Available online at www.sciencedirect.com
ScienceDirect
Tetrahedron
Tetrahedron 63 (2007) 6182–6184

A genetically encoded metallocene containing amino acid

Eric M. Tippmann 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, CA 92037, USA

Received 11 January 2007; revised 28 February 2007; accepted 28 February 2007
Available online 3 March 2007

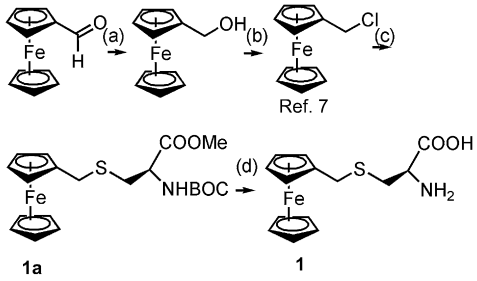
Abstract—Redox active amino acids, cofactors, and metal ions are involved in a large number of catalytic, electron transfer, and regulatory processes in biology. Consequently, the ability to engineer redox active centers at defined sites in proteins would facilitate both the study and manipulation of a wide range of biological processes. Recently, we demonstrated that the redox active amino acid 3,4-dihydroxyphenylalanine could be efficiently and selectively incorporated into proteins in *Escherichia coli* using a nonsense codon and a corresponding orthogonal tRNA/aminoacyl-tRNA synthetase pair. We now report that ferrocene derivative **1** can be genetically encoded in *Saccharomyces cerevisiae* (*S. cerevisiae*) in good yield in response to the amber codon, TAG.

1. Introduction

Redox active amino acids, cofactors, and metal ions are involved in a large number of catalytic, electron transfer, and regulatory processes in biology.[1] Consequently, the ability to engineer redox active centers at defined sites in proteins would facilitate both the study and manipulation of a wide range of biological processes. Because redox active cofactors or metal ion binding sites involve multiple precisely positioned protein side chains, they are relatively difficult to introduce at specific sites in proteins.[2] Chemical methods can be used to modify proteins with transition metal complexes or cofactors, but typically these reactions are relatively unselective, require denaturing conditions, or proceed in low yield.[3,4] Recently, we demonstrated that the redox active amino acid 3,4-dihydroxyphenylalanine could be efficiently and selectively incorporated into proteins in *Escherichia coli* using a nonsense codon and a corresponding orthogonal tRNA/aminoacyl-tRNA synthetase pair.[5] We now report that ferrocene derivative **1** can be genetically encoded in *Saccharomyces cerevisiae* (*S. cerevisiae*) in good yield in response to the amber codon, TAG.

2. Results and discussion

The unnatural amino acid **1** was chosen for a number of reasons—it acts as a fast, reversible, one-electron redox couple, is nontoxic to bacteria and yeast, and is stable under physiological conditions.[6] Compound **1** was synthesized in five steps in an overall yield of 18% from ferrocenecarboxaldehyde according to Scheme 1.



Scheme 1. Synthesis of amino acid **1**. Reagents and conditions: (a) $NaBH_4$ 1/3 equiv, THF/MeOH 50:1, 12 h (99%); (b) oxalyl chloride 1.5 equiv, CH_2Cl_2, 4 h (99%), Ref. 7; (c) Cs_2CO_3/*t*-BOC-(L)-cysteine methylester, DMF, 12 h (40%); (d) 4 M HCl in dioxane, THF, 40 °C, 6–12 h, LiOH in THF/H_2O (20–45%).

To genetically encode **1** in *S. cerevisiae*, a previously reported orthogonal *E. coli* amber suppressor $tRNA^{Leu5}_{CUA}$/leucyl-tRNA synthetase (LeuRS) pair was used.[8,9] Neither this tRNA nor the aminoacyl-tRNA synthetase (aaRS) cross-reacts with the endogenous tRNAs or aaRSs in yeast. A library of 10^7 mutants of LeuRS was generated by randomizing residues Met40, Leu41, Tyr499, Tyr527, and His537 in the leucine-binding site.[10] To evolve a LeuRS specific for **1**, this library was then subjected to a selection scheme in which suppression of amber codons at positions T44 and R110 (T44 and R100 are permissive sites) in the gene for the transcriptional activator GAL4 (in the MaV203:pGADGAL4(2TAG)

Keywords: Unnatural amino acid; Organometallic; Protein expression; hSOD.
* Corresponding author. Tel.: +1 858 784 9300; fax: +1 858 784 9440; e-mail: schultz@scripps.edu

0040–4020/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tet.2007.02.125

yeast strain) leads to production of functional full-length GAL4. GAL4 in turn upregulates expression of *his3* and *ura3* reporter genes.[11] These gene products complement histidine and uracil auxotrophy, allowing clones harboring active synthetase mutants that aminoacylate endogenous amino acids or **1** to be selected on synthetic dropout medium (SD medium containing dextrose) supplemented with 1 mM amino acid **1**. Negative selection of synthetases that accept endogenous amino acids was performed by growth on SD medium lacking **1** but containing 0.1% 5-fluoroorotic acid, which is converted into a toxic product by URA3 resulting in the removal of these clones.[12]

After five rounds of selection (positive, negative, positive, etc.), the growth rates of 96 isolated clones were assayed individually on selective media in the presence or absence of **1**. More than 80 clones had the desirable phenotype, that is growth in the presence of **1** and no growth in the absence of **1**. However, when sequenced these clones corresponded to only two different clones represented by LeuRS32 and LeuRS35. The ability of these two clones to selectively incorporate **1** into proteins was tested by suppression of an amber mutant (Trp33→TAG) in C-terminal His6-tagged human superoxide dismutase (hSOD-33TAG-His6). hSOD is a bimetallic Cu/Zn enzyme that catalyzes the disproportionation of superoxide to molecular oxygen and hydrogen peroxide. It has an overall Greek key β-barrel fold. Trp33 is a surface residue at the center of β-strand 3 (the first N-terminal strand of Greek key 1), approximately 14 Å from the active site Cu/Zn ion center. Substitution of **1** at this site is therefore not expected to perturb protein structure or activity.[13,14]

Protein expression was performed with yeast strain SCY4 (*MATa*, *ade2-101*, *ura3-1*, *leu2-3*, *112trpI*, *his3-11*, *CYB::Kan* [cir0]) transformed with the hSOD expression plasmid pC1SOD-33TAG-His6 and the respective $tRNA^{Leu5}_{CUA}$/mutant LeuRS encoding plasmid (pLeuRS32 or pLeuRS35).[6] Cultures were grown to saturation at 30 °C in SD media with a Leu and Trp dropout media and diluted into SD media (dextrose, −Leu, −Trp) containing 1 mM of **1** and grown for 18–24 h. Cells were lysed with Y-PER lysis reagent (Pierce)/Complete Protease Inhibitor (-EDTA, Roche) and protein purified on Ni–NTA resin (Qiagen) with 250 mM imidazole eluent. Only in the presence of 1 mM amino acid **1** was full-length protein observed by SDS–PAGE analysis with Coomassie staining (Supplementary data, Fig. S1) indicating that the mutant aaRS is specific for **1**. ESI and MALDI TOF mass spectral analysis of the purified protein revealed a mass of 16,782, in excellent agreement with the calculated mass (16,782) (Fig. 1 and Fig. S2), and additionally some minor peaks are also observed in expression of wildtype (wt) His6-tagged hSOD. Furthermore, ESI analysis indicated that if any leucine is incorporated, it is less than 2% (MW hSOD-33Leu=16,595).

Both LeuRS32 and LeuRS35 afforded approximately 0.25–0.5 mg/L protein; the yield of wt hSOD under the same conditions was 1.0–1.5 mg/L. Sequencing revealed the following mutations: Met40Gly, Leu41Lys, Tyr499Ser,

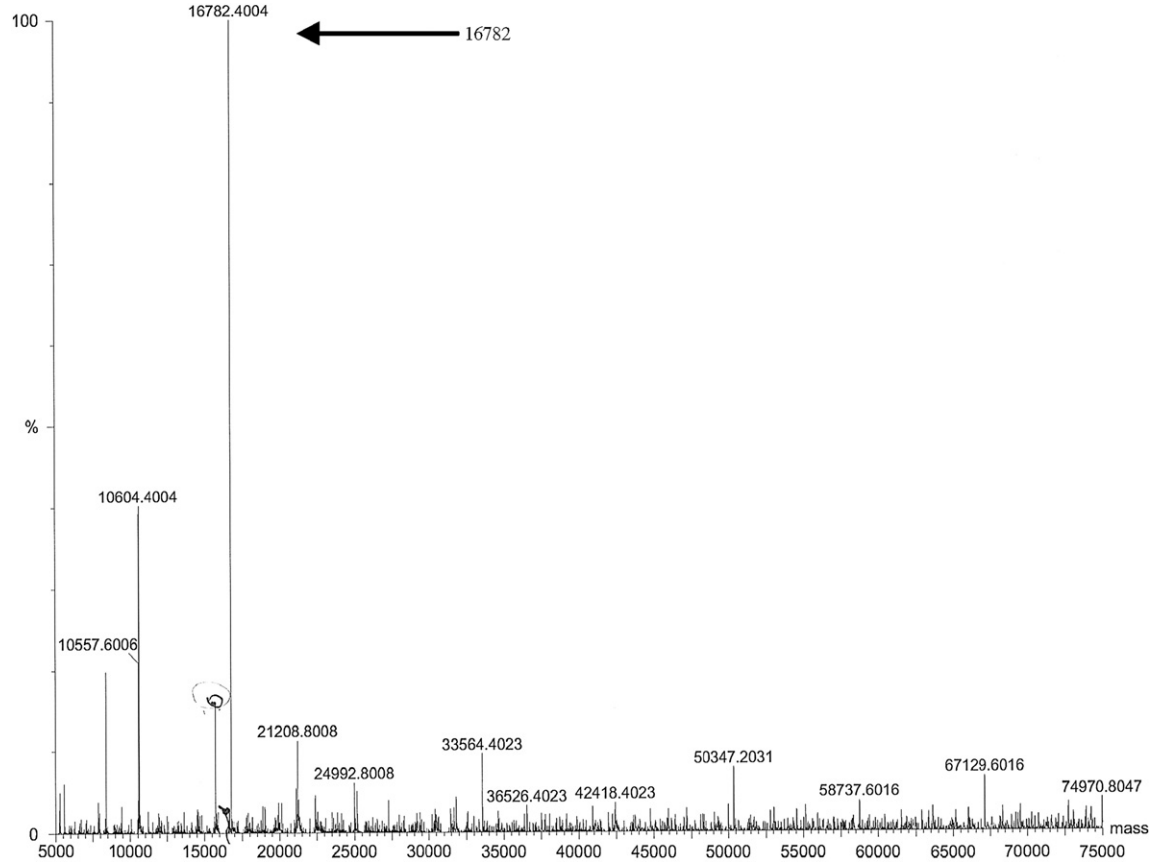


Figure 1. ESI MS shows correct mass (16,782) for hSOD with ferrocenylcysteine **1** incorporated at residue 33.

Tyr527Ala, and His537Gly (LeuRS32); and Met40Gly, Leu41Glu, Tyr499Arg, Tyr527Ile, and His537Gly (LeuRS35). Both clones share Met40Gly and His537Gly mutations, which likely provide space for the bulky metallocene side chain.

3. Conclusion

We have evolved an orthogonal aminoacyl-tRNA synthetase that incorporates **1** with high fidelity and good efficiency. This redox active amino acid should facilitate electron transfer (ET) processes in proteins including ET from electrode surfaces to catalytic sites. In addition, these experiments suggest that it will be possible to genetically encode other metal binding amino acids into proteins for applications in catalysis, imaging, or radiotherapy.

4. Experimental

4.1. General

NMR spectra were recorded on a Bruker 400 MHz instrument. Ultra-pure solvents were purchased from Sigma and dried using the column method of activated alumina and copper catalysts. Reagents were purchased from Acros (oxalyl chloride) or Sigma (Cs_2CO_3, *N*-(*tert*-butoxy)-(L)-cysteine methyl ester, 4 M HCl in dioxane).

4.1.1. Preparation of 1a. Chloromethylferrocene (7 g, 0.03 mol) was prepared as a solution in 150 mL CH_2Cl_2/DMF (1:5). Cs_2CO_3 (15 g) was added followed by 1 equiv of neat *N*-(*tert*-butoxy)-(L)-cysteine methyl ester. The reaction was stirred under nitrogen at 40 °C and the reaction monitored by LC/MS. The reaction was quenched by pouring into 300 mL of 0.1 M HCl. The organic layer was repeatedly washed with water and then dried over $MgSO_4$. The product was purified by flash chromatography on silica (CH_2Cl_2 elution) to yield 5.5 g of **1a**. ^{1}H NMR ($CHCl_3$) δ 1.54 (s, 9H), 2.98 (dd, J=9, 42 Hz, 2H), 3.58, (s, 2H), 3.82 (s, 3H), 4.21 (m, 7H), 4.27 (s, 2H), 4.58 (d, J=13 Hz, 1H), 5.37 (d, J=15 Hz, 1H). ^{13}C NMR ($CHCl_3$) δ 28.77, 32.75, 34.43, 52.94, 53.58, 68.20, 69.32, 80.60, 85.01, 99.81, 155.64, 172.06. HR ESI TOF MS Na^+ calcd $C_{20}H_{27}SNO_4Fe$ 456.0902, found 456.0900.

4.1.2. Preparation of ferrocenylcysteine 1. A solution of protected ferrocenylcysteine **1a** (5 g, 0.012 mol) in 200 mL dry THF was treated with 1 equiv 4 M HCl in dioxane in portions and the reaction heated to 40 °C until LC/MS analysis showed complete deprotection. The pH was adjusted to 8 with 6 M NaOH and the crude product was washed with 150 mL ether. The pH was then adjusted to pH 5 and the product extracted into 150 mL ethyl acetate. The solvent was evaporated and the residue taken up in 40 mL THF/H_2O (10:1), and 6 M LiOH (~1.5 equiv) was added in portions until LC/MS showed complete deprotection. The final product was purified by prep-reverse phase HPLC (AcCN/H_2O 0–50% gradient AcCN) and lyophilised to afford 1.6 g of **1** (45% from **1a**). ^{1}H NMR (DMSO) δ 2.92 (dd, J=12, 21 Hz, 2H), 3.61 (s, 2H), 4.13 (m, 3H), 4.16 (s, 5H), 4.23 (q, J=4 Hz, 2H), 8.42 (br). ^{13}C NMR (DMSO) δ 31.31, 31.45, 51.78, 67.82, 68.64, 84.22, 100.01, 169.80. HR ESI TOF MS H^+ calcd $C_{14}H_{17}O_2NSFe$ 320.0402, found 320.0391.

4.2. General procedure for protein expression and analysis

This information, including the sequence of hSOD used, is available in the Supplementary data.

Acknowledgements

The authors thank Dr. Ansgar Brock for help in acquiring ESI spectra. This work was supported by DOE grant DE-FG02-03ER46051 and the Skaggs Institute for Chemical Biology.

Supplementary data

It includes protein mass spectra and expression analysis. Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.tet.2007.02.125.

References and notes

1. (a) *Bioinorganic Chemistry*; Bertini, I., Gray, H. B., Lippard, S. J., Valentine, S. J., Eds.; University Science Books: Mill Valley, CA, 1994; (b) Holm, R. H.; Kennepohl, P.; Solomon, E. I. *Chem. Rev.* **1996**, *96*, 2239.
2. Arnold, F. H. *Nature* **2001**, *409*, 253.
3. Degani, Y.; Heller, A. *J. Am. Chem. Soc.* **1988**, *110*, 2615.
4. (a) Skov, L. K.; Pascher, T.; Winkler, J. R.; Gray, H. B. *J. Am. Chem. Soc.* **1998**, *120*, 1102; (b) Udit, A. K.; Arnold, F. H.; Gray, H. B. *J. Inorg. Biochem.* **2004**, *98*, 1547.
5. Alfonta, L.; Zhang, Z.; Uryu, S.; Loo, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 14662.
6. van Staveren, D. R.; Metzler-Nolte, N. *Chem. Rev.* **2004**, *104*, 5931.
7. Bellouard, F.; Chuburu, F.; Yaouanc, J.-J.; Handel, H.; Le Mest, Y. *New J. Chem.* **1999**, *23*, 1133.
8. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964.
9. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306.
10. Cusack, S.; Yaremchuk, A.; Tukalo, M. *EMBO J.* **2000**, *19*, 2351.
11. Keegan, L.; Gill, G.; Ptashne, M. *Science* **1986**, *231*, 699.
12. Boeke, J. D.; LaCroute, F.; Fink, G. R. *Mol. Gen. Genet.* **1984**, *197*, 345.
13. DiDonato, M., et al. *J. Mol. Biol.* **2003**, *332*, 601.
14. Assfalg, M.; Banci, L.; Bertini, I.; Turano, P.; Vasos, P. R. *J. Mol. Biol.* **2003**, *330*, 145.