Nonnatural Amino Acids

DOI: 10.1002/anie.200701990

Structural Basis for the Recognition of *para*-Benzoyl-L-phenylalanine by Evolved Aminoacyl-tRNA Synthetases**

*Wenshe Liu, Lital Alfonta, Antha V. Mack, and Peter G. Schultz**

A wide array of amino acids with novel chemical and biological properties have been genetically encoded in both prokaryotic and eukaryotic organisms,[1–3] which include the efficient photo-cross-linker *para*-benzoyl-L-phenylalanine (pBpa, Figure 1). Orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs that selectively recognize pBpa have been evolved from both *Methanococcus jannaschii* (*Mj*) and *Escherichia coli* (*Ec*) tyrosyl-tRNA synthetases (TyrRS) in bacteria and yeast, respectively.[4,5] To understand the structural basis for selective recognition of the large benzophenone side chain by these mutant enzymes, we now report the X-ray crystal structure of the mutant *Mj*TyrRS-pBpa complex. A model of the corresponding mutant *Ec*TyrRS-pBpa complex was also generated and shares many features with the *M. jannaschii* structure. In contrast to previous structural studies of evolved aminoacyl-tRNA synthetases, these mutant enzymes bind the relatively large side chain of pBpa in a deep hydrophobic cavity with relatively little change in the polypeptide backbone.

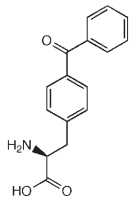

Figure 1. Structure of *para*-benzoyl-L-phenylalanine.

To evolve a pBpa-specific *Mj* aminoacyl-tRNA synthetase (*Mj*pBpaRS) in *E. coli*, a pool of *Mj*TyrRS variants was generated with random mutations at five active-site amino acid residues (Y32, E107, D158, I159, and L162). Alternating rounds of positive and negative selection were then used to identify mutant aaRSs that aminoacylate an amber suppressor $tRNA_{CUA}$ with pBpa, but not with endogenous host amino acids.[6,7] After five rounds of selection, six *Mj*pBpaRSs were identified that showed sequence convergence. In most of these clones, Y32 was mutated to alanine or glycine; D158 was mutated to threonine; I159 was mutated to serine; and L162 was conserved; Y32G/D158T/I159S emerged as the consensus set of mutations. A similar method was used to evolve an *Ec*TyrRS that selectively charges its cognate suppressor $tRNA_{CUA}$ with pBpa in yeast (*Ec*pBpaRS). Positive and negative selections of a library that contains random mutations at five active-site residues (Y37, N126, D182, F183, and L186) afforded two clones selective for pBpa. In both clones, Y37 was mutated to glycine, N126 was conserved, D182 was mutated to glycine, F183 was either conserved or mutated into tyrosine, and L186 was mutated to alanine or methionine. Given the structural homology between archael and bacterial tyrosyl-tRNA synthetases, and the fact that the mutations to each are quite similar, it is likely that *Mj*pBpaRSs and *Ec*pBpaRSs bind the side chain of pBpa in a similar fashion. To address this question, the X-ray crystal structure of the *Mj*pBpaRS (Y32G/D158T/I159S)-pBpa complex was determined, and the corresponding structure of the *Ec*pBpaRS (Y37G/D182G/L186A)-pBpa complex was modeled.

*Mj*pBpaRS was crystallized in the presence of 1 mM pBpa by the hanging-drop vapor-diffusion method. The protein crystals belong to the space group $P4_32_12$ and contain one molecule per asymmetric unit (Table 1). There is one molecule of pBpa per protein bound at the active site as evidenced by the strong Fo–Fc electron density in the active site when pBpa is omitted from structure refinement (Figure 2a). *Mj*pBpaRS is composed of five regions: the Rossmann-fold catalytic domain, the short N-terminal region, the connective polypeptide 1 domain, the C-terminal domain, and the KMSKS loop. Its overall structure is very close to that of the wild-type *Mj*TyrRS-tyrosine complex[8] with only a 0.47 Å root-mean-square (rms) deviation in Cα. The only significant changes are in the KMSKS motif loop from G203 to G210. This loop region is disordered in the wild-type structure, but traceable in *Mj*pBpaRS. The active-site superposition of *Mj*pBpaRS and wild-type *Mj*TyrRS is shown in Figure 2b. The two structures are almost identical to each other with the exception of the side chains of the three mutated residues.

Although the overall structure of *Mj*pBpaRS is very close to wild-type *Mj*TyrRS, their substrate-binding interactions are quite distinct. The three active-site mutations alter hydrogen-bonding and packing interactions with the bound amino acid. In wild-type *Mj*TyrRS, Y32 and D158 form two hydrogen bonds with the phenolic oxygen atom of the tyrosine substrate. These two hydrogen bonds are eliminated in *Mj*pBpaRS by the Y32G and D158T mutations, consistent with the loss in selectivity for tyrosine. The Y32G mutation also deepens the substrate side-chain binding pocket to accommodate the *para*-benzoyl group of pBpa—this substituent occupies the same site as the phenyl group of Y32 in wild-type *Mj*TyrRS (Figure 2b). In the mutant *Mj*pBpaRS

[*] Dr. W. Liu, Dr. L. Alfonta, A. V. Mack, 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

[**] Part of this work is based on experiments conducted at the Advanced Light Source (ALS). The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Material Sciences Division of the US Department of Energy under Contract DE-AC03-76SF00098 at the Lawrence Berkeley National Laboratory. This work is supported by grants from the National Institutes of Health (GM62159) to P.G.S., and this is manuscript 18906 of The Scripps Research Institute.

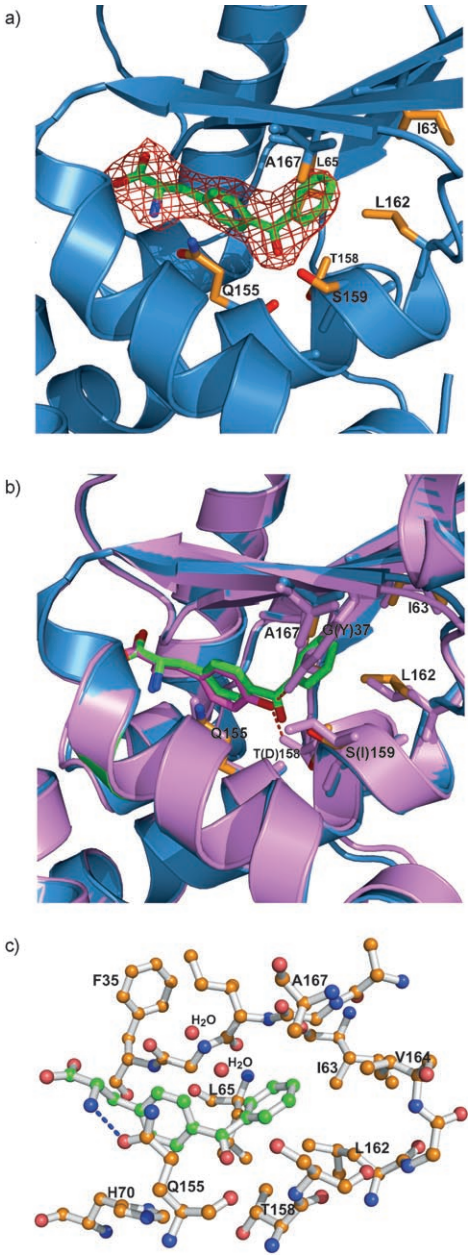

Figure 2. a) Structure of *Mj*pBpaRS with pBpa bound in the active site; the Fo–Fc electron density map of pBpa was contoured at 3σ. b) Superposition of the active sites of *Mj*TyrRS (purple) and *Mj*pBpaRS (blue); amino acids in wild-type *Mj*TyrRS that are mutated in *Mj*pBpaRS are shown in parentheses; 2 dashed lines show hydrogen bonds formed between the tyrosine phenolic oxygen atom and the side chains of Y32 and D158 in *Mj*TyrRS. c) Amino acid residues in the active site of *Mj*pBpaRS. For the stick or ball-and-stick representations of amino acid residues of *Mj*pBpaRS, N: blue, O: red, and C: orange, and their corresponding amino acids in wild-type *Mj*TyrRS are in the same color as the backbone. Carbon atoms in pBpa are shown in green.

structure (Figure 2c), the side chains of I63, L65, S159, L162, V164, and A167, and main-chain groups from G32 to G34 form a deep hydrophobic pocket that selectively binds the large pBpa side chain; no other apparent significant interactions are indicated by the structure. The benzoyl oxygen

Table 1: Data collection and refinement statistics of the MjpBpaRS-pBpa complex.[a]

unit-cell dimensions [Å]	$a=b=101.6$, $c=71.5$
space group	$P4_32_12$
X-ray radiation wavelength [Å]	1.0
resolution range [Å]	71.8–2.50
R_{sym}[b] (highest resolution shell)	0.098 (0.268)
$\langle I\rangle/\sigma\langle I\rangle$	47.0
number of reflections	13361
redundancy	7.7
completeness [%] (highest shell)	99.3 (99.9)
R_{cryst} (R_{free})[c]	0.21 (0.27)
number of protein atoms	2450
number of heterogenic atoms	20
number of water molecules	132
mean isotropic B value [Å^2]	46.5
rms bond distances [Å]	0.0065
rms bond angle [°]	1.25

[a] The coordinates have been deposited in the Protein Data Bank as entry 2HGZ. [b] $R_{sym}=\Sigma\Sigma_j|I_j(hkl)-\langle I(hkl)\rangle|/\Sigma\Sigma_j|I(hkl)|$, where I_j is the measured intensity of reflection j and $\langle I\rangle$ is the mean intensity over j reflections. [c] $R_{cryst}=\Sigma||F_{obs}(hkl)|-|F_{calc}(hkl)||/\Sigma|F_{obs}(hkl)|$, where F_{obs} and F_{calc} are observed and calculated structure factors, respectively. R_{free} is defined the same as that for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from structure refinement.

atom of pBpa is within van der Waals distance of the D158T side chain, and the γ-carbon and carbonyl oxygen atoms of Q155. Given the distance (2.96 Å) between the benzoyl carbonyl group and the main-chain carbonyl group of Q155, a strong dipole–dipole repulsion is likely; however, this repulsive interaction must be compensated by other attractive interactions formed between the benzoyl group of pBpa and surrounding protein residues. The side-chain oxygen atom of D158T is involved in two hydrogen-bonding interactions with the phenolic oxygen atom of Y114 and the amide nitrogen atom of Q109. These two hydrogen bonds are also made by a side-chain oxygen atom of D158 in wild-type *Mj*TyrRS, which suggests that the D158T mutation likely maintains structural stability in the mutant enzyme. The I159S mutation may minimize a possible steric clash between the side chain of isoleucine and the *para*-benzoyl group of pBpa.

Given the similarity between the mutations in *Ec*pBpaRS and *Mj*pBpaRS, and the homology between the two enzymes, molecular docking studies of *Ec*pBpaRS-pBpa complex were carried out. The three-dimensional coordinates of *Ec*TyrRS were obtained from the PDB (code 1X8X).[9] Mutation of Y37, D182, and L186 in *Ec*pBpaRS was performed using the program PyMol and all bound water molecules and the heteroatoms were removed. The AutoDock 4.0 program was used to perform automated molecular docking,[10] and the Lamarckian genetic algorithm was used to model the interaction between *Ec*pBpaRS and pBpa. The simulated model in which pBpa has the lowest binding free energy and high geometric quality was then used to perform energy minimization and used for structure analysis.

Comparison of the docking model of *Ec*pBpaRS (Figure 3a) with the crystal structure of *Mj*pBpaRS indicates that the substrate specificity in both enzymes arises from similar interactions with pBpa. The main chain groups of G37, C38, and G39 and side chains of V69, L71, F183, A186, and N195 in

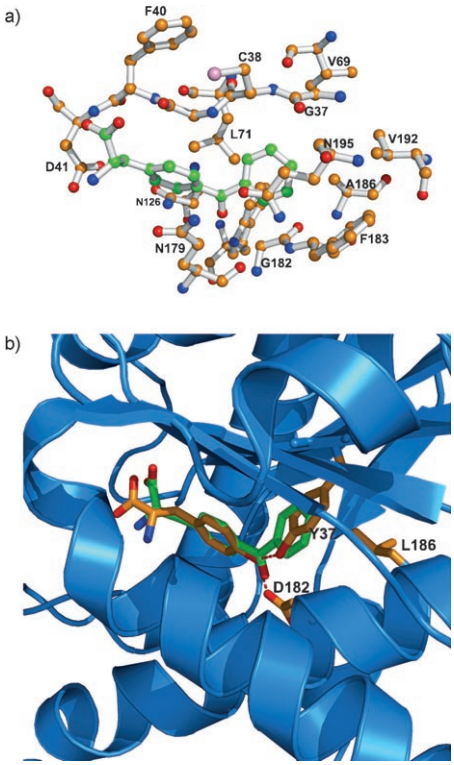

Figure 3. a) Ball-and-stick representation of amino acid residues in the active site of *Ec*pBpaRS. b) The location of pBpa in *Ec*pBpaRS relative to wild-type *Ec*TyrRS. The 3 amino acid residues of *Ec*TyrRS that are mutated in *Ec*pBpaRS and tyrosine are shown as a stick diagram. Two dashed lines represent hydrogen bonds formed between the substrate tyrosine phenolic oxygen atom and the side chains of Y37 and D182. For the stick and ball-and-stick representations, N: blue, O: red, C: orange, and S: purple. Carbon atoms in pBpa are shown in green.

*Ec*pBpaRS form a deep hydrophobic pocket that binds the large side chain of pBpa. Again the two hydrogen bonds to the substrate phenolic oxygen atom (from the side chains of Y37 and D182) are removed. Similarly, Tyr37 is mutated into glycine, and the benzoyl substituent of pBpa occupies the position of the phenyl group of Y37, (similar to the Y32G mutation in *Mj*pBpaRS). Finally, the benzoyl carbonyl group of pBpa is again involved in a repulsive interaction with the carbonyl group of Q179 which must be compensated by other van der Waals attractive interactions between the benzoyl group of pBpa and surrounding protein residues.

In both mutant enzymes, no significant changes in the polypeptide backbone are required to accommodate the large pBpa substrate. This is in contrast to the structures of three substrate-bound mutant *Mj*TyrRSs published previously (specific for naphthylalanine, *para*-acetyl-L-phenylalanine and *para*-bromo-L-phenylalanine),[11,12] and a recently determined *para*-trifluoromethyl-L-phenylalanyl-tRNA synthetase structure.[13] In all four structures, an alpha helix adjacent to the active site is disrupted. Altered substrate specificity is due to both side-chain and backbone rearrangements within the active site that modify hydrogen bonds and packing interactions with substrate. This is a somewhat surprising result since, among the mutant aaRS structures that have been determined to date, *Mj*pBpaRS has the largest side-chain binding pocket. In aggregate, these studies indicate that aminoacyl-tRNA synthetase active sites have a remarkable ability to accommodate structurally diverse substrates, but suggest that the mutations may be difficult to predict computationally. Finally, it is likely that additional libraries can be generated based on the *Mj*pBpaRS structure which further enlarge or deepen the side-chain binding pocket to accommodate amino acids with even larger side chains (for example, lipids, spin labels, fluorophores, and so on).

Experimental Section

The expression and purification of *Mj*pBpaRS are similar to methods described for *para*-acetyl-L-phenylalanyl-tRNA synthetase with slight modifications.[11] *Mj*pBpaRS crystals were grown by the hanging-drop vapor-diffusion method with a 1:1 mixture of around 15 mg mL^{-1} protein and mother liquor containing 5% PEG8K, 8–20% PEG300, 9% glycerol, 1 mM pBpa, and 100 mM Tris-HCl at pH 7.5–9.0. X-ray diffraction data were collected at beamline 5.0.3 of ALS at a wavelength of 1.0 Å. The data were reduced and scaled using HKL2000. The structure was solved by molecular replacement using *Mj*TyrRS (Protein Data Bank entry 1J1U) as a search model and processed by Amore.[14] The following structure refinement was carried out using the combination of the Refmac5 and CNS program suites.[14,15]

Received: May 4, 2007
Published online: July 12, 2007

Keywords: amino acids · enzymes · hydrophobic interactions · photo-cross-linker · protein structures

[1] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225.
[2] J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775.
[3] W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, *Nat. Methods* **2007**, *4*, 239.
[4] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883.
[5] J. W. Chin, T. A. Cropp, S. Chu, E. Meggers, P. G. Schultz, *Chem. Biol.* **2003**, *10*, 511.
[6] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
[7] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020.
[8] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nat. Struct. Biol.* **2003**, *10*, 425.
[9] T. Kobayashi, T. Takimura, R. Sekine, V. P. Kelly, K. Kamata, K. Sakamoto, S. Nishimura, S. Yokoyama, *J. Mol. Biol.* **2005**, *346*, 105.
[10] D. S. Goodsell, G. M. Morris, A. J. Olson, *J. Mol. Recognit.* **1996**, *9*, 1.
[11] J. M. Turner, J. Graziano, G. Spraggon, P. G. Schultz, *J. Am. Chem. Soc.* **2005**, *127*, 14976.
[12] J. M. Turner, J. Graziano, G. Spraggon, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6483.
[13] W. Liu, P. G. Schultz, unpublished data.
[14] Collaborative Computational Project, Number 4, *Acta Crystallogr. Sect. D* **1994**, *50*, 760.
[15] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. Sect. D* **1998**, *54*, 905.

