Structural basis for lysidine formation by ATP pyrophosphatase accompanied by a lysine-specific loop and a tRNA-recognition domain

Kotaro Nakanishi*, Shuya Fukai*, Yoshiho Ikeuchi†, Akiko Soma‡, Yasuhiko Sekine‡, Tsutomu Suzuki†, and Osamu Nureki*§¶‖

*Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan; †Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan; ‡Department of Life Science, College of Science, Rikkyo (St. Paul's) University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan; §Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan; and ¶RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

Edited by Dieter Söll, Yale University, New Haven, CT, and approved April 5, 2005 (received for review February 7, 2005)

Lysidine, a lysine-combined modified cytidine, is exclusively located at the anticodon wobble position (position 34) of eubacterial $tRNA^{Ile}_2$ and not only converts the codon specificity from AUG to AUA, but also converts the aminoacylation specificity from recognition by methionyl-tRNA synthetase to that by isoleucyl-tRNA synthetase (IleRS). Here, we report the crystal structure of lysidine synthetase (TilS) from *Aquifex aeolicus* at 2.42-Å resolution. TilS forms a homodimer, and each subunit consists of the N-terminal dinucleotide-binding fold domain (NTD), with a characteristic central hole, and the C-terminal globular domain (CTD) connected by a long α-helical linker. The NTD shares striking structural similarity with the ATP-pyrophosphatase domain of GMP synthetase, which reminds us of the two-step reaction by TilS: adenylation of C34 and lysine attack on the C2 carbon. Conserved amino acid residues are clustered around the NTD central hole. Kinetic analyses of the conserved residues' mutants indicated that C34 of $tRNA^{Ile}_2$ is adenylated by an ATP lying across the NTD central hole and that a lysine, which is activated at a loop appended to the NTD, nucleophilically attacks the C2 carbon from the rear. *Escherichia coli* TilS (called MesJ) has an additional CTD, which may recognize the $tRNA^{Ile}_2$ acceptor stem. In contrast, a mutational study revealed that *A. aeolicus* TilS does not recognize the tRNA acceptor stem but recognizes the C29·G41 base pair in the anticodon stem. Thus, the two TilS enzymes discriminate $tRNA^{Ile}_2$ from $tRNA^{Met}$ by strategies similar to that used by IleRS, but in distinct manners.

crystal structure | isoleucyl-tRNA synthetase | mutagenesis | RNA modification

The accuracy of protein biosynthesis is ensured by two processes involving tRNA: One is the correct attachment of the amino acid to the 3′ terminus of tRNA, which is catalyzed by 20 aminoacyl-tRNA synthetases, and the other is codon recognition by the tRNA anticodon. The posttranscriptional modification of the anticodon wobble position (position 34) establishes the codon–anticodon pairing rule to prevent the anticodon from misreading a noncognate codon (1). Several wobble modifications are known to alter the codon specificity for precise decoding (2–5). Lysidine is a lysine-containing cytidine derivative that occurs at the wobble position of eubacterial and some organellar AUA codon-specific isoleucine tRNAs ($tRNA^{Ile}_2$) (6–8). The premodified $tRNA^{Ile}_2$ has a CAU anticodon, which also appears in premodified methionine tRNA ($tRNA^{Met}$) specific to the AUG codon. The premodified $tRNA^{Ile}_2$ reads the AUG codon and is aminoacylated by methionyl-tRNA synthetase and thus behaves as a methionine tRNA. A previous *in vitro* microsurgery experiment revealed that the replacement of lysidine 34 by an unmodified C34 in *Escherichia coli* $tRNA^{Ile}_2$ drastically reduced the isoleucine-accepting activity but increased the methionine-accepting activity (9). Thus, once C34 is modified to lysidine, the mature $tRNA^{Ile}_2$ reads the AUA codon, is aminoacylated by isoleucyl-tRNA synthetase (IleRS), and functions as an isoleucine tRNA. Therefore, the single posttranscriptional modification of C34 to lysidine 34 in $tRNA^{Ile}_2$ concomitantly converts the codon specificity from AUG to AUA and the amino acid specificity from methionine to isoleucine.

The gene encoding the enzyme that directly introduces lysine to C34 of $tRNA^{Ile}_2$ was recently identified as *tilS* ($tRNA^{Ile}$-lysidine synthetase) by genetics and proteomics techniques (10). However, the mechanism by which lysidine synthetase (TilS) synthesizes lysidine has not yet been elucidated. Here, we present the crystal structure of TilS from a thermophilic eubacterium, *Aquifex aeolicus*, at 2.42 Å. To our knowledge, this study provides the first reported structure of TilS; the structure of the *E. coli* enzyme was only deposited in the Protein Data Bank as a function-unknown protein named MesJ. The structure of the catalytic domain of *A. aeolicus* TilS closely resembles that of an "N-type" ATP pyrophosphatase (PPase), which suggests a possible molecular evolution pathway and a putative catalytic mechanism of TilS. Extensive mutant analyses based on the atomic structure provide the structural basis for the mechanism of lysidine formation by TilS. Structural and functional comparisons with *E. coli* TilS revealed that the two TilS enzymes discriminate premodified $tRNA^{Ile}_2$ from premodified $tRNA^{Met}$ by strategies similar to that used by IleRS, but in distinct manners.

Materials and Methods

Crystallization. The expression vector, pET15-TilS, was constructed by cloning the *A. aeolicus tilS* gene into the plasmid pET-15b (Novagen). A culture of *E. coli* strain BL21 codon plus (DE3), transformed with pET15-TilS, was grown to an OD_{600} of 0.5, and then the expression of TilS was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG), followed by 20 h of cultivation at 37°C. After centrifugation at 5,000 × *g* for 30 min, the harvested cells were suspended in 100 mM Tris·HCl buffer (pH 7.8) containing 10 mM KCl, 8 mM $MgCl_2$, 10 mM DTT, and 1 mM PMSF and were gently sonicated. After centrifugation at 20,000 × *g* for 30 min, the supernatant was heat-treated at 70°C for 30 min, and then the recombinant TilS was purified by chromatography on SP Sepharose, Phenyl Superose, and Resource S columns (Amersham Biosciences). The purified TilS was dialyzed against a crystallization solution composed of 10 mM Tris·HCl (pH 8.0) and 10 mM $MgCl_2$ and was concentrated. Crystals of *A. aeolicus* TilS were grown

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: IleRS, isoleucyl-tRNA synthetase; TilS, lysidine synthetase; NTD, N-terminal dinucleotide-binding fold domain; CTD, C-terminal globular domain; PPase, pyrophosphatase; rmsd, rms deviation.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1WY5).

‖To whom correspondence should be addressed. E-mail: onureki@bio.titech.ac.jp.

within 2 days at 20°C by the hanging-drop vapor diffusion method. Drops were prepared by mixing equal volumes of a 5 mg/ml TilS solution and the reservoir solution, containing 0.1 M sodium acetate (pH 4.6) and 3.5 M sodium formate. Selenomethionine-labeled TilS was prepared by the conventional method and was purified in the same manner as the WT.

Structure Determination and Refinement. The TilS crystals were harvested in 0.12 M sodium acetate buffer (pH 4.6) containing 4.2 M sodium formate and then were sequentially transferred into the buffers containing 10% and 20% glycerol at 10-min intervals for cryoprotection. The crystals were flash-cooled in a nitrogen stream at 100 K. All diffraction data sets were collected at the NW12 station at the Photon Factory (Tsukuba, Japan) and were processed with the HKL2000 suite (HKL Research, Charlottesville, VA). The data set at the peak wavelength of the multiple wavelength anomalous dispersion (MAD) experiment up to 2.63 Å was used to search for the locations of the selenium atoms, by using the program SNB (11). Several trials by SNB revealed all eight consistent peaks expected in the asymmetric unit. Subsequent phase refinements were performed with the program SHARP (12), and the model was manually built into the electron density maps by using the program O (13). The model was refined against the reflections from the native crystal up to 2.4-Å resolution by using the program CNS (14), as shown in Table 2, which is published as supporting information on the PNAS web site. Structure representations were prepared with PYMOL (15) and CUEMOL (http://cuemol.sourceforge.jp/en).

Kinetic Analyses of TilS Mutants. All of the mutations were introduced by PCR into the *A. aeolicus tilS* gene cloned in pET-15b. The TilS mutants were prepared as the WT enzyme (described above). After dialysis against 50 mM Tris·HCl (pH 7.8), 25 mM $MgCl_2$, and 2-mercaptoethanol, the WT and mutant TilS proteins were concentrated and were stored at −25°C as 50% glycerol stocks before the kinetic assays. The gene encoding *A. aeolicus* $tRNA^{Ile}_2$ was amplified by PCR and cloned into the plasmid pUC119 (Takara Shuzo, Kyoto) by using the BamHI and HindIII sites. *A. aeolicus* $tRNA^{Ile}_2$ was transcribed *in vitro* by T7 RNA polymerase at 37°C in a reaction mixture composed of 80 mM Hepes (pH 8.1)/12 mM spermidine/100 mM DTT/100 mM KCl/0.1 mg/ml BSA/10 mM each NTP/8 mM GMP/100 mM $MgCl_2$/0.24 units/μl RNase inhibitor (Takara Shuzo)/0.05 units/μl PPase (Sigma). The transcript was subjected to phenol/chloroform extraction and isopropanol precipitation and then dried. The tRNA transcript was purified by denaturing (urea) 10% PAGE, followed by HPLC using a Resource Q column (Shimadzu).

The lysidine formation assay was performed at 60°C in a reaction buffer containing 100 mM Tris·HCl (pH 7.8), 10 mM $MgCl_2$, 10 mM KCl, and 10 mM DTT. The final concentrations of *A. aeolicus* TilS in all reaction mixtures were 0.25 μM for WT and 1, 2, and 4 μM for the mutants. To calculate the kinetic parameters, the concentrations of two of the three substrates, ATP (2 mM), $tRNA^{Ile}_2$ transcript (40 μM), and L-lysine (4 mM), were fixed, and the concentrations of the other substrates were varied (40–640 μM ATP, 2.5–156.8 μM $tRNA^{Ile}_2$ transcript, and 0.25–4 mM L-lysine). The reaction was terminated by spotting an aliquot on a Whatman 3-mm paper filter, which was previously soaked with 200 μl of 5% trichloroacetic acid (TCA) and dried. The filters were washed three times with ice-cold 5% TCA and twice with ethanol and were dried, and their radioactivities were counted by a scintillation counter. The assay was carried out at least twice for each mutant. Kinetic parameters were calculated by fitting to a Lineweaver–Burk plot. In this kinetic analysis, the concentration of L-lysine could not be increased to >4 mM, because the concentration of the commercially available [U-^{14}C] L-lysine is 165 μM, which is lower than the K_m value (629 μM) for L-lysine of the WT enzyme. Accordingly, the commercially available [U-^{14}C] L-lysine had to be diluted with unlabeled L-lysine to the extent that the resulting scintillation counts are not below the threshold level.

Lysidine Formation Assay of $tRNA^{Ile}_2$ Mutants. The genes encoding the *A. aeolicus* $tRNA^{Ile}_2$ mutants were synthesized by PCR with mutated primers. The subsequent purification steps were the same as those used for the WT *A. aeolicus* $tRNA^{Ile}_2$ transcript. The lysidine-forming activity of the $tRNA^{Ile}_2$ mutants were measured in 50 μl of reaction buffer (100 mM Tris·HCl, pH 7.8/10 mM $MgCl_2$/10 mM KCl/10 mM DTT) containing 2 mM ATP, 10 μM $tRNA^{Ile}_2$ transcript, 250–4,000 μM [U-^{14}C] L-lysine, and 0.125 μM *A. aeolicus* TilS. After 2, 5, and 10 min, a 15-μl aliquot was removed. The assay procedures were the same as those used for the aforementioned kinetic analyses.

Results and Discussion

Overall Structure. We solved the crystal structure of *A. aeolicus* TilS (317 aa) and refined the model to 2.42-Å resolution with an R_{cryst} of 21.8% and an R_{free} of 25.3% (Fig. 5, which is published as supporting information on the PNAS web site). The crystals belong to the $P4_12_12$ space group, with unit cell parameters of $a = b = 81.96$ Å and $c = 302.6$ Å (Table 2). The asymmetric unit contained the TilS dimer (Fig. 1*A*). The gel filtration and ultracentrifugation analyses revealed that *A. aeolicus* TilS behaves as a protein with a M_r of 57–62 kDa, corresponding to 1.50–1.65 times the M_r of the monomer (Fig. 6, which is published as supporting information on the PNAS web site). Each subunit of *A. aeolicus* TilS has a hammer-like shape composed of the N-terminal domain, referred to as the NTD (amino acid residues 1–217; blue and green in Fig. 1 *A* and *B*), and the C-terminal domain, referred to as the CTD (amino acid residues 249–317; red in Fig. 1 *A* and *B*), which are connected by a long α-helical linker (orange in Fig. 1 *A* and *B*). The NTD consists of a core of six stranded β-sheet flanked by nine α-helices, forming a dinucleotide-binding fold with a characteristic central hole; the CTD is composed of antiparallel β-sheets and three α-helices (Fig. 1 *A*–*C*).

Recently, the crystal structure of *E. coli* MesJ protein was determined as a function-unknown protein (PDB ID code 1NI5) (Fig. 1 *D* and *E*). A recent proteomics analysis has revealed that the MesJ functions in lysidine formation both in *E. coli* and *Bacillus subtilis*, and thus the gene was renamed as *tilS*. Although the coordinates of *E. coli* TilS were deposited in the Protein Data Bank as a monomer, a symmetrical operation of the *E. coli* TilS structure implies that the *E. coli* enzyme can also form a dimer (Fig. 1*E*) in a similar manner as *A. aeolicus* TilS. Actually, our recent gel filtration analysis indicated that *E. coli* TilS forms a dimer (data not shown). The solvent-accessible surface areas buried in the dimer interface are 1,800.5 Å^2 and 1,775.0 Å^2 for *A. aeolicus* and *E. coli* TilSs, respectively. The structures of the NTD and CTD of *A. aeolicus* TilS are quite similar to those of the NTD and CTD1, respectively, of the *E. coli* enzyme [rms deviation (rmsd) = 1.19 Å over 159 Cα atoms for the NTD and rmsd = 1.19 Å over 50 Cα atoms for the CTD] (Fig. 1 *A* and *E*). However, the orientations against the NTD and the surface electronic potentials are quite different between the CTD of *A. aeolicus* TilS and the CTD1 of the *E. coli* enzyme (described below). Furthermore, *E. coli* TilS has an additional CTD, CTD2, after CTD1, compared with the *A. aeolicus* enzyme (Fig. 1*D*). These structural differences may reflect the distinct manners of tRNA recognition by the two TilS enzymes, as described below.

Catalytic NTD Derived from the N-type ATP-PPase. The NTD of *A. aeolicus* TilS has the amino acid sequence ^{32}SGGVDS^{37}. This sequence is identical to the fingerprint sequence, SGGXDss (where X is any hydrophobic amino acid and lowercase letters are highly conserved amino acids), of the ATP-binding P-loop in the N-type ATP-PPase subfamily, such as GMP synthetase, NAD synthetase, asparagine synthetase, and argininosuccinate synthetase, which

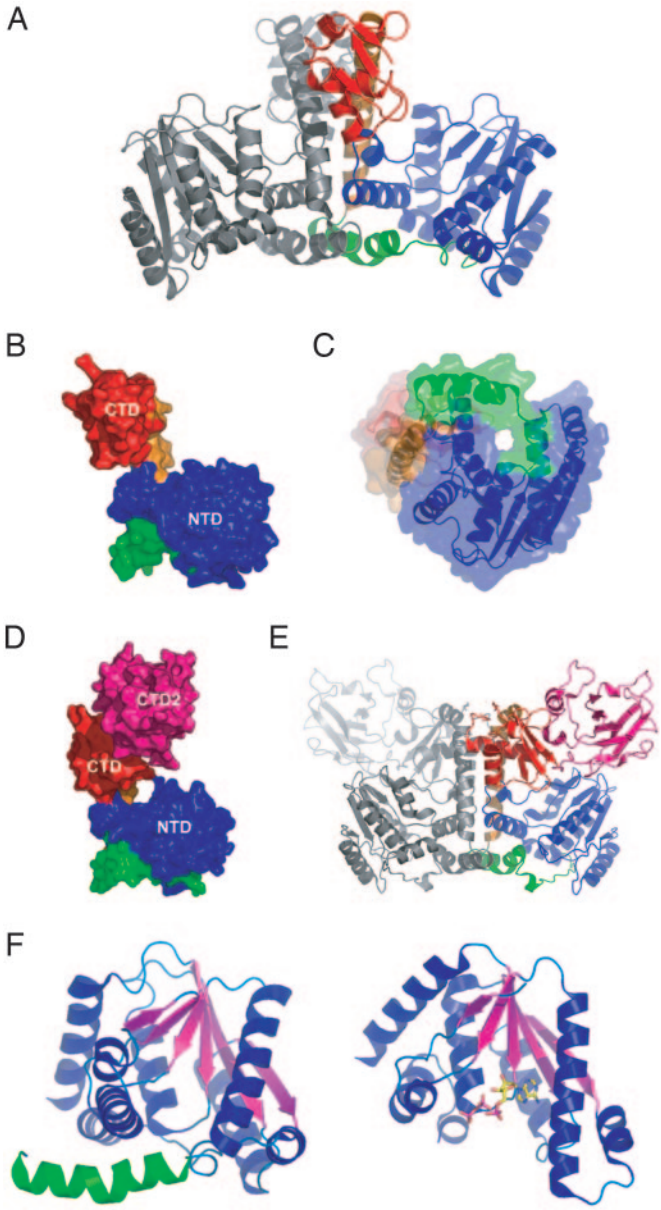

Fig. 1. TilS structures. (*A*) Overall structure of the *A. aeolicus* TilS dimer. The NTD, the CTD, and the long α-helical linker are colored blue, red, and orange, respectively. The subdomain appended to the NTD is colored green. (*B*) A contact surface representation of the *A. aeolicus* TilS monomer. The color code is the same as in *A*. (*C*) Characteristic central hole in the TilS NTD. (*D*) A contact surface representation of the *E. coli* TilS monomer. The additional CTD2 domain is colored magenta. The other color codes are the same as in *A*. (*E*) The structure of the *E. coli* TilS dimer. The color code is the same as in *D*. (*F*) Structural comparison of the NTD of *A. aeolicus* TilS (*Left*) and the ATP-PPase domain of *E. coli* GMP synthetase (*Right*). The AMP and PP_i bound to the P-loop of *E. coli* GMP synthetase are represented by stick models.

commonly adenylates a substrate and then attacks the adenylated substrate with a nucleophilic nitrogen from a second substrate (16–19). Furthermore, the catalytic domain of the ATP-PPase subfamily has been shown to adopt a dinucleotide-binding fold, which indicates that the TilS NTD was derived from the N-type ATP-PPase. Particularly, the TilS NTD shares the highest structural homology with the ATP-PPase domain of GMP synthetase (rmsd = 2.397 Å over 106 Cα atoms) (Fig. 1*F*) (16). Furthermore, the TilS NTD shares a high sequence similarity with the ATP-PPase domain of GMP synthetase (Fig. 7, which is published as supporting information on the PNAS web site). GMP synthetase includes an amidotransferase domain appended to the N-type ATP-PPase domain, which allows the synthesis of guanine 5′-monophosphate from a xanthosine 5′-monophosphate (XMP) by means of two steps, the adenylation of the C2 carbon of XMP and the nucleophilic attack by an ammonium derived from the amide group of glutamine (20, 21). By analogy, it is suggested that TilS first adenylates the C2 carbon of cytosine 34 of $tRNA^{Ile}_2$ and then catalyzes the nucleophilic attack of the adenylated C2 carbon by the activated terminal amino group of the L-lysine substrate (Fig. 2*A*). Because the soaking or cocrystallization experiment did not allow us to see the electron density corresponding to ATP and L-lysine, we superposed the TilS NTD on the AMP/PP_i-bound ATP-PPase domain of *E. coli* GMP synthetase (Fig. 2*B*) (16)

Intriguingly, in our docking model, the TilS NTD interacts with AMP and PP_i in the same manner as the ATP-PPase domain of GMP synthetase. The backbone carbonyl and amide groups of Phe-61 hydrogen-bond to the N6 amino group and the N1 atom, respectively, of the adenine ring of AMP, and the hydroxyl groups of Ser-32 and Ser-37 hydrogen-bond to the phosphate oxygens of PP_i. The side chain of Arg-113 in one monomer of *A. aeolicus* TilS was bent to stack with the adenine ring of AMP, as the side chain of Phe-315 of GMP synthetase, and the hydroxyl group of Tyr-114 hydrogen-bonds to O4′ of the AMP ribose. In *E. coli* TilS, the corresponding Arg-97 stacks with the ribose ring as well. Thus, our docking model suggests that some amino acid residues comprising the central hole of NTD, such as Ser-32, Asp-36, Ser-37, Phe-61, Arg-113, Tyr-114, and His-133, could participate in ATP binding. Therefore, we prepared the Ala mutants of Ser-32, Asp-36, Ser-37, Arg-113, Tyr-114, and His-133 and analyzed their kinetic parameters for ATP, L-lysine, and the *A. aeolicus* $tRNA^{Ile}_2$ transcript (Table 1). We tried to prepare the S32A mutant, but it was not expressed well in *E. coli*. The hydroxyl group of Ser-32 forms an intramolecular hydrogen bond with the backbone nitrogen of Gly-34 to stabilize the sharp turn of the P-loop, which forms an inner core of the TilS NTD. This interaction may explain why the S32A mutant was not properly folded and not expressed in *E. coli*. The other five mutants either completely impaired the lysidine-forming activity or increased the K_m for ATP 7- to 14-fold, although the H133A mutant decreased the K_m for ATP 2.5-fold, and the other parameters remained essentially unchanged. His-133 may participate in the release of AMP from the ATP-binding pocket after lysidine formation for efficient enzymatic turnover. These results support our AMP/PP_i docking model. Furthermore, the K_m values for L-lysine of the S37A and Y114A mutants increased 6- and 5-fold, respectively, concomitantly with the increase of the K_m for ATP, which implies that the central hole of NTD places the L-lysine-binding site near the ATP-binding site. It was reported that the closely related GMP synthetase exhibits a sequential and ordered reaction: ATP, xanthosine 5′-monophosphate (XMP), and the nitrogen source bind to the enzyme in this order, and the XMP adenylation occurs in the absence of an ammonium (21). Therefore, ATP, $tRNA^{Ile}_2$, and L-lysine are likely to sequentially bind to TilS in this order to fulfill the base-modification reaction.

A subdomain composed of a loop and a helix is appended to the C terminus of the TilS NTD (green in Fig. 1). This subdomain is absent from the other N-type ATP-PPase and is specific to TilS (Fig. 7). The subdomain includes a conserved sequence, ^{190}eDXS/TNXXXXJXRnXJR^{205} (where J is any hydrophobic amino acid), which forms an edge of the aforementioned characteristic central hole. In contrast, in the other N-type ATP-PPase domain, the central hole is replaced by a valley-shaped pocket. Therefore, this appended subdomain may have essential functions specific to TilS.

Catalytic Site for C34 Adenylation. In addition to the aforementioned residues involved in the ATP binding, several charged or polar amino acid residues that are strictly conserved in the TilS enzymes are clustered at the edge of the central hole of the NTD, suggesting

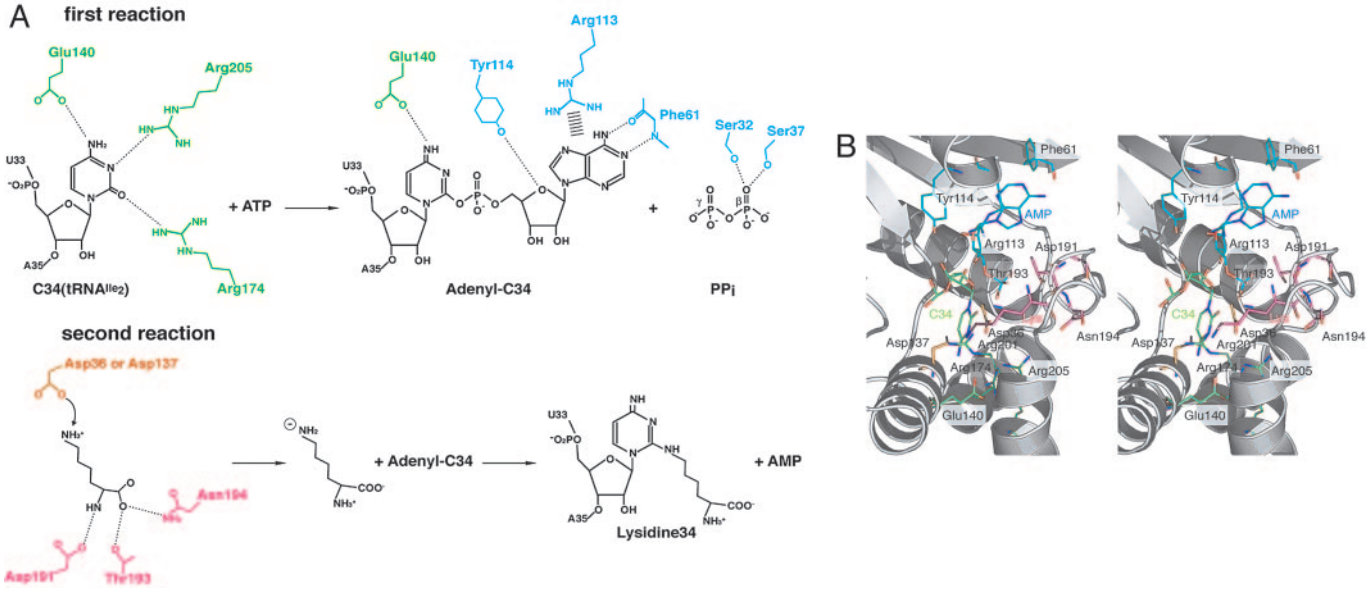

Fig. 2. Mechanism of lysidine formation by TilS. (*A*) Putative mechanism of the two-step reaction by TilS. The amino acid side chains that are proposed to participate in the substrate recognition and the catalysis are indicated in the same color code as in *B*. (*B*) The docking model of AMP (blue), C34 of $tRNA^{Ile}_2$ (green), and L-lysine (magenta) (stereoview). Amino acid residues that seem to be involved in the recognition of ATP, C34, and L-lysine, as suggested by the mutant analysis, are shown in blue, green, and magenta, respectively. For clarity, the modeled PP_i is not shown.

that the reaction step after the ATP binding could occur near the central hole. Mutant analyses of these residues revealed their involvement not only in the binding of ATP and L-lysine, but also in the recognition of C34 of $tRNA^{Ile}_2$ (Table 1). The substitution of Ala for Glu-140 and Arg-205 similarly affected the kinetic parameters. The E140A mutant increased the K_m values for ATP and tRNA by 7- and 5-fold, respectively, and decreased the k_{cat} by 6-fold, whereas it barely affected the K_m for L-lysine (Table 1). The R205A mutant also showed 15-fold and 4-fold higher K_m values for ATP and tRNA, respectively, and a 16-fold lower k_{cat}, than the WT, with the K_m for lysine substantially unchanged (Table 1). Therefore, Glu-140 and Arg-205 may be involved in the first-step reaction, the C34 adenylation (ATP binding, tRNA binding, and/or transition-state stabilization). The R174A mutant showed a 6-fold increase in the K_m for tRNA, whereas all of the other kinetic parameters remained essentially unchanged (Table 1), suggesting that Arg-174 is involved in the tRNA recognition. In contrast, the mutation of Glu-106 to Ala caused 2-fold and 18-fold reductions in the K_m values for tRNA and k_{cat}, respectively; the K_m values for L-lysine and ATP remained unchanged (Table 1). These results suggest that Glu-106 functions in the release of the lysidine-modified $tRNA^{Ile}_2$ product, thus contributing to the effective turnover of the TilS reaction. On the basis of these kinetic analyses, we propose a docking model for C34 of $tRNA^{Ile}_2$ (Fig. 2*B*). In this docking model, the base moiety of C34 is recognized by Glu-140, Arg-174, and Arg-205. The side chain of Arg-174 is fixed by the carboxyl group of Asp-36 (Fig. 2*B*) and may also hydrogen-bond to the PP_i product, by analogy to the fact that, in the closely related GMP synthetase, the side chain of Lys-381 corresponding to Arg-174 is oriented by the equivalent of Asp-36 and forms a hydrogen bond to the PP_i

Table 1. Kinetics analyses of the *A. aeolicus* TilS mutants

Protein	$k_{cat} \times 10^3$, (s^{-1})	K_m (μM)		
		Lysidine	ATP	tRNA
WT	156.3 ± 1.2	629.4 ± 38.2	19.4 ± 2.1	3.34 ± 0.27
D36A	N.D.	N.D.	N.D.	N.D.
S37A	>20.5 ± 2.1	>4,086.7 ± 967.1	255.0 ± 36.5	<4.66 ± 0.96
E106A	8.5 ± 0.5	1,398.6 ± 273.3	38.5 ± 4.6	1.90 ± 0.19
R113A	N.D.	N.D.	N.D.	N.D.
Y114A	51.7 ± 0.6	3,052.7 ± 75.7	155.1 ± 32.1	7.36 ± 1.06
H133A	125 ± 3.0	1,037.5 ± 92.6	7.67 ± 1.2	3.54 ± 0.31
D137A	N.D.	N.D.	N.D.	N.D.
E140A	27.7 ± 3.4	1,363.6 ± 20.7	134.6 ± 43.7	15.67 ± 2.86
R174A	173.8 ± 5.2	1,147.1 ± 141.6	41.2 ± 1.1	20.20 ± 0.54
W188A	>166.6 ± 17.2	>5,190.5 ± 132.1	97.0 ± 2.1	<10.46 ± 0.71
D191A	N.D.	N.D.	N.D.	N.D.
T193A	>22.1 ± 4.0	>6,121.8 ± 2,768.1	<41.9 ± 1.9	<5.62 ± 1.15
N194A	N.D.	N.D.	N.D.	N.D.
R201A	N.D.	N.D.	N.D.	N.D.
R205A	9.57 ± 2.7	1,013.0 ± 73.0	287.2 ± 144.9	11.90 ± 1.38

N.D., not detectable.

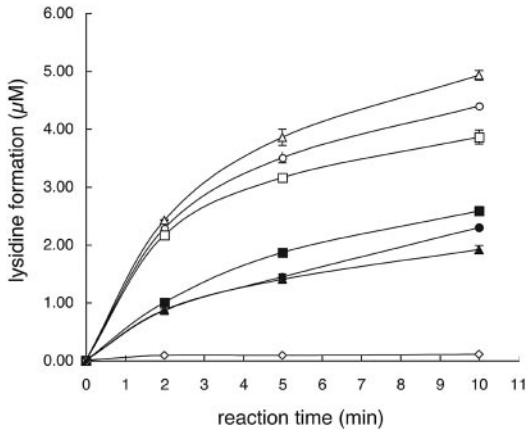

Fig. 3. Time course of lysidine formation of $tRNA^{Ile}_2$ (○), C3G70 mutant (△), G4C69 mutant (□), C20U mutant (▲), U20 inserted mutant (■), G29C41 mutant (◇), and $tRNA^{Met}_m$ (●) by *A. aeolicus* TilS. The assays were repeated at least three times.

product. Arg-201, which when mutated to Ala abolished the lysidine-forming activity, might have a role in flipping and fixing the C34 base, which stacks within the anticodon loop in the canonical tRNA.

The Appended Subdomain Essential for L-Lysine Recognition. As described above, the NTD, which is involved in the C34 adenylation, is followed by a subdomain that consists of a loop and a helix. This subdomain is characteristic of TilS and may have TilS-specific functions. Actually, Arg-201 and Arg-205 are supposed to have TilS-specific functions, including interacting with and recognizing C34 of $tRNA^{Ile}_2$ (see the previous paragraph). Therefore, we analyzed the kinetic parameters of the other conserved residues, Asp-191, Thr-193, and Asn-194, on the subdomain loop (Table 1). The substitution of Ala for the absolutely conserved Asp-191 and Asn-194 abolished the lysidine-forming activity (Table 1). The replacement of Thr-193 with Ala increased the K_m for L-lysine by 10-fold and reduced the k_{cat} by 7-fold but hardly affected the K_m values for ATP and $tRNA^{Ile}_2$ (Table 1). The maximum concentration of L-lysine used in the kinetic experiment was 4 mM (see *Materials and Methods*). Because the K_m value (6.1 mM) of the T193A mutant is >4 mM, the real k_{cat} value of T193A might be higher under appropriate experimental conditions with a sufficient L-lysine concentration. Therefore, we propose that Thr-193, and probably Asp-191 and Asn-194, are involved in L-lysine recognition.

The substitution of Ala for Trp-188, located just upstream of the appended subdomain, increased the K_m values for L-lysine, ATP, and $tRNA^{Ile}_2$ 8-fold, 5-fold, and 3-fold, respectively, whereas it did not affect the k_{cat} (Table 1). However, because the K_m value for L-lysine (5.2 mM) of W188A is >4 mM, the K_m values for ATP and $tRNA^{Ile}_2$ might be lower under appropriate experimental conditions, as in the case of the T193A mutant. Trp-188 forms a hydrophobic core, together with Val-38, Val-39, Tyr-171, and Ala-181. This hydrophobic core fixes the spatial arrangement of the second α-helix of the NTD, which bears the P-loop, and the α-helix of the subdomain, which follows the subdomain loop with Asp-191, Thr-193, and Asn-194 (Fig. 2). Also, in *E. coli* TilS, the corresponding Trp-174 forms a hydrophobic core in the same way. Nevertheless, the rmsd of the lysine-binding loop between the *E.* coli and *A. aeolicus* TilS structures is as large as 7 Å (data not shown), suggesting that the lysine-binding loop is quite flexible and that the conserved Trp residue is structurally important for the formation of the L-lysine-binding site as well as the C34 adenylation site. The mutation of Trp-188 may indirectly reduce the lysine-binding activity by augmenting the flexibility of the lysine-binding loop so it cannot bind L-lysine efficiently. The substantial flexibility of the lysine-specific loop may explain both the high K_m value (630 μM) for L-lysine and why no electron density for L-lysine was observed in the soaking experiment.

A sequence alignment of TilS and GMP synthetase revealed that eight residues are absolutely conserved (Fig. 7). Among them, Asp-36, Asp-137, and Glu-140 (residue numbering according to *A. aeolicus* TilS) are located near the phosphate group of the superposed AMP and the C2 carbon of the docked C34 of $tRNA^{Ile}_2$ (Fig. 2*B*), and, particularly, Asp-36 has been suggested to play a catalytic role in the amino transfer reaction in GMP synthetase (16). The mutations of Asp-36 and Asp-137 abolished the lysidine-forming activity, whereas the E140A mutant retained some activity (Table 1). Therefore, these structural and kinetic results suggest that Asp-36 and/or Asp-137 are involved in deprotonation of the ε-amino group of L-lysine. Based on these results, we constructed a lysine-docking model where Asp-191, Thr-193, and Asn-194 are involved in the recognition of the α-amino and α-carboxyl groups and Asp-36 and Asp-137 hydrogen-bond to the terminal amino group of the L-lysine substrate (Fig. 2*B*).

Recognition and Discrimination of $tRNA^{Ile}_2$. The posttranscriptional modification from C34 to lysidine 34 in $tRNA^{Ile}_2$ not only changes the codon specificity from AUG to AUA, but also converts the aminoacylation specificity from recognition by methionyl-tRNA synthetase to that by IleRS (9). Therefore, TilS must discriminate unmodified $tRNA^{Ile}_2$ from unmodified $tRNA^{Met}$ species bearing the anticodon CAU. When we compare the nucleotide sequence of $tRNA^{Ile}_2$ with those of $tRNA^{Met}$ isoacceptors (initiator $tRNA^{Met}_f$ and elongator $tRNA^{Met}_m$) in *A. aeolicus*, only the G3·C70 and C4·G69 base pairs in the acceptor stem and the C29·G41base pair in the anticodon stem are specific to $tRNA^{Ile}_2$, which are replaced by the C3·G70, G4·C69, and G29·C41 base pairs, respectively, in two $tRNA^{Met}$ isoacceptors (Fig. 8, which is published as supporting information on the PNAS web site). To identify the nucleotide

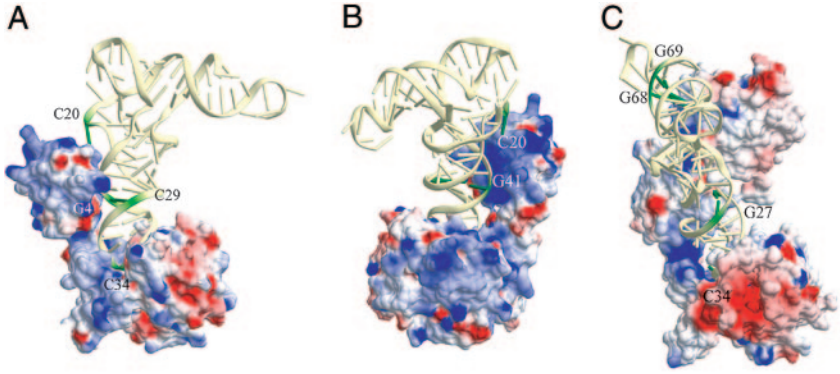

Fig. 4. Docking analysis of tRNA to the TilS structures from *A. aeolicus* (*A* and *B*) and *E. coli* (*C*). The protein structures are represented by the solvent-accessible surfaces with the electrostatic potentials. The nucleotides recognized by each TilS are indicated in green.

elements for TilS discrimination of $tRNA^{Ile}_2$ from $tRNA^{Met}$ species, the lysidine-forming activities were investigated for C3G70, G4C69, and G29C41 mutants of $tRNA^{Ile}_2$ (Fig. 3). The C3G70 and G4C69 mutants of $tRNA^{Ile}_2$ exhibited lysidine-forming activity similar to that of the WT $tRNA^{Ile}_2$, whereas the mutation of the C29·G41 base pair abolished the lysidine-forming activity (Fig. 3). These results revealed that the C29·G41 base pair is crucial for the discrimination of $tRNA^{Ile}_2$ by *A. aeolicus* TilS. Therefore, we manually docked the *Thermus thermophilus* $tRNA^{Glu}$ complexed with the cognate glutamyl-tRNA synthetase (PDB ID code 1N77) onto the *A. aeolicus* TilS so that the base pair corresponding to C29·G41 can bind to the positively charged area of the CTD and C34 can enter into the catalytic pocket in the NTD (Fig. 4 *A* and *B*). The used $tRNA^{Glu}$ retains canonical L-shape and is suitable for the present docking analysis. The sequence difference between $tRNA^{Ile}_2$ and $tRNA^{Glu}$ is not critical because the aim of the model construction is only to evaluate the distribution of the nucleotide elements recognized by TilS. The tRNA fits well to the molecular surface of *A. aeolicus* TilS. This docking model shows that C20 approaches closely so it can be recognized by the TilS CTD (Fig. 4 *A* and *B*). Therefore, we analyzed the lysidine-forming activity for the C20U mutant and $tRNA^{Ile}_2$ transcript with an additional uridine inserted between G19 and C20, as in the cases of $tRNA^{Met}_f$ and $tRNA^{Met}_m$, respectively (Fig. 8). Both of these mutations reduced the lysidine-forming activity to the level of $tRNA^{Met}_m$ (described below), which suggests that C20 is essential for *A. aeolicus* TilS to distinguish $tRNA^{Ile}_2$ from two $tRNA^{Met}$ isoacceptors and strongly supports the integrity of our docking model (Fig. 4 *A* and *B*). The efficiency of the lysidine-forming activity for $tRNA^{Met}_m$ was less than half of that for $tRNA^{Ile}_2$ at their plateau levels (Fig. 3), which indicates that unmodified $tRNA^{Met}_m$ was loosely recognized by *A. aeolicus* TilS *in vitro*. However, *in vivo*, because the anticodon wobble position of $tRNA^{Met}_m$ is modified into ac^4C, competition with the modification enzyme may prevent $tRNA^{Met}_m$ from being modified by TilS (22, 23). Therefore, these results suggest that TilS specifically modifies $tRNA^{Ile}_2$ *in vivo*.

In *E. coli*, it has recently been revealed that the C4·G69, C5·G68 (in the acceptor stem), and G27·U43 (in the anticodon stem) base pairs, which are specific to $tRNA^{Ile}_2$ as compared with $tRNA^{Met}$, play a key role in recognition by the cognate TilS (Y.I., A.S., T. Ote, J. Kato, Y.S., and T.S., unpublished data). Whereas the electrostatic potential is limited to one part of the CTD in *A. aeolicus* TilS, *E. coli* TilS has two positively charged areas, in a distinct part of the CTD1 and in the α-helical linker connecting the NTD and the CTD1 (Fig. 4). On the basis of these data, we also built the tRNA-docking model of *E. coli* TilS (Fig. 4*C*). As described above, *E. coli* TilS has CTD2 (equivalent to CTD of *A. aeolicus* TilS) (Fig. 1 *B* and *D*). The positively charged interdomain cleft between CTD1 and CTD2 grasps the acceptor stem of *E. coli* $tRNA^{Ile}_2$ (Fig. 4*C*). These functional and structural analyses clearly indicate that the tRNA recognition manners fundamentally differ between the *A. aeolicus* and *E. coli* TilS enzymes; *A. aeolicus* TilS does not recognize the acceptor stem of $tRNA^{Ile}_2$, in striking contrast to the *E. coli* enzyme, but strictly recognizes the C29·G41 base pair in the anticodon stem. These differences in the tRNA recognition mechanism between the two TilS enzymes are consistent with the fact that *A. aeolicus* TilS efficiently modifies *E. coli* $tRNA^{Ile}_2$ with the C29·G41 base pair but that *E. coli* TilS does not modify *A. aeolicus* $tRNA^{Ile}_2$ with G5·C68 instead of the C5·G68 base pair (Fig. 8) (data not shown). *E. coli* IleRS recognizes the C4·G69 and C29·G41 base pairs as well as the U12·A23 base pair, the anticodon trinucleotides, t^6A37, and A73 (the discriminator base) of $tRNA^{Ile}$ to discriminate it from the other tRNA species (24, 25). Particularly, the C29·G41 base pair is one of the strongest recognition elements (24), whereas C4·G69 serves as an antideterminant against methionyl-tRNA synthetase (26). Therefore, the *E. coli* and *A. aeolicus* TilS enzymes discriminate $tRNA^{Ile}_2$ from $tRNA^{Met}$ by a strategy similar to that used by IleRS, but in individually distinct manners. Elucidation of the precise tRNA recognition mechanism should be facilitated by the structure determination of the TilS·$tRNA^{Ile}_2$ complex.

We thank S. Wakatsuki, N. Igarashi, N. Matsugaki, and K. Demura for their help in data collection at KEK (Tsukuba, Japan) and F. Arisaka for the ultracentrifugation analysis. This work was supported by a grant from the Ministry of Education, Culture, Sports, Science, and Technology (to O.N.), a Precursory Research for Embryonic Science and Technology program grant from the Japan Science and Technology Agency, and Sumitomo and Naito Foundation grants (to O.N.).

1. Bjork, G. (1995) *tRNA: Structure, Biosynthesis, and Function*, eds. Soll, D. & RajBhandary, U. L. (American Society for Microbiology, Washington, D.C.)
2. Schimmel, P. R. (1979) in *Transfer RNA: Structure, Properties, and Recognition*, eds. Schimmel, P. R., Soll, D. & Abelson, J. N. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 297–310.
3. Ozeki, H., Inokuchi, H., Yamao, F., Kodaira, M., Sakano, H., Ikemura, T. & Shimura, Y. (1980) in *Transfer RNA: Biological Aspects*, eds. Soll, D., Abelson, J. N. & Schimmel, P. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 341–362.
4. Yokoyama, S., Yamaizumi, Z., Nishimura, S. & Miyazawa, T. (1979) *Nucleic Acids Res.* **7,** 2611–2626.
5. Gerber, A. P. & Keller, W. (1999) *Science* **286,** 1146–1149.
6. Harada, F. & Nishimura, S. (1974) *Biochemistry* **13,** 300–307.
7. Muramatsu, T., Yokoyama, S., Horie, N., Matsuda, A., Ueda, T., Yamaizumi, Z., Kuchino, Y., Nishimura, S. & Miyazawa, T. (1988) *J. Biol. Chem.* **263,** 9261–9267.
8. Matsugi, J., Murao, K. & Ishikura, H. (1996) *J. Biochem. (Tokyo)* **119,** 811–816.
9. Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T. & Yokoyama, S. (1988) *Nature* **336,** 179–181.
10. Soma, A., Ikeuchi, Y., Kanemasa, S., Kobayashi, K., Ogasawara, N., Ote, T., Kato, J., Watanabe, K., Sekine, Y. & Suzuki, T. (2003) *Mol. Cell* **12,** 689–698.
11. Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994) *J. Appl. Crystallogr.* **27,** 613–621.
12. de La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494.
13. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
14. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54**, 905–921.
15. DeLano, W. L. (2002) The PYMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).
16. Tesmer, J. J., Klem, T. J., Deras, M. L., Davisson, V. J. & Smith, J. L. (1996) *Nat. Struct. Biol.* **3,** 74–86.
17. Rizzi, M., Bolognesi, M. & Coda, A. (1998) *Structure* **15,** 1129–1140.
18. Lemke, C. T. & Howell, P. L. (2001) *Structure (Cambridge, MA)* **9,** 153–164.
19. Savage, H., Montoya, G., Svensson, C., Schwenn, J. D. & Sinning, I. (1997) *Structure* **15,** 895–906.
20. Fukuyama, T. & Moyed, H. S. (1964) *Biochemistry* **20,** 1488–1492.
21. von der Saal, W., Crysler, C. S. & Villafranca, J. J. (1985) *Biochemistry* **24,** 5343–5350.
22. Parthasarathy, R., Ginell, S. L., De, N. C. & Chheda, G. B. (1978) *Biochem. Biophys. Res. Commun.* **28,** 657–663.
23. Oashi, Z., Murano, K., Yahagi, T., Bon, Minden, D. L., McCloskey, J. A. & Nishimura, S. (1972) *Biochim. Biophys. Acta* **262,** 209–213.
24. Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., Giege, R. & Yokoyama, S. (1994) *J. Mol. Biol.* **236,** 710–724.
25. Silvian, L. F., Wang, J. & Steitz, T. A. (1999) *Science* **285,** 1074–1077.
26. Martin, S. A. & Schimmel, P. (1993) *J. Biol. Chem.* **268,** 6069–6072.