Structural and mutational studies of the recognition of the arginine tRNA-specific major identity element, A20, by arginyl-tRNA synthetase

Atsushi Shimada*†, Osamu Nureki*‡, Mie Goto§¶, Seizo Takahashi§, and Shigeyuki Yokoyama*‡‖**

*Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan; ‡Cellular Signaling Laboratory, Institute of Physical and Chemical Research (RIKEN) Harima Institute, 1-1-1 Kohto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan; §Faculty of Science, Japan Women's University, Tokyo 112-0015, Japan; and ‖Institute of Physical and Chemical Research (RIKEN) Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama-shi, Kanagawa 230-0045, Japan

Edited by Paul R. Schimmel, The Scripps Research Institute, La Jolla, CA, and approved September 18, 2001 (received for review May 29, 2001)

Arginyl-tRNA synthetase (ArgRS) recognizes two major identity elements of $tRNA^{Arg}$: A20, located at the outside corner of the L-shaped tRNA, and C35, the second letter of the anticodon. Only a few exceptional organisms, such as the yeast *Saccharomyces cerevisiae*, lack A20 in $tRNA^{Arg}$. In the present study, we solved the crystal structure of a typical A20-recognizing ArgRS from *Thermus thermophilus* at 2.3 Å resolution. The structure of the *T. thermophilus* ArgRS was found to be similar to that of the previously reported *S. cerevisiae* ArgRS, except for short insertions and a concomitant conformational change in the N-terminal domain. The structure of the yeast ArgRS·$tRNA^{Arg}$ complex suggested that two residues in the unique N-terminal domain, Tyr^{77} and Asn^{79}, which are phylogenetically invariant in the ArgRSs from all organisms with A20 in $tRNA^{Arg}$s, are involved in A20 recognition. However, in a docking model constructed based on the yeast ArgRS·$tRNA^{Arg}$ and *T. thermophilus* ArgRS structures, Tyr^{77} and Asn^{79} are not close enough to make direct contact with A20, because of the conformational change in the N-terminal domain. Nevertheless, the replacement of Tyr^{77} or Asn^{79} by Ala severely reduced the arginylation efficiency. Therefore, some conformational change around A20 is necessary for the recognition. Surprisingly, the N79D mutant equally recognized A20 and G20, with only a slight reduction in the arginylation efficiency as compared with the wild-type enzyme. Other mutants of Asn^{79} also exhibited broader specificity for the nucleotide at position 20 of $tRNA^{Arg}$. We propose a model of A20 recognition by the ArgRS that is consistent with the present results of the mutational analyses.

Aminoacyl-tRNA synthetases (aaRSs) catalyze the esterification of their cognate tRNAs with specific amino acids. Strict recognition of both the amino acid and the tRNA by the aaRS ensures the correct translation of the genetic code. To discriminate the cognate tRNA from the non-cognate tRNAs, the aaRS recognizes characteristic features of the cognate tRNA, which are usually a small number of nucleotides, called identity elements, that comprise the identity sets (1). In most tRNAs, the identity elements are concentrated in the anticodon and/or the acceptor stem (1). This recognition mode can be explained by the fact that most synthetases interact with the inner side of the L-shaped tRNA, so that they can easily access the two extremities of the tRNA (2–7). However, in the cases of arginine, leucine, and serine, where six codons are assigned to a single amino acid in the genetic code, the anticodons of the cognate tRNAs share only one or no nucleotide. Consequently, an additional major identity element may be required in a region other than the anticodon. The leucine and serine tRNAs possess characteristically long variable arms, which are used as important identity elements in most organisms (8–13). In contrast, $tRNA^{Arg}$ utilizes an adenosine at position 20 for the non-anticodon major identity element, which is completely conserved among the $tRNA^{Arg}$ species in most organisms, and is missing in other tRNA species (14, 15). Actually, genetic and biochemical studies of *Escherichia coli* $tRNA^{Arg}$ demonstrated that A20 is one of the major identity elements of $tRNA^{Arg}$, with a very high discrimination rate of three orders of magnitude on average (14, 16–19). A20 is located in a small, single-stranded region in the D loop at the outside corner of the L-shaped tRNA structure, which is far from both the anticodon and the acceptor stem. No other tRNAs specific to the other amino acids have a major identity element with such a high discrimination rate in this position.

On the other hand, the exceptional organisms that lack A20 in their $tRNA^{Arg}$ sequences are *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Neurospora crassa*, in addition to most of the mitochondria of animals and single cell eukaryotes (15). In *S. cerevisiae*, the $tRNA^{Arg}$ species have U, dihydrouridine (D), or C at position 20 (15), and the replacement of C20 by A in a $tRNA^{Arg}$ species caused no reduction in the arginine-accepting activity (18). The crystal structure of the *S. cerevisiae* arginyl-tRNA synthetase (ArgRS) was reported at 2.85 Å resolution (20), and recently that of the complex between the *S. cerevisiae* ArgRS and $tRNA^{Arg}$ was reported at 2.2 Å resolution (21). In the complex structure, the D20 of the $tRNA^{Arg}$ is recognized by Asn^{106}, Phe^{109}, and Gln^{111} in the N-terminal domain characterizing ArgRS (21). Thus, Cavarelli and colleagues proposed a model for the canonical recognition of A20 by Phe/Tyr and Asn of other ArgRSs, corresponding to Phe^{109} and Gln^{111}, respectively, of the *S. cerevisiae* ArgRS (21). In the present study, we determined the 2.3 Å resolution crystal structure of the *Thermus thermophilus* ArgRS, which recognizes the canonical A20. Furthermore, we carried out site-directed mutagenesis and kinetic analyses at the predicted crucial residues for the recognition of A20 to investigate their precise recognition mode.

Materials and Methods

Structure Determination. The *T. thermophilus* ArgRS gene (accession number: AJ278815) was overexpressed in the *E. coli* strain JM109(DE3) (22). The recombinant protein was purified and

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

Abbreviations: aaRS, aminoacyl-tRNA synthetase; ArgRS, arginyl-tRNA synthetase.

Data deposition: The atomic coordinates of the enzyme and the docking model with tRNA have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 1IQ0 and 1IR4, respectively).

See commentary on page 13473.

†Present address: Max-Planck-Institut für Molekulare Physiologie, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany.

¶Present address: Institute of Physical and Chemical Research (RIKEN) Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama-shi, Kanagawa 230-0045, Japan.

**To whom reprint requests should be addressed. E-mail: yokoyama@biochem.s.u-tokyo.ac.jp.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

BIOCHEMISTRY

Table 1. Refinement statistics

Resolution, Å	50.0–2.3
R_{work}, %	21.5 (23.3)
R_{free}, %	24.2 (26.3)
rms deviation from ideal values	
Bond length, Å	0.0058
Bond angles, °	1.16
Dihedral angles, °	21.78
Improper angles, °	0.86
Average *B*-factor, Å^2	38.18

Values in parentheses are for the highest resolution shell (2.38–2.30 Å). $R_{work} = \Sigma|F_O - F_C|/\Sigma|F_O|$ for the working set reflections (95% of the data) used for the refinement. $R_{free} = \Sigma|F_O - F_C|/\Sigma|F_O|$ for the test set reflections (5% of the data) excluded from the refinement.

crystallized, and the crystal structure was solved by multiple isomorphous replacement augmented with anomalous scattering (MIRAS) (22). The atomic model was built in the electron density map by the program O (23). Crystallographic positional and slow-cooling refinements were carried out by the programs X-PLOR (24) and CNS (25) against the 2.3-Å data set collected from the beamline 41XU at SPring-8 (Harima, Japan). Refinement statistics are shown in Table 1.

Construction of a Docking Model of ArgRS and tRNA. A docking model of *T. thermophilus* ArgRS and tRNA was constructed according to the following simple procedures. The structure of the complex of *S. cerevisiae* ArgRS and tRNA^{Arg} (21) was superposed on that of the *T. thermophilus* ArgRS, based on the Rossmann-fold and α-helix bundle domains, by the program LSQKAB (26). Subsequently, the atomic coordinates of the *T. thermophilus* ArgRS and those of the tRNA^{Arg} were merged. No additional operations were performed.

Cloning of the *T. thermophilus* Major tRNA^{Arg} Gene. A mixture of *T. thermophilus* tRNAs was extracted from *T. thermophilus* HB8 cells according to Zubay's method (27), and was fractionated by chromatography on a column of DEAE-Sephadex A50 (pH 7.5). The tRNAs of each fraction were further separated by electrophoresis on denaturing 15% polyacrylamide gels. The arginine-accepting activity of each fraction was monitored with recombinant *T. thermophilus* ArgRS throughout. The purified *T. thermophilus* major tRNA^{Arg} was labeled with ^{32}P at its 3′-terminus, and was used as a probe for Southern hybridization to *T. thermophilus* chromosomal DNA digested with several restriction enzymes, resulting in marked hybridization to a *Pvu*II fragment of about 3.3 kb. The 3.3-kb *Pvu*II fragment was cloned into the *Sma*I site of plasmid pUC118 by colony hybridization. The insert was sequenced and found to include the major portion of the gene encoding the *T. thermophilus* major tRNA^{Arg} (nucleotides 16–76). Because the tRNA^{Arg} gene includes a *Pvu*II site, the 5′ portion corresponding to nucleotides 1–15 is missing. The complete sequence of the gene for the major tRNA^{Arg} was deduced based on the complementary nature of its acceptor and D stems and the sequence homology with its *E. coli* homolog, and was later verified by *T. thermophilus* HB8 genome sequencing (data not shown).

Site-Directed Mutagenesis and Kinetic Analyses. The ArgRS mutant genes were made by oligonucleotide-directed mutagenesis, and were ligated into the vector pK7 (28). The nucleotide sequences of the genes were confirmed by dideoxy sequencing. All of the ArgRS mutants could be purified by the same procedure as that used for the wild-type enzyme (22). *The* DNA fragment containing the gene for the *T. thermophilus* major tRNA^{Arg} and the T7 promoter sequence, with an artificial *Bam*HI site at the 5′ end

Fig. 1. Ribbon diagram displaying the overall folding of *T. thermophilus* ArgRS (stereo view). N and C indicate the amino and carboxyl termini, respectively. All of the secondary structures are numbered. All of the graphic figures in the present paper were drawn by using MOLSCRIPT (31) and RASTER3D (32).

and a *Hin*dIII site at the 3′ end, was amplified by PCR by using the *T. thermophilus* chromosomal DNA as the template. The amplified DNA fragment was ligated into the appropriate sites of the vector pUC 118. The *in vitro* transcription systems for the tRNA^{Arg} variants were constructed in a similar way. The pUC118 vectors harboring the genes for the major tRNA^{Arg} and its variants were digested with *Mva*I, and were used for runoff transcription with T7 RNA polymerase. The transcripts were further purified by electrophoresis on denaturing 15% polyacrylamide gels.

The kinetic parameters of the arginylation were determined at 65°C in 100 mM Hepes-NaOH buffer (pH 7.5) containing 30 mM KCl, 5 mM magnesium acetate, 2 mM ATP, 100 μM L-[^{14}C]arginine, 0.05–1.5 μM of the wild-type and mutant ArgRSs, and 1–80 μM of the major tRNA^{Arg} transcripts and its variants. Because the background level of L-[^{14}C]arginine incorporation was relatively high and increased depending on the L-[^{14}C]arginine concentration, it was difficult to determine the correct K_m value for L-arginine. Therefore, in the present measurement, the L-[^{14}C]arginine concentration was fixed to 100 μM, which is an estimated physiological L-arginine concentration and is higher than that used thus far for the measurement of arginylation kinetics for other ArgRS species.

Because the purification step included a heat treatment at 75°C for 30 min to denature the *E. coli* proteins, the amount of the *E. coli* chromosomal ArgRS in the preparation of the recombinant *T. thermophilus* ArgRS was negligible. Furthermore, the arginylation assay of *T. thermophilus* ArgRS was carried out at 65°C, and, under these conditions, the aminoacylation activity of *E. coli* ArgRS was not detectable. Therefore, the contributions of the *E. coli* chromosomal ArgRS to the aminoacylation activities measured for the present recombinant *T. thermophilus* ArgRS samples were estimated to be less than 0.001%.

Results and Discussion

Overall Structure of *T. thermophilus* ArgRS. The current model comprises residues 1–397 and 402–592. Residues 398–401, which are in a loop near the "KMSKS" signature motif, were poorly ordered. The size of the *T. thermophilus* ArgRS is ≈85 Å by 65 Å by 25 Å (Fig. 1). The final model has an R_{work} of 21.5% and an R_{free} of 24.2% at 2.3 Å resolution, and shows very good geometry, as determined by the program PROCHECK (29): 92.7% and 7.1% of the residues have ϕ/ψ angles in the "most favored" and "additional allowed regions," respectively. The structure can be divided into seven regions, which each contain a single domain: the N-terminal α/β globular domain (colored green in

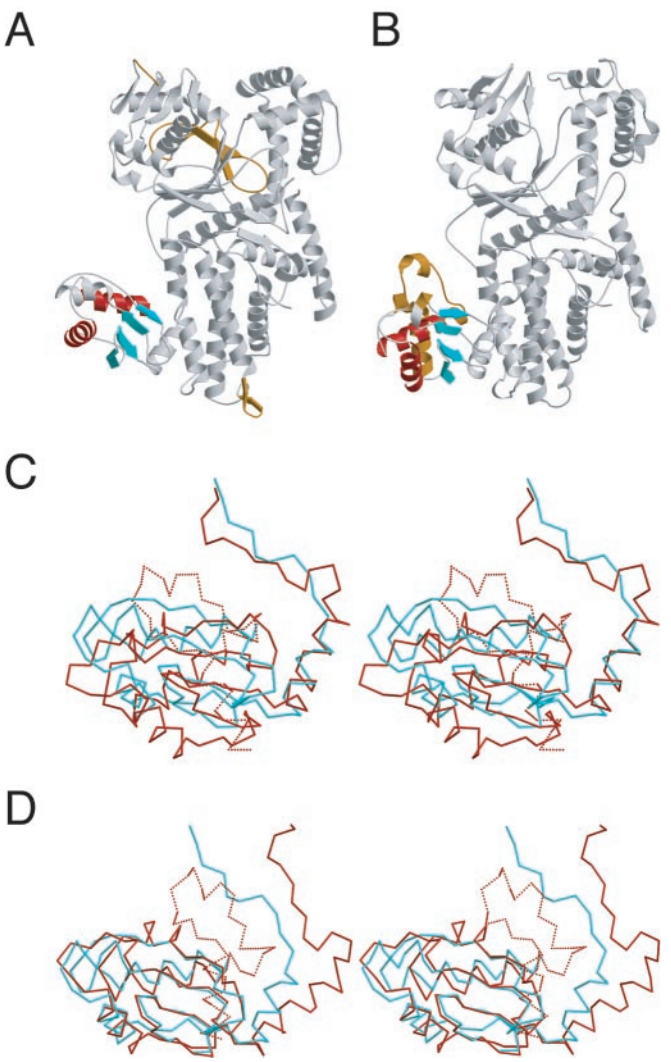


Fig. 2. (*A* and *B*) Ribbon diagrams of *T. thermophilus* ArgRS (*A*) and *S. cerevisiae* ArgRS (*B*). The common antiparallel β-sheets with the lining α helices of the N-terminal domain of the *T. thermophilus* and *S. cerevisiae* ArgRSs are colored red for α-helices and cyan for β-sheets, respectively. The additional N-terminal extension of the *S. cerevisiae* ArgRS and the specific insertions of the *T. thermophilus* ArgRS are colored brown. (*C*) Different orientations of the antiparallel β-sheet with the three lining α helices of the N-terminal domain between the *T. thermophilus* ArgRS (cyan) and the *S. cerevisiae* ArgRS (red), with the rest of the molecules superimposed (stereo view). The N-terminal extension specific to the *S. cerevisiae* ArgRS is indicated as dashed lines. (*D*) Superposition of the antiparallel β-sheet of the N-terminal domain between the *T. thermophilus* ArgRS (cyan) and the *S. cerevisiae* ArgRS (red; stereo view).

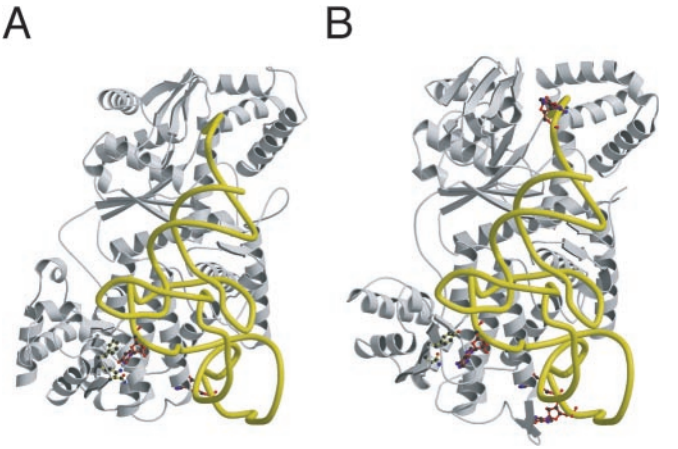
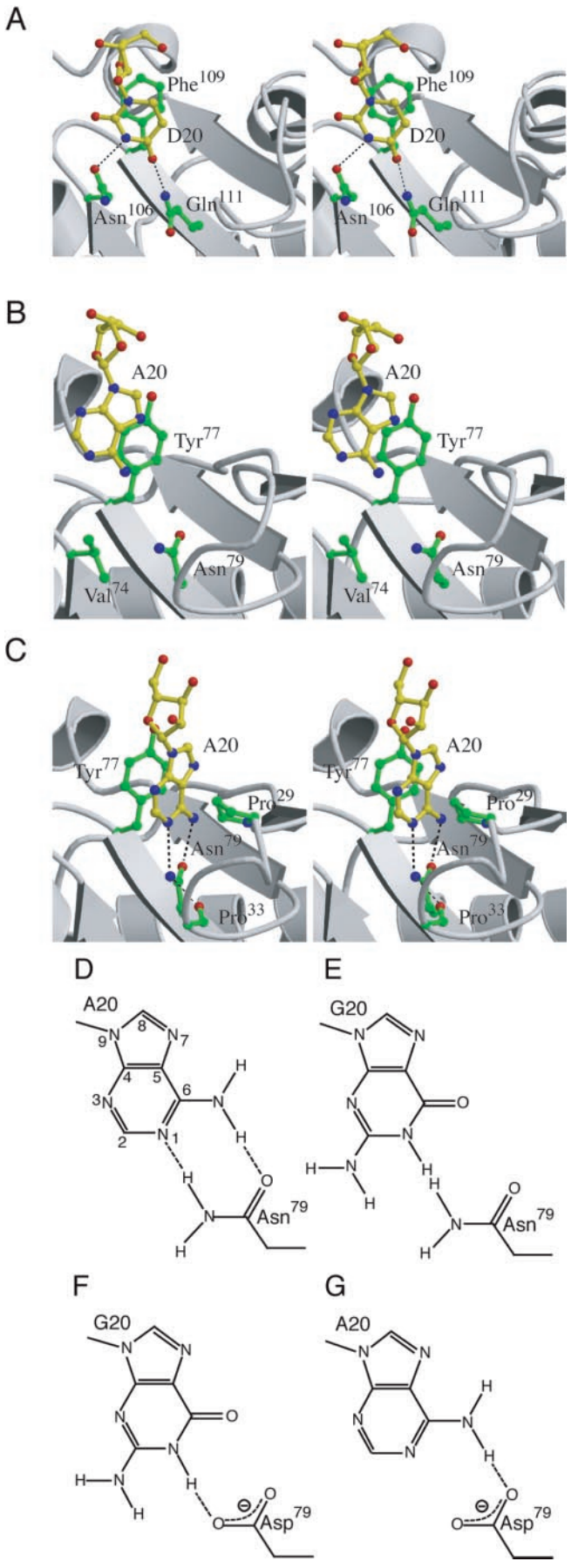


Fig. 3. (*A*) The structure of the *S. cerevisiae* ArgRS complexed with $tRNA^{Arg}$. The tRNA is indicated as a yellow tube. The side chains of Asn^{106}, Phe^{109}, and Gln^{111}, and D20 and C34 are depicted as a ball and stick representation. (*B*) A docking model of the *T. thermophilus* ArgRS and tRNA. The side chains of Tyr^{77} and Asn^{79}, and G34, C35, the discriminator G73, and the predicted A20 are depicted as a ball and stick representation.

Fig. 1), the class-I-specific Rossmann-fold domain (colored brown in Fig. 1), three domains that intervene into the Rossmann-fold domain (colored blue, cyan, and gray, respectively, in Fig. 1), the ''stem contact fold (SC-fold)'' domain (ref. 30; colored purple in Fig. 1), and the α-helix bundle domain in the C terminus (colored red in Fig. 1).

Structural Comparison Between the *T. thermophilus* and *S. cerevisiae* ArgRSs. The overall structure of the *T. thermophilus* ArgRS is similar to that of the *S. cerevisiae* ArgRS, except for several insertions and domains as described below (Figs. 2 and 3). On the basis of the *S. cerevisiae* ArgRS·$tRNA^{Arg}$ complex structure (21), a docking model of the *T. thermophilus* ArgRS and tRNA was constructed (Fig. 3*B*). In this docking model, both C35 (the other major identity determinant) and the 3′ adenosine of $tRNA^{Arg}$ fit well into the putative recognition pockets on the *T. thermophilus* ArgRS structure, in the same manner as the *S. cerevisiae* ArgRS·$tRNA^{Arg}$ complex (Fig. 3). On the other hand, there are three specific insertions in the *T. thermophilus* ArgRS structure. First, a short antiparallel β-sheet (βs and βt) is inserted at the end of α19 in the α-helix bundle domain of the *T. thermophilus* ArgRS (Figs. 1 and 2*A*). The docking model suggests that the short antiparallel β-sheet of the *T. thermophilus* ArgRS comes close to position 34 (the first letter of the anticodon) of $tRNA^{Arg}$ (Fig. 3*B*). Second, in the CP domain, which intervenes into the Rossmann-fold domain (31), the loop between βh and βi is elongated by five amino acid residues in the *T. thermophilus* ArgRS (Fig. 2*A*). This elongated loop of the *T. thermophilus* ArgRS is thought to be close to position 73 (the discriminator base) of $tRNA^{Arg}$ (Fig. 3*B*), whereas the discriminator base is a weak identity determinant of *E. coli* $tRNA^{Arg}$ (17). Specific recognition of the discriminator base was not described for the *S. cerevisiae* ArgRS·$tRNA^{Arg}$ complex (21). The orientation of the CP domain relative to the Rossmann-fold domain in the *T. thermophilus* ArgRS structure is different from that in the *S. cerevisiae* ArgRS·$tRNA^{Arg}$ complex (Fig. 2 *A* and *B*), which is ascribed to the crystal packing effect in the *T. thermophilus* ArgRS crystals (data not shown). The third insertion of the *T. thermophilus* ArgRS intervenes between α11 and βo (colored gray in Fig. 1), and extends the β-sheet of the Rossmann fold by a two-stranded antiparallel β-sheet (βl, βm, and βn) at the back of the catalytic site (Fig. 2*A*). This insertion is species-specific to the *T. thermophilus* and *Pyrococcus horikoshii* ArgRSs (data not shown), and may contribute to the thermostability of the catalytic domain. Finally, the N-terminal domain of the *T. thermophilus* ArgRS has a significantly different structure from that of the *S. cerevisiae* ArgRS, as described below (Fig. 2).

Amino Acid Residues Essential for A20 Recognition. The N-terminal domain (residues 1–102, colored green in Fig. 1) is the most characteristic domain of ArgRS, and is missing in other class-I aaRS structures (20). The *S. cerevisiae* ArgRS also has a homologous N-terminal domain, designated as an additional domain 1 (Add 1; ref. 20). In the structure of the complex of *S. cerevisiae* ArgRS and $tRNA^{Arg}$, Add 1 contacts D20 of the cognate $tRNA^{Arg}$. Phe^{109} of the *S. cerevisiae* ArgRS is involved in a stacking-type interaction with D20, and Asn^{106} and Gln^{111} each



interact with D20 by a single hydrogen bond (ref. 21; Figs. 3*A* and 4*A*). The sequence alignment of the ArgRSs indicates that the A20-recognizing ArgRSs have the invariant Phe/Tyr and Asn residues (Tyr^{77} and Asn^{79} of *T. thermophilus* ArgRS) at the positions corresponding to Phe^{109} and Gln^{111}, respectively, of the *S. cerevisiae* ArgRS (21). By analogy, this invariant Asn residue was suggested to be involved in the interaction with A20 (21).

The N-terminal domain of the *T. thermophilus* ArgRS folds into a globular structure comprising a four-stranded antiparallel β-sheet (βa to -d) with four lining α-helices (α1 to -4; Fig. 1). This N-terminal globule is covalently connected to the Rossmann-fold domain by an extended nine-residue linker, and the linker and helix α4 are fixed on the C-terminal α-helix bundle domain through hydrophobic interactions (Fig. 1). On the other hand, the corresponding N-terminal domain of the *S. cerevisiae* ArgRS has an extension of 33 aa residues at the N terminus. This *S. cerevisiae* ArgRS-specific extension is extensively involved in the hydrophobic core of the N-terminal domain (Fig. 2 *C* and *D*). Because of this significant difference in the N-terminal-domain architecture, the *S. cerevisiae* ArgRS positions the antiparallel β-sheet with the three α helices, corresponding to the *T. thermophilus* α1, α2, and α3, in a quite different orientation relative to the last α-helix, corresponding to the *T. thermophilus* α4 (Fig. 2 *C* and *D*). Nevertheless, the former part, consisting of α1–α3 and βa–βd, and the latter part, consisting of α4 and the linker of the *T. thermophilus* N-terminal domain, separately superpose well on the corresponding parts of the *S. cerevisiae* ArgRS (Fig. 2 *D* and *C*, respectively). Consequently, the orientation of the former part of the N-terminal domain relative to the rest of the entire protein is significantly different between the two ArgRSs, because the latter part of the N-terminal domain is hydrophobically fixed on the C-terminal α-helix bundle domain. This structural difference observed in the N-terminal domain between the *T. thermophilus* and *S. cerevisiae* ArgRSs is not likely to be due to the crystal packing.

The two amino acid residues discussed above with respect to the interaction with the nucleotide at position 20 (Tyr^{77}/Asn^{79} in *T. thermophilus* and Phe^{109}/Gln^{111} in *S. cerevisiae*) are located on one surface of the antiparallel β-sheet (βc and βd in *T. thermophilus* ArgRS, Fig. 1). In the *S. cerevisiae* ArgRS·tRNA^{Arg} complex, D20 directly interacts not only with the two residues Phe^{109} and Gln^{111} but also with Asn^{106} (Fig. 4*A*). In the docking model of *T. thermophilus* ArgRS and tRNA, the invariant Tyr^{77} and Asn^{79} residues were indeed located in the proximity of A20 (Fig. 3*B*). However, A20 is not close enough to contact Tyr^{77} and Asn^{79} (Figs. 3*B* and 4*B*), because of the architectural difference of the N-terminal domain, as described above. If the proposed involvement of the two invariant residues is the case, then either tRNA^{Arg} or ArgRS, or both, should change the conformation to make the two residues interact directly with A20. Because the N-terminal domain is tightly folded and fixed to the neighboring α-helix bundle domain, it is likely that the D-loop, which bears A20 in a flipped out orientation, changes its conformation on tRNA^{Arg} binding to the enzyme. The putative conformational change should translate and rotate A20 for the proper recognition by Tyr^{77} and Asn^{79}. Other than the side chain of Asn^{79},

Fig. 4. (*A*) Interaction of the *S. cerevisiae* ArgRS with D20. Hydrogen bonds between D20 and Asn^{106}, and that between D20 and Gln^{111}, are indicated as dashed lines. (*B*) The docking model of the *T. thermophilus* ArgRS and tRNA. A20 and the side chains of Val^{74}, Tyr^{77}, and Asn^{79} are depicted as a ball and stick representation. (*C*) A model of A20 recognition by Tyr^{77} and Asn^{79}, based on the present mutagenesis analyses. Putative hydrogen bonds formed between the modeled A20 and Asn^{79}, and that between Asn^{79} and Pro^{33}, are indicated as dashed lines. (*D*, *E*, *F*, and *G*) Schematic representations of possible interactions between A20 and Asn^{79} (*D*), G20 and Asn^{79} (*E*), G20 and Asp^{79} (*F*), and A20 and Asp^{79} (*G*). Possible hydrogen bonds are indicated as dashed lines.

Table 2. Arginylation activities of the wild-type and mutant ArgRS enzymes

Protein	tRNA^{Arg} (A20)			tRNA^{Arg} (G20)			tRNA^{Arg} (C20)			tRNA^{Arg} (U20)		
	K_m, μM	V_{max}, s^{-1}	V_{max}/K_m, relative	K_m, μM	V_{max}, s^{-1}	V_{max}/K_m, relative	K_m, μM	V_{max}, s^{-1}	V_{max}/K_m, relative	K_m, μM	V_{max}, s^{-1}	V_{max}/K_m, relative
WT	9.8	0.67	1	ND*	ND	<0.001†	ND	ND	<0.001	ND	ND	<0.001
Y77F	20.4	0.66	0.47	ND	ND	<0.001	ND	ND	<0.001	ND	ND	<0.001
Y77A	ND	ND	<0.001	ND	ND	<0.001	ND	ND	<0.001	ND	ND	<0.001
N79A	20.0	0.019	0.013	ND	ND	<0.001	ND	ND	<0.001	ND	ND	<0.001
N79D	7.5	0.38	0.74	9.5	0.33	0.51	ND	ND	<0.001	ND	ND	<0.001
N79Q	25.2	0.034	0.02	19.6	0.047	0.035	ND	ND	<0.001	ND	ND	<0.001
N79E	12.5	0.14	0.17	9.0	0.061	0.099	ND	ND	<0.001	19.6	0.030	0.022
N79K	8.3	0.0024	0.0042	ND	ND	<0.001	ND	ND	<0.001	58.0	0.029	0.0073
N79R	21.7	0.0069	0.0046	16.5	0.0036	0.0032	ND	ND	<0.001	9.3	0.0037	0.0058

*ND, Not determined.
†The arginylation could not be detected even when the concentration of the enzyme was increased to 5 μM.

there is no other polar group that might interact with A20. Val^{74} of the *T. thermophilus* ArgRS, which is located at the position corresponding to Asn^{106} of the *S. cerevisiae* ArgRS, is too far from A20 to be involved in the recognition (Fig. 4*B*).

Mutational Analyses of the ArgRS. To validate the proposed model of the A20 recognition by the canonical ArgRSs (21), and to investigate the recognition modes of Tyr^{77} and Asn^{79} further, we mutated the two invariant residues in the N-terminal domain. The wild-type *T. thermophilus* ArgRS arginylated the tRNA^{Arg} harboring A20, whereas it did not arginylate the tRNA^{Arg} variants with G20, C20, or U20 (Table 2). This result confirmed that A20 actually acts as a major identity element of the *T. thermophilus* tRNA^{Arg}.

The replacement of Tyr^{77} by Ala resulted in a significant reduction in the arginylation (Table 2), whereas the replacement of Tyr^{77} by Phe had no serious effect on the arginylation (Table 2). These results indicate that Tyr^{77} makes a stacking interaction with A20. In addition, the replacement of Asn^{79} by Ala reduced the arginylation activity for the tRNA^{Arg} (Table 2), indicating that Asn^{79} plays an important role in A20 recognition.

Surprisingly, the N79D mutant equally arginylated tRNA^{Arg} transcripts harboring A20 and G20, and the relative arginylation activity was almost the same as that of the wild-type ArgRS for the tRNA^{Arg} harboring A20 (Table 2). The mutants of Asn^{79}, such as the N79D mutant, might be highly detrimental to cell growth, because they might arginylate non-cognate tRNAs. This detrimental effect may have acted as the major selective pressure to conserve the invariant Asn residue. The success in the overproduction of these potential misarginylating mutants might be due to the low activity of the *T. thermophilus* enzymes at the growth temperature of the host *E. coli* cells (37°C). The arginylation activities of the N79D mutant for the tRNA^{Arg} harboring A20 and G20 were significantly higher than that of the N79A mutant. These results indicate that the N79D mutant acquired the ability to interact with G20, instead of having lost the nucleotide specificity for A20.

The replacement of Asn^{79} by other polar amino acids reduced the arginylation activity for the tRNA^{Arg} without exception (Table 2). The amount of the arginylation activity reduction, however, depended on the mutation. For example, the N79R and N79K mutations remarkably reduced the arginylation efficiency, to a rate even lower than that caused by the N79A mutation. The bulky and positively charged side chains of the Arg and Lys residues may destroy the local conformation of the recognition pocket for A20.

It should be noted that the replacement of Tyr^{77} or Asn^{79} by other amino acids predominantly affected the V_{max}, rather than the K_m, in all cases. This tendency corresponds well to the fact that the A20 mutations of *E. coli* tRNA^{Arg} mostly affected V_{max}, rather than K_m, in the arginylation (16). These results suggest that the A20 recognition mainly contributes to the "enzyme activation," which may include changes in the spatial arrangement of important residues in the catalytic cleft and/or the CCA terminus of the tRNA.

The Recognition Pocket for A20. On the basis of the mutational analyses, we constructed a local structural model of the interaction of A20 with Tyr^{77} and Asn^{79}, by moving A20 toward the two residues (Fig. 4*C*). The aromatic ring of Tyr^{77} was thought to make a stacking interaction with the base of A20, and, therefore, the adenosine could be modeled within an appropriate distance for a stacking interaction with Tyr^{77} (≈3.5 Å distant, Fig. 4*C*). On the other hand, based on the mutational results of the base-specific interaction between Asn^{79} and A20, the γ-NH$_2$ and γ-CO groups of Asn^{79} may form bipartite hydrogen bonds with the N1 and N6, respectively, of A20 (Fig. 4 *C* and *D*). Except for Tyr^{77} and Asn^{79}, the side chain of Pro^{29} is located at an appropriate distance from the modeled adenosine for a van der Waals interaction (≈3.5 Å distant, Fig. 4*C*). Therefore, in this case, A20 is probably sandwiched between Tyr^{77} and Pro^{29}.

This model can explain how A20 is discriminated from guanosine. As shown in Fig. 4*E*, the γ-NH$_2$ group of Asn^{79} would clash with the N1 of guanosine, thus precluding the guanosine from being recognized. The side chain of Asn^{79} may be fixed by a hydrogen bond between the γ-NH$_2$ group of Asn^{79} and the α-CO group of Pro^{33} (≈3.2 Å distant, Fig. 4*C*). Therefore, hydrogen bonds cannot be formed between Asn^{79} and the guanosine by the rotation of the Asn^{79} side chain on its C_β-C_γ axis. Our model of A20 recognition is also consistent with the fact that the N79D mutant efficiently arginylated tRNA^{Arg} with G20. The replacement of Asn^{79} by Asp may remove the steric hindrance caused by the γ-NH$_2$ group, and yet the δ2-O atom of Asp could serve as another hydrogen-bond acceptor for the N1 of the guanosine (Fig. 4*F*). On the other hand, the replacement of Asn^{79} by Asp had only a small effect on the arginylation (Table 2). This fact may be ascribed to the distorted orientation of Asn^{79}. The γ-NH$_2$ group of Asn^{79} is not on the plane of the modeled adenine base, and thus deviates from the ideal position to provide a hydrogen bond with the N1 of the nucleotide (Fig. 4*C*). Therefore, the γ-NH$_2$ group of Asn^{79} may have a minor role in the arginylation, as compared with the γ-CO group of Asn^{79} (Fig. 4 *D* and *G*).

The recognition mode of A20 described above is compared with the interaction between D20 and the *S. cerevisiae* ArgRS (Fig. 4 *A* and *C*). In the *T. thermophilus* case, A20 forms bipartite hydrogen bonds, analogous to Watson-Crick base pairing, with the Asn^{79} side chain (Fig. 4 *C* and *D*). By contrast, in the complex

BIOCHEMISTRY

of *S. cerevisiae* ArgRS and $tRNA^{Arg}$ (Fig. 4*A*), D20 forms only one hydrogen bond with Gln^{111}, which corresponds to the *T. thermophilus* Asn^{79}, but forms another hydrogen bond with Asn^{106}, which is replaced by Val^{74} in the *T. thermophilus* ArgRS. Therefore, the orientation of the A20 base relative to Asn^{79} of the *T. thermophilus* ArgRS is slightly different from that of the D base relative to Gln^{111} of the *S. cerevisiae* ArgRS. The D20-recognition pocket of the *S. cerevisiae* ArgRS may recognize C20 by concomitant flipping of the side chain extremities of Asn^{106} and Gln^{111}, to fulfill the hydrogen bond scheme (21). Furthermore, it appears to be possible for the *S. cerevisiae* pocket to adopt even A20 in a similar manner, whereas the *S. cerevisiae* $tRNA^{Arg}$ species have U, dihydrouridine (D), or C, but not A, at position 20 (15). Actually, the arginine-accepting activity was not decreased by the replacement of C20 by A in a *S. cerevisiae* $tRNA^{Arg}$ species (18). Therefore, it might be possible that the *S. cerevisiae* ArgRS has multiple specificities toward position 20 of the tRNA. In contrast, the canonical ArgRSs can recognize only A at this position, because the side chain of the invariant Asn residue (Asn^{79}) is prevented from rotating by a hydrogen bond between the side chain NH_2 group and the main chain CO group (of Pro^{33}, as described above; Fig. 4*C*).

A major unsolved problem is the mechanism by which A20 recognition leads to the enzyme activation that causes the arginylation with such a high discrimination rate. The ongoing structure determination of the complex of *T. thermophilus* ArgRS and $tRNA^{Arg}$ will provide a structural basis for the mechanism of enzyme activation after A20 recognition, as well as a detailed mechanism of A20 recognition.

We thank Masayoshi Nakasako, Michiko Konno, and Masami Horikoshi for data collection. We are greatly indebted to Nobuo Kamiya (Institute of Physical and Chemical Research) and Masahide Kawamoto (Japan Synchrotron Radiation Research Institute) for their help in data collection at SPring-8. We thank Jean Cavarelli for providing us the coordinates of the *S. cerevisiae* ArgRS·$tRNA^{Arg}$ complex. We also thank Kiyotaka Shiba, Shun-ichi Sekine, Takaho Terada, and Kyoko Saito for helpful discussions. We are grateful to Kouichiro Kodama, Takanori Kigawa, Hitoshi Kurumizaka, and Mikako Shirouzu for their help in the measurement of arginylation kinetics. This work was supported in part by Grants-in-Aid for Science Research on Priority Areas (09278101 and 11169204) to S.Y. and O.N., respectively, from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by a grant from the Japan Society for the Promotion of Science (JSPS) Research Fellowships for Young Scientists (A.S.).

1. Giegé, R., Sissler, M. & Florentz, C. (1998) *Nucleic Acids Res.* **26,** 5017–5035.
2. Cavarelli, J., Rees, B., Ruff, M., Thierry, J. C. & Moras, D. (1993) *Nature (London)* **362,** 181–184.
3. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991) *Science* **252,** 1682–1689.
4. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) *Science* **246,** 1135–1142.
5. Rould, M. A., Perona, J. J. & Steitz, T. A. (1991) *Nature (London)* **352,** 213–218.
6. Silvian, L. F., Wang, J. & Steitz, T. A. (1999) *Science* **285,** 1074–1077.
7. Sankaranarayanan, R., Dock-Bregeon, A. C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B. & Moras, D. (1999) *Cell* **97,** 371–381.
8. Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994) *Science* **263,** 1404–1410.
9. Soma, A., Uchiyama, K., Sakamoto, T., Maeda, M. & Himeno, H. (1999) *J. Mol. Biol.* **293,** 1029–1038.
10. Asahara, H., Himeno, H., Tamura, K., Hasegawa, T., Watanabe, K. & Shimizu, M. (1993) *J. Mol. Biol.* **231,** 219–229.
11. Breitschopf, K., Achsel, T., Busch, K. & Gross, H. J. (1995) *Nucleic Acids Res.* **23,** 3633–3637.
12. Normanly, J., Ogden, R. C., Horvath, S. J. & Abelson, J. (1986) *Nature (London)* **321,** 213–219.
13. Normanly, J., Ollick, T. & Abelson, J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5680–5684.
14. McClain, W. H. & Foss, K. (1988) *Science* **241,** 1804–1807.
15. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998) *Nucleic Acids Res.* **26,** 148–153.
16. Schulman, L. H. & Pelka, H. (1989) *Science* **246,** 1595–1597.
17. Tamura, K., Himeno, H., Asahara, H., Hasegawa, T. & Shimizu, M. (1992) *Nucleic Acids Res.* **20,** 2335–2339.
18. Liu, W., Huang, Y.-W., Eriani, G., Gangloff, J., Wang, E.-D. & Wang, Y.-L. *(1999) Biochim. Biophys. Acta* **1473,** 356–362.
19. McClain, W. H., Foss, K., Jenkins, R. A. & Schneider, J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9260–9264.
20. Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J. & Moras, D. (1998) *EMBO J.* **17,** 5438–5448.
21. Delagoutte, B., Moras, D. & Cavarelli, J. (2000) *EMBO J.* **19,** 5599–5610.
22. Shimada, A., Nureki, O., Dohmae, N., Takio, K. & Yokoyama, S. (2001) *Acta Crystallogr. D* **57,** 272–275.
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119.
24. Brünger, A. T. (1992) *X-PLOR Version 3.1 Manual* (Yale Univ. Press, London).
25. 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.
26. Collaborative Computational Project, Number 4. (1994) *Acta Crystallogr. D* **50,** 760–763.
27. Zubay, G. (1962) *J. Mol. Biol.* **4,** 347–356.
28. Kigawa, T., Muto, Y. & Yokoyama, S. (1995) *J. Biomol. NMR* **6,** 129–134.
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291.
30. Sugiura, I., Nureki, O., Ugaji-Yoshikawa, Y., Kuwabara, S., Shimada, A., Tateno, M., Lorber, B., Giegé, R., Moras, D., Yokoyama, S. & Konno, M. (2000) *Struct. Fold. Des.* **8,** 197–208.
31. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
32. Merrit, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.