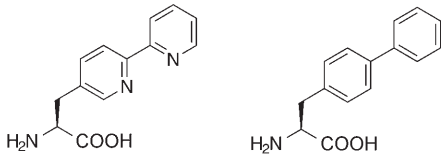
Nonnatural Amino Acids

DOI: 10.1002/anie.200703397

# A Genetically Encoded Bidentate, Metal-Binding Amino Acid**

*Jianming Xie, Wenshe Liu, and Peter G. Schultz**

Metal ions play important roles in the structure and function of many proteins: they serve as important structural elements, regulate biological activity, and act as cofactors in a wide array of catalytic and electron-transfer processes. Accordingly, there is considerable interest in the rational design of metalloproteins.[1,2] However, it remains a challenge to engineer a cluster of precisely oriented inner- and outer-shell residues that selectively coordinate a metal ion at a defined site in a protein. The ability to genetically encode[3,4] multidentate, metal-binding amino acids would simplify the design of metalloproteins by reducing the number of residues required to bind a desired metal ion. One such amino acid (2,2′-bipyridin-5-yl)alanine (BpyAla) bears an *N*,*N*′-bidentate




Bipyridylalanine (BpyAla) Biphenylalanine (BipAla)

side chain that strongly chelates transition-metal ions such as $Fe^{2+/3+}$, $Cu^{2+}$, $Co^{2+/3+}$, and $Ru^{2+/3+}$ and is able to form dimeric or trimeric metal ion complexes that can template the formation of polypeptide oligomers and conjugates.[5] Herein, we demonstrate that BpyAla can be cotranslationally incorporated into proteins in *E. coli* in response to an amber nonsense codon with excellent fidelity and yield. The structural basis for selective recognition of this novel amino acid by its cognate aminoacyl-tRNA synthetase (aaRS) has also been determined by X-ray crystallography.

BpyAla was synthesized by using a previously reported method with minor modifications.[6,7] For the selective incorporation of BpyAla into proteins in *E. coli*, an orthogonal tRNA/aaRS pair was generated from a *Methanococcus jannaschii* amber suppressor tyrosyl-tRNA ($Mj$tRNA$^{Tyr}_{CUA}$)/tyrosyl-tRNA synthetase ($Mj$TyrRS) pair by using previously reported directed evolution methodology.[8,9] Based on the structure of the $Mj$TyrRS/tyrosine complex,[10] two libraries of $Mj$TyrRS active-site mutants were generated in which residues within 6.5 Å of the tyrosine substrate (Figure 1a; library-1: Y32, L65, H70, Q155, D158, and L162; library-2: Y32, L65, F108, Q109, D158, and L162) were randomized by overlap extension polymerase chain reaction using synthetic oligonucleotide primers in which the intended randomization was encoded by NNK (N = A + T + C + G; K = T + G). These libraries were then passed through rounds of alternating positive and negative selections.[8,9] The positive selection is based on resistance to chloramphenicol, which is conferred by suppressing a permissive amber mutation in the chloramphenicol acetyl transferase gene in the presence of 1 mM BpyAla. The negative selection utilizes the toxic barnase gene with amber mutations at two permissive sites and is carried out in the absence of BpyAla. Cells containing $Mj$TyrRS variants that acylate $Mj$tRNA$^{Tyr}_{CUA}$ with BpyAla but not any endogenous amino acids survive both positive and negative selections, whereas cells containing $Mj$TyrRS variants that acylate $Mj$tRNA$^{Tyr}_{CUA}$ with endogenous amino acids express barnase and die in the negative selection. Unfortunately, no desired $Mj$TyrRS mutant was isolated in the first attempt.

To overcome this obstacle, we attempted a two-step evolution strategy in which the same selection scheme was employed first to identity $Mj$TyrRS variants that utilize biphenylalanine (BipAla), a hydrophobic isosteric analogue of BpyAla. Subsequent mutation of the BipAla-specific synthetase might then lead to selective recognition of BpyAla through additional interactions with the pyridine nitrogen atoms. Five rounds of selection afforded 20 clones that survive on chloramphenicol only in the presence of BipAla. DNA sequencing revealed seven unique mutants (Table 1). One of the most active mutants (BipAlaRS1) was further characterized. When the BipAlaRS1/$Mj$tRNA$^{Tyr}_{CUA}$ pair was used to suppress an amber codon at position 7 in Z-domain protein[11] in the presence of 0.2 mM BipAla, full-length Z-domain protein was produced (Figure 1b). Expression in minimal medium yielded 0.9 mg $L^{-1}$ full-length Z-domain. MALDI-TOF mass spectrometry (MS) analysis of the purified protein confirmed the incorporation of BipAla (Figure S1 in the Supporting Information). In contrast, no full-length Z-domain protein was detected when BipAla or BipAlaRS1 was absent, or if BpyAla was substituted for BipAla in the growth medium. These results indicate that the BipAlaRS1/$Mj$tRNA$^{Tyr}_{CUA}$ pair specifies BipAla but not BpyAla or any endogenous amino acids.

Next, a second-generation library of $Mj$TyrRS mutants was generated based on the sequences of the seven BipAlaRS

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

[+] Current address:
Department of Chemistry
Texas A&M University, College Station, TX 77843

[**] We thank Dr. Lubica Supekova, Chang Liu, Eric Brustad, Hyunsoo Lee, and Emily Remba for experimental assistance and helpful discussions. This work is supported by a grant from the Department of Energy ER46051 and the Skaggs Institute for Chemical Biology.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

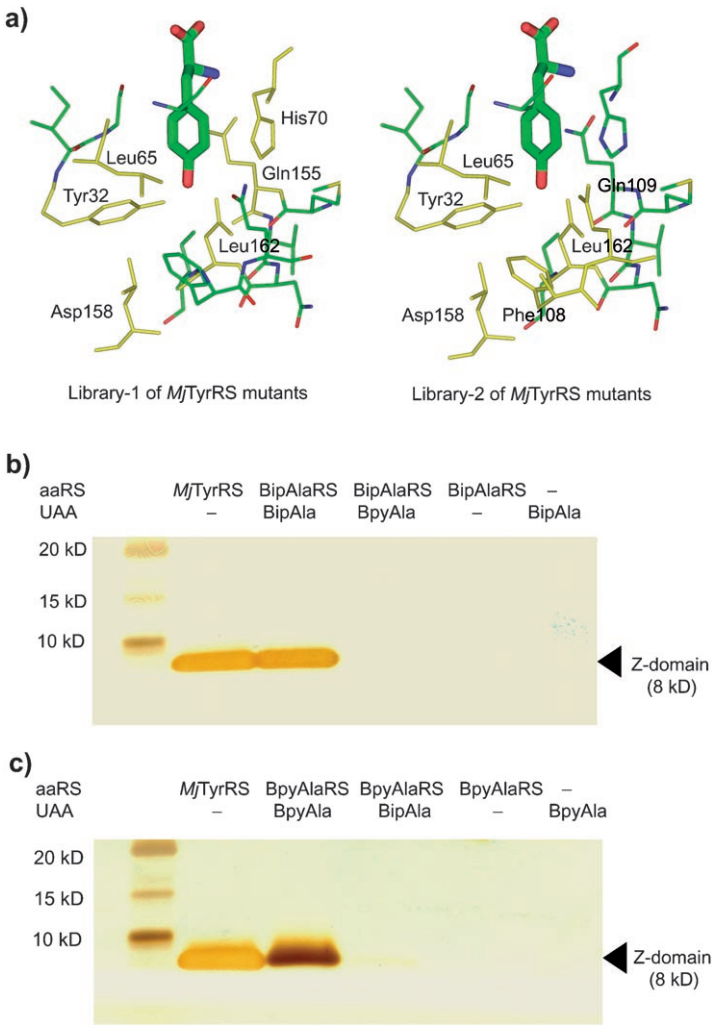



***Figure 1.*** Genetic incorporation of BipAla and BpyAla into proteins in *E. coli*. a) Two libraries of *Mj*TyrRS mutants were generated by randomizing six active-site residues (in yellow). Tyrosine substrate is shown in bold sticks; b) specificity and efficiency of BipAla incorporation by the BipAlaRS1/*Mj*tRNA$^{Tyr}_{CUA}$ pair; and c) specificity and efficiency of BpyAla incorporation by the BpyAlaRS1/*Mj*tRNA$^{Tyr}_{CUA}$ pair. BipAla and BpyAla were incorporated in response to an amber codon at position 7 in Z-domain and analyzed by SDS-PAGE and silver staining. The wild-type protein was expressed as a positive control, in which the *Mj*TyrRS/*Mj*tRNA$^{Tyr}_{CUA}$ pair was used to suppress the amber codon with tyrosine. Proteins were purified by $Ni^{2+}$ affinity chromatography.

variants. D158 was mutated to a mixture of Gly, Ser, and Val; F108 and Q109 were substituted with a mixture of the following dipeptides: TrpMet, AlaAsp, SerLys, ArgGlu, ArgPro, SerHis, and PheGln; and L162 was substituted with a mixture of Lys, Arg, Ser, Leu, Glu, and His. The side chains of residues Y32, L65, H70, Q155, and I159, which line the substrate-binding pocket of *Mj*TyrRS, were then randomized (NNK) to introduce new interactions with the bipyridyl moiety. Subsequent rounds of positive and negative selections with the second-generation library ($10^9$ diversity) in the presence of 1 mM BpyAla afforded 11 viable clones representing 2 unique sequences (Table 2). When the more active BpyAlaRS1 was used to express the amber-mutant Z-domain protein in the presence of 1 mM BpyAla, full-length protein was produced at 1.5 mg $L^{-1}$ from minimal medium (Figure 1c). Specific incorporation of BpyAla was confirmed by MALDI-TOF-MS analysis of the mutant protein (Figure S1 in the Supporting Information). Furthermore, no expressed full-length Z-domain protein was detectable in the absence of either BpyAlaRS1 or BpyAla; substitution of BpyAla with BipAla in the growth medium afforded only a small amount of full-length Z-domain protein. These results indicate a high fidelity for the incorporation of BpyAla into proteins by BpyAlaRS1.

The structural basis for the altered amino acid specificity of BpyAlaRS1 was then determined by X-ray crystallographic analysis (Figure 2). The structure of the BpyAlaRS1/BpyAla complex was solved to a resolution of 1.97 Å, with $R_{cryst} = 0.19$ and $R_{free} = 0.24$ (Table 3). The backbone conformation of BpyAlaRS1 is almost identical to that of the wild-type *Mj*TyrRS. However, the size and electrostatic properties of the active site are altered to favor the binding of BpyAla relative to tyrosine. The BpyAla substrate is bound with dihedral angle of 21° between the two pyridyl rings. The Y32G and D158G mutations remove two hydrogen bonds to the phenolic oxygen atom of bound tyrosine in the wild-type structure, which is consistent with the loss of affinity for tyrosine and similar to mutations observed in other mutant aaRS crystal structures.[12–14] These two mutations, together with H70A and L162S, also markedly expand the active site to accommodate the bulky side chain of BpyAla. Furthermore, an extensive hydrogen-bonding network is formed which involves the two pyridine nitrogen atoms of BpyAla, four water molecules, the side chain amide group of Q109, the backbone carbonyl group of Y151, and the α-amino group of A70. It should be noted that two

***Table 1:*** The 20 selected BipAla-specific *Mj*TyrRS mutants represent 7 unique mutants.

| *Mj*TyrRS | Frequency | Y32 | L65 | H70 | F108 | Q109 | Q155 | D158 | L162 |
|---|---|---|---|---|---|---|---|---|---|
| BipAlaRS1 | 1 clone | H | H | H | W | M | Q | G | K |
| BipAlaRS2 | 8 clones | G | V | H | A | D | Q | S | R |
| BipAlaRS3 | 4 clones | H | V | H | S | K | Q | G | S |
| BipAlaRS4 | 3 clones | A | S | H | R | E | Q | S | S |
| BipAlaRS5 | 2 clones | H | V | H | R | P | Q | G | L |
| BipAlaRS6 | 1 clone | A | V | H | S | H | Q | G | E |
| BipAlaRS7 | 1 clone | L | T | S | F | Q | G | V | H |

***Table 2:*** The 11 selected BpyAla-specific *Mj*TyrRS mutants represent 2 unique mutants.

| *Mj*TyrRS | Frequency | Y32 | L65 | H70 | F108 | Q109 | Q155 | D158 | I159 | L162 |
|---|---|---|---|---|---|---|---|---|---|---|
| BpyAlaRS1 | 10 clones | G | Y | A | F | Q | E | G | W | S |
| BpyAlaRS2 | 1 clone | E | H | H | W | M | Q | G | H | H |

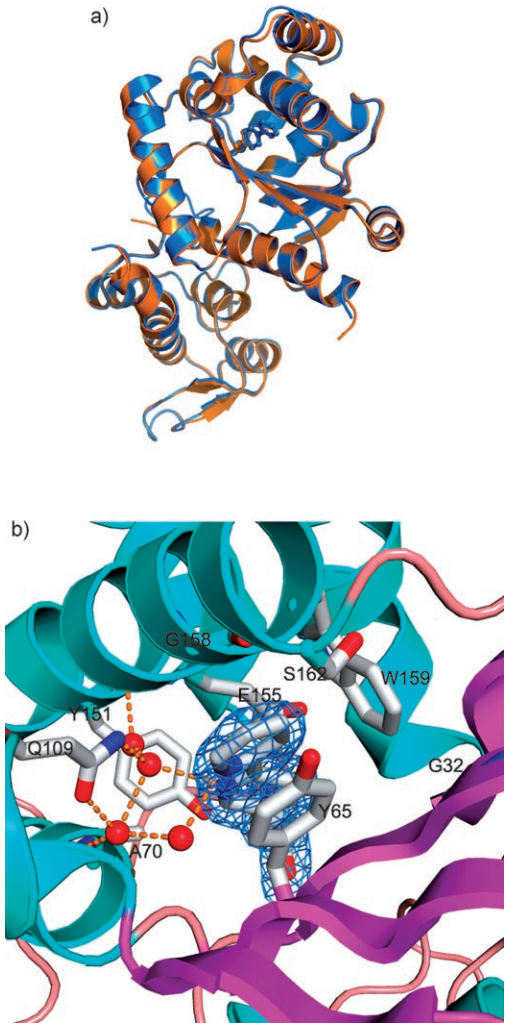



***Figure 2.*** a) Comparison of the overall structures of the *Mj*TyrRS/tyrosine complex (orange) and the BpyAlaRS1/BpyAla complex (blue) reveals identical backbone conformations; b) Mutations in BpyAlaRS1 markedly expand the active site to accommodate BpyAla and lead to a hydrogen-bond network and aromatic interactions that favor the binding of BpyAla. The electron density covering BpyAla is contoured at 1.5 σ. Dashed lines shown in orange are hydrogen bonds. Water molecules are presented in red spheres. The structure of BpyAlaRS1/BpyAla complex is deposited into the Protein Data Bank with PDB entry 2PXH.

***Table 3:*** Data collection and refinement statistics of BpyAlaRS1 bound with BpyAla.

| | |
|---|---|
| unit cell dimensions [Å] | $a=b=100.84$, $c=71.07$ |
| space group | $P4_32_12$ (tetragonal) |
| X-ray radiation wavelength [Å] | 1.0 |
| resolution range [Å] | 71.25–1.97 |
| $R_{sym}$[a] (highest resolution shell) | 0.082 (0.496) |
| $\langle I\rangle/\sigma\langle I\rangle$ | 26.2 |
| no. of reflections | 25063 |
| redundancy | 7.7 |
| completeness [%] (highest shell) | 99.7 (98.2) |
| $R_{cryst}$ ($R_{free}$)[b] | 0.19 (0.24) |
| no. of protein atoms | 2453 |
| no. of heterogenic atoms | 18 |
| no. of water molecules | 278 |
| mean isotropic B-value [Å$^2$] | 30.8 |
| rmsd bond distances [Å] | 0.017 |
| rmsd bond angle [deg.] | 1.69 |

[a] $R_{sym}=\Sigma\Sigma_j|I_j(hkl)-<I(hkl)>|/\Sigma\Sigma_j|I(hkl)|$; $I_j$ is the measured intensity of reflection $j$ and $\langle I\rangle$ is the mean intensity over $j$ reflections. [b] $R_{cryst}=\Sigma||F_{obsd}(hkl)|-|F_{calcd}(hkl)||/\Sigma|F_{obsd}(hkl)|$; $F_{obsd}$ and $F_{calcd}$ are observed and calculated structure factors, respectively. $R_{free}$ is defined in the same way as for $R_{cryst}$, but for 5.0% of the total reflections chosen at random and omitted from structure refinement.

pyridine nitrogen atoms cannot be rigorously assigned from the electron density map, because of the similarity of the carbon and nitrogen atoms. The configuration was therefore assigned on the basis of hydrogen-bonding interactions with the nitrogen lone pairs versus the aryl C–H bonds. Finally, the L65Y and I159W mutations appear to form aromatic interactions with BpyAla. The Y65 phenyl ring is parallel to the second pyridyl ring of BpyAla, likely forming a face-to-face π interaction; the W159 indole ring is perpendicular to the same pyridyl ring, forming an edge-to-face π interaction.[15,16]

Finally, the ability of BpyAla to bind metal ions when incorporated into proteins was assessed. BpyAla and Tyr were substituted for A82, a surface residue of bacteriophage T4 lysozyme. The expressed T4 lysozyme mutants were purified to homogeneity by cation-exchange chromatography and size-exclusion chromatography (Figure S2 in the Supporting Information). Both mutants were verified by MALDI-TOF-MS analysis: A82BpyAla, $M_{exptl}=18758$ Da, $M_{theor}=18756$ Da; A82Y, $M_{exptl}=18693$ Da, $M_{theor}=18694$ Da (Figure S3 in the Supporting Information). The proteins (8.0 nM) in Tris buffer (5.0 mM, pH 7.5) were then mixed with $CuCl_2$ solution (20 nM) and subjected to MS analysis again (Figure S3). Only one major mass peak at $m/z$ 18694 was identified from the sample containing the A82Y mutant, corresponding to the $Cu^{2+}$-free protein. In contrast, two mass peaks ($m/z$ 18755 and 18819) were identified from the sample containing the A82BpyAla mutant, corresponding to $Cu^{2+}$-free and $Cu^{2+}$-bound proteins, respectively. When two other residues (S44 and E128) were chosen for mutation and the resulting mutants were premixed with $CuCl_2$, mass signals corresponding to the $Cu^{2+}$-bound species were also observed only when the residue was mutated to BpyAla, but not when it was mutated to Tyr (Figures S4 and S5 in the Supporting Information). The metalated protein was further characterized by UV/Vis absorption spectroscopic analysis. Upon the addition of $CuCl_2$, the UV/Vis spectrum of the A82BpyAla mutant of T4 lysozyme displayed two new absorption bands at 317 and 304 nm, along with a decreased absorption at 283 nm. This result is consistent with the red shift of the bipyridyl π-π* transition upon chelation of $Cu^{2+}$ ion. For comparison, the A82Y mutant of T4 lysozyme did not show such spectral changes upon addition of $CuCl_2$ (Figure S6 in the Supporting Information).

In conclusion, an orthogonal *Mj*TyrRS/*Mj*tRNA$^{Tyr}_{CUA}$ pair was evolved in a stepwise fashion that allows the selective incorporation of BpyAla into proteins in response to an amber codon in *E. coli* with high fidelity. This two-step selection strategy may be useful for the evolution of aaRSs specific for other structurally diverse amino acids. Also, the

structural basis for recognition of BpyAla by the mutant aaRS was determined by X-ray crystallography; this structure will likely facilitate the design of new aaRS libraries with distinct active sites. Finally, the genetic incorporation of BpyAla and other metal-chelating amino acids into proteins provides a novel approach towards the de novo engineering of metalloproteins with a range of novel structures and functions, including oligomeric polypeptides (e.g., helical bundles[17–19]) or proteins with redox-active, radioactive, or fluorescent metal ion complexes.[20–24]

Received: July 27, 2007
Published online: September 24, 2007

**Keywords:** metalloproteins · N ligands · nonnatural amino acids · protein engineering

[1] W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri, A. Lombardi, *Annu. Rev. Biochem.* **1999**, *68*, 779–819.
[2] Y. Lu, S. M. Berry, T. D. Pfister, *Chem. Rev.* **2001**, *101*, 3047–3080.
[3] L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34–68; *Angew. Chem. Int. Ed.* **2005**, *44*, 34–66.
[4] A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.
[5] A. J. Doerr, G. L. McLendon, *Inorg. Chem.* **2004**, *43*, 7916–7925.
[6] S. A. Savage, A. P. Smith, C. L. Fraser, *J. Org. Chem.* **1998**, *63*, 10048–10051.
[7] B. Imperiali, T. J. Prins, S. L. Fisher, *J. Org. Chem.* **1993**, *58*, 1613–1616.
[8] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500.
[9] J. Xie, P. G. Schultz, *Methods* **2005**, *36*, 227–238.
[10] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nat. Struct. Biol.* **2003**, *10*, 425–432.
[11] Z. Zhang, L. Wang, A. Brock, P. G. Schultz, *Angew. Chem.* **2002**, *114*, 2964–2966; *Angew. Chem. Int. Ed.* **2002**, *41*, 2840–2842.
[12] J. M. Turner, J. Graziano, G. Spraggon, P. G. Schultz, *J. Am. Chem. Soc.* **2005**, *127*, 14976–14977.
[13] J. M. Turner, J. Graziano, G. Spraggon, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6483–6488.
[14] W. Liu, L. Alfonta, A. V. Mack, P. G. Schultz, *Angew. Chem.* **2007**, *119*, 6185–6187; *Angew. Chem. Int. Ed.* **2007**, *46*, 6073–6075.
[15] S. K. Burley, G. A. Petsko, *Science* **1985**, *229*, 23–28.
[16] S. K. Burley, G. A. Petsko, *Adv. Protein Chem.* **1988**, *39*, 125–189.
[17] M. Lieberman, T. Sasaki, *J. Am. Chem. Soc.* **1991**, *113*, 1470–1471.
[18] M. Z. Ghadiri, C. Soares, C. Choi, *J. Am. Chem. Soc.* **1992**, *114*, 825–831.
[19] M. Gochin, R. K. Guy, M. A. Case, *Angew. Chem.* **2003**, *115*, 5483–5486; *Angew. Chem. Int. Ed.* **2003**, *42*, 5325–5328.
[20] D. S. Wuttke, H. B. Gray, S. L. Fisher, B. Imperiali, *J. Am. Chem. Soc.* **1993**, *115*, 8455–8456.
[21] H. K. Rau, N. DeJonge, W. Haehnel, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11526–11531.
[22] L. S. Zuckier, G. L. DeNardo, *Semin. Nucl. Med.* **1997**, *27*, 10–29.
[23] R. T. Maguire, V. L. Pascucci, A. N. Maroli, J. V. Gulfo, *Cancer* **1993**, *72*, 3453–3462.
[24] R. P. Cheng, S. L. Fisher, B. Imperiali, *J. Am. Chem. Soc.* **1996**, *118*, 11349–11356.