The structural basis of cysteine aminoacylation of $tRNA^{Pro}$ by prolyl-tRNA synthetases

Satwik Kamtekar[†‡§], W. Dexter Kennedy[†§¶], Jimin Wang[†], Constantinos Stathopoulos[†], Dieter Söll[†‖], and Thomas A. Steitz[†‡‖]**

Departments of [†]Molecular Biophysics and Biochemistry and [‖]Chemistry, Yale University, [¶]Department of Pediatrics, Yale University School of Medicine, and [‡]Howard Hughes Medical Institute, 266 Whitney Avenue, New Haven, CT 06520-8114

Contributed by Thomas A. Steitz, December 24, 2002

Cysteinyl-tRNA synthetase is an essential enzyme required for protein synthesis. Genes encoding this protein have not been identified in *Methanocaldococcus jannaschii*, *Methanothermobacter thermautotrophicus*, or *Methanopyrus kandleri*. It has previously been proposed that the prolyl-tRNA synthetase (ProRS) enzymes in these organisms recognize either proline or cysteine and can aminoacylate their cognate tRNAs through a dual-specificity mechanism. We report five crystal structures at resolutions between 2.6 and 3.2 Å: apo *M. jannaschii* ProRS, and *M. thermautotrophicus* ProRS in apo form and in complex with cysteinyl-sulfamoyl-, prolyl-sulfamoyl-, and alanyl-sulfamoyl-adenylates. These aminoacyl-adenylate analogues bind to a single active-site pocket and induce an identical set of conformational changes in loops around the active site when compared with the ligand-free conformation of ProRS. The cysteinyl- and prolyl-adenylate analogues have similar, nanomolar affinities for *M. thermautotrophicus* ProRS. Homology modeling of tRNA onto these adenylate complexes places the 3′-OH of A76 in an appropriate position for the transfer of any of the three amino acids to tRNA. Thus, these structures explain recent biochemical experiments showing that *M. jannaschii* ProRS misacylates $tRNA^{Pro}$ with cysteine, and argue against the proposal that these archaeal ProRS enzymes possess the dual capacity to aminoacylate both $tRNA^{Pro}$ and $tRNA^{Cys}$ with their cognate amino acids.

Aminoacyl-tRNA synthetases (aaRS) play a central role in translation by providing the aminoacyl-tRNAs used in protein biosynthesis. These synthetases show exquisite specificity for both their amino acid and tRNA substrates; overall misincorporation in polypeptide synthesis appears to be on the order of 1 in 10^4 (1). The specificity of aaRS for their cognate amino acids is ensured in a variety of ways, including shape complementarity of binding pockets, hydrogen bonding, chelation by metals, and the hydrolysis (editing) of mischarged products (2–4). On the basis of the topology of their ATP-binding sites, the aaRS are divided into two distinct structural classes (5–7). Putative aaRS genes can be identified by virtue of the signature sequences associated with these structural motifs.

The complete genome sequences of the thermophilic methanogens *Methanocaldococcus jannaschii* (8), *Methanothermobacter thermautotrophicus* (9), and *Methanopyrus kandleri* (10) do not contain genes with recognizable cysteinyl-tRNA synthetase (CysRS) motifs. Because Cys-$tRNA^{Cys}$ is required for protein synthesis *in vivo*, its formation is an essential cellular function. Various ways in which these organisms may make Cys-$tRNA^{Cys}$ have been proposed.

One possibility is that Cys-$tRNA^{Cys}$ may be formed through an indirect pathway, in which the amino acid on a mischarged $tRNA^{Cys}$ is converted to cysteine. Although such a possibility has precedents in the cases of other amino acids (1), biochemical studies have so far not supported an indirect pathway for Cys-$tRNA^{Cys}$ formation (11). A second alternative is that the protein responsible for cysteine charging is so highly diverged from canonical CysRS that it cannot be recognized from sequence alone. A third alternative is that it could belong to a structurally novel class (12). A fourth alternative is that the specific charging of Cys onto $tRNA^{Cys}$ is an additional function of prolyl-tRNA synthetase (ProRS) in these organisms.

This fourth alternative was proposed after the purification of a Cys charging activity in *M. jannaschii*. The purification led to the isolation of a protein that was identified as a canonical ProRS on the basis of N-terminal amino acid sequencing. Subsequent cloning and overexpression of the archaeal *proS* gene product in *Escherichia coli* showed that *M. jannaschii* ProRS could indeed charge cysteine onto *M. jannaschii* unfractionated tRNA [with ≈5- (13) or 36-fold (14) lower efficiency than its charging of proline]. These results, along with complementation experiments of a temperature-sensitive *E. coli cysS* strain, led to the proposal that *M. jannaschii* ProRS functioned *in vivo* as both a CysRS and a ProRS and thus had dual specificity (14, 15).

If a ProRS is to act as a dual specificity enzyme *in vivo*, it must satisfy severe structural constraints. Its active site must be capable of binding and activating both cysteine and proline to form their respective adenylates. The enzyme must also bind to both $tRNA^{Cys}$ and to $tRNA^{Pro}$, and must specifically transfer the activated amino acid only to the cognate tRNA. Failure to make the proper discrimination would lead to mischarged tRNA and errors in protein translation. In addition, all of these reactions must be performed at rates comparable to the needs of the cell for protein synthesis.

Here we describe the structures of *M. jannaschii* ProRS, and *M. thermautotrophicus* ProRS in apo form and in complex with cysteine, proline, and alanine adenylate analogues. These structures show that these adenylates bind in the same orientation within a single active site and induce virtually identical protein conformations. Because the cysteinyl- and prolyl-adenylate analogues bind with comparable, nanomolar affinity, the archaeal ProRS enzymes do not appear to have dual specificity during amino acid activation with ATP. An additional structural consequence of identical adenylate bound conformations is that the enzyme presents indistinguishable intermediates to the 3′ end of a bound tRNA. Thus, it appears that these ProRS enzymes both charge and mischarge tRNA, rather than being dual-specific tRNA synthetases. Consequently, our structural study is compatible with recent biochemical experiments, which indicate that these and other ProRS enzymes efficiently mischarge $tRNA^{Pro}$ with cysteine but do not charge $tRNA^{Cys}$ (16, 17).

Methods

***M. jannaschii* ProRS Purification and Crystallization.** pET15b MJproS was used to transform *E. coli* strain BL21-CodonPlus-RIL (Stratagene). After a 4-h induction, cells were sonicated in 50 mM Tris·HCl (pH 8.0), 500 mM NaCl, 5 mM $MgCl_2$, 0.1% Triton X-100, 5 mM 2-mercaptoethanol, 1 mM benzamidine, 10 mg/ml

Abbreviations: ProRS, prolyl-tRNA synthetase; CysRS, cysteinyl-tRNA synthetase; rmsd, rms deviation.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 1NJ1, 1NJ2, 1NJ5, 1NJ6, and 1NJ8).

[§]S.K. and W.D.K. contributed equally to this work.

**To whom correspondence should be addressed. E-mail: eatherton@csb.yale.edu.

BIOPHYSICS

Table 1. Crystallographic statistics

	Data collection						Model refinement				
Data set*	Resolution, Å†	R_{merge}‡, %	I/σ§	Completeness, %	Unique refl.	Red¶	R_{cryst}‖, %	R_{free}‖, %	rmsd bond length, Å	rmsd bond angle	PDB ID code
Mj Apo	20.0–3.2 (3.31–3.20)	10.2	7.9 (2.2)	86.9 (80.3)	30,271	2.5	23.0	30.1	0.008	1.4	1NJ8
Mt Apo	30.0–3.1 (3.21–3.10)	6.6	22.8 (2.0)	99.6 (99.7)	35,918**	4.7	23.9	27.3	0.009	1.4	1NJ2
Mt CysAMS	50.0–2.55 (2.64–2.55)	6.1	48.9 (2.6)	100 (100)	33,200	14.8	22.5	25.7	0.007	1.3	1NJ1
Mt ProAMS	20.0–2.8 (2.90–2.80)	7.9	24.9 (2.1)	99.8 (99.5)	25,241	7.6	22.2	25.4	0.008	1.3	1NJ5
Mt AlaAMS	30.0–2.85 (2.95–2.85)	12.4	14.4 (2.4)	99.7 (99.6)	23,693	7.0	22.3	25.6	0.007	1.3	1NJ6

Mj, M. jannaschii; Mt, *M. thermautotrophicus*; Cys-, Pro-, and AlaAMS, cysteinyl-sulfamoyl-, prolyl-sulfamoyl-, and alanyl-sulfamoyl-adenylate.
†Values in parentheses are calculated for the highest resolution bin in the data.
‡$R_{merge} = \{\Sigma_{hkl}\Sigma|I_i(hkl) - \langle I\rangle(hkl)|\}/\{\Sigma_{hkl}\Sigma_i I_i(hkl)\}$ where $I_i(hkl)$ is the ith measured intensity and $\langle I\rangle(hkl)$ is the mean intensity measured for Miller index (hkl).
§Intensity/error.
¶Redundancy.
‖$R_{cryst} = \{\Sigma_{hkl}||F_{obs}(hkl)| - |F_{calc}(hkl)||\}/\{\Sigma_{hkl}|F_{obs}(hkl)|\}$; $R_{free} = R_{cryst}$ calculated over a test set of reflections not included in refinement.
**Friedel pairs of reflections are unmerged.

lysozyme, and 10% glycerol and centrifuged. The supernatant was then heat treated at 55°C for 30 min. After centrifugation, the overexpressed protein was purified over a nickel affinity column. It was then dialyzed into 20 mM potassium phosphate buffer (pH 6.8) and 10% glycerol and loaded onto a Mono S column. To remove the N-terminal tag, the enzyme was digested with biotinylated thrombin. Streptavidin agarose was used to remove the thrombin, and nickel affinity resin removed residual uncut synthetase. Several changes of buffer during concentration yielded a final solution of ≈5 mg/ml protein in 20 mM Tris (pH 8.5), 200 mM NaCl, and 5 mM 2-mercaptoethanol. Crystals were obtained in vapor-diffusion experiments by mixing protein with an equal volume of well solution containing 0.2 M ammonium sulfate, 20–30% polyethylene glycol (PEG) 4000, 5 mM 2-mercaptoethanol, and 0.1 M NaOAc (pH 4.5). Crystals grew within a few days at 12–25°C as very thin plates. The unit cell dimensions were $a = 108.94$ Å, $b = 104.85$ Å, and $c = 91.75$ Å, with $\beta = 93.63°$, in space group $P2_1$. The crystals were irreproducible, and only grew from a single protein preparation. Crystals were soaked in 100 mM NaOAc pH (4.5), 50 mM ammonium sulfate, 50 mM NaCl, 25% PEG 4000, 5 mM 2-mercaptoethanol, and 20% ethylene glycol for 5 min before freezing in propane.

***M. thermautotrophicus* ProRS Purification and Crystallization.** *M. thermautotrophicus* ProRS was cloned into a pET15b vector. Cells were lysed in 50 mM Tris (pH 7.5), 500 mM NaCl, and 1 mM 2-mercaptoethanol supplemented with complete EDTA-free Protease Inhibitors (Roche Diagnostics). After centrifugation, the supernatant was incubated at 55°C for 15 min. The supernatant from this heat treatment was loaded onto a nickel affinity column. The column was then washed with 10 mM Tris (pH 7.5), 500 mM NaCl, and 1 mM 2-mercaptoethanol, and the protein was eluted with a buffer containing 750 mM imidazole. Peak fractions were dialyzed, without stirring, overnight into 10 mM Tris (pH 7.5), 100 mM NaCl, and 5 mM 2-mercaptoethanol. Under these conditions, some of the protein precipitated to form microcrystals. The salt concentration in the slurry was raised to 1 M to partially resolubilize the protein (to a concentration of 5.7 mg/ml). The protein solution was clarified using a benchtop centrifuge before crystallization trials. In some experiments, the protein was mixed with 2 mM 5′-*O*-[*N*-(L-prolyl)sulfamoyl]-adenosine (prolyl-sulfamoyl-adenylate), 2 mM 5′-*O*-[*N*-(L-cysteinyl)sulfamoyl]adenosine (cysteinyl-sulfamoyl-adenylate), or 10 mM 5′-*O*-[*N*-(L-alanyl)sulfamoyl]adenosine (alanyl-sulfamoyl-adenylate), which are isosteric but nonhydrolyzable analogues of aminoacyl adenylates (RNA-TEC, Leuven, Belgium). To prevent rapid nucleation in the presence of some of these analogues, the protein solution was further diluted to 3.8 mg/ml by the addition of NaCl to 2 M. Equal volumes of protein solution and precipitant buffer (20 mM Tris, pH 7.5/5 mM 2-mercaptoethanol/50–250 mM $MgCl_2$) were mixed and equilibrated against precipitant buffer by vapor diffusion at 20°C. Hexagonal rods appeared, typically in 12–48 h. The majority of these were in space group $p6_522$. The unit cell dimensions vary slightly between crystals (those of the cysteinyl-sulfamoyl-adenylate cocrystal are $a = b = 143.75$ Å and $c = 163.89$ Å); crystals were also obtained in space groups $p3_221$ (usually with a hexagonal bipyramidal morphology) and $P6_5$, with very similar cell dimensions. Crystals were flash frozen in propane after incubation for 5 min in 20 mM Tris (pH 7.5), 0.1 M $MgCl_2$, 0.2 M NaCl, and 37.5–40% ethylene glycol.

Structure Determination and Refinement. Data were processed using DENZO or HKL2000 software and scaled with SCALEPACK (18). Starting with the coordinates of ProRS from *Thermus thermophilus* (19), the structure of *M. jannaschii* ProRS was solved by molecular replacement using AMORE (20). The asymmetric unit contains two dimers, which are related by a noncrystallographic twofold axis parallel to the c* axis. CNS was used for refinement (21). The two dimers are nearly identical, with a Cα rms deviation (rmsd) of 0.02 Å. The monomers within a dimer show a larger Cα rmsd of 0.56 Å. As a result, fourfold noncrystallographic restraints were useful early in the refinement but were then gradually eliminated; reduced twofold restraints between dimers remained essential given the limited resolution of the data. Cycles of positional, individual B factor, and simulated annealing refinement, together with manual rebuilding in O (22), resulted in a final model containing two N-terminal tag residues in addition to all residues from Glu-2 to Tyr-455 (15,088 atoms) and 93 water molecules.

Starting with a newly refined model of *M. jannaschii* ProRS, the structure of *M. thermautotrophicus* ProRS was solved by molecular replacement using AMORE (20). The asymmetric unit contains a monomer of synthetase. The models were rebuilt where necessary by using O (22) and refined using CNS (21). The final models contain all residues from Glu-19 to Tyr-481, except in the case of the apo enzyme, where the residues Lys-85 to Asp-95 were also disordered. These models contain 3,692–3,782 protein atoms, zinc atoms, and inhibitors, and 13–30 water molecules.

The data collection and refinement statistics for these structures are summarized in Table 1. Renderings of the structural models were generated using MOLSCRIPT (23) and SPOCK (24).

Activity Assays. ATP-PP_i exchange reaction assays were carried out under previously described conditions (17), except that

reactions were run at 35°C and in the presence of varying amounts of either proline or cysteine [60 nM *M. thermautotrophicus* ProRS/100 mM Tris (pH 7.5)/10 mM $MgCl_2$/250 mM NaCl/2 mM KF/2 mM ATP/10 mM 2-mercaptoethanol/2 mM $^{32}PP_i$]. Reaction products were bound to activated charcoal (acid-washed, Norit, Atlanta) and collected onto Whatman GF/C filter paper. ATP-PP_i exchange reactions in the presence of either prolyl-sulfamoyl- or cysteinyl-sulfamoyl-adenylates were performed under the same conditions. Aliquots of the reactions were collected from 0 to 10 min to determine the initial rate measurements for the reactions. Double-reciprocal Lineweaver–Burk plots were used to estimate V_{max} and K_i values for the adenylate analogues, and estimates were verified by nonlinear regression of velocity curves.

Aminoacylation assays of $tRNA^{Pro}$ were performed at 35°C in the presence of either [^{35}S]cysteine (50 Ci/mmol, Amersham Pharmacia; 1 Ci = 37 GBq) or [^{3}H]proline (10 Ci/mmol, Amersham Pharmacia). The reaction buffer consisted of 2 mM ATP, 5 mM 2-mercaptoethanol, 20 μM labeled amino acid, 10 mM $MgCl_2$, 50 mM Tris (pH 7.5), 20 nM *M. thermautotrophicus* $tRNA^{Pro}$ transcribed *in vitro*, and 20–200 μg/ml purified, recombinant *M. thermautotrophicus* ProRS in a total volume of 50 μl (15). After 30 min, 10-μl aliquots were spotted on Whatman 3MM filter discs, washed three times with ice-cold 5% trichloroacetic acid containing 0.05% proline, washed once with cold 95% ethanol, and dried and transferred to vials for determination of radioactive counts.

Results and Discussion

Five structures are reported here (*i*) apo *M. jannaschii* ProRS, (*ii*) apo *M. thermautotrophicus* ProRS and *M. thermautotrophicus* ProRS cocrystallized with isosteric but nonhydrolyzable analogues (25) of (*iii*) prolyl-, (*iv*) cysteinyl-, or (*v*) alanyl-adenylates. These structures were solved by molecular replacement using the coordinates of *T. thermophilus* ProRS (19). The *M. jannaschii* ProRS crystallizes in space group $p2_1$ with two dimers in each asymmetric unit and diffracts to 3.2 Å. The *M. thermautotrophicus* structures crystallized in space group $p6_522$ with a monomer in the asymmetric unit and diffract to between 2.55 and 3.1 Å resolution (Table 1). In solution, *M. thermautotrophicus* ProRS is dimeric, and the dimer interface is observed in the crystal structure as a consequence of a crystallographic twofold axis.

Comparison of ProRS Structures. On the basis of primary sequence similarities, signature peptide motifs, and domain organization, ProRS enzymes are class IIa tRNA synthetases (26). ProRS sequences have been further subdivided into two groups: (*i*) archaeal/eukaryotic or (*ii*) prokaryotic. Both of these groups share a core consisting of a catalytic domain and an anticodon-binding domain (Fig. 1). In addition to these core regions, the archaeal/eukaryotic group also possesses a C-terminal domain, whereas the prokaryotic group contains an insertion domain between motifs 2 and 3 of the catalytic domain.

The structure of *T. thermophilus* ProRS confirmed that this eubacterial organism contains an archaeal/eukaryotic type of ProRS (19, 27). The sequences of *T. thermophilus*, *M. jannaschii*, and *M. thermautotrophicus* ProRS are thus homologous: whereas the ProRS molecules from *T. thermophilus* and *M. jannaschii* share 39% sequence identity, *T. thermophilus* and *M. jannaschii* are 41% identical and *M. jannaschii* and *M. thermautotrophicus* are 60% identical. As anticipated, their structures are also very similar (Fig. 1); the rmsd of Cα positions between *T. thermophilus* ProRS and *M. jannaschii* ProRS is 1.6 Å (for 418 of 456 Cα positions), that between *T. thermophilus* ProRS and *M. thermautotrophicus* ProRS is 2.0 Å (for 423 of 452 Cα positions), and that between *M. thermautotrophicus* ProRS and *M. jannaschii* ProRS is 1.4 Å (for 428 of 452 Cα positions).

The largest structural differences among these three enzymes

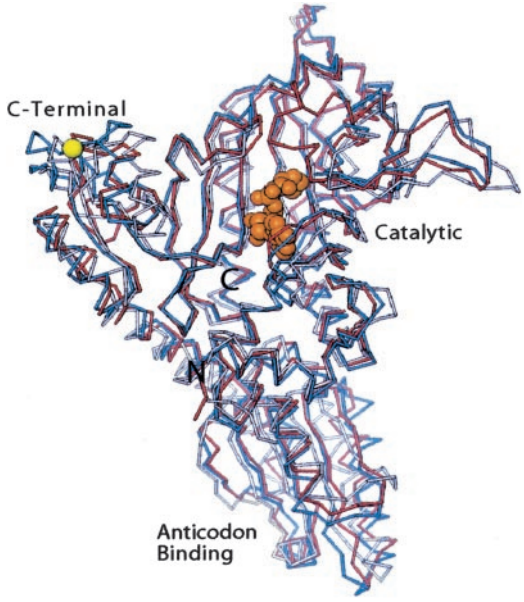

Fig. 1. Cα traces of the superimposed structures of apo *M. thermautotrophicus* ProRS (cyan), apo *M. jannaschii* ProRS (red), and apo *T. thermophilus* ProRS (gray; PDB ID code 1HC7). The catalytic, anticodon-binding, and C-terminal domains are indicated. The N- and C-terminal residues of the *M. thermautotrophicus* ProRS structure are labeled N and C, respectively. The active site is marked by an orange, nonhydrolyzable analogue of prolyl-adenylate taken from its cocrystal structure with *M. thermautotrophicus* ProRS. The zinc bound by the C-terminal domain of *M. thermautotrophicus* ProRS is colored yellow.

are in their C-terminal domains. *T. thermophilus* and *M. thermautotrophicus* ProRS both contain a zinc atom chelated by four cysteine residues in their C terminus, whereas *M. jannaschii* ProRS, which does not bind zinc, lacks these residues. However, these differences do not appear to have functional consequences for aminoacylation because the zinc-binding regions of the C-terminal domains are distant from both the active site and the expected tRNA-binding surface. Furthermore, these C-terminal domains appear not to be involved in editing of mischarged amino acids because crystals grown in the presence of millimolar concentrations of either cysteinyl-sulfamoyl-adenylate or alanyl-sulfamoyl-adenylate show that these analogues bind only in one location, the active site, and that no secondary binding sites are observed as in the case of synthetases with separate editing domains (28–33).

Binding of Prolyl- and Cysteinyl-Adenylate Analogues. Prolyl-sulfamoyl-adenylate binds to *M. thermautotrophicus* ProRS in identically the same manner as it does *T. thermophilus* ProRS (34). All of the residues shown in Fig. 2*a* are identical in *T. thermophilus*, *M. thermautotrophicus*, and *M. jannaschii* ProRS, except for three conservative substitutions in *T. thermophilus* (Phe-166 to Trp-158, Cys-265 to Ser-258, and Tyr-266 to Trp-259). The interactions between these residues and prolyl-adenylates also appear to be conserved. The adenylate-bound structures are consistent with biochemical experiments, which show that the *M. jannaschii* ProRS Glu-103 to Ala mutation completely abolishes prolylation activity and significantly reduces cysteinylation activity (Fig. 2 *a* and *b*: the corresponding residue is Glu-119 in *M. thermautotrophicus*; ref. 13). The shapes and sizes of the prolyl-binding pockets also appear to be conserved across all three species.

A number of loops near the active site of the *M. thermautotrophicus* apo ProRS structure move toward or become or-

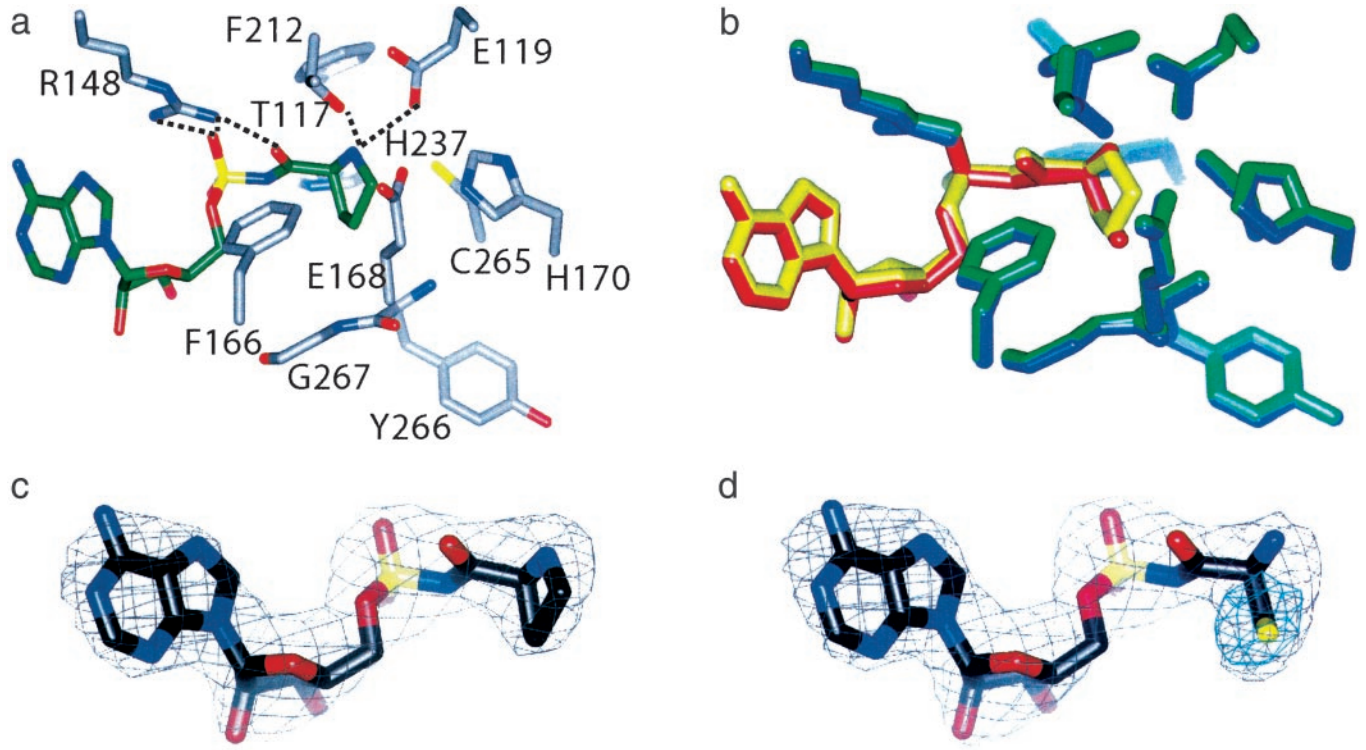

Fig. 2. (*a*) Prolyl-sulfamoyl-adenylate bound to the active site of *M. thermautotrophicus* ProRS. All residues within 4 Å of the amino acid moieties of the inhibitors are shown (Phe-166, Glu-168, His-170, Phe-212, His-237, Cys-265, Tyr-266, and Gly-267). Hydrogen bonds are indicated as black dotted lines. (*b*) The active-site structures of *M. thermautotrophicus* ProRS with prolyl-sulfamoyl-adenylate (ligand in yellow and protein in green) and cysteinyl-sulfamoyl-adenylate (ligand in red and protein in dark blue). The superimpositions were made by aligning the structures over all Cα atoms. Unbiased maps showing electron density for the bound inhibitors: prolyl-sulfamoyl-adenylate (*c*) and cysteinyl-sulfamoyl-adenylate (*d*). The gray maps are $F_{obs} - F_{calc}$ maps made by omitting the inhibitors and contoured at the level of 4 standard deviations. To confirm the location of the thiol moiety in the cysteinyl analogue structure, data were collected on a crystal containing alanyl-sulfamoyl-adenylate. A difference map between the cysteinyl and alanyl data is shown in cyan [F_{obs}CysAMS $-$ F_{obs}AlaAMS (Cys- and AlaAMS, cysteinyl-sulfamoyl- and alanyl-sulfamoyl-adenylate), using phases calculated from the alanyl structure and contoured at the level of 15 standard deviations].

dered on binding prolyl-sulfamoyl-adenylate. Similar changes have been described in the *T. thermophilus* ProRS:prolyl-sulfamoyl-adenylate structure (34). Thus, the main chain of a loop containing *M. thermautotrophicus* residues 211–217 moves by up to 1.5 Å toward the intermediate. A second loop, homologous to the highly conserved "TXE" loop of class II synthetases (35), also moves a similar distance toward the intermediate. The largest conformational changes occur within residues 148–154, with main chain movements of up to 4.6 Å toward the active site and bound sulfamoyl-adenylate. The ordering loop (residues 85–95), which is disordered in the apo crystal structure, assumes a helical conformation in the presence of adenylate analogue, albeit with high B factors. The conformational changes induced by the binding of prolyl-adenylate therefore appear to be conserved between these two species.

It has been proposed that the 3′ end of tRNA can be bound productively by *T. thermophilus* ProRS only when the ordering loop is in a helical conformation (34). Because the helical conformation was observed only in the presence of prolyl-adenylate, this provided a plausible means for specificity: only when the prolyl-adenylate was formed would tRNA assume a catalytically competent conformation (34). However, alanyl-, cysteinyl-, and prolyl-adenylates all appear to induce the ordering of this loop in *M. thermautotrophicus* ProRS. Thus, the assumption of a helical conformation by this loop does not signal that cognate aminoacyl-adenylate has been bound by *M. thermautotrophicus* ProRS, but rather only that an aminoacyl-adenylate has bound. In addition, in *M. jannaschii* ProRS this loop appears to be at least partly ordered in the absence of any substrate.

The interactions between the cysteinyl-adenylate analogue and *M. thermautotrophicus* ProRS are also remarkably similar to those between the prolyl-adenylate analogue and *M. thermautotrophicus* ProRS (Fig. 2*b*). The rmsd between all corresponding protein atoms shown in Fig. 2*b* is only 0.17 Å. The only amino acid side chain that appears to contact one of the analogues, and not the other, is His 170. This residue contacts the Cδ of the prolyl-adenylate, but is too far to contact any of the cysteinyl-adenylate atoms. This structure also eliminates the possibility that the high affinity of *M. thermautotrophicus* ProRS for cysteine adenylate is due to interactions between the thiols of the aminoacyl-adenylate and Cys-265, because these atoms are not in contact with each other. Thus, *M. thermautotrophicus* ProRS contains only one active site for these intermediates in the activation process.

Affinity of *M. thermautotrophicus* ProRS for Cysteinyl- and Prolyl-Adenylate Analogues. The selectivity of tRNA synthetases is, in part, determined by the relative affinities with which they bind different amino acids, as well as the cellular concentrations of those amino acids. The *in vivo* concentrations of alanine, cysteine, or proline are currently unknown in most prokaryotes. The K_d values of amino acids for archaeal ProRS enzymes have not been measured, although K_m values have been determined. In the ATP-PP$_i$ exchange reaction, the previously determined K_m values for proline and cysteine are 285 and 90 μM, respectively, for *M. jannaschii* (13) and 260 and 50 μM, respectively, for

Table 2. Kinetic constants for ATP-PP_i exchange reaction

	K_m or K_{app}, μM	K_i, nM	V_{max}, nM·min^{-1}
Proline	67 ± 11		2.1
ProAMS	110 ± 22	50 ± 18	2.0 ± 0.3
CysAMS	128 ± 28	25 ± 4	2.1 ± 0.4

M. thermautotrophicus (16). Alanine, perhaps because of its smaller size, appears to have much higher K_m values [with K_m values for various ProRS enzymes between 31 and 500 mM (16)]. This suggests that much of the high selectivity of ProRS enzymes against alanine [6,800-fold in *M. jannaschii* (36)] occurs during amino acid activation.

We used the ATP-PP_i exchange assay for amino acid activation (37) to determine the affinities of prolyl- and cysteinyl-sulfamoyl adenylates for *M. thermautotrophicus* ProRS (Table 2). As run in the reverse ATP-PP_i exchange reaction, the cysteinyl-sulfamoyl- and prolyl-sulfamoyl-adenylates are competitive inhibitors of proline activation, with K_i values of 25 and 50 nM, respectively. Thus, these two analogues bind tightly to *M. thermautotrophicus* ProRS with only 2-fold selectivity between them.

Homology Modeling of tRNA onto *M. thermautotrophicus* ProRS. This single mode of binding by alanyl-, cysteinyl-, and prolyl-adenylates to *M. thermautotrophicus* ProRS appears to represent a competent intermediate state for the subsequent aminoacylation of the acceptor end of tRNA. The 3′-terminal CCA is not ordered in the structures of *T. thermophilus* ProRS bound to tRNA. However, a homologous structure of *E. coli* threonyl-tRNA synthetase (ThrRS) bound to tRNA and AMP has been determined (31). Superimposing this ternary structure on the corresponding atoms of our binary *M. thermautotrophicus* ProRS structures yields an overall Cα rmsd of 1.9 Å (using 283–402 Cα positions and omitting the editing domain of ThrRS, which has no counterpart in archaeal ProRS). There are few steric clashes between this noncognate, homology-modeled tRNA and *M. thermautotrophicus* ProRS. In particular, no clashes between the main chain of ProRS and the tRNA occur except in the anticodon loop. The modeled $tRNA^{Thr}$ has the 3′ hydroxyl of its terminal adenosine ribose poised for nucleophilic attack on the carbonyl carbon of the prolyl moiety (Fig. 3). The distance between these two atoms in the model is only 3.6 Å. This homology model therefore suggests that no major conformational changes are required in the active site to accommodate the acceptor stem of tRNA. Because the cysteinyl- and alanyl-adenylate analogues bind to ProRS nearly identically, as would the tRNA, an identical displacement would be expected for their charging of tRNA. Thus, charging of $tRNA^{Pro}$ with cysteine should be as facile as its aminoacylation with proline.

Implications for Specific Dual Charging of $tRNA^{Cys}$ and $tRNA^{Pro}$. The ProRS enzymes described above appear unable to discriminate between cysteine and proline. The chemistry of proline and cysteine are dissimilar, but the two amino acids are similar in size. The van der Waal's volume of cysteine is 86 $Å^3$ and that of proline is 90 $Å^3$ (38). The two differ by less than the volume of a methylene group. Significant substrate discrimination between these two amino acids would not be expected on the basis of size alone, but would likely exploit their differences in shape and in chemistry. For example, *E. coli* CysRS specifically coordinates the sulfhydryl of cysteine by using an active-site zinc metal ion (4). As the cocrystal structures described above show, cysteinyl- and prolyl-adenylates can assume nearly identical conformations and shapes, and both can be accommodated by the active-site pocket of *M. thermautotrophicus* ProRS. This active-site pocket

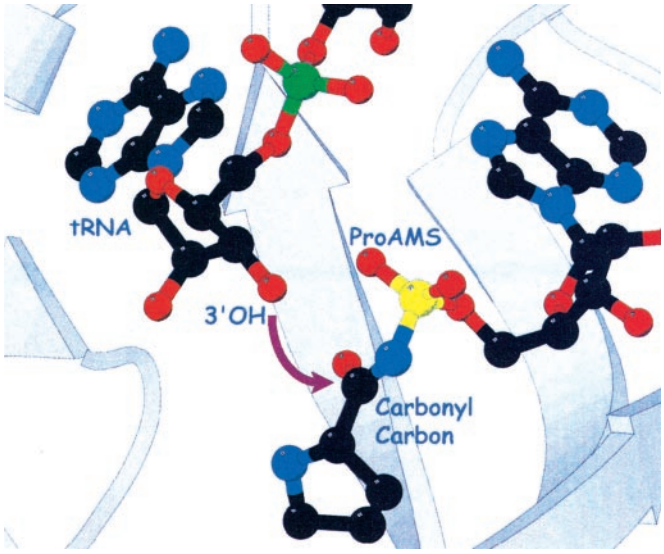

Fig. 3. View of the ProRS active site complexed with prolyl-sulfamoyl-adenylate and docked tRNA. The homology-modeled structure of the tRNA is based on the superposition of the Cα atoms of *E. coli* ThrRS (in ternary complex with AMP and $tRNA^{Thr}$) with those of *M. thermautotrophicus* ProRS. The 3′ hydroxyl of the terminal adenosine is 3.6 Å away from the carbonyl carbon of the prolyl-sulfamoyl-adenylate, and positioned for nucleophilic attack. Part of the active-site protein structure has been removed for this rendition in MOLSCRIPT.

does not contain a metal ion and, surprisingly, does not appear to chemically distinguish between the two amino acids through specific recognition of imino or sulfhydryl moieties. The structures are thus compatible with biochemical experiments that showed that these archaeal ProRS enzymes can efficiently charge either cysteine or proline onto $tRNA^{Pro}$ (15–17).

These intermediate-bound structures of *M. thermautotrophicus* ProRS do not support the hypothesis of dual specific charging. An essential element of this hypothesis is that ProRS should charge $tRNA^{Pro}$ and $tRNA^{Cys}$ only with their cognate amino acids and not generate mischarged products. This specificity could arise during amino acid activation, tRNA charging, or, if mischarging occurs, by a process of editing.

Biochemical experiments have shown that both prolyl activation and cysteinyl activation can occur efficiently in the absence of tRNA (39). The activation step of the ProRS reaction thus cannot be a major determinant for dual specificity. Our results, consistent with these biochemical observations, show that both cysteinyl-sulfamoyl- and prolyl-sulfamoyl-adenylate bind to the same active site, with comparable affinity, and induce nearly identical conformational changes in the enzyme. Furthermore, no cysteine editing activity has been reported in any ProRS to date, and in the cocrystal structures described above, all of the analogues bind only at the active site, suggesting there is no distinct editing domain in archaeal ProRS enzymes. In bacterial replicative polymerases editing (exonuclease), domains can be found either on the same peptide as synthetic (polymerase) domains or can be found on separate polypeptides (40). By analogy, it is possible that an editing activity associated with archaeal ProRS enzymes might be the product of a separate gene.

The structures presented here do not support the hypothesis that tRNA specificity is determined by the identity of the bound aminoacyl-adenylate during the charging step of aminoacylation. The protein conformation of *M. thermautotrophicus* ProRS is identical (to within experimental error) whether cysteinyl- or prolyl-adenylate is bound (the rmsd between the two is 0.38 Å

over all C_α atoms). Homology modeling indicates that no significant structural change is required for tRNA to be bound in a catalytically competent conformation. Taken together, these two structural observations provide evidence against specific tRNA charging (the incoming $tRNA^{Pro}$ would not be able to distinguish which aminoacyl-adenylate was bound by the enzyme). These structural observations are also consistent with recent biochemical results, which show that *M. jannaschii* ProRS does in fact mischarge $tRNA^{Pro}$ with cysteine and can not charge native $tRNA^{Cys}$ *in vitro* (16, 17).

The original experiments with *M. jannaschii* ProRS that formed the basis for the dual tRNA specificity hypothesis can be entirely explained by its efficient mischarging of $tRNA^{Pro}$ with cysteine. The purified enzyme was shown to aminoacylate unfractionated *M. jannaschii* tRNA with cysteine (15). Although it was assumed at the time that this tRNA was $tRNA^{Cys}$, it seems very likely, given the facile mischarging of $tRNA^{Pro}$ with cysteine (16, 17), that the tRNA charged with cysteine was in fact $tRNA^{Pro}$. How Cys-$tRNA^{Cys}$ is produced in these archaeal hyperthermophiles and how they avoid producing misacylated Cys-$tRNA^{Pro}$ *in vivo* thus remain intriguing questions.

Data for the apo *M. thermautotrophicus* and alanyl-sulfamoyl-adenylate data sets were collected at Cornell High Energy Synchrotron Source (CHESS; Cornell University, Ithaca, NY) beamline A1, the cysteinyl-sulfamoyl-adenylate data were collected at Advanced Photon Source (APS; Argonne National Laboratory, University of Chicago) ID19, the prolyl-sulfamoyl-adenylate data were collected at APS ID14, and the apo *M. jannaschii* data were collected at Advanced Light Source (Berkeley Lab, Berkeley, CA) 5.0.1 and CHESS F1. We thank Stephen Cusack for the coordinates of *T. thermophilus* ProRS before Protein Data Bank deposition; Ivan Ahel, Alexandre Ambrogelly, and Whitney Yin for helpful discussions and activity assays; and Janice Pata for comments on the manuscript. This work was supported in part by National Institutes of Health Grants GM22778 (to T.A.S.) and GM22854 (to D.S.).

1. Ibba, M. & Söll, D. (2000) *Annu. Rev. Biochem.* **69,** 617–650.
2. Fersht, A. R. & Dingwall, C. (1979) *Biochemistry* **18,** 2627–2631.
3. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science* (Freeman, New York).
4. Newberry, K. J., Hou, Y. M. & Perona, J. J. (2002) *EMBO J.* **21,** 2778–2787.
5. Moras, D. (1992) *Trends Biochem. Sci.* **17,** 159–164.
6. Moras, D. (1993) *Biochimie* **75,** 651–657.
7. Cusack, S. (1997) *Curr. Opin. Struct. Biol.* **7,** 881–889.
8. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996) *Science* **273,** 1058–1073.
9. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997) *J. Bacteriol.* **179,** 7135–7155.
10. Slesarev, A. I., Mezhevaya, K. V., Makarova, K. S., Polushin, N. N., Scherbinina, O. V., Shakhova, V. V., Belova, G. I., Aravind, L., Natale, D. A., Rogozin, I. B., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 4644–4649.
11. Kim, H. S., Vothknecht, U. C., Hedderich, R., Celic, I. & Söll, D. (1998) *J. Bacteriol.* **180,** 6446–6449.
12. Fabrega, C., Farrow, M. A., Mukhopadhyay, B., de Crecy-Lagard, V., Ortiz, A. R. & Schimmel, P. (2001) *Nature* **411,** 110–114.
13. Stathopoulos, C., Jacquin-Becker, C., Becker, H. D., Li, T., Ambrogelly, A., Longman, R. & Söll, D. (2001) *Biochemistry* **40,** 46–52.
14. Lipman, R. S. A., Sowers, K. R. & Hou, Y. M. (2000) *Biochemistry* **39,** 7792–7798.
15. Stathopoulos, C., Li, T., Longman, R., Vothknecht, U. C., Becker, H. D., Ibba, M. & Söll, D. (2000) *Science* **287,** 479–482.
16. Ahel, I., Stathopoulos, C., Ambrogelly, A., Sauerwald, A., Toogood, H., Hartsch, T. & Söll, D. (2002) *J. Biol. Chem.* **277,** 34743–34748.
17. Ambrogelly, A., Ahel, I., Polycarpo, C., Bunjun-Srihari, S., Krett, B., Jacquin-Becker, C., Ruan, B., Köhrer, C., Stathopoulos, C., RajBhandary, U. L. & Söll, D. (2002) *J. Biol. Chem.* **277,** 34749–34754.
18. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
19. Yaremchuk, A., Cusack, S. & Tukalo, M. (2000) *EMBO J.* **19,** 4745–4758.
20. Navaza, J. (2001) *Acta Crystallogr. D* **57,** 1367–1372.
21. Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstlev, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
22. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
23. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
24. Christopher, J. A. (1998) SPOCK: The Structural Properties Observation and Calculation Kit, Program Manual (Texas A&M Univ., College Station).
25. Ueda, H., Shoku, Y., Hayashi, N., Mitsunaga, J., In, Y., Doi, M., Inoue, M. & Ishida, T. (1991) *Biochim. Biophys. Acta* **1080,** 126–134.
26. Cusack, S. (1995) *Nat. Struct. Biol.* **2,** 824–831.
27. Cusack, S., Yaremchuk, A., Krikliviy, I. & Tukalo, M. (1998) *Structure (London)* **6,** 101–108.
28. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P. & Yokoyama, S. (1998) *Science* **280,** 578–582.
29. Silvian, L. F., Wang, J. & Steitz, T. A. (1999) *Science* **285,** 1074–1077.
30. Cusack, S., Yaremchuk, A. & Tukalo, M. (2000) *EMBO J.* **19,** 2351–2361.
31. Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C. & Moras, D. (2000) *Cell* **103,** 877–884.
32. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G. & Yokoyama, S. (2000) *Cell* **103,** 793–803.
33. Alexander, R. W. & Schimmel, P. (2001) *Prog. Nucleic Acid Res. Mol. Biol.* **69,** 317–349.
34. Yaremchuk, A., Tukalo, M., Grøtli, M. & Cusack, S. (2001) *J. Mol. Biol.* **309,** 989–1002.
35. Belrhali, H., Yaremchuk, A., Tukalo, M., Berthet-Colominas, C., Rasmussen, B., Bosecke, P., Diat, O. & Cusack, S. (1995) *Structure (London)* **3,** 341–352.
36. Beuning, P. J. & Musier-Forsyth, K. (2001) *J. Biol. Chem.* **276,** 30779–30785.
37. Yarus, M. & Berg, P. (1969) *J. Mol. Biol.* **42,** 171–189.
38. Creighton, T. E. (1996) *Proteins Structure and Molecular Properties* (Freeman, New York).
39. Lipman, R. S. A., Beuning, P. J., Musier-Forsyth, K. & Hou, Y. M. (2002) *J. Mol. Biol.* **316,** 421–427.
40. Huang, Y. P. & Ito, J. (1999) *J. Mol. Evol.* **48,** 756–769.