J|A|C|S
COMMUNICATIONS

Published on Web 10/12/2006

A Genetically Encoded Infrared Probe

Kathryn C. Schultz, Lubica Supekova, Youngha Ryu, Jianming Xie, Roshan Perera,* and Peter G. Schultz*

Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received June 6, 2006; E-mail: perera@scripps.edu; schultz@scripps.edu

The genetic incorporation of unnatural amino acids into proteins has proven a powerful technique for investigating structure and function in biological systems.[1−3] The ability to site-specifically introduce infrared (IR) probes at defined sites in proteins in both prokaryotic and eukaryotic organisms would further expand the utility of this method. In particular, the C≡N stretching vibration is a useful probe since it is small and has a strong absorption in a region of the spectrum (~2200 cm^{-1}) that is free of other vibrational absorptions typically present in proteins. Furthermore, the C≡N stretching frequency is sensitive to subtle changes in local environment, as well as changes due to nitrogen interaction with metal ions.[4] Indeed, the nitrile group has been used as a probe of ligand binding, local environment, and electric field in proteins.[5] Here we report that *para*-cyano-L-phenylalanine (*p*CNPhe) can be efficiently and selectively introduced into proteins in bacteria using an orthogonal nonsense suppressor tRNA/aminoacyl-tRNA synthetase pair derived from the *Methanococcus janannaschi* (*Mj*) tyrosyl tRNA/aminoacyl-tRNA synthetase pair.[1] In addition, we show that *p*CNPhe can be used as a probe of ligand binding to myoglobin (Mb).

To genetically encode *p*CNPhe, a library of 10^9 independent active site mutants of *Mj* tyrosyl-tRNA synthetase (in which Tyr32, Leu65, Phe108, Gln109, Asp158, and Ile159 were randomized) was constructed in the plasmid pBK-lib3D.[1,6] *E. coli* DH10B cells harboring pREP(2)/YC-J17 expressing an Asp112TAG mutant of chloramphenicol acetyl transferase (CAT) and mutRNA$^{Tyr}_{CUA}$ were used as the host strain to positively select for TyrRS mutants that aminoacylate *p*CNPhe[3,6] (Supporting Information). Selected clones were then retransformed into *E. coli* cells harboring the negative selection plasmid pLWJ17B3 in order to remove TyrRS mutants that aminoacylate mutRNA$^{Tyr}_{CUA}$ with endogenous amino acids (Supporting Information). pLWJ17B3 encodes mutRNA$^{Tyr}_{CUA}$ and the toxic protein barnase with three amber mutations (at Gln2, Asp44, and Gly65) under control of the *ara* promoter. After three positive and two negative rounds of selection, one clone (*p*CNPhe RS1) was isolated that survives at 100 μg/mL chloramphenicol when supplemented with 2 mM *p*CNPhe, but not at 20 μg/mL in the absence of *p*CNPhe.

To confirm that *p*CNPhe RS1, together with mutRNA$^{Tyr}_{CUA}$, selectively incorporate *p*CNPhe, a TAG amber codon was introduced at the seventh position (lysine in wild-type protein) in the Z-domain gene. SDS−PAGE analysis revealed that full length Z-domain was produced only in the presence of 2 mM *p*CNPhe, indicating that no endogenous amino acid is inserted at residue 7. Mass spectral analysis of the mutant protein afforded a mass of 7814 Da, as predicted for the site-specific incorporation of *p*CNPhe (Supporting Information, Figure S-4). DNA sequencing of *p*CNPhe RS1 revealed the following active site mutations: Tyr32 → Leu; Leu65 → Val; Phe108 → Trp; Gln109 → Met; Asp158 → Gly; and Ile159 → Ala. Both Tyr32 and Asp158, which hydrogen bond to the *p*-hydroxy group of tyrosine in the wild-type (wt) TyrRS, are mutated to smaller non-hydrogen bonding residues, consistent with replacement of the *p*-hydroxy substituent with a nitrile group.

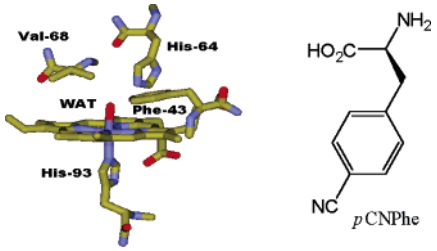

Figure 1. (Left) Schematic representations of the ligand-bound active sites of Mb. The structure was derived from the crystallographic coordinates from pdb file number 2MBW.[7] (Right) Structure of *para*-cyano-L-phenylalanine (*p*CNPhe).

Next we asked whether *p*CNPhe could be used as a probe of ligand binding to protein active sites. Myoglobin was used as a model system since it is a well characterized heme protein that binds a variety of ligands in the ferrous, ferric, and ferryl oxidation states without significant changes in active site structure.[7] *p*CNPhe was substituted for His64, which is located on the distal face of the bound heme group in close proximity to the iron center (Figure 1).[8] The gene encoding Mb with a C-terminal hexahistidine tag was inserted into a pBAD expression vector, and the histidine codon (CAT) at residue 64 of the Mb gene was mutated to an amber (TAG) codon by site-directed mutagenesis to generate pBAD-Mb-(H64TAG). The *p*CNPhe RS1 gene was inserted into a pSup plasmid to generate pSup-*p*CNPheRS-6TRN[9] (Supporting Information). The H64*p*CNPhe Mb mutant was expressed in the presence of 2 mM *p*CNPhe and purified by Ni^{2+} affinity chromatography in an isolated yield of 30 mg/L (Supporting Information); no protein was found by SDS−PAGE analysis in the absence of *p*CNPhe (Figure S-6). Mass spectral analysis (MALDI-TOF) afforded a parent ion mass of 18 391 Da as expected for the H64*p*CNPhe mutant Mb. Mb samples (1 mM) for Fourier transform infrared spectroscopy (FTIR) were prepared by adding sodium cyanide to ferric protein; ferrous adducts were prepared by bubbling CO, NO, and O_2 gases through anaerobically reduced Mb as previously reported.[7] FTIR and electronic absorption (EA) measurements and data processing were carried out as described elsewhere.[10]

The native form of Fe(III) Mb at pH 7.0 is ligated to a water molecule in the distal site with Fe(III) in a high-spin ($S = {}^5/_2$) electronic configuration. The overall electronic absorption spectra of ferric wild-type and ferric H64*p*CNPhe Mb are quite similar (Figure 2a). Addition of the anionic donor ligand cyanide to ferric Mb, which results in a low-spin ($S = {}^1/_2$) six-coordinate state, also does not result in significant differences between the absorption spectra of the mutant and wt ferric cyanide Mb complexes (Figure 2a). These results suggest that substitution of His64 with *p*CNPhe does not significantly affect the active site structure of Mb nor the

10.1021/ja0636690 CCC: $33.50

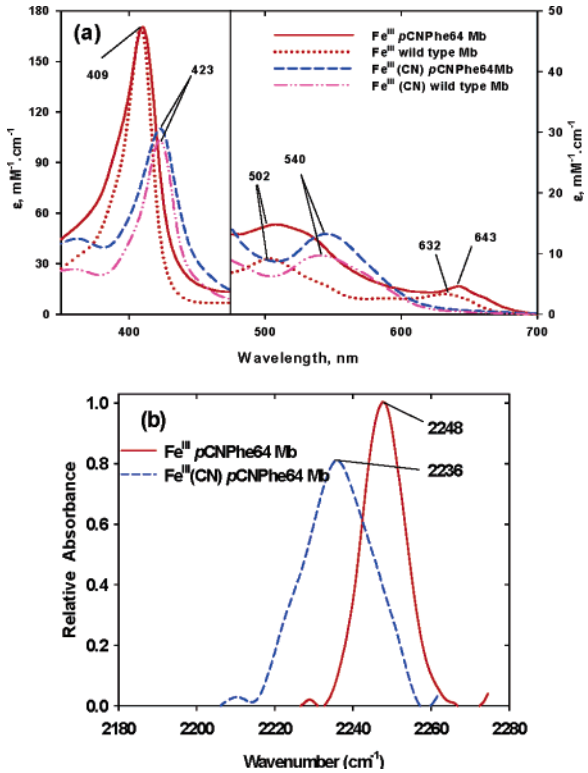

Figure 2. (a) The electronic absorption spectra of the ferric-met (solid red line) and ferric CN (blue dashed line) His64*p*CNPhe mutant Mb complexes are plotted against ferric (red dotted line) and ferric CN (red dashed−dotted line) wild-type Mb. The spectra were recorded at 4 °C with Mb (~10 μM) in 0.1 M potassium phosphate buffer (pH 7.0) in the absence and in the presence of 10 mM sodium cyanide. (b) FTIR spectra (in C≡N stretching region) of ferric-met (solid red line) and ferric CN bound (blue dashed line) complex of His64*p*CNPhe mutant Mb. The spectra were measured at 25 °C in 100 mM potassium phosphate buffer, pH 7.0 with 200 μM Mb. Addition of 10 mM sodium cyanide to free *p*CNPhe in buffer resulted in no IR shift.

overall fold of the protein. The same is true for the electronic absorption spectra of the CO, NO, and O_2 bound six-coordinate, low-spin ($S = 0$) states of the ferrous forms of wt and the H64*p*CNPhe mutant Mb (Supporting Information, Figures S-2 and S-3). Interestingly, a comparison of ferrous wt and H64*p*CNPhe mutant Mb, in which Fe(II) is five coordinate ($S = 2$) and bound only to the proximal heme, does show differences in the electronic absorption spectra (Supporting Information, Figure S-2). The H64*p*CNPhe mutant has a Soret peak at 426 nm as well as maxima at 530 and 559 nm, whereas wt Mb has maxima at 434 (Soret) and 556 nm. Thus in the absence of bound ligand, the nitrile group of *p*CNPhe may be interacting with Fe(II) when fixed in close proximity to the heme.

The FTIR spectra were then measured for the ligated forms of the H64*p*CNPhe mutant Mb and wt Mb to determine whether the C≡N stretch frequency is sensitive to the nature of the ligand bound in the distal site. Previous work has shown that ν_{CN} is highly sensitive to solvent polarity and hydrogen bonding (ν_{CN} of *p*CNPhe is 2237 cm^{-1} in H_2O versus 2228 cm^{-1} in tetrahydrofuran).[4,5] The ν_{CN} of the water ligated mutant ferric Mb (200 μM) is 2248 cm^{-1} (Figure 2b). This represents a significant shift (11 cm^{-1}) relative to that of free *p*CNPhe in 100 mM potassium phosphate buffer (2237.5 cm^{-1}) or water (2237 cm^{-1}),[5] consistent with the increased polarity of water bound to the Fe(III) ion in Mb.[8] A lower stretch frequency ($\nu_{CN} = 2236$ cm^{-1}) is observed for the mutant Fe(III)-CN complex, indicating a less polar active site which is consistent with a weakly π-back-bonded linear ferric CN complex. Fe(II) can act as an effective dπ-donor, and Fe(II) readily coordinates with π-acid ligands, such as CO, NO, and O_2, to form six-coordinate, low-spin ($S = 0$) complexes.[11,12] Both the Fe(II) NO and O_2 adducts of the H64*p*CNPhe mutant Mb show a ν_{CN} stretching band at 2230 cm^{-1} (Supporting Information, Figure S-1). This is consistent with isoelectronic, bent conformations for both the Fe(II)NO and Fe(II)O_2 complexes. In contrast, the linear Fe(II)CO complex shows a ν_{CN} absorption at 2239 cm^{-1} [in which the CO−Fe(II) heme interaction involves significant back-bonding], consistent with decreased electron density on the carbonyl oxygen relative to NO and O_2. These results taken together demonstrate that the nitrile group of *p*CNPhe, when site-specifically introduced into the active site of Mb, is an excellent probe of ligand binding and local electronic environment. In conclusion, we have shown that one can genetically encode the infrared probe *p*CNPhe in bacteria. This unnatural amino acid should also be a useful site-specific probe of protein folding, conformational changes, and biomolecular interactions and may be a useful probe of local electric field in proteins.[5,13]

Acknowledgment. We acknowledge support from the Department of Energy (DE-FG02-03ER46051) and the Skaggs Institute for Chemical Biology. We also thank Johnathan Chittuluru, Dr. Matthew E. Cremeens, and Dr. Floyd E. Romesberg for their help. This is manuscript number 18265 of The Scripps Research Institute.

Supporting Information Available: Electronic and infrared absorption and mass spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Wang, L.; Xie J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225−249. (b) Wang, L.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2004**, *44*, 34−66.
(2) Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 435−462.
(3) (a) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498−500. (b) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964−967.
(4) (a) Reimers, J. R.; Hall, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 3730−3744. (b) Pace, E. L.; Noe, L. J. *J. Chem. Phys.* **1968**, *49*, 5317−5325. (c) Eaton, G.; Pena-Nunez, A. S.; Symons, M. C. R. *J. Chem. Soc., Faraday Trans. 1* **1988**, *84*, 2181−2193. (d) Fawcett, W. R.; Liu, G. J.; Kessler, T. E. *J. Phys. Chem.* **1993**, *97*, 9293−9298. (e) Nyquist, R. A. *Appl. Spectrosc.* **1990**, *44*, 1405−1407.
(5) (a) Yoshikawa, S.; O'Keeffe, D. H.; Caughey, W. S. *J. Biol. Chem.* **1985**, *260*, 3518−3528. (b) Getahun, Z.; Huang, C. Y.; Wang, T.; De Leon, B.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125*, 405−411. (c) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science* **2006**, *313*, 200−204.
(6) Xie, J.; Schultz, P. G. *Methods* **2005**, *36*, 227−238.
(7) (a) Perera, R.; Sono, M.; Sigman, J. A.; Pfister, T. D.; Lu, Y.; Dawson, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3641−3646. (b) Perera, R.; Dawson, J. H. *J. Porphyrins Phthalocyanins* **2004**, *8*, 246−254.
(8) (a) Brucker, E. A.; Olson, J. S.; Phillips, G. N., Jr.; Dou, Y.; Ikeda-Saito, M. *J. Biol. Chem.* **1996**, *271*, 25419−25422. (b) Evans, S. V.; Brayer, G. D. *J. Biol. Chem.* **1988**, *263*, 4263−4268.
(9) Ryu, Y.; Schultz, P. G. *Nat. Methods* **2006**, *3*, 263−265.
(10) (a) Chin, J. K.; Jimenez, R.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1846−1847. (b) Cremeens, M. E.; Fujisaki, H.; Zhanf, Y.; Zimmermann, J.; Sagle, L. B.; Matsuda, S.; Dawson, P. E.; Straub, J. E.; Romesberg, F. E. *J. Am. Chem. Soc.* **2006**, *128*, 6028−6029.
(11) (a) Carver, T. E.; Rohlfs, R. J.; Olson, J. S.; Gibson, Q. H.; Blackmore, R. S.; Springer, B. A.; Sligar, S. G. *J. Biol. Chem.* **1990**, *265*, 20007−20020. (b) Doster, W.; Beece, D.; Bowne, S. F.; Dilorio, E. E.; Eisenstein, L.; Frauenfelder, H.; Reinisch, L.; Shyamsunder, E.; Winterhalter, K. H.; Yue, K. T. *Biochemistry* **1982**, *21*, 4831−4839. (c) Gibson, Q. H.; Regan, R.; Elbers, R.; Olson, J. S.; Carver, T. E. *J. Biol. Chem.* **1992**, *267*, 22022−22034. (d) Quillin, M. L.; Li, T.; Olson, J. S.; Phillips, G. N., Jr.; Duo, Y.; Ikeda-Saito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; Li, H.; Elbe, R. *J. Mol. Biol.* **1995**, 245, 416−436.
(12) (a) Storhoff, B. N.; Lewis, H. C., Jr. *Coord. Chem. Rev.* **1997**, *23*, 1−29. (b) Corain, B. *Coord. Chem. Rev.* **1982**, *47*, 165−200. (c) Miller, L. M.; Pedraza, A. J.; Chance, M. R. *Biochemistry* **1997**, *36*, 12199−12207.
(13) (a) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2000**, *104*, 11853−11863. (b) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2002**, *106*, 469−477.

JA0636690