**Supplemental Information**

# Substrate-mediated fidelity mechanism ensures accurate decoding of proline codons

Byung Ran So[1,4], Songon An[5,‡], Sandeep Kumar[1,4], Mom Das[1,3,4], Daniel A. Turner[1], Christopher M. Hadad[1], and Karin Musier-Forsyth[1,2,3,4]

Departments of [1]Chemistry and [2]Biochemistry, [3]The Ohio State Biochemistry Program, [4]Center for RNA Biology, The Ohio State University, Columbus, OH 43210; [5]Department of Chemistry, University of Minnesota, Minneapolis, MN 55455

Correspondence should be addressed to: Karin Musier-Forsyth, Department of Chemistry, The Ohio State University, 100 West 18th Ave., Columbus, OH 43210, USA

Tel.: 614-292-2021; Fax: 614-688-5402; E-mail: musier@chemistry.ohio-state.edu

‡Current address: Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA

## SUPPLEMENTAL METHODS

**Protein and tRNA preparation.** All tRNAs used in this study were prepared by *in vitro* transcription as described (1,2). His-tagged *E. coli* ProRS (3), CysRS (4), WT AlaRS (5), C666A/Q584H AlaRS (AlaRS-CQ) (5) and *E. coli* tRNA nucleotidyltransferase (NTase) (6) were purified using the Talon cobalt affinity resin (Clontech) as previously described. *H. influenzae* YbaK protein was purified from *E. coli* B834 cells containing plasmid pCYB2_HI1434 using the IMPACT™ I system (New England Biolabs) as described (7). All of the YbaK point mutants were created from plasmid pET15b_HI1434 encoding the WT *H. influenzae* YbaK protein using the QuikChange site-directed mutagenesis kit (Stratagene). All YbaK mutants were overexpressed and purified as described previously (8). The concentrations of *E. coli* AlaRS-CQ, tRNA NTase and all YbaK proteins were determined by the Bradford assay (9). The concentrations of *E. coli* ProRS, AlaRS and CysRS were determined by active site titration (10).

**3'-End-modified tRNA preparation.** To prepare 3'-deoxy-A76 tRNA variants, the tRNA NTase-catalyzed nucleotide exchange reaction was performed in 20 mM Tris-HCl (pH 8.5), 20 mM $MgCl_2$, 1 mM sodium pyrophosphate, 5 mM ATP analogs (2'-H-ATP and 3'-H-ATP), 15 μM *E. coli* $tRNA^{Cys}$, and 5 μM tRNA NTase. After a 4 hr incubation at 37 °C, the reactions were subjected to phenol:chloroform extraction and ethanol precipitation. To remove residual amounts of the ATP analogs, dialysis against diethyl pyrocarbonate (DEPC)-treated water was carried out at room temperature. To eliminate the possibility of residual WT tRNA contamination, the sample was oxidized using 4 mM $NaIO_4$ in 50 mM sodium acetate (pH 5.0) for 1 hr at room temperature as described previously (11). The sample was dialyzed against DEPC-treated water and the final product containing the modified tRNAs was obtained by ethanol precipitation. Aminoacylation reactions were performed as described (2) to compare charging levels of deoxy-tRNAs to those of WT $tRNA^{Cys}$ and to optimize conditions for Cys-tRNA preparation. 3'-[$^{32}$P]-labeled tRNA$^{Pro}$ for aminoacylation assays was prepared as previously described (12).

**Visualization of reaction products of YbaK-catalyzed deacylation of [$^{35}$S]-Cys-tRNA$^{Pro}$.** To view the reaction products, some [$^{35}$S]-Cys-tRNA$^{Pro}$ deacylation assays were analyzed using cellulose TLC according to the published conditions (13)**.** Briefly, reactions containing 0.3 μM [$^{35}$S]-Cys-tRNA$^{Pro}$ in deacylation buffer were initiated with 0.1 μM WT YbaK. At each time point, 2 μL of reaction mixture was quenched with 8 μL of 6 mM dithiothreitol in ethanol at 4 $^{o}$C**.** The quenched mixture (1 μL) was spotted on a cellulose TLC plate pre-run with water, developed in butanol: acetic acid: water (4:1:1 (v/v/v)), and exposed on phosphor-screens for 8-12 hrs. Screens were scanned using a Typhoon phosphorimager (GE healthcare). Deacylation assays for ESI-MS and LC/MS analysis were performed in similar fashion except 50 μM non-radioactive Cys-tRNA$^{Pro}$ and 10 μM YbaK were used.

**ESI-MS and LC/MS analysis.** High-resolution ESI-MS was performed on a Bruker BioTOF II mass spectrometer (University of Minnesota). Crude reaction mixtures were mixed with equal volumes of methanol and data were recorded in the ESI mode. LC/MS was performed using multimode ESI-APCI ionization on an Agilent Technologies 110 series LC equipped with G1956B LC/MSD SL mass selective detector (University of Minnesota). A C8 column (150 mm x 5 μm) was used and the solvent elution conditions were as follows: (a) water/MeOH (0.1% formic acid) (20:80) isocratic elution or (b) water/MeOH (0.1% formic acid) (0-100%) linear gradient over 20 min using a 0.5 mL/min flow rate.

**Computational analysis of cysteine and selenocysteine ring strain.** The homodesmotic model (14) was utilized to accurately calculate the difference in ring strain between cysteine thiolactone and selenocysteine selenolactone. Unlike isodesmic calculations, the homodesmotic model accounts for the hybridization of each carbon atom on both sides of the reaction, as well as the same number of hydrogen atoms attached to the carbon. Additionally, the lactone analog was computed to explore various trends in the three group VIA elements. Geometry optimizations were performed on all the homodesmotic model structures using both density functional theory (B3LYP/6-311+G(2df,2pd)) and the Complete Basis Set method (CBS-QB3). The Gaussian 09 (15) software package was chosen for the geometry optimizations. The frequencies of all optimized structures were computed and verified to have no imaginary frequencies.

## SUPPLEMENTAL RESULTS

### Effect of mutations and non-cognate substrates on binding of YbaK to CCA-Cys

To determine the factors controlling substrate binding by YbaK, we calculated the free energy of binding of YbaK and several mutants to 5′-CCA-Cys (cognate), as well as binding of WT YbaK to non-cognate CCA-Ala, CCA-Pro, and CCA-Ser substrates, employing the molecular mechanics Poisson−Boltzmann surface area (MM-PBSA) method. All *in silico* mutants of YbaK, as well as the non-cognate substrates, were obtained by manually modifying the structure of the YbaK:CCA-Cys complex, followed by energy minimization. These structures were subjected to 15 ns of MD simulation in explicit solvent and 250 snapshots of the protein, ligand, and the complex (without solvent molecules) were extracted at equal intervals. The total energy of each of the components was calculated in implicit solvent using the MM-PBSA module in AMBER 9. The free energy of binding was calculated as the difference between the total energy of the complex and that of the free ligand and the protein.

Our results show that while the binding free energy of the CCA-Pro substrate is moderately reduced compared to the CCA-Cys substrate due to the loss of H-bonding interactions of the amine group, the overall orientation of the CCA-Pro is preserved. In case of CCA-Ala and CCA-Ser, the binding free energy is similar to that of the cognate substrate (**Supplemental Table 2**). This implies that the CCA-Pro substrate can bind YbaK with an orientation similar to that of CCA-Cys, albeit with less affinity, whereas CCA-Ala and CCA-Ser substrates are not discriminated at the level of binding. These results provide further support for the hypothesis that deacylation by YbaK involves cysteine thiol chemistry. We also performed similar MD and MM-PBSA analyses for several YbaK mutants. As expected, the K46A mutation reduced the CCA-Cys binding free energy by 5-fold and perturbed the substrate orientation significantly, whereas CCA-Cys binding to T47A, S129A, and Y20A mutants remained relatively unperturbed (**Supplemental Table 2**). These data support the role of Lys46 in substrate binding and the role of Tyr20, Thr47, and Ser129 in subsequent catalysis.

**Computational analysis of cysteine and selenocysteine ring strain**

Using the homodesmotic model, the ring strain was calculated for the cysteine thiolactone, selenocysteine selenolactone and lactone analog. For both B3LYP/6-311+G(2df,2pd) and CBS-QB3 calculations, the selenolactone was calculated to have the least amount of ring strain, followed by thiolactone and lactone. Using B3LYP, the difference in ring strain energy for selenocysteine when compared to cysteine was 2.8 kcal/mol and was 3.6 kcal/mol with CBS-QB3 (**Supplemental Table 3**). Analysis of the lactone, thiolactone and selenolactone geometries shows that as the size of the group VIA elements increases, the bond length between the two carbon atoms and the group VIA element increases, which increases the angle opposite to the group VIA element. This increase in angle relieves some ring strain and can explain the difference between the thiolactone and selenolactone (**Supplemental Table 4**).

**Supplemental Table 1. Relative deacylation activities of YbaK against Ala-, Ser and Pro-$tRNA^{Pro}$ substrates in the presence of high concentrations of *H. influenzae* YbaK.**

| Substrate | Ala-$tRNA^{Pro}$ | | Ser-$tRNA^{Pro}$ | | Pro-$tRNA^{Pro}$ | |
|---|---|---|---|---|---|---|
| Mutant | Relative activity | S.D. | Relative activity | S.D. | Relative activity | S.D. |
| WT | 1.00 | 0.12 | 1.00 | 0.06 | 1.00 | 0.03 |
| Y20A | 0.48 | 0.04 | 0.26 | 0.13 | 0.20 | 0.13 |
| F29A | 0.09 | 0.10 | 0.02 | 0.08 | 0.08 | 0.02 |
| K46A | 0.10 | 0.12 | 0.18 | 0.08 | 0.17 | 0.06 |
| K46R | -0.04 | 0.05 | 0.20 | 0.02 | 0.27 | 0.04 |
| K46I | 0.35 | 0.1 | 0.41 | 0.05 | 0.39 | 0.07 |
| T47A | 0.12 | 0.06 | 0.16 | 0.09 | 0.30 | 0.05 |
| G101A | 0.12 | 0.05 | 0.36 | 0.25 | 0.21 | 0.06 |
| S129A | 0.04 | 0.03 | 0.05 | 0.02 | 0.21 | 0.03 |
| G131A | 0.49 | 0.08 | 0.80 | 0.28 | 0.56 | 0.17 |
| G134A | 1.04 | 0.22 | 0.91 | 0.21 | 1.10 | 0.02 |
| S136A | 0.29 | 0.05 | 0.20 | 0.22 | 0.70 | 0.32 |
| S136H | 0.30 | 0.06 | 1.12 | 0.11 | 0.45 | 0.05 |

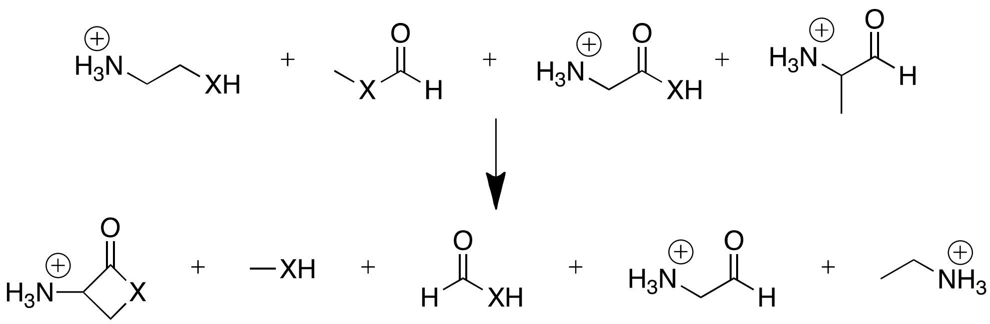
Ala-, Ser- and Pro-$tRNA^{Pro}$ (1.0 μM) deacylation by YbaK variants (21 μM) was performed at room temperature. The $k_{obs}$ ($min^{-1}$) values were calculated by fitting the percent Cys-$tRNA^{Pro}$ remaining as a function of time to a linear equation, $y = k_{obs} \times t + A$. Relative activity was calculated by comparing the $k_{obs}$ of deacylation with that of WT YbaK, which was set to 1.0 for each substrate tested. Results are the average of three trials with the standard deviation (S.D.) indicated. Negative values indicate substrates were protected from hydrolysis. WT YbaK deacylates Ala-, Ser-, and Pro-$tRNA^{Pro}$ with 240, 572 and 466-fold reduced efficiency relative to Cys-$tRNA^{Pro}$.

**Supplemental Table 2. Substrate binding energies of WT and mutant *H. influenzae* YbaK to CCA-AA substrates calculated using MM-PBSA.**

| YbaK | Ligand | Binding energy (kcal/mol) |
|---|---|---|
| WT | CCA-Cys | -47.15 ± 7.73 |
| | CCA-Ala | -47.09 ± 6.00 |
| | CCA-Pro | -19.88 ± 10.06 |
| | CCA-Ser | -36.87 ± 6.82 |
| Y20A | CCA-Cys | -43.31 ± 7.48 |
| K46A | CCA-Cys | -9.47 ± 6.37 |
| T47A | CCA-Cys | -41.06 ± 6.27 |
| S129A | CCA-Cys | -39.72 ± 6.36 |

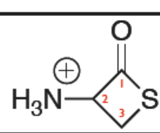
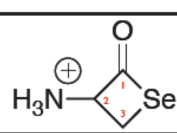
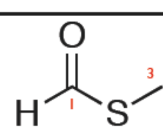
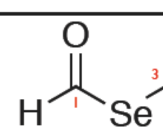
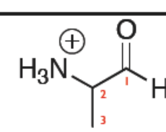
All *in silico* mutants of YbaK, as well as the non-cognate substrates, were obtained by manually modifying the structure of the YbaK:CCA-Cys complex, followed by energy minimization. These structures were subjected to 15 ns of MD simulation in explicit solvent and 250 snapshots of the protein, ligand and the complex (without solvent molecules) were extracted at equal intervals. The total energy of each of the components was calculated in implicit solvent using the MM-PBSA module in AMBER 9. The free energy of binding was calculated as the difference between the total energy of the complex and that of the free ligand and the protein.

**Supplemental Table 3. Calculated ring strain energy values (in kcal/mol) for cysteine, selenocysteine and the oxygen analog using the homodesmotic model (shown below, where X=O, S, Se).**

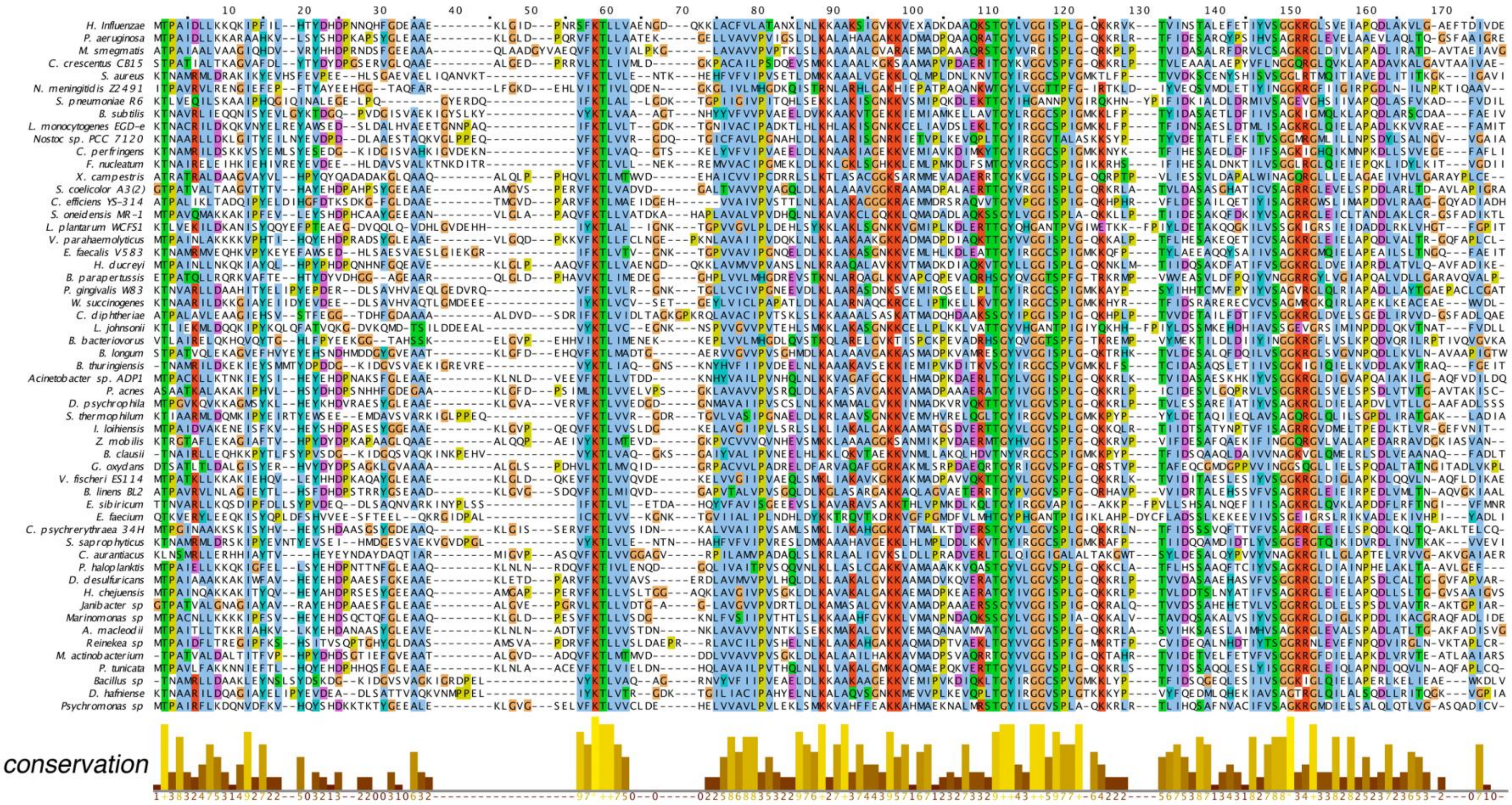


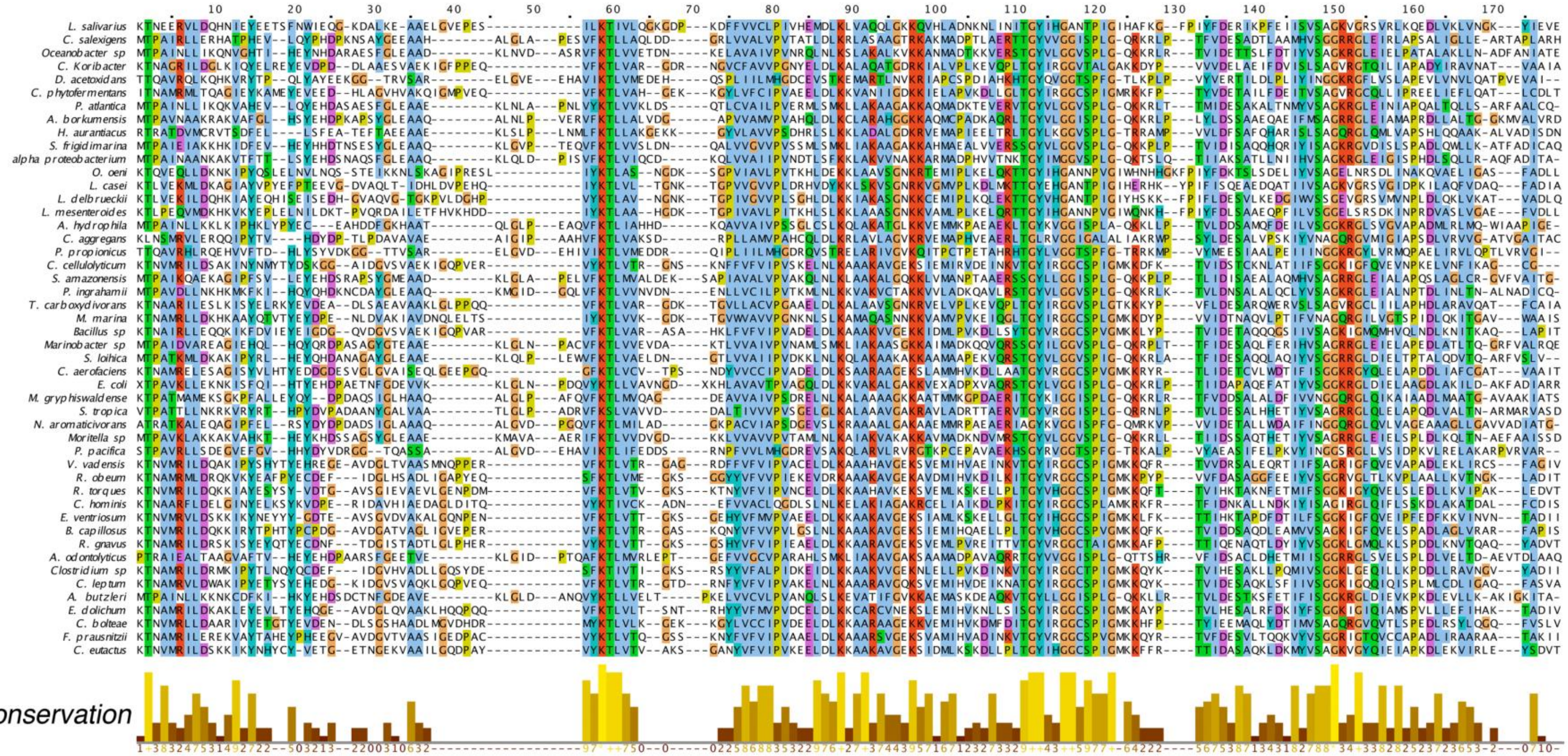
| Homodesmotic Model Compound | B3LYP/ 6-311+G(2df,2pd) | CBS-QB3 |
|---|---|---|
| lactone (X=O) | 36.4 | 36.8 |
| cysteine thiolactone (X=S) | 20.4 | 23.6 |
| selenocysteine selenolactone (X=Se) | 17.6 | 20.0 |

**Supplemental Table 4. Geometric comparison of CBS-QB3 optimized $NH_3^+$ based thio- and selenolactones (bond lengths in Å, angles in degrees).**

| Parameters |  |  |  |  |  |
|---|---|---|---|---|---|
| C3-S(Se) | 1.86 | 2.01 | 1.82 | 1.96 | |
| C1-S(Se) | 1.78 | 1.94 | 1.78 | 1.94 | |
| C1-S(Se)-C3 | 77.2 | 72.6 | 99.2 | 95.7 | |
| C1-C2-C3 | 95.6 | 99.2 | | | **113.0** |

**Supplemental Figure 1. Multiple sequence alignment of YbaK subfamily using CLUSTALW** (16)

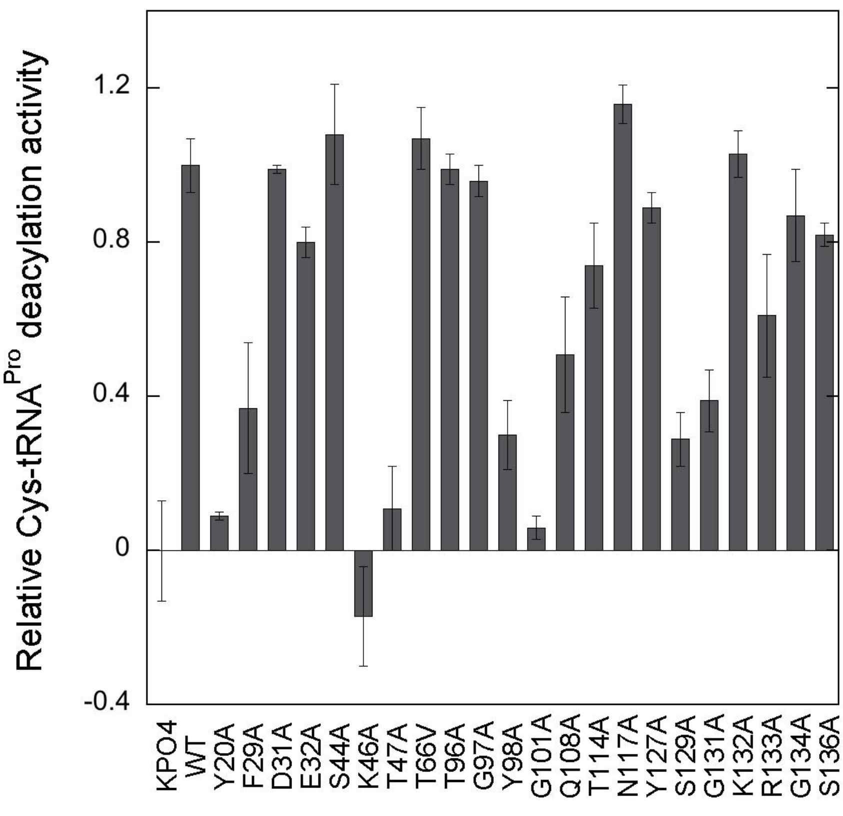



L. salivarius
C. salexigens
Oceanobacter sp
C. Koribacter
D. acetoxidans
C. phytofermentans
P. atlantica
A. borkumensis
H. aurantiacus
S. frigidimarina
alpha proteobacterium
O. oeni
L. casei
L. delbrueckii
L. mesenteroides
A. hydrophila
C. aggregans
P. propionicus
C. cellulolyticum
S. amazonensis
P. ingrahamii
T. carboxydivorans
M. marina
Bacillus sp
Marinobacter sp
S. loihica
C. aerofaciens
E. coli
M. gryphiswaldense
S. tropica
N. aromaticivorans
Moritella sp
P. pacifica
V. vadensis
R. obeum
R. torques
C. hominis
E. ventriosum
B. capillosus
R. gnavus
A. odontolyticus
Clostridium sp
C. leptum
A. butzleri
E. dolichum
C. bolteae
F. prausnitzii
C. eutactus
conservation

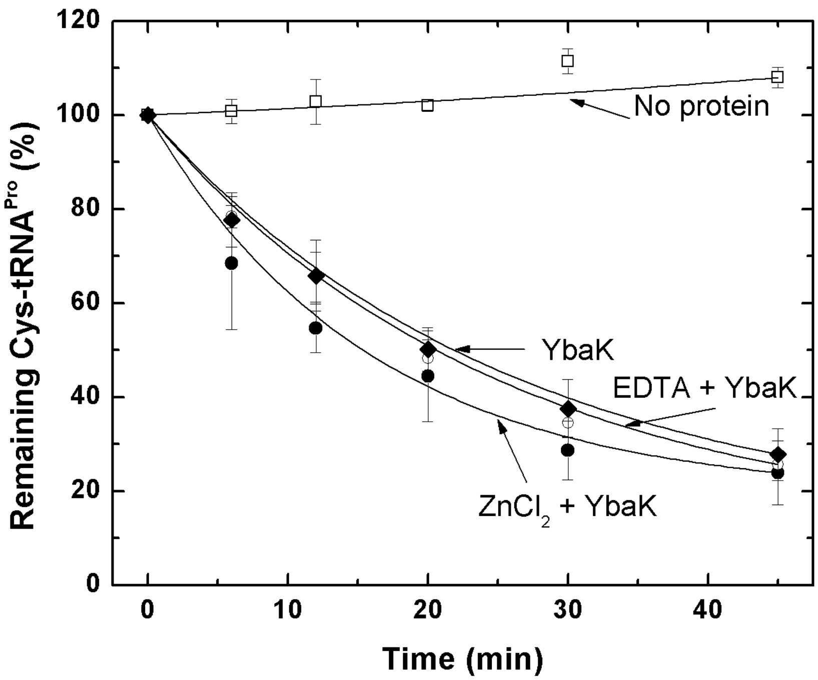
The abbreviations used are:

*H. Influenzae Hi1434, Haemophilus influenzae; P. aeruginosa, Pseudomonas aeruginosa; M. smegmatis, Mycobacterium smegmatis; C. crescentus CB15, Caulobacter crescentus CB15; S. aureus, Staphylococcus aureus; N. meningitidis Z2491, Neisseria meningitidis Z2491; S. pneumoniae R6, Streptococcus pneumoniae R6; B. subtilis, Bacillus subtilis; L. monocytogenes EGD-e, Listeria monocytogenes EGD-e; C. perfringens, Clostridium perfringens str. 13; F. nucleatum, Fusobacterium nucleatum; X. campestris, Xanthomonas campestris; S. coelicolor A3(2), Streptomyces coelicolor A3(2); C. efficiens YS-314, Corynebacterium efficiens YS-314; S. oneidensis MR-1, Shewanella oneidensis MR-1; L. plantarum WCFS1, Lactobacillus plantarum WCFS1; V. parahaemolyticus, Vibrio parahaemolyticus; E. faecalis V583, Enterococcus faecalis V583; H. ducreyi, Haemophilus ducreyi; B. parapertussis, Bordetella parapertussis; Porphyromonas gingivalis W83, Porphyromonas gingivalis W83; W. succinogenes, Wolinella succinogenes DSM 1740; C. diphtheriae, Corynebacterium diphtheriae; L. johnsonii, Lactobacillus johnsonii; B. bacteriovorus, Bdellovibrio bacteriovorus; B. longum, Bifidobacterium longum; B. thuringiensis, Bacillus thuringiensis; P. acnes, Propionibacterium acnes; D. psychrophila, Desulfotalea psychrophila; S. thermophilum, Symbiobacterium thermophilum; I. loihiensis, Idiomarina loihiensis; Z. mobilis, Zymomonas mobilis; B. clausii, Bacillus clausii; G. oxydans, Gluconobacter oxydans; V. fischeri ES114, Vibrio fischeri ES114; B. linens BL2, Brevibacterium linens BL2; E. sibiricum, Exiguobacterium sibiricum; E. faecium, Enterococcus faecium; C. psychrerythraea 34H, Colwellia psychrerythraea 34H; S. saprophyticus, Staphylococcus saprophyticus; C. aurantiacus, Chloroflexus aurantiacus; P. haloplanktis, Pseudoalteromonas haloplanktis; D. desulfuricans, Desulfovibrio desulfuricans; H. chejuensis, Hahella chejuensis; A. macleodii, Alteromonas macleodii; M. actinobacterium, Marine actinobacterium; P. tunicata, Pseudoalteromonas tunicata; D. hafniense, Desulfitobacterium hafniense; L. salivarius, Lactobacillus salivarius; C. Koribacter, Candidatus Koribacter; D. acetoxidans, Desulfuromonas acetoxidans; C. phytofermentans, Clostridium phytofermentans; P. atlantica, Pseudoalteromonas atlantica T6c; A. borkumensis, Alcanivorax borkumensis SK2; H. aurantiacus, Herpetosiphon aurantiacus; S. frigidimarina, Shewanella frigidimarina; O. oeni, Oenococcus oeni; L. casei, Lactobacillus casei; L. mesenteroides, Leuconostoc mesenteroides; A. hydrophila, Aeromonas hydrophila; C. aggregans, Chloroflexus aggregans; P. propionicus, Pelobacter propionicus; C. cellulolyticum, Clostridium cellulolyticum; S. amazonensis, Shewanella amazonensis; P. ingrahamii, Psychromonas ingrahamii; T. carboxydivorans, Thermosinus carboxydivorans Nor1; M. marina, Microscilla marina; S. loihica, Shewanella loihica; C. aerofaciens, Collinsella aerofaciens; E. coli,* *Escherichia coli; M. gryphiswaldense, Magnetospirillum gryphiswaldense; S. tropica, Salinispora tropica; N. aromaticivorans, Novosphingobium aromaticivorans; P. pacifica, Plesiocystis pacifica; V. vadensis, Victivallis vadensis; R. obeum, Ruminococcus obeum; R. torques, Ruminococcus torques; C. hominis, Campylobacter hominis; B. capillosus, Bacteroides capillosus; R. gnavus, Ruminococcus gnavus; A. odontolyticus, Actinomyces odontolyticus; Clostridium sp, Clostridium sp. L2-50; C. leptum, Clostridium leptum; A. butzleri, Arcobacter butzleri RM4018; E. dolichum, Eubacterium dolichum; F. prausnitzii, Faecalibacterium prausnitzii; C. eutactus, Coprococcus eutactus*

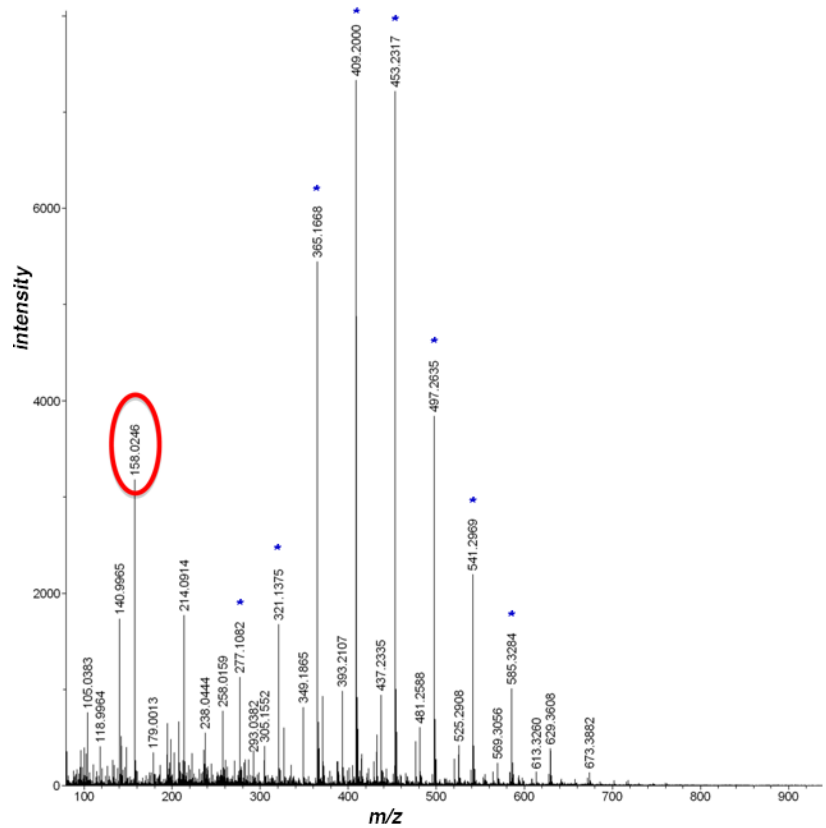
**Supplemental Figure 2. Relative Cys-tRNA$^{Pro}$ editing activity of *H. influenzae* YbaK mutants.** Cys-tRNA$^{Pro}$ (1.0 μM) deacylation by YbaK variants (1.0 μM) was performed at room temperature. Relative activity was calculated by comparing the deacylation level with that of WT YbaK at 30 min. $KPO_4$ refers to deacylation performed in the presence of 150 mM $KPO_4$, pH 7.0. The negative level of relative deacylation activity by the K46A variant was due to higher levels of Cys-tRNA$^{Pro}$ remaining relative to buffer hydrolysis. This is likely due to protection of the substrate by K46A YbaK.



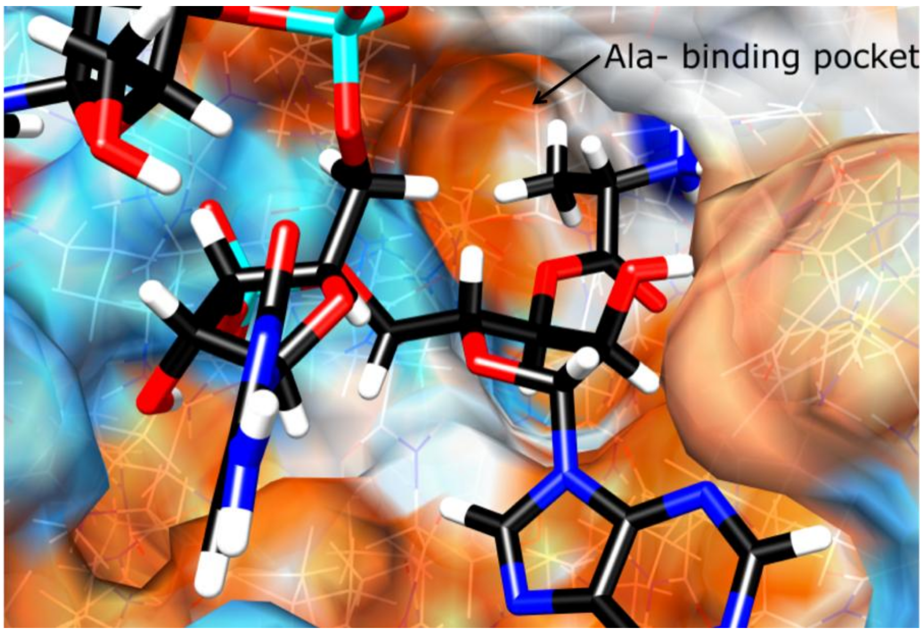
**Supplemental Figure 3. Cys-tRNA$^{Pro}$ deacylation by YbaK in the presence of EDTA or $ZnCl_2$.** Cys-tRNA$^{Pro}$ (0.5 μM) deacylation by WT YbaK (0.1 μM) in presence of EDTA or $ZnCl_2$ (5mM) was performed at 37°C. The plots represent the average of three trials with standard deviation indicated.



**Supplemental Figure 4. ESI-MS spectrum of crude reaction mixture following YbaK-catalyzed Cys-tRNA$^{Pro}$ deacylation.** The data were collected in positive ionization mode. Methanol was used as the carrier solvent and PEG-300 was used as a reference for accurate mass calibration. The peak identified as cysteine methyl ester is highlighted in red and peaks due to the PEG-300 reference are marked as with an asterisk (*).



**Supplemental Figure 5. Homology model of the editing (INS) domain of *E. faecalis* ProRS bound to CCA-Ala.** The model was generated by first superimposing the *E. faecalis* ProRS structure (2J3L) (17) on our model of *H. influenzae* YbaK bound to CCA-Cys. Next, substrate cysteine was mutated to alanine, followed by energy minimization using AMBER9. The hydrophobicity surface shown is colored blue for hydrophilic residues and orange for hydrophobic residues. The methyl sidechain of substrate alanine binds in a small hydrophobic pocket that would not accommodate larger amino acids such as cysteine or proline.



**SUPPLEMENTAL REFERENCES**

## Additional References for Computational Methods Described in the Manuscript:

**CLUSTALW:**

Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680 (1994).

**AUTODOCK 4:**

Huey, R., Morris, G.M., Olson, A.J. & Goodsell, D.S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **28**, 1145-1152 (2007).

**AMBER 9:**

Case, D.A. *et al. AMBER 9* (University of California, San Francisco, 2006).

**ArchPRED** (loop structure prediction server):

Fernandez-Fuentes, N., Zhai, J. & Fiser, A. ArchPRED: a template based loop structure prediction server. *Nucleic Acids Res.* **34**, W173-176 (2006).

**PropKa** (Prediction of protonation states of residues based on protein structure):

Jensen, J.H., Li, H., Robertson, A.D. & Molina, P.A. Prediction and rationalization of protein pKa values using QM and QM/MM methods. *J. Phys. Chem. A* **109**, 6634-6643 (2005).

**TIP3P (3-site model for water molecule):**

Jorgensen, W.L. *et al.* Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).

**Adaptive Poisson-Boltzmann Solver (APBS):**

Baker, N.A. *et al.* Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041 (2001).

1. Beuning, P. J., and Musier-Forsyth, K. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8916-8920

2. Shitivelband, S., and Hou, Y. M. (2005) *J. Mol. Biol.* **348**, 513-521

3. Wong, F. C., Beuning, P. J., Nagan, M., Shiba, K., and Musier-Forsyth, K. (2002) *Biochemistry* **41**, 7108-7115

4. Newberry, K. J., Hou, Y. M., and Perona, J. J. (2002) *EMBO J.* **21**, 2778-2787

5. Beebe, K., Ribas De Pouplana, L., and Schimmel, P. (2003) *EMBO J.* **22**, 668-675

6. Nordin, B. E., and Schimmel, P. (2002) *J. Biol. Chem.* **277**, 20510-20517

7. Zhang, H., Huang, K., Li, Z., Banerjei, L., Fisher, K. E., Grishin, N. V., Eisenstein, E., and Herzberg, O. (2000) *Proteins: Struct. Funct. Bioinform.* **40**, 86-97

8. An, S., and Musier-Forsyth, K. (2004) *J. Biol. Chem.* **279**, 42359-42362

9. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248-254

10. Fersht, A. R., Ashford, J. S., Bruton, C. J., Jakes, R., Koch, G. L., and Hartley, B. S. (1975) *Biochemistry* **14**, 1-4

11. Stathopoulos, C., Jacquin-Becker, C., Becker, H. D., Li, T., Ambrogelly, A., Longman, R., and Söll, D. (2001) *Biochemistry* **40**, 46-52

12. Wolfson, A. D., and Uhlenbeck, O. C. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5965-5970

13. Jakubowski, H. (1994) *Nucleic Acids Res.* **22**, 1155-1160

14. George, P., Trachtman, M., Bock, C. W., and Brett, A. M. (1975) *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **38**, 121-129

15. Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. (2009) Gaussian, Inc., Gaussian 09, Revision A.01 Ed., Wallingford, CT

16. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680

17. Crepin, T., Yaremchuk, A., Tukalo, M., and Cusack, S. (2006) *Structure* **14**, 1511-1525