases with a dodecameric composition that feature zinc-active sites within their cavities.[15,16]

The cooperative network of zinc–protein interactions formed in the Zn_{30}:CFMC-1_{12} crystal lattice are sufficiently strong to expose the engineered hydrophobic residues as well as the uncoordinated H73 side chain toward the inside of the cage cavity. We envisioned that the resulting hydrophobic patches and H73 near the large threefold NCS pore could together provide a small microenvironment to immobilize a microperoxidase and determine its crystal structure. Microperoxidases (MPs) are small, highly soluble proteolytic digest products of *c*-type cytochromes that are generally 8–11 amino acids long.[17] MPs contain the heme group attached to a conserved Cys-X-X-Cys-His motif, whereby the two Cys residues form covalent thioether bonds (*c*-type linkages) to the heme vinyl groups and the His side chain acts as an axial ligand to iron. Owing to their small sizes and high solubilities, MPs have been widely utilized as structural and chemical models for the inner-sphere coordination environments of heme proteins.[17] Despite their extensive use, however, no full atomic-resolution structure for any MP is available owing to the conformational flexibility of their peptide portion. To obtain an MP, which could be used as a guest inside Zn_{30}:CFMC-1_{12}, we subjected cytochrome cb_{562} to tryptic digestion, which yielded a 9-amino-acid-long peptide fragment ($MP9_{cb562}$; K_{95}TTCNACHQ$_{103}$; Figure 3 a) attached to the heme cofactor. Characteristic of all microperoxidases, the visible spectrum of $MP9_{cb562}$ indicates a high-spin iron(III) center axially coordinated to a His and a solvent molecule (H_2O or OH^-), which becomes low-spin upon addition of excess imidazole to occupy the second axial position (Supporting Information, Figure S6a).[18]

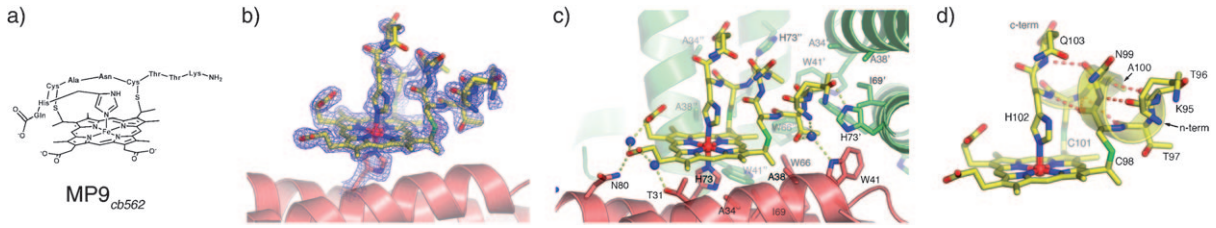

Figure 3. a) Molecular connectivity of $MP9_{cb562}$. b) $2F_o-F_c$ electron density map corresponding to $MP9_{cb562}$ anchored onto His73 inside the Zn_{30}:CFMC-1_{12} cavity (blue: 1.1 σ, purple: 4 σ). c) Interactions of $MP9_{cb562}$ with the cavity surface. H-bonding interactions are shown as yellow dashed lines; ordered water molecules are shown as blue spheres. d) Intramolecular H-bonding interactions (red dashed lines) within $MP9_{cb562}$.

CFMC-1 was cocrystallized with equimolar zinc and substoichiometric amounts $MP9_{cb562}$ to ensure that the latter does not capture the histidines (other than H73) that are involved in interprotein zinc coordination. The resulting crystals maintain the original rhombohedral spacegroup with similar unit cell dimensions indicating the formation of the Zn_{30}:CFMC-1_{12} cage (Supporting Information, Table S1). The 1.9 Å-resolution structure clearly shows the presence of three symmetry-related $MP9_{cb562}$ molecules (one per asymmetric unit) within the cage (Figure 2b; PDB ID: 3M4C). The quality of the data has allowed us to unambiguously place the heme group and all 9 amino acids in the electron density with the exception of the side chains of the N-terminal K95 and the C-terminal Q103 (Figure 3b).

As intended, each $MP9_{cb562}$ is anchored firmly to the protein surface through the Fe–H73 ligation and tucks into a threefold NCS corner of the cage (Supporting Information, Figure S7). While the average temperature factor for $MP9_{cb562}$ ((32 ± 5) Å^2 for all atoms, 25 Å^2 for the iron center as a reference) is somewhat higher than that of the cage ((22 ± 3) Å^2 for all atoms, (16 ± 3) Å^2 for the iron centers of internal hemes), this most likely reflects a less-than-unity occupancy rather than disorder. The best evidence for the lack of disorder is provided by one of the heme propionate groups, which is found in a distinctly bent rather than extended conformation owing to H-bonding to an ordered water molecule shared with the other propionate group (Figure 3c).

$MP9_{cb562}$ adopts a unique orientation due to a number of interactions with the inner cage surface (Figure 3c; see also the Supporting Information, Figure S8, for a detailed map of interactions). Nearly the entire plane of the heme group facing away from the *c*-type linkages is situated on a bed of hydrophobic side chains (A34, A38, W41, W66, I69, L76), of which the first five had been engineered. The heme group, along with the peptide portion of $MP9_{cb562}$, are further involved in hydrophobic contacts with a subset of the same residues from the crystallographically related protein monomers. Indicative of these interactions, the three W66 side chains, which point toward the threefold NCS pore but are disordered in the free cage structure, become ordered upon $MP9_{cb562}$ binding. Of particular importance for its unique orientation, both ends of $MP9_{cb562}$ are involved in H-bonds with the cage surface, including those between a heme propionate and N80, $T96_{MP9}$ hydroxy and $H73_{symm1}$, and $A100_{MP9}$ and $H73_{symm2}$.

Despite the interactions with the cage surface that enable its specific orientation, $MP9_{cb562}$ is in an exposed environment (Supporting Information, Figure S7), particularly when compared to the interior of a protein. With this low-energy structure in hand, we set out to analyze the isolated effects of the *c*-type linkages on the heme and the attached peptide structure. The structure of $MP9_{cb562}$ indicates that the planarity of the heme group is not significantly perturbed by *c*-type linkages. It was suggested that the H-bonds between backbone carbonyl of the first Cys in the CXXCH motif and the amide nitrogen atoms of the second Cys and the His residue may be important contributors to heme ruffling,[19] which in turn may influence heme redox potentials.[20] Comparison of H-bonding interactions in $MP9_{cb562}$ with those from the corresponding fragments in the other *c*-type cytochromes (Supporting Information, Table S2) indicates that there is little variance between the distances of the aforementioned H-bonds, despite significant differences in the extent of heme ruffling (Supporting Information, Figure S9).

Another possible effect of *c*-type linkages is on the orientation of the coordinating His imidazole with respect to the porphyrin ring, which was proposed to contribute to heme electronic asymmetry,[21] and modulate electronic coupling to the heme iron.[22] Paramagnetic NMR studies on the CN^--adduct of Fe^{III}-microperoxidase-8 (MP8) from cytochrome *c* showed that imidazole orientation in this fragment is the same as that in the holo-protein (orange structure in Figure 4b),[23]

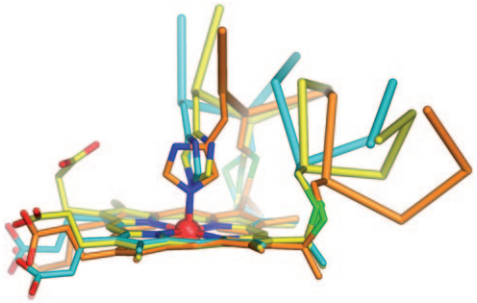

Figure 4. Comparison of $MP9_{cb652}$ structure (yellow) with those of the corresponding heme fragments found within cyt cb_{562} (cyan; PDB ID: 3M4C), and tuna cyt *c* (orange; PDB ID: 3CYT, 1.8 Å resolution).

which is nearly perpendicular to the orientation observed in the $MP9_{cb562}$ crystal structure (yellow structure in Figure 4a). In the cases of both MP8 and $MP9_{cb562}$, a possible *trans*-directing effect of the other axial ligand (cyanide for the former and imidazole for the latter) cannot be discounted. Nevertheless, it appears that in the absence of H-bonding interactions between the His δ-N and the protein backbone or an ordered water (frequently the case in cytochrome *c* structures), the lowest energy orientation is that observed in $MP9_{cb562}$; that is, the imidazole plane roughly parallel to the axis between the two *c*-type linkages.

Aside from their possible influences on heme architecture, *c*-type linkages can also have profound effects on the attached polypeptide. For example, the stability of cytochrome b_{562}, the parent species for cytochrome cb_{562}, was found to increase by more than 12 kJ mol^{-1} upon the engineering of one or both these covalent links.[24,25] The structure of $MP9_{cb562}$ reveals an α-helical peptide conformation throughout its length (Figure 3d) that is slightly more expanded than that of the corresponding segment within the holo-protein (compare yellow and cyan backbones in Figure 4). Given that free $MP9_{cb562}$ in solution assumes a predominantly random-coil conformation (see Supporting Information, Figure S6b, for a far-UV circular dichroism spectrum), the interactions between $MP9_{cb562}$ and the Zn_{30}:CFMC-1_{12} surface clearly contribute to the stabilization of the α-helical structure of $MP9_{cb562}$ and its capture in a discrete state. Nevertheless, the fact that a nine-residue long fragment can assume a helical architecture outside the context of a folded protein suggests that the *c*-type linkages may provide some bias the formation of an α-helical structure toward both the N- and C-terminal ends. This would be consistent with the increased stability of cytochrome cb_{562} compared to the isostructural cytochrome b_{562}, which lacks the *c*-type linkages.

In summary, we were able to immobilize an microperoxidase inside a protein lattice cage, which allowed the first crystallographic structure determination for any member of the microperoxidase family. The structure of $MP9_{cb562}$ confirms previous suggestions that *c*-type linkages likely have little effect on the heme conformation and therefore may primarily serve other functions.[20,26,27] More generally, our study highlights the potential of cage-like protein architectures in hosting large and flexible targets for structural interrogation. A particular advantage of the Zn_{30}:CFMC-1_{12} cage in this regard is that it is formed within an ordered lattice, which would enable the structural characterization of any guest molecule provided that they are properly immobilized within the cage cavity through careful surface design. We can also envision how this strategy can be extended to natural protein cages like ferritins and virus capsids, which are readily crystallized[28,29] and whose cavity surfaces can withstand significant modifications[30,31] toward designing specific interactions with guest molecules. We predict that because targeted biological guests likely will occupy a considerably smaller volume than what is available inside these large hosts, strategies for the covalent or metal-mediated anchoring of the guests on the host surface, in addition to noncovalent interactions, will be necessary for success. This is clearly the case for the $MP9_{cb562}\subset Zn_{30}$:CFMC-$1_{12}$ complex, where the three guests take up only a fraction (ca. 20%) of the internal cage volume.

Received: March 12, 2010
Revised: June 17, 2010
Published online: August 18, 2010

Keywords: host–guest systems · microperoxidases · peptides · protein cages · structure determination

[1] K. Moffat, *Chem. Rev.* **2001**, *101*, 1569–1582.
[2] M. Yoshizawa, J. K. Klosterman, M. Fujita, *Angew. Chem.* **2009**, *121*, 3470–3490; *Angew. Chem. Int. Ed.* **2009**, *48*, 3418–3438.
[3] C. Schmuck, *Angew. Chem.* **2007**, *119*, 5932–5935; *Angew. Chem. Int. Ed.* **2007**, *46*, 5830–5833.
[4] D. J. Cram, M. E. Tanner, R. Thomas, *Angew. Chem.* **1991**, *103*, 1048–1051; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1024–1027.
[5] R. Warmuth, M. A. Marvel, *Angew. Chem.* **2000**, *112*, 1168–1171; *Angew. Chem. Int. Ed.* **2000**, *39*, 1117–1119.
[6] D. Ajami, J. Rebek, *Nat. Chem.* **2009**, *1*, 87–90.
[7] D. Fiedler, R. B. Bergman, K. N. Raymond, *Angew. Chem.* **2006**, *118*, 759–762; *Angew. Chem. Int. Ed.* **2006**, *45*, 745–748.
[8] M. Kawano, Y. Kobayashi, T. Ozeki, M. Fujita, *J. Am. Chem. Soc.* **2006**, *128*, 6558–6559.
[9] S. Tashiro, M. Tominaga, Y. Yamaguchi, K. Kato, M. Fujita, *Angew. Chem.* **2006**, *118*, 247–250; *Angew. Chem. Int. Ed.* **2006**, *45*, 241–244.
[10] S. Tashiro, M. Kobayashi, M. Fujita, *J. Am. Chem. Soc.* **2006**, *128*, 9280–9281.
[11] Y. Hatakeyama, T. Sawada, M. Kawano, M. Fujita, *Angew. Chem.* **2009**, *121*, 8851–8854; *Angew. Chem. Int. Ed.* **2009**, *48*, 8695–8698.
[12] E. N. Salgado, J. Faraone-Mennella, F. A. Tezcan, *J. Am. Chem. Soc.* **2007**, *129*, 13374–13375.
[13] E. N. Salgado, X. I. Ambroggio, J. D. Brodin, R. A. Lewis, B. Kuhlman, F. A. Tezcan, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1827–1832.
[14] There are two other zinc ions involved in twofold symmetric interactions, with one coordinated to Glu57 in chain A (and its symmetry-related partner) and the other to Glu4 (and its symmetry-related partner). These zinc ions are not observed in the second cage-like structure (Zn_{30}:CFMC-1_{12} containing $MP9_{cb562}$). Judging from this observation, and the limited amount of interactions that they are involved in, these zinc ions are deemed not critical for the formation of the cage.
[15] B. Franzetti, G. Schoehn, J. F. Hernandez, M. Jaquinod, R. W. H. Ruigrok, G. Zaccai, *EMBO J.* **2002**, *21*, 2132–2138.
[16] S. Russo, U. Baumann, *J. Biol. Chem.* **2004**, *279*, 51275–51281.
[17] P. A. Adams, D. A. Baldwin, H. M. Marques in *Cytochrome c—A Multidisciplinary Approach* (Eds.: R. A. Scott, A. G. Mauk), University Science Books, Sausalito, **1996**, pp. 635–692.
[18] F. A. Tezcan, J. R. Winkler, H. B. Gray, *J. Am. Chem. Soc.* **1998**, *120*, 13383–13388.
[19] J. G. Ma, M. Laberge, X. Z. Song, W. Jentzen, S. L. Jia, J. Zhang, J. M. Vanderkooi, J. A. Shelnutt, *Biochemistry* **1998**, *37*, 5118–5128.
[20] S. E. J. Bowman, K. L. Bren, *Nat. Prod. Rep.* **2008**, *25*, 1118–1130.
[21] D. L. Turner, C. A. Salgueiro, P. Schenkels, J. Legall, A. V. Xavier, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1995**, *1246*, 24–28.
[22] A. A. Stuchebrukhov, R. A. Marcus, *J. Phys. Chem.* **1995**, *99*, 7581–7590.

[23] D. W. Low, H. B. Gray, J. O. Duus, *J. Am. Chem. Soc.* **1997**, *119*, 1 – 5.
[24] F. Arnesano, L. Banci, I. Bertini, S. Ciofi-Baffoni, T. D. Woodyear, C. M. Johnson, P. D. Barker, *Biochemistry* **2000**, *39*, 1499 – 1514.
[25] J. Faraone-Mennella, F. A. Tezcan, H. B. Gray, J. R. Winkler, *Biochemistry* **2006**, *45*, 10504 – 10511.
[26] P. D. Barker, S. J. Ferguson, *Structure* **1999**, *7*, R281 – R290.
[27] J. M. Stevens, O. Daltrop, J. W. A. Allen, S. J. Ferguson, *Acc. Chem. Res.* **2004**, *37*, 999 – 1007.
[28] D. M. Lawson, P. J. Artymiuk, S. J. Yewdall, J. M. A. Smith, J. C. Livingstone, A. Treffry, A. Luzzago, S. Levi, P. Arosio, G. Cesareni, C. D. Thomas, W. V. Shaw, P. M. Harrison, *Nature* **1991**, *349*, 541 – 544.
[29] R. Golmohammadi, K. Valegard, K. Fridborg, L. Liljas, *J. Mol. Biol.* **1993**, *234*, 620 – 639.
[30] J. Swift, W. A. Wehbi, B. D. Kelly, X. F. Stowell, J. G. Saven, I. J. Dmochowski, *J. Am. Chem. Soc.* **2006**, *128*, 6611 – 6619.
[31] M. Uchida, M. T. Klem, M. Allen, P. Suci, M. Flenniken, E. Gillitzer, Z. Varpness, L. O. Liepold, M. Young, T. Douglas, *Adv. Mater.* **2007**, *19*, 1025 – 1042.

