Substrate Specificity of Bacterial Prolyl-tRNA Synthetase Editing Domain Is Controlled by a Tunable Hydrophobic Pocket*5

Sandeep Kumar151, Mom Das15*, Christopher M. Hadad6, and Karin Musier-Forsyth15%2

From the 1Department of Chemistry and Biochemistry, 2The Ohio State Biochemistry Program, and the 3Center for RNA Biology, The Ohio State University, Columbus, Ohio 43210

Background: Proofreading by aminoacyl-tRNA synthetases is a key translational quality control step.

Results: The alanine specificity of the prolyl-tRNA synthetase editing domain can be modulated by mutations within the conserved hydrophobic binding pocket.

Conclusion: Modulation of the editing domain specificity by mutation confirms a size exclusion-based mechanism.

Significance: Distinct editing mechanisms are required to clear mischarged tRNA species and ensure accuracy of translation.

Aminoacyl-tRNA synthetases catalyze the covalent attachment of amino acids onto their cognate tRNAs. High fidelity in this reaction is crucial to the accurate decoding of genetic information and is ensured, in part, by proofreading of the newly synthesized aminoacyl-tRNAs. Prolyl-tRNA synthetases (ProRS) mischarge tRNAPro with alanine or cysteine due to their smaller or similar sizes relative to cognate proline. Mischarged Ala-tRNAPro is hydrolyzed by an editing domain (INS) present in most bacterial ProRSs. In contrast, the INS domain is unable to deacylate Cys-tRNAPro, which is hydrolyzed exclusively by a freestanding trans-editing domain known as YbaK. Here, we have used computational and experimental approaches to probe the molecular basis of INS domain alanine specificity. We show that the methyl side chain of alanine binds in a well defined hydrophobic pocket characterized by conserved residues Ile-263, Leu-266, and Lys-279 and partially conserved residue Thr-277 in Escherichia coli ProRS. Site-specific mutation of these residues leads to a significant loss in Ala-tRNAPro hydrolysis, (ProRS) mischarge tRNAPro with alanine or cysteine due to their synthesized aminoacyl-tRNAs. Prolyl-tRNA synthetases formation and is ensured, in part, by proofreading of the newly this reaction is crucial to the accurate decoding of genetic information. AaRSs can confuse cognate amino acids with isosteric noncognate amino acids. For example, prolyl-tRNA synthetase (ProRS) mischarges tRNAPro with alanine and cysteine, in addition to the cognate proline. Mistakes by aaRSs are generally cleared either by hydrolysis or selective release of the aminoacyl-adenylate intermediate (pre-transfer editing) or by hydrolysis of the mischarged aminoacyl-tRNA (post-transfer editing). Post-transfer editing is catalyzed by a distinct editing domain that is either appended to the ancient catalytic core of the aaRS or encoded by a freestanding domain.

AaRSs catalyze aminoacyl-tRNA synthesis in a two-step reaction. In the first step, amino acids are activated by ATP to form an aminoacyl-adenylate intermediate, and in the second step, the activated amino acid is transferred from the adenylate to the 3’ terminus of the tRNA. AaRSs can confuse cognate amino acids with isosteric noncognate amino acids. For example, prolyl-tRNA synthetase (ProRS) mischarges tRNAPro with alanine and cysteine, in addition to the cognate proline. Mistakes by aaRSs are generally cleared either by hydrolysis or selective release of the aminoacyl-adenylate intermediate (pre-transfer editing) or by hydrolysis of the mischarged aminoacyl-tRNA (post-transfer editing). Post-transfer editing is catalyzed by a distinct editing domain that is either appended to the ancient catalytic core of the aaRS or encoded by a freestanding domain.

In the ProRS system, several distinct homologous domains play a role in post-transfer editing. An editing domain (INS) that clears mischarged Ala-tRNAPro is appended to most bacterial ProRSs, whereas Cys-tRNAPro is cleared by a freestanding editing domain known as YbaK (12–14). Some species, including those that lack an INS domain, encode a freestanding domain known as PrdX that hydrolyzes Ala-tRNAPro (14, 15). At least three additional freestanding editing domain homologs of unknown function (YeaK, ProX, and PA2301) have been identified based on sequence similarity to INS and YbaK (16).

Errors in protein synthesis often lead to misfolded or nonfunctional proteins that can be deleterious to cells (1). Several checkpoints exist to minimize such errors, including proofreading by aminoacyl-tRNA synthetases (aaRSs) (2–5), resampling of the aminoacyl-tRNA before its transport to the ribosome by elongation factors (6), and premature termination of the polypeptide chain by the ribosome (7). Because the errors associated with each step in translation are cumulative, uncorrected mistakes made during aaRS catalysis are more likely to get misincorporated into proteins. Hence, editing activities play critical roles in vivo (8, 9), and reduced proofreading by aaRSs is correlated with disease (10, 11).

AaRSs catalyze aminoacyl-tRNA synthesis in a two-step reaction. In the first step, amino acids are activated by ATP to form an aminoacyl-adenylate intermediate, and in the second step, the activated amino acid is transferred from the adenylate to the 3’ terminus of the tRNA. AaRSs can confuse cognate amino acids with isosteric noncognate amino acids. For example, prolyl-tRNA synthetase (ProRS) mischarges tRNAPro with alanine and cysteine, in addition to the cognate proline. Mistakes by aaRSs are generally cleared either by hydrolysis or selective release of the aminoacyl-adenylate intermediate (pre-transfer editing) or by hydrolysis of the mischarged aminoacyl-tRNA (post-transfer editing). Post-transfer editing is catalyzed by a distinct editing domain that is either appended to the ancient catalytic core of the aaRS or encoded by a freestanding domain.

In the ProRS system, several distinct homologous domains play a role in post-transfer editing. An editing domain (INS) that clears mischarged Ala-tRNAPro is appended to most bacterial ProRSs, whereas Cys-tRNAPro is cleared by a freestanding editing domain known as YbaK (12–14). Some species, including those that lack an INS domain, encode a freestanding domain known as PrdX that hydrolyzes Ala-tRNAPro (14, 15). At least three additional freestanding editing domain homologs of unknown function (YeaK, ProX, and PA2301) have been identified based on sequence similarity to INS and YbaK (16).

# This work was supported, in whole or in part, by National Institutes of Health Grant GM049928 (to K. M.-F.).

* This article contains supplemental Figs. 1–5 and additional references.

1 Present address: C-PCS, Chemistry Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

2 To whom correspondence should be addressed: Dept. of Chemistry and Biochemistry, The Ohio State University, 100 West 18th Ave., Columbus, OH 43210. Tel.: 614-292-2021; Fax: 614-688-5402; E-mail: musier@chemistry.ohio-state.edu.

3 The abbreviations used are: aaRS, aminoacyl-tRNA synthetase; 2-Abu, 2-aminobutyric acid; AlaRS, alanyl-tRNA synthetase; INS, editing domain of bacterial ProRS; MD, molecular dynamics; PDB, Protein Data Bank; ProRS, prolyl-tRNA synthetase.
Substrate Specificity of ProRS INS Domain

Although freestanding editing domains with homology to the editing domains of alanyl-tRNA synthetase (AlaRS) and threonyl-tRNA synthetase have also been identified, YbaK is the only known editing domain with a distinct substrate specificity relative to the homologous synthetase domain. We recently showed that YbaK uses a unique substrate-assisted mechanism of catalysis involving thiol-specific chemistry to cleave Cys-tRNAPro\(^{16}\). We hypothesize that due to the similar sizes of cysteine and proline, the ProRS INS domain could not evolve a substrate binding pocket that hydrolyzed both Ala- and Cys-tRNAPro\(^{pro}\) while excluding Pro-tRNAPro. Here, we explore the proposed steric exclusion mechanism of substrate discrimination by the INS domain using computational modeling and biochemical approaches. We identify a hydrophobic substrate binding pocket that is optimized for Ala-tRNAPro\(^{pro}\) hydrolysis but can be engineered to allow larger amino acids such as cysteine to bind and be hydrolyzed efficiently. The tunable nature of the substrate binding pocket is consistent with a steric exclusion mechanism. Remarkably, an INS variant that hydrolyzed Cys-tRNAPro but not Pro-tRNAPro also failed to hydrolyze Ala-tRNAPro\(^{pro}\), in accordance with the evolution of distinct editing domains that act upon each mischarged substrate in vivo.

EXPERIMENTAL PROCEDURES

Materials—All amino acids and chemicals were purchased from Sigma unless otherwise noted. \(^3\)H]Alanine (54 Ci/mmol), \(^3\)H]proline (99 Ci/mmol), \(^3\)H]serine (33 Ci/mmol), and \([\alpha-\beta\)^32]P]ATP were from Amersham Biosciences and \[^{35}\]S]cysteine (1075 Ci/mmol) was from PerkinElmer Life Sciences.

Molecular Dynamics (MD) Simulation of Enterococcus faecalis INS Domain—The crystal structure of E. faecalis ProRS (Protein Data Bank (PDB) code 2J3L, chain B, residues 224–407) was used as the starting structure for MD simulation. The protonation states of the titratable residues were determined by Propka\(^{17}\) at a pH of 7. The protein was charge-neutralized by addition of Na\(^+\) counterions, and an octahedral box of TIP3\(^{18}\) water molecules within 8 Å of the protein was created using the xleap package in AMBER10\(^{19}\). The all-atom MD simulations with explicit water molecules were performed using AMBER10 with the ff03 force field\(^{20}\). The system was energy-minimized for 2000 steps, keeping the protein rigid, followed by a 5000-step minimization for the entire system. In a preproduction MD simulation of 100 ps, the temperature of the system was raised from 0 to 300 K with a small force constant on the protein to restrict any drastic changes. Final production MD simulations were performed for 15 ns under constant temperature and pressure. The MD trajectory was analyzed for stability using Visual Molecular Dynamics software. The atomic fluctuations were calculated using the ptraj module of AMBER10.

Molecular Modeling of 5′-CCA-Ala Bound to E. faecalis INS Domain—Following MD simulation of the free protein, 25 snapshots of the protein structure were extracted at equal time intervals across the 15-ns MD simulation and were used for molecular docking. The structure of the 5′-CCA-Ala ligand was generated using the xleap module of AMBER10. This ligand was docked onto the 25 structures of INS using AutoDock 4.0\(^{21}\). Conformation of all of the ligand torsions were sampled during docking using a genetic algorithm. Side chain torsions of several residues in the binding pocket predicted from existing structural and biochemical data as well as MD simulation of the free protein, including Ile-263, Lys-279, Val-281, Glu-316, Phe-330, Val-333, and His-366, were allowed to change while the rest of the protein torsions were fixed. Each docking simulation involved generation of 200 different conformers, which were then clustered using a root mean square deviation cutoff of 2.0 Å. Resulting clusters were manually inspected, and four docked structures, deemed probable complex structures based on biochemical data, were chosen for further analysis. These structures were minimized using the ff03 force field in AMBER10, followed by 15 ns of MD simulation, as described above. A binding energy for each ligand was calculated using the Molecular Mechanic-Poisson Boltzmann Surface Area module of AMBER10. The complex structure displaying the most stable trajectory and the best binding energy was further minimized and was used as the final model for the INS-CCA-Ala complex.

Models of WT INS with noncognate substrates and H366A INS with cognate substrate were generated by manually mutating the residue of interest followed by energy minimization.

Protein Expression and Purification—His-tagged Escherichia coli wild-type (WT) ProRS\(^{22}\), WT AlaRS\(^{23}\), C666A/Q584H AlaRS (AlaRS-CQ)\(^{23}\), and E. coli tRNA nucleotidyltransferase\(^{24}\) were purified using the Talon cobalt affinity resin (Clontech) as described previously. The editing active site variants of E. coli ProRS were generated using the QuikChange site-directed mutagenesis kit (Stratagene). Mutations were confirmed by DNA sequencing (Genewiz). All E. coli ProRS variants were overexpressed in BL21(DE3) cells and purified as described for the WT protein. The concentration of E. coli AlaRS-CQ and tRNA nucleotidyltransferase was determined by the Bradford assay\(^{25}\). The concentrations of WT and mutant E. coli ProRS and AlaRS were determined by active site titration\(^{26}\).

Preparation of tRNA and Aminoacyl-tRNA Substrates—WT E. coli tRNA\(^{pro}\) and G1:C72/U70 E. coli tRNA\(^{pro}\) were prepared by in vitro transcription as described\(^{27}\). Preparation of radio-labeled Ala-, Ser-, Cys-, Pro-, and 2-amino butyric acid (2-Abu)-tRNA\(^{pro}\) was carried out as described\(^{12,16}\). Briefly, \[^{3}\text{H}]	ext{Ala-tRNA}^{pro} (G1:C72/U70 variant to facilitate charging by AlaRS) was prepared by incubating 2 \(\mu\)M E. coli WT AlaRS, 8 \(\mu\)M G1:C72/U70 tRNA\(^{pro}\), and \[^{3}\text{H}]	ext{alanine (12.9 }\mu\text{M) in buffer A (50 mM HEPES (pH 7.5), 4 mM ATP, 20 mM KCl, 20 mM }\beta\text{-mercaptoethanol, 25 mM MgCl}_{2}, and 0.1 mg/ml BSA) for 4 h at 25 °C. \[^{3}\text{H}]	ext{Ser-tRNA}^{pro} (G1:C72/U70 variant) was prepared by incubating 4 \(\mu\)M E. coli AlaRS-CQ, 10 \(\mu\)M E. coli G1:C72/U70 tRNA\(^{pro}\) in reaction buffer A containing \[^{3}\text{H}]	ext{serine (19 }\mu\text{M) for 1.5 h at 25 °C.}^{35}\text{S}]	ext{Cys-tRNA}^{pro} was prepared by incubating E. coli ProRS (8 \(\mu\)M) and E. coli tRNA\(^{pro}\) (8 \(\mu\)M) with \[^{35}\text{S}]	ext{cysteine (0.9 }\mu\text{M) and unlabeled cysteine (50 }\mu\text{M) in a reaction buffer containing 20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM MgCl}_{2}, 25 mM dithiothreitol, and 2 mM ATP for 1 h at 37 °C.}^{3}\text{H}]	ext{Pro-tRNA}^{pro} was prepared by incubating 2 \(\mu\)M E. coli WT ProRS, 8 \(\mu\)M tRNA\(^{pro}\), and \[^{3}\text{H}]	ext{proline (8.8 }\mu\text{M) in buffer A for 2 h at 25 °C.}^{32}\text{P}]	ext{tRNA}^{pro} was prepared by incubating 8 \(\mu\)M E. coli ProRS, 8 \(\mu\)M 3'\(\text{[}^{32}\text{P}]\) tRNA\(^{pro}\), and 450 mM 2-Abu in buffer A for 1.5 h at 37 °C.}^{3}\text{[}^{32}\text{P}]\) tRNA\(^{pro}\) was prepared
using E. coli tRNA nucleotidyltransferase and WT tRNA<sub>Pro</sub> as described previously (28). Following aminoacylation reactions, the substrate aminoacyl-tRNAs were phenol-chloroform-extracted, ethanol-precipitated, and stored at −80 °C for use in deacylation assays.

**Deacylation Assays**—To compare the loss in deacylation activity of WT and mutant ProRS, 5 μM enzyme concentration was used in all assays. Deacylation of Ala-, Ser-, Cys-, Pro-, and 2-Abu-tRNA<sup>Pro</sup> was carried out at 25 °C, whereas deacylation of Cys-tRNA<sup>Pro</sup> was carried out at 25 °C or 37 °C according to published protocols (13, 16). Reactions were performed using ~1 μM aminoacyl-tRNA in 150 mM KPO4 reaction buffer at pH 7.0. A background reaction was carried out in which enzyme was omitted. For [<sup>3</sup>H]Ala-, [<sup>3</sup>H]Ser-, [<sup>35</sup>S]Cys-, and [<sup>3</sup>H]Pro-tRNA<sup>Pro</sup>, the reactions were monitored by precipitating the tRNA on Whatman filter pads followed by scintillation counting (12). For 2-Abu-3′-[<sup>32</sup>P]tRNA<sup>Pro</sup>, reactions were quenched by adding 0.5 unit/μL P1 nuclease (Sigma-Aldrich) in 200 mM NaOAc (pH 5.0) on ice. After digestion for 20 min at room temperature, 1 μL of quenched mixture was spotted onto a PEI-cellulose TLC plate pre-run with water. Separation of 2-Abu-AMP and AMP was accomplished by developing the TLC plates in 0.1 M ammonium chloride and 5% acetic acid. The radioactivity was analyzed using a Typhoon PhosphorImager, and data were analyzed using ImageQuant TL software (GE Healthcare). The level of deacylation was determined from the ratio of 2-Abu-AMP remaining over the total AMP plus 2-Abu-AMP, as described previously (28). The fraction of aminoacyl-tRNA remaining was plotted as a function of time following subtraction of the buffer-only hydrolysis background reaction and fitted to a single-exponential equation to obtain <i>k</i><sub>obs</sub>. All assays were carried out in triplicate.

**RESULTS**

Several homologous domains, including the INS domain of bacterial ProRS and the freestanding editing domains, YbaK and PrdX, have been shown to enhance the fidelity of proline codon translation by editing mischarged tRNA<sup>Pro</sup>. To understand the molecular basis of substrate selection and discrimination by the INS domain, we first used computational approaches to generate a model of the bacterial ProRS INS domain bound to an Ala-tRNA substrate analog, 5′-CCA-Ala.

**MD Simulation of the Free INS Domain**—Crystal structures represent one of many possible conformations that proteins may adopt and are often not well suited for the molecular docking of a large substrate. MD simulations of the free protein not only allow the sampling of these dynamic motions, but also provide information on the flexible regions of the protein that may be involved in substrate binding. We performed a 15-ns all-atom MD simulation of the E. faecalis ProRS INS domain (PDB code 2J3L, residues 224–407). Atomic fluctuations of backbone atom C, Cα, and N for each residue, calculated as the root mean square deviation from the starting structure over the course of the simulation, identified several flexible regions of the protein (Fig. 1A). In addition to the linker regions of the INS domain (Fig. 1B), flexible regions were generally localized around the Lys-279 residue, earlier identified as critical for catalysis (22, 29). Residue Gly-331, part of the highly conserved GXXXP motif hypothesized to play an analogous role to that of an oxyanion hole (30), was also found to be highly flexible. Residue His-366, which is equivalent to residue His-369 in E. coli ProRS and previously implicated in Ala specificity of the INS domain (22), does not show high flexibility but lies close to more flexible regions of the protein. The congruence between the flexible regions of the protein and the catalytically important residues is consistent with putative substrate binding in this region of the protein.

**Modeling of the INS-CCA-Ala Complex**—A truncated substrate analog of Ala-tRNA, 5′-CCA-Ala, was docked into the putative active site of the protein. Twenty-five snapshots of the protein structure were extracted at equal time intervals from the MD trajectory of the free INS domain. A large grid, encompassing all of the flexible regions of the protein, was used for the

---

**FIGURE 1.** Relative flexibility of residues in the INS domain of *E. faecalis* ProRS. A, root mean square of the atomic fluctuation of the backbone atoms for each residue of the INS domain over 15-ns MD simulation. Horizontal line represents the average fluctuation for all residues. Residues with greater than average fluctuation are shown in magenta for the linker region and in light blue for the others. B, ribbon representation of the INS domain with flexible residues colored as in A. Catalytically important residues Lys-279, Gly-331 (part of GXXXP loop), and His-366 are shown as stick models. First residues of some flexible regions are labeled.
docking of the substrate. All torsions of the substrate and side chain torsions of several bulky residues in the flexible regions of the protein, including Ile-263, Lys-279, Val-281, Glu-316, Phe-330, Val-333, and His-366, were conformationally sampled. The substrate generally clustered near the flexible regions, but the computed binding energy did not show a significant preference for any conformation. Because previous studies have shown Lys-279 and His-366 (His-369 in E. coli ProRS) residues to be important in catalysis and substrate specificity, respectively, a weak constraint that the Cα carbon atom of the substrate alanine should be within 10 Å of these residues was applied. Four unique conformations of the substrate were selected for further analysis (supplemental Fig. 1). These structures of the INS-CCA- Ala complex were energy minimized and subjected to 5 ns of equilibration MD and 10 ns of production MD. Complex A (supplemental Fig. 1) showed a stable trajectory during the MD simulation, as well as the best binding energy calculated by Mechanic-Poisson Boltzmann Surface Area, and was further optimized and selected as the final structure of the INS-CCA- Ala complex.

The model of the INS-CCA- Ala complex (Fig. 2A) reveals several key interactions of the substrate with the protein (Fig. 2, B–D). The A76 phosphate group interacts with the Lys-279 residue via charge-charge interaction. This interaction may be crucial for the positioning of the 3′-end of the tRNA in the INS active site and would explain the essential role of Lys-279 residue in catalysis. The adenine base of the substrate is predicted to form three H-bonds with the protein, one between N7 and the Leu-282 backbone amine and two between the N6 hydrogens and the Leu-282 and Ala-317 carbonyl groups. The adenine base also interacts with residue Phe-330 via a stacking interaction (Fig. 2B). These interactions may play an important role in orienting the substrate in a catalytically competent manner in the active site. The C75 and C74 bases bind on the surface of the INS domain and H-bond to residues Glu-316 and Asp-299, respectively. The 5′-end of the substrate remains exposed.
to the solvent in a manner that would be compatible with binding of full-length tRNA.

The substrate alanine binds in a hydrophobic cleft in the INS domain. The amine group makes H-bonds with the backbone carbonyls of residues Gly-331 and Ile-333, as well as with the His-366 side chain. The substrate carbonyl group is directed toward the positively charged GXXXP loop (Figs. 2C and 3A). The methyl side chain of the alanine substrate binds in a hydrophobic pocket formed by the side chains of highly conserved residues Lys-279, Ile-263, and Val-266 (Leu-266 in E. coli ProRS) (Fig. 2D and supplemental Fig. 2). The partially conserved residue Ile-277 (Thr-277 in E. coli ProRS) interacts with

---

**FIGURE 3.** Optimized structures of WT and H366A E. faecalis ProRS INS domain bound to cognate and noncognate substrates. WT ProRS INS domain in complex with CCA-Ala (A), CCA-Cys (B), CCA-Pro (C), CCA-Ser (D), and H366A ProRS INS domain in complex with CCA-Pro (E). Distances from the substrate carbonyl oxygen to the backbone N atoms of Gly-331, Ser-332, and Ile-333 of the GXXXP loop and from the substrate amino group to His-366 (A–D only) are indicated. H-bonding interactions of two active site water molecules are also shown. C74 and C75 nucleotides are not shown for clarity. Substrates are labeled in green and red for active and inactive substrates, respectively. Direction and magnitude of displacement of Cα atom of substrates, relative to the Cα of Ala substrate in WT ProRS, are shown in magenta. A small displacement in computational models for Ser deacylation by WT ProRS and Pro deacylation by H366A ProRS is consistent with their slightly reduced deacylation in experimental studies.
Substrate Specificity of ProRS INS Domain

Ile-263 and Val-266 and completes the hydrophobic pocket into which the substrate alanine binds; however, it does not interact directly with the substrate. Overall, this alanine binding pocket is relatively large, and based on the docking model, we predict that this pocket would be able to fit slightly larger hydrophobic substrates.

Model of INS Domain Bound to Noncognate Substrates—Biochemical data show that the INS domain selectively hydrolyzes Ala-tRNAPro but not Cys-tRNAPro or Pro-tRNAPro (12, 27). To understand the molecular basis of this selectivity, we generated models of the INS domain bound to 5′-CCA-Cys (Fig. 3B) and 5′-CCA-Pro (Fig. 3C) by changing the alanine residue of the docked substrate to cysteine or proline, followed by energy minimization. The model of INS-CCA-Ala (Fig. 3A) shows that the carbonyl group of the substrate points to the backbone amine groups of the GXXXP loop (residues 331–333). The oxygen anionic tetrahedral intermediate formed during catalysis is likely stabilized by this low electron density oxyanion hole. As mentioned above, this orientation of the substrate is further stabilized by H-bonds of the substrate amine group (Fig. 2B).

There are two water molecules in our models that may play catalytic or structural roles. One water is situated close to the Ca of the substrate alanine and forms H-bonds with the 2′-OH (Fig. 3A) and Gly-261 carbonyl (data not shown). This water may play a catalytic role as it is positioned for nucleophilic attack on the carbonyl center of the substrate. Another water molecule on the active site floor (Fig. 3), which is also present in the crystal structure of E. faecalis ProRS and is highly stable in MD simulations, does coordinate with the Ser-280 side chain and with backbone atoms of Ile-278 and Gly-359. This water likely plays a structural role and is important in maintaining the overall configuration of the INS active site.

Upon substitution of alanine with cysteine, the large sulphhydril side chain pushes the substrate 0.6 Å away from the floor of the binding pocket (Fig. 3B), resulting in a change in orientation of the substrate carbonyl relative to the GXXXP loop, as well as altering the position of the putative catalytic water molecule relative to the 2′-OH of the substrate. These relatively minor structural changes may result in destabilization of the transition state. Larger structural changes occur upon substitution with proline (Fig. 3C), resulting in a loss of one H-bond of the substrate amine, as well as loss of interaction of the substrate carbonyl with the GXXXP loop. The Ca of the substrate proline is also displaced by 0.7 Å relative to the Ca of the alanine substrate. These structural perturbations would not only result in weaker binding of the substrate but also a destabilized transition state.

Residue His-366 has been previously implicated in Ala specificity of the INS domain. An H366A variant of E. coli ProRS displays significantly reduced Ala-tRNAPro deacylation activity and gains Pro-tRNAPro deacylation capability (22). Loss of the H-bond between His-366 and the substrate amine may explain the loss in alanine deacylation activity. To understand the basis of Pro-tRNAPro deacylation by the mutant, we generated a model of H366A INS bound to CCA-Pro (Fig. 3E). This shows that proline is easily accommodated in the active site of the mutant enzyme, whereas the critical interactions between the substrate and the INS active site are not perturbed as they are in the case of WT INS (Fig. 3C). In WT ProRS, there is strong interaction between the substrate Ala-tRNAPro and the His-366 residue and a weaker interaction with Val-266. In contrast, in the case of the H366A mutant, the substrate Pro-tRNAPro lacks a close interaction with Ala-366, but appears to interact with the Val-266 residue strongly (supplemental Fig. 3). The better fit of proline in the mutated active site pocket together with this new interaction may explain the hydrolysis of Pro-tRNAPro by the H366A mutant.

Substrate Specificity of the WT INS Domain—The model of INS-CCA-Ala described above suggests that the INS domain might accommodate and hydrolyze aminoacyl-tRNA substrates larger than alanine. To test this hypothesis, we examined post-transfer editing of Ser- and 2-Abu-tRNAPro by E. coli ProRS. As expected, E. coli ProRS can readily hydrolyze Ala-tRNAPro (kobs = 0.088 ± 0.0004 min⁻¹) but cannot hydrolyze Cys-tRNAPro (kobs = 0.014 ± 0.005 min⁻¹) and Ser-tRNAPro (kobs = 0.012 ± 0.001 min⁻¹) is observed, albeit with 6–7-fold lower efficiency. Hydrolysis of Ser- and 2-Abu, a nonnatural amino acid with an additional methyl group compared with alanine, is consistent with the expansive hydrophobic substrate binding pocket observed in our model. A model of INS-CCA-Ser (Fig. 3D), generated as described for CCA-Cys and CCA-Pro substrate analogs, suggests that the hydroxyl side chain of serine is stabilized by H-bonding to the structural active site water molecule (coordinated to Ser-280 residue) without any significant disruption of the carbonyl or amine geometries (compared with Fig. 3A). These results suggest that the larger size and less polar character of the cysteine sulphydryl group may be responsible for the lack of hydrolytic activity against Cys-tRNAPro.

Altered Substrate Specificities of INS Domain Mutants—Our model of INS-CCA-Ala predicts that mutation of the bulky hydrophobic residues comprising the substrate binding pocket...
Substrate Specificity of ProRS INS Domain

to smaller residues, such as alanine, would result in significant loss of interaction with the active substrates and thus reduced deacylation activity. At the same time, these changes may allow larger substrates to bind in the active site and be deacylated. To test these predictions, we generated several site-directed mutants of *E. coli* ProRS and tested their activity against Ala-, 2-Abu-, Ser-, Cys-, and Pro-tRNA<sup>Pro</sup>, Ile-263, Val-266, and Ile-277 residues in *E. faecalis* ProRS correspond to Ile-263, Leu-266, and Thr-277 residues in *e. coli* ProRS, respectively (Fig. 2D). The structure of *E. coli* ProRS generated by homology modeling using SWISS-MODEL (data not shown) suggests that Thr-277, although polar in nature, contributes to the hydrophobic pocket via its terminal methyl group. Accordingly, we generated single *E. coli* ProRS mutants I263A, L266A, and T277S for our analysis, as well as a I263A/L266A double mutant. In the case of nonconserved residue Thr-277, a change to serine rather than alanine was made based on inspection of an *E. coli* ProRS homology model generated using 3D-JIGSAW (data not shown) (31). This model predicts an intramolecular interaction of the Thr-277 hydroxyl group, which points away from the active site pocket, whereas the methyl side chain points into the proposed active site. Thus, Thr-277 was mutated to serine to preserve the predicted interaction of the hydroxyl group while altering the active site pocket size.

Consistent with our prediction, Ala-tRNA<sup>Pro</sup> (Fig. 5A) and 2-Abu-tRNA<sup>Pro</sup> (Fig. 5B), both substrates for WT ProRS, were either deacylated very poorly or not deacylated by the mutants examined. In contrast Ser-tRNA<sup>Pro</sup>, another substrate of WT ProRS, was hydrolyzed by all the active site variants, in some cases more efficiently than the WT enzyme (Fig. 5C). Strikingly, the I263A/L266A variant displayed an ~8-fold increased rate of hydrolysis (k<sub>obs</sub> = 0.091 ± 0.006 min<sup>-1</sup>) relative to WT ProRS. This behavior can be explained by the model of INS-CCA-Ser (Fig. 3D), which shows that the hydroxyl group of serine forms an H-bond with a structural active site water. Mutations in the hydrophobic pocket would generate more space for additional water molecules to enter the active site and form an H-bonding network, thus allowing the stabilization of the active site water coordinated to Ser-280, as well as the bound serine substrate. Consistent with this idea, single mutations in residues closer to this water (e.g. Thr-277 and Leu-266) result in enhanced Ser-tRNA<sup>Pro</sup> hydrolysis, whereas changing a residue farther away (e.g. Ile-263) does not (Fig. 5C).

Cys-tRNA<sup>Pro</sup> deacylation by the INS mutants was also tested (Fig. 5D). Interestingly, efficient deacylation of Cys-tRNA<sup>Pro</sup> by the I263A variant (k<sub>obs</sub> = 0.068 ± 0.005 min<sup>-1</sup>) was observed, and weak deacylation by the L266A variant was also detected (k<sub>obs</sub> = 0.01 ± 0.001 min<sup>-1</sup>) at 37 °C, the temperature typically used for Cys-tRNA deacylation by YbaK (16). To compare the observed rates for Cys-deacylation with the rates obtained for other substrates at room temperature, cysteine deacylation assays using the I263A and L266A variants were also carried at room temperature (supplemental Fig. 4). The k<sub>obs</sub> of 0.07 ± 0.007 min<sup>-1</sup> was similar to that obtained at 37 °C for the I263A mutant, whereas L266A showed negligible deacylation at room temperature (k<sub>obs</sub> of 0.002 ± 0.0005 min<sup>-1</sup>). These same variants failed to deacylate Ala-tRNA<sup>Pro</sup> (Fig. 5A). The complete switch of substrate specificity from Ala- to Cys-tRNA<sup>Pro</sup> suggests that the size of the active site pocket was successfully engineered to accommodate the larger cysteine side chain in a manner that mutually excludes Ala-tRNA<sup>Pro</sup> deacylation (Table 1).

Similar to WT ProRS, none of the mutants showed significant deacylation activity against Pro-tRNA<sup>Pro</sup> (Fig. 5E and Table 1). Binding of CCA-Pro to WT INS (Fig. 3C) suggests that the bulky proline ring results in disruption of the catalytically competent orientation of the substrate carbonyl. The observed lack of proline deacylation by the INS variants suggests that merely increasing the binding pocket size is not sufficient to allow binding of a catalytically competent Pro-tRNA<sup>Pro</sup> substrate.

**DISCUSSION**

Three-dimensional crystal structures of post-transfer editing substrate analogs bound to the ProRS editing domain or to freestanding homologous domains are not yet available. Our recent model of YbaK bound to a 5'-CCA-Cys substrate analog (16) and the model of the INS domain bound to 5'-CCA-Ala presented in this study together with new biochemical data provide the first structure-based understanding of the distinct substrate specificities of these editing domains. The only available structures with bound substrate are those of *Thermus thermophilus* ProX, a homologous domain of unknown function, bound to pre-transfer editing substrate analogs 5'-O-[N-(L-alanyl)-sulfamoyl]adenosine and 5'-O-[N-(L-cysteinyl)-sulfamoyl]adenosine (PDB codes 2Z0K and 2Z0X). Although these substrates differ significantly from the post-transfer substrate analogs modeled here, these structures are consistent with aspects of our models. A comparison of our INS-CCA-Ala model with the *T. thermophilus* ProX-Ala-SA structure (supplemental Fig. 5) shows that the substrate alanine occupies a similar active site pocket in both the domains. In addition, the substrate carbonyl points toward the GXXXP loop in both structures. However, the ribose moiety is flipped by ~180°, presumably due to differences in connectivity of the adenine base to alanine (via the ribose in CCA-Ala and via the phosphate in 5'-O-[N-(L-alanyl)-sulfamoyl]adenosine), causing the adenine base and the phosphate group of the substrate to occupy distinct positions and orientations in the two structures.

Based on the INS-CCA-Ala model, we predicted correctly that the INS domain could hydrolyze somewhat larger substrates such as 2-Abu in addition to alanine. Ser-tRNA<sup>Pro</sup> was also hydrolyzed by WT ProRS, and by increasing the size of the active site pocket, the efficiency of this reaction could be increased. Mutant forms of *E. coli* ProRS that result in a complete switch in substrate specificity from alanine to cysteine (I263A) or to serine (I263A/L266A) (Table 1) were also generated.

Several different mechanisms have been proposed for discrimination of substrates by post-transfer editing domains. For example, isoleucyl-tRNA synthetase uses the so-called “double-sieve” mechanism to activate isoleucine and smaller valine in the synthetic active site ("coarse" sieve), but uses size exclusion to reject cognate Ile-tRNA in the editing active site ("fine" sieve) (32, 33). Discrimination between isosteric substrates valine and threonine by valyl-tRNA synthetase is dependent on the chem-

JANUARY 27, 2012 • VOLUME 287 • NUMBER 5 JOURNAL OF BIOLOGICAL CHEMISTRY 3181
ically distinct properties of their side chains (34). Recently, it was shown that a coordinated zinc ion in the editing domain of AlaRS may help provide specificity for noncognate serine, which is larger than alanine (35, 36). Structural studies of the ThrRS editing domain suggest that the discrimination against threonine in the editing domain is achieved by positioning of the substrate rather than steric exclusion (37). We recently showed that YbaK, which only cleaves Cys-tRNAs, achieves sub-

FIGURE 5. Deacylation substrate specificity of WT and mutant E. coli ProRSs. Deacylations of Ala-tRNAPro (A), 2-Abu-tRNAPro (B), Ser-tRNAPro (C), Cys-tRNAPro (D), and Pro-tRNAPro (E) are shown. Deacylations by WT (●), I263A (○), L266A (■), T277S (○), and I263A/L266A (▲) E. coli ProRSs are shown. All experiments were performed at 25 °C except for Cys deacylation, which was at 37 °C. Substrate aa-tRNA and enzyme concentrations were 1 μM and 5 μM, respectively. Data points and error bars represent average and S.D., respectively, of three independent experiments. The extent of hydrolysis of each substrate by buffer (0.15 M KPO4, pH 7.0) was subtracted prior to plotting the data.
substrate discrimination by taking advantage of the unique side chain chemistry of the substrate cysteine sulfhydryl moiety (16). The results presented in this study suggest that cysteine discrimination by the INS domain is achieved via size exclusion, which can be counteracted by mutations in the substrate binding pocket. Discrimination against proline, on the other hand, is largely dependent on functional orientation of the substrate, in addition to size exclusion. For example, H369A E. coli ProRS does hydrolyze Pro-tRNAPro, albeit with significantly reduced efficiency relative to Ala-tRNAPro hydrolysis by WT ProRS (22).

Taken together, the results of this study shed new insights into substrate selection and specificity of the bacterial ProRS INS domain. Although the INS domain can be engineered to deacylate Cys-tRNAPro, the mechanism of catalysis is unlikely to be the same as that of YbaK. The latter is believed to exclude water from the active site and uses a substrate-assisted mechanism of catalysis (16). Consistent with this mechanism, selenocysteine-tRNA is a substrate for YbaK, whereas Ser-tRNA is a competitive inhibitor of Cys-tRNA deacylation (16). In addition, manipulating the size of the active site pocket of YbaK fails to alter the substrate specificity.4 In contrast, cleavage in the INS active site is likely initiated by an active site water molecule, and substrate specificity can readily be modulated (Table 1).

The fact that the engineered I263A INS domain deacylates Cys-but not Ala-tRNAPro is consistent with the need for distinct editing domains to perform these proofreading functions in vivo.

Acknowledgments—We thank Jo Marie Bacusmo for help in purification of E. coli ProRS editing domain mutants. Generous computational resources were provided by the Ohio Supercomputer Center.

REFERENCES

1. Reynolds, N. M., Lazazzera, B. A., and Ibba, M. (2010) Cellular mechanisms that control mistranslation. Nat. Rev. Microbiol. 8, 849–856
2. Jakubowski, H., and Goldman, E. (1992) Microbiol. Rev. 56, 412–429
3. Hendrickson, T. L., de Crécy-Lagard, V., and Schimmel, P. (2004) Incorporation of nonnatural amino acids into proteins. Annu. Rev. Biochem. 73, 147–176
4. Ding, J., Reynolds, N., and Ibba, M. (2009) Aminoacyl-tRNA synthesis and translational quality control. Annu. Rev. Biochem. 74, 61–78
5. Mascarenhas, A. P., An, S., Rosen, A. E., Martinis, S. A., and Musier-Forsyth, K. (2008) Fidelity mechanisms of the aminoacyl-tRNA synthetases. Protein Engineering (Köhler, C., and RajBhandary, U. L., eds) pp. 155–203, Springer-Verlag, Berlin
6. Ling, J., So, B. R., Yadavalli, S. S., Roy, H., Shoji, S., Frederick, K., Musier-Forsyth, K., and Ibba, M. (2009) Resampling and editing of mischarged tRNA prior to translation elongation. Mol. Cell 33, 654–660
7. Zaher, H. S., and Green, R. (2009) Quality control by the ribosome following peptide bond formation. Nature 457, 161–166
8. Döring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crécy-Lagard, V., Schimmel, P., and Marlére, P. (2001) Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. Science 292, 501–504
9. Roy, H., Ling, J., Irniv, M., and Ibba, M. (2004) Post-transfer editing in vivo and in vitro by the β-subunit of phenylalanyl-tRNA synthetase. EMBO J. 23, 4639–4648
10. Lee, J. W., Beebe, K., Nangle, L. A., Jang, J., Longo-Guess, C. M., Cook, S. A., Davison, M. T., Sundberg, J. P., Schimmel, P., and Ackerman, S. L. (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature 443, 50–55
11. Nangle, L. A., Motta, C. M., and Schimmel, P. (2006) Global effects of mistranslation from an editing defect in mammalian cells. Chem. Biol. 13, 1091–1100
12. An, S., and Musier-Forsyth, K. (2004) Trans-editing of Cys-tRNAPro by Haemophilus influenzae YbaK protein. J. Biol. Chem. 279, 42359–42362
13. An, S., and Musier-Forsyth, K. (2005) Cys-tRNA(Pro) editing by Haemophilus influenzae YbaK via a novel synthetase-YbaKtRNA ternary complex. J. Biol. Chem. 280, 34465–34472
14. Ruan, B., and Söll, D. (2005) The bacterial YbaK protein is a Cys-tRNAPro and Cys-tRNA Cys deacylase. J. Biol. Chem. 280, 25887–25891
15. Abel, I., Korencic, D., Ibba, M., and Söll, D. (2003) Trans-editing of mischarged tRNAs. Proc. Natl. Acad. Sci. U.S.A. 100, 15422–15427
16. So, B. R., An, S., Kumar, S., Das, M., Turner, D. A., Hadad, C. M., and Musier-Forsyth, K. (2011) Substrate-mediated fidelity mechanism ensures accurate decoding of proline codons. J. Biol. Chem. 286, 31810–31820
17. Jensen, J. H., Li, H., Robertson, A. D., and Molina, P. A. (2005) Prediction and rationalization of protein pKa values using QM and QM/MM methods. J. Phys. Chem. A 109, 6634–6643
18. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) J. Chem. Phys. 79, 926–935
19. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvary, L., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Golik, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagu, C., Babin, V., and Kollman, P. A. (2008) AMBER 10, University of California, San Francisco
20. Duan, Y., Wu, C., Choudhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24, 1999–2012
21. Huey, R., Morris, G. M., Olson, A. J., and Goodsell, D. S. (2007) A semempirical free energy force field with charge-based desolvation. J. Comput. Chem. 28, 1145–1152
22. Wong, F. C., Beuning, P. J., Nagan, M., Shibahara, K., and Musier-Forsyth, K. (2002) Functional role of the prokaryotic proline-tRNA synthetase insertion domain in amino acid editing. Biochemistry 41, 7108–7115
23. Beebe, K., Ribas De Pouplana, L., and Schimmel, P. (2003) Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. EMBO J. 22, 668–675
24. Nordin, B. E., and Schimmel, P. (2002) Plasticity of recognition of the 3’-end of mischarged tRNA by class I aminoacyl-tRNA synthetases. J. Biol. Chem. 277, 20510–20517
25. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254
26. Fersht, A. R., Ashford, J. S., Brutton, C. J., Jakes, R., Koch, G. L., and Hartley, B. S. (1975) Active site titration and aminoacyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. Biochemistry 14, 1–4
27. Beuning, P. J., and Musier-Forsyth, K. (2000) Hydrolytic editing by a class
Substrate Specificity of ProRS INS Domain

II aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8916–8920

28. Wolfson, A. D., and Uhlenbeck, O. C. (2002) Modulation of tRNAAla identity by inorganic pyrophosphatase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5965–5970

29. Crepin, T., Yaremchuk, A., Tukalo, M., and Cusack, S. (2006) Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain. *Structure* 14, 1511–1525

30. Zhang, H., Huang, K., Li, Z., Banerjei, L., Fisher, K. E., Grishin, N. V., Eisenstein, E., and Herzberg, O. (2000) Crystal structure of YbaK protein from *Haemophilus influenzae* (HI1434) at 1.8 Å resolution: functional implications. *Proteins* 40, 86–97

31. Bates, P. A., Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins Suppl.* 5, 39–46

32. Fersht, A. R. (1998) Sieves in sequence. *Science* 280, 541

33. Nureki, O., Fukai, S., Sekine, S., Shimada, A., Terada, T., Nakama, T., Shirouzu, M., Vassylyev, D. G., and Yokoyama, S. (2001) Structural basis for amino acid and tRNA recognition by class I aminoacyl-tRNA synthetases. *Cold Spring Harbor Symp. Quant. Biol.* 66, 167–173

34. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G., and Yokoyama, S. (2000) Structural basis for double-sieve discrimination of l-valine from l-isoleucine and l-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell* 103, 793–803

35. Sokabe, M., Okada, A., Yao, M., Nakashima, T., and Tanaka, I. (2005) Molecular basis of alanine discrimination in editing site. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11669–11674

36. Sokabe, M., Ose, T., Nakamura, A., Tokunaga, K., Nureki, O., Yao, M., and Tanaka, I. (2009) The structure of alanyl-tRNA synthetase with editing domain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11028–11033

37. Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C., and Moras, D. (2000) Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. *Cell* 103, 877–884