Site-Specific in vivo Labeling of Proteins for NMR Studies

Alexander Deiters,[b] Bernhard H. Geierstanger,[c] and Peter G. Schultz*[a]

Studies of biological macromolecules by NMR spectroscopy become increasingly difficult as the molecular weight of the molecule of interest increases, due to signal overlap and signal reduction resulting from faster transverse relaxation. Partial and uniform ^{2}H-, ^{13}C-, and ^{15}N-labeling of proteins combined with heteronuclear, multidimensional NMR experiments[1] can overcome these problems to some extent and has allowed the structure elucidation of proteins with molecular weights of

[a] *Prof. P. G. Schultz*
Department of Chemistry and the Skaggs Institute for Chemical Biology
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-9440
E-mail: schultz@scripps.edu

[b] *Prof. A. Deiters*
Present address:
Department of Chemistry, North Carolina State University
Raleigh, NC 27695-8204 (USA)

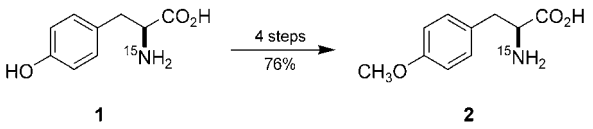
[c] *Dr. B. H. Geierstanger*
Genomics Institute of the Novartis Research Foundation
10675 John Jay Hopkins Drive, San Diego, CA 92121 (USA)

 DOI: 10.1002/cbic.200400319

30 kDa.[1,2] The development of transverse relaxation optimized spectroscopy (TROSY)[3] has extended the limit of solution NMR studies to systems as large as 900 kDa.[4,5] Ultimately the resonances in large proteins become impossible to resolve even at the highest-available magnetic fields. In many cases, including solution studies of protein–ligand interactions, protein conformational changes, and enzyme catalysis, it would be sufficient to only assign residues of an active site or a ligand-binding site by using, for example, the SEA-TROSY experiment.[6] The ability to introduce one or several site-specific NMR labels at defined locations in a protein would greatly simplify the assignment problem.

To obtain sufficient quantities for NMR measurements, most isotopically labeled proteins are recombinantly expressed in *E. coli* by using minimal media in combination with ^{13}C glucose, ^{15}N ammonium salts, and deuterium oxide. Strategies for the more selective incorporation of isotopes include feeding experiments with labeled amino acids in defined media,[1b] often by utilizing auxotrophic bacterial expression strains, "reverse isotope" labeling,[7] segmental labeling by *trans*-splicing,[8] total and semisynthesis by chemical ligation,[9] and cell-free expression systems with chemically aminoacylated suppressor tRNAs.[10] Although site-specific incorporation of isotopic labels into a protein has been demonstrated by the last method,[10b] the production of milligram quantities sufficient for NMR measurements can be difficult. Here, we present a solution to this problem based on unnatural amino acid mutagenesis[11] that allows for the site-specific incorporation of isotopically-labeled amino acids in vivo.

Previously, we demonstrated that a *Methanococcus jannaschii* tyrosyl tRNA/tRNA-synthetase pair (MjTyrRS/$tRNA_{CUA}$) is orthogonal in *E. coli*, that is, neither the tRNA nor the synthetase cross reacts with endogenous *E. coli* tRNAs or synthetases. The specificity of this and other orthogonal tRNA-synthetase pairs has been evolved to allow the selective and efficient incorporation of a number of unnatural amino acids in response to nonsense and frameshift codons, including keto, sugar, azido, alkynyl, and photocrosslinking amino acids.[12] In order to selectively introduce an isotopically labeled amino acid into a protein in *E. coli* by this method, the amino acid must be structurally distinguishable from the common 20 amino acids by the mutant aminoacyl tRNA synthetase. This difference cannot rely on the isotope itself, since the wild-type synthetase for any particular common amino acid would incorporate the corresponding isotopically substituted amino acid throughout the protein. However, a methoxy group is sufficient for the translational machinery of *E. coli* to differentiate it from phenylalanine, tyrosine, and other natural amino acids, yet it is small enough to minimize structural perturbations within the protein of interest. Therefore a methylated ^{15}N-labeled tyrosine derivative **2** was initially used and synthesized from commercially available material **1** in four steps and an overall yield of 76% (Scheme 1). The reaction sequence consists of a Boc-protection of the amino group (Boc_2O, Et_3N, dioxane/H_2O), simultaneous methylation of the hydroxy and the carboxy group (MeI, K_2CO_3, DMF), removal of the Boc group (HCl, MeOH), and a subsequent saponification of the ester (NaOH, MeOH/H_2O).



1 **2**

Scheme 1. *Synthesis of ^{15}N-labeled* p*-methoxyphenylalanine (**2**).*

To incorporate **2** into proteins at unique sites, we employed an orthogonal TyrRS/$tRNA_{CUA}$ pair that genetically encodes *p*-methoxylphenylalanine in *E. coli*. This tRNA-synthetase pair has been used to incorporate *p*-methoxylphenylalanine into dihydrofolate reductase with high fidelity and efficiency.[13] Here, we use this $tRNA_{CUA}$/TyrRS pair to selectively incorporate isotopically labeled **2** into sperm whale myoglobin, a monomeric 153-residue heme that has been the focus of a large number of structural, mechanistic, and protein-folding studies.[14] Myoglobin is therefore an attractive model system to test this approach. To produce site-specifically ^{15}N-labeled myoglobin, the fourth codon (Ser4) was mutated to the amber stop codon TAG and a C-terminal 6×His tag was added. In the presence of the mutant MjTyrRS, $tRNA_{CUA}$, and **2** (1 mM in liquid minimal media), full-length myoglobin was produced with a yield of 1 mg L^{-1} after purification by Ni-affinity chromatography, and judged to be >90% homogeneous by SDS-Page and Gelcode Blue staining. In the absence of **2** no myoglobin was visible; the fidelity for the incorporation of **2** is >99% (Figure 1).

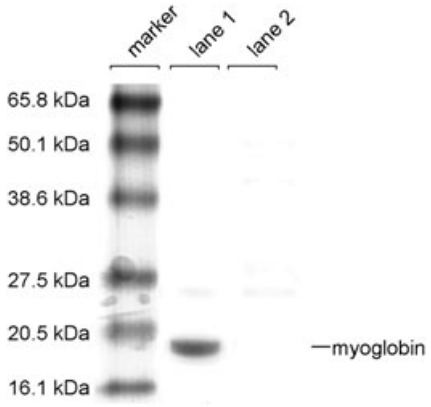


Figure 1. *Gelcode Blue-stained SDS-PAGE gel of purified ^{15}N-MeOPhe-myoglobin. Lane 1 contains protein expressed in minimal media in the presence of 1 mM **2**; lane 2 contains a sample expressed in the absence of **2**.*

The purified protein was dialysed against 50 mM phosphate buffer (pH 5.6) and concentrated to give 0.5 mL of a 55 μM NMR sample (H_2O/D_2O 90%:10%).[10b] A similar sample was prepared with unlabeled *p*-methoxyphenylalanine. Both samples were used in $^1H,^{15}N$ HSQC experiments that were acquired with 64 t_1 increments and 512 scans per increment on a Bruker Avance 600 at 300 K. The spectrum of the ^{15}N-labeled protein shows a single amide correlation peak at 8.86 ppm (^1H chemical shift) for the amide proton and 120.6 ppm (^{15}N chemical shift) for the amide nitrogen resonance. The same region of a $^1H,^{15}N$ HSQC experiment acquired under identical conditions

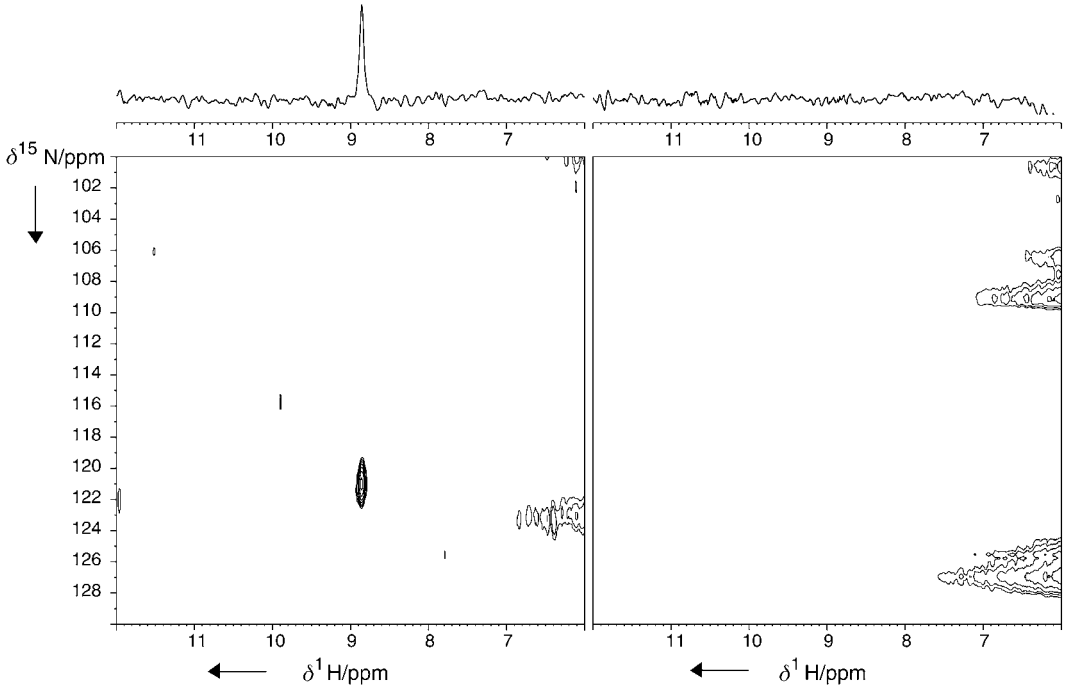


Figure 2. *^{1}H,^{15}N HSQC NMR spectra of ^{15}N-MeOPhe-labeled (left) and unlabeled myoglobin (right). Cross sections along the nitrogen chemical shift of 120.6 ppm are shown above the 2D contour plots.*

for the unlabeled myoglobin sample shows no correlation peak (Figure 2).

In summary, we have demonstrated that genetically encoded isotopically labeled amino acids can be used to obtain amounts of site-specifically labeled proteins sufficient for NMR studies. Since our in vivo expression system uses defined minimal media, it is expected that, in addition to incorporation of the ^{15}N label, fully or partially deuterated or ^{13}C-labeled protein samples of large proteins can be produced. One can also isotopically label additional positions in *p*-methoxyphenylalanine or other unnatural amino acids. The production of site-specifically labeled proteins should also be possible in yeast.[15] Ultimately, this methodology should facilitate detailed studies of larger proteins, protein–ligand interactions, protein conformational changes, and mechanisms of enzyme catalysis. Moreover, this in vivo labeling technique may allow in-cell NMR applications by enabling the observation of a particular protein in the context of other macromolecules.[16]

Experimental Section

To express protein, plasmid pBAD/JYAMB-4TAG containing the mutant sperm whale myoglobin gene (constructed in our laboratory) on an arabinose promoter and a rrnB terminator, tyrosyl $tRNA_{CUA}$ on a lpp promoter and a rrnC terminator, and a tetracycline resistance marker were cotransformed with a pBK vector expressing the mutant synthethase and a kanamycin resistance gene into DH10B *E. coli*. Cells were amplified in LB media (5 mL) supplemented with tetracycline (25 mg L^{-1}) and kanamycin (30 mg L^{-1}), washed with PBS, and being used to inoculate GMML (500 mL) with the appropriate antibiotics, **2** (1 mM), and arabinose (0.002%). Cells were grown to saturation and then harvested by centrifugation. The protein was purified by Ni-affinity chromatography, yielding ^{15}N-labeled myoglobin (~0.5 mg).

Acknowledgements

A.D. gratefully acknowledges a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft. This work is supported by funding from the National Institutes of Health (GM62159) and is manuscript CH-16657 of The Scripps Research Institute.

Keywords: amino acids · isotopic labeling · mutagenesis · NMR spectroscopy · protein expression

[1] a) N. K. Goto, L. E. Kay, *Curr. Opin. Struct. Biol.* **2000**, *10*, 585; b) K. H. Gardner, L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357.
[2] a) K. Wüthrich, *Angew. Chem.* **2003**, *115*, 3462; *Angew. Chem. Int. Ed.* **2003**, *42*, 3340; b) A. Bax, *Curr. Opin. Struct. Biol.* **1994**, *4*, 738.
[3] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366.
[4] J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich, *Nature* **2002**, *418*, 207.
[5] C. Fernandez, G. Wider, *Curr. Opin. Struct. Biol.* **2003**, *13*, 570.
[6] M. Pellecchia, D. Meininger, A. L. Shen, R. Jack, C. B. Kasper, D. S. Sem, *J. Am. Chem. Soc.* **2001**, *123*, 4633.
[7] a) G. W. Vuister, S. J. Kim, C. Wu, A. Bax, *J. Am. Chem. Soc.* **1994**, *116*, 9206; b) M. J. S. Kelly, C. Krieger, L. J. Ball, Y. H. Yu, G. Richter, P. Schmieder, A. Bacher, H. Oschkinat, *J. Biomol. NMR* **1999**, *14*, 79.
[8] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, H. Nakamura, *J. Am. Chem. Soc.* **1998**, *120*, 5591.
[9] R. Xu, B. Ayers, D. Cowburn, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388.
[10] a) T. Yabuki, T. Kigawa, N. Dohmae, K. Takio, T. Terada, Y. Ito, E. D. Laue, J. A. Cooper, *J. Biomol. NMR* **1998**, *11*, 295; b) J. A. Ellman, B. F. Volkman,

D. Mendel, P. G. Schultz, D. E. Wemmer, *J. Am. Chem. Soc.* **1992**, *114*, 7959.
[11] L. Wang, P. G. Schultz, *Chem. Commun.* **2002**, *1*, 1.
[12] a) L. Alfonta, Z. Zhang, S. Uryu, J. A. Loo, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 14662; b) A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782; c) Z. Zhang, B. A. Smith, L. Wang, A. Brock, C. Cho, P. G. Schultz, *Biochemistry* **2003**, *42*, 6735; d) J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020.
[13] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
[14] a) C. J. Reedy, B. R. Gibney, *Chem. Rev.* **2004**, *104*, 617, and references therein; b) T. Uzawa, S. Akiyama, T. Kimura, S. Takahashi, K. Ishimori, I. Morishima, T. Fujisawa, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1171, and references therein; c) P. E. Wright, R. L. Baldwin in *Frontiers in Molecular Biology: Mechanisms of Protein Folding*, (Ed.: R. Pain), Oxford University Press, London, **2000**, p. 309.
[15] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, *301*, 964.
[16] Z. Serber, W. Straub, L. Corsini, A. M. Nomura, N. Shimba, C. S. Craik, P. Ortiz de Montellano, V. Dötsch, *J. Am. Chem. Soc.* **2004**, *126*, 7119, and references therein.

Received: September 9, 2004
Early View Article
Published online on November 17, 2004

