ABSTRACT

In many prokaryotes the biosynthesis of the amide aminoacyl-tRNAs, Gln-tRNA\textsuperscript{Gln} and Asn-tRNA\textsuperscript{Asn}, proceeds by an indirect route in which mischarged Glu-tRNA\textsuperscript{Gln} or Asp-tRNA\textsuperscript{Asn} is amidated to the correct aminoacyl-tRNA catalyzed by a tRNA-dependent amidotransferase (AdT). Two types of AdTs exist: bacteria, archaea and organelles possess heterotrimeric GatCAB, while heterodimeric GatDE occurs exclusively in archaea. Bacterial GatCAB and GatDE recognize the first base pair of the acceptor stem and the D-loop of their tRNA substrates, while archaeal GatCAB recognizes the tertiary core of the tRNA, but not the first base pair. Here, we present the crystal structure of the full-length \textit{Staphylococcus aureus} GatCAB. Its GatB tail domain possesses a conserved Lys rich motif that is situated close to the variable loop in a GatCAB:tRNA\textsuperscript{Gln} docking model. This motif is also conserved in the tail domain of archaeal GatCAB, suggesting this basic region may recognize the tRNA variable loop to discriminate Asp-tRNA\textsuperscript{Asn} from Asp-tRNA\textsuperscript{Asp} in archaea. Furthermore, we identified a 3\(_{10}\) turn in GatB that permits the bacterial GatCAB to distinguish a U1–A72 base pair from a G1–C72 pair; the absence of this element in archaeal GatCAB enables the latter enzyme to recognize aminoacyl-tRNAs with G1–C72 base pairs.

INTRODUCTION

Correct pairing of an amino acid with its cognate tRNA is an essential step to maintain the accuracy of translation. This is usually accomplished by aminoacyl-tRNA synthetases (aaRSs) that catalyze the direct attachment of an amino acid to its cognate tRNA (1). However, glutaminyl-tRNA synthetase (GlnRS) is absent in the majority of bacteria, in all known archaea, and in many organelles (2). These organisms utilize an indirect pathway for Gln-tRNA\textsuperscript{Gln} formation where a non-discriminating glutamyl-tRNA synthetase (ND-GluRS) synthesizes mischarged Glu-tRNA\textsuperscript{Gln} (3) that is then amidated to the cognate Gln-tRNA\textsuperscript{Gln} by glutamyl-tRNA\textsuperscript{Gln} amidotransferase (Glu-AdT) (2,4). Similarly, many prokaryotes lacking an asparaginyl-tRNA synthetase (AsnRS) generate Asn-tRNA\textsuperscript{Asn} by the combined actions of a non-discriminating aspartyl-tRNA synthetase (ND-AspRS) and an aspartyl-tRNA\textsuperscript{Asn} amidotransferase (Asp-AdT) (5,6).

Two types of AdTs exist: the heterotrimeric GatCAB (7) present in bacteria, archaea and organelles (2), and the heterodimeric GatDE found exclusively in archaea (8). GatDE specifically converts Glu-tRNA\textsuperscript{Gln} (8), while bacterial GatCAB acts as both a Glu-AdT and an Asp-AdT \textit{in vitro} (2). \textit{The in vivo} role of bacterial GatCAB is defined by the nature of ND-aaRS (ND-GluRS and/or ND-AspRS) present in the cell (2). In archaea that lack AsnRS, GatCAB is encoded (9). This enzyme (e.g. from \textit{Methanothermobacter thermautotrophicus}) \textit{in vitro} strongly prefers Asp-tRNA\textsuperscript{Asn} over the homologous Glu-tRNA\textsuperscript{Gln} (10); thus archaeal GatCAB may act \textit{in vivo} as an Asp-AdT.

AdTs accurately distinguish their mischarged aa-tRNA substrates (Glu-tRNA\textsuperscript{Glu} and/or Asp-tRNA\textsuperscript{Asp}) from the cognate Glu-tRNA\textsuperscript{Glu} and Asp-tRNA\textsuperscript{Asp} species. Bacterial GatCAB and GatDE achieve this by recognizing the first base pair of the acceptor stem and the D-loop of their tRNA substrates (11–13). In contrast, archaeal GatCAB does not recognize the first base pair of Asp-tRNA\textsuperscript{Asn} (13,14). Instead the \textit{M. thermautotrophicus} GatCAB makes use of the D-loop, the nucleotide in position 49, and to a lesser extent of the length of the variable loop to distinguish Asp-tRNA\textsuperscript{Asn} from Gln-tRNA\textsuperscript{Gln}.
Asp-tRNA\textsuperscript{Asp} (14), while the \textit{ Methanosarcina barkeri } GatCAB appears to use primarily the length of the variable loop for the same task (13).

Previous work implicated the tail domain of the GatB and GatE in D-loop recognition of their respective aa-tRNA substrates (11,12). These tail domains share homology with the standalone YqeY proteins of \textit{Escherichia coli} (10). Here, we present a full-length domain of AdTs (11,12,17,18) or of crystallographic studies did not resolve the YqeY-like homology with the standalone YqeY proteins of \textit{Staphylococcus aureus} (15). Here, we present a full-length \textit{Staphylococcus aureus} GatCAB structure (containing the \textit{S. aureus} GlnRS (15). Here, we present a full-length domain of AdTs (11,12,17,18) or of crystallographic studies did not resolve the YqeY-like domain of AdTs (11,12,17,18) or of \textit{Deinococcus radiodurans} GlnRS (15). Here, we present a full-length \textit{Staphylococcus aureus} GatCAB structure (containing the YqeY domain) resolved at 1.9 \AA. The structure reveals a Lys rich motif in the tail domain of GatB which is absent in GatE. We also identified by biochemical studies a bacterial GatCAB-specific element that enables this enzyme to distinguish a U1–A72 base pair from a G1–C72 base pair.

**MATERIALS AND METHODS**

**Preparation of \textit{S. aureus} GatCAB**

The enzyme was over-produced in an \textit{Escherichia coli} B834 strain and purified over a HiTrap HP column (GE Healthcare) as described (11). The sample was then diluted 5-fold with buffer \textit{A} [50 mM Tris–HCl pH 7.5, 10 mM MgCl\textsubscript{2}, 1 mM DTT and 10% (v/v) glycerol], and applied to a HiTrap Heparin HP column (GE Healthcare) equilibrated with buffer \textit{A}. The column was washed with buffer \textit{A} containing 50 mM NaCl, and proteins were eluted with a linear gradient of 50–500 mM NaCl. The enzyme eluted at \approx 250 mM NaCl. The enzyme fractions were then loaded onto a HiLoad 26/60 Superdex 200 pg column (GE Healthcare) equilibrated with buffer \textit{B} [20 mM Tris–HCl pH 7.5, 10 mM MgCl\textsubscript{2}, 1 mM DTT and 10% (v/v) glycerol]. Pooled fractions were concentrated by ultrafiltration using Vivaspin devices (VIVASCIENCE) to a final concentration of 18 mg/ml. The enzyme was purified by a Lys rich motif in the tail domain of GatB which is absent in GatE. We also identified by biochemical studies a bacterial GatCAB-specific element that enables this enzyme to distinguish a U1–A72 base pair from a G1–C72 base pair.

**Preparation of \textit{M. thermautotrophicus} GatCAB and ND-AspRS**

The over-production and purifications were as described (10). \textit{Methanothermobacter thermautotrophicus} GatCAB mutants were generated as described above.

**Preparation of \textit{S. aureus} ND-GluRS**

The gene encoding the ND-GluRS from \textit{S. aureus} Mu50 was amplified by polymerase chain reaction (PCR). The \textit{S. aureus} ND-GluRS is toxic for an \textit{E. coli}, which does not possess GatCAB, therefore the gene was cloned into NcoI/XhoI site of a pTip vector, which is used to protein expression in \textit{Rhodococcus erythropolis}, and then the \textit{S. aureus} ND-GluRS protein was expressed in \textit{R. erythropolis} as described (19). Cells were harvested (4000 x g, 15 min at 4°C) and disrupted using sonication in buffer \textit{C} [50 mM Tris–HCl pH 7.5, 300 mM NaCl, 5 mM MgCl\textsubscript{2}, 10% (v/v) glycerol, 0.5 mg/ml lysozyme and 0.1 mg/ml DNase I]. All the following purification processes were carried out at 4°C. Cell debris was removed by centrifugation (40 000 x g, 1 h), and clarified supernatant was applied to a HisTrap HP column as described (11). Pooled fractions were loaded onto a HiLoad 26/60 Superdex 200 pg column equilibrated with buffer \textit{D} [20 mM HEPES-K pH 7.6, 5 mM MgCl\textsubscript{2}, 100 mM KCl, 1 mM DTT and 10% (v/v) glycerol]. The purified \textit{S. aureus} ND-GluRS was concentrated by ultrafiltration to a final concentration of 18 mg/ml, and then diluted 2-fold with 100% (v/v) glycerol, and stored at \textit{–}30°C.

**Preparation of \textit{S. aureus} tRNA\textsuperscript{Gln}, and \textit{M. thermautotrophicus} and \textit{C. trachomatis} tRNA\textsuperscript{Asn}**

The tRNA isoacceptors were in vitro transcribed and purified as described (10,11).

**Preparation of aminoacyl-[\textsuperscript{32}P] labeled tRNA**

The tRNA isoacceptors were [\textsuperscript{32}P]-labeled and aminoacylated as described (10,20) with minor modifications. For glutamylation of \textit{S. aureus} tRNA\textsuperscript{Gln}, 5 \textmu M \textit{S. aureus} ND-GluRS and tRNA\textsuperscript{Gln} were added in the aminoacylation reaction.

**Crystallization and structure determination**

The high quality single crystals of \textit{S. aureus} GatCAB were obtained by using the micro-seeding technique (21) from hanging drops set up in a 1:1:0.1 ratio from protein, reservoir solution [25% (w/v) PEG 600, 5 mM MgCl\textsubscript{2}, 50 mM HEPES–NaOH pH 7.2 and 3% (v/v) 2-methyl-2,4-pentanediol (MPD)] and a micro-seeds stock solution (21). The crystal of \textit{S. aureus} GatCAB was rapidly soaked through the reservoir containing 50 mM MES-Na, pH 6.4, 25% (w/v) PEG 600, 5 mM MgCl\textsubscript{2}, 3% (v/v) MPD and 10% (v/v) glycerol as a cryoprotectant, and then a data set was collected to 1.9 \AA resolution at SPring-8 beamline 41XU (Hyogo, Japan) under cryogenic condition (\textit{–}173°C). The data set was processed and scaled using the HKL2000 package (22). The structure of \textit{S. aureus} GatCAB was solved by molecular replacement using AMoRe (23), with the refined model of the previous (11) \textit{S. aureus} GatCAB structure (PDB ID: 2G5H) as a search model. The model of \textit{S. aureus} GatCAB was rebuilt by automatic refinement program LA\textit{FIRE} (24,25) running with CNS (26), and modified manually by using Coot (27) followed yielded the final model with the crystallographic \textit{R}/\textit{R}_{\text{free}} factor of 19.5/21.4%. The summary of data statistics is presented in Table 1. All figures were generated by PyMol (28).

**Small-angle X-ray scattering**

For preparation of the tRNA\textsuperscript{Gln}-bound \textit{S. aureus} GatCAB, the \textit{S. aureus} GatCAB was mixed with tRNA\textsuperscript{Gln} in a molar ratio of 1:4, and then purified by a
Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB ID</td>
<td>3IP4</td>
<td></td>
</tr>
<tr>
<td>Beaml ine</td>
<td>SPring-8 BL41XU</td>
<td></td>
</tr>
<tr>
<td>Wavelength</td>
<td>1.00 Å</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2&lt;sub&gt;1&lt;/sub&gt;2&lt;sub&gt;1&lt;/sub&gt;2&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>Cell dimensions a, b, c (Å)</td>
<td>71.1, 92.7, 180.4</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00–1.90 (1.97–1.90)</td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>0.070 (0.457)</td>
<td></td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>23.2 (3.5)</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100)</td>
<td></td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20.0–1.90</td>
<td></td>
</tr>
<tr>
<td>No. reflections</td>
<td>94598</td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;, R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.195/0.214</td>
<td></td>
</tr>
<tr>
<td>Number of atoms (protein/water/others)</td>
<td>8299/801/1</td>
<td></td>
</tr>
<tr>
<td>Average B factors (Å&lt;sup&gt;2&lt;/sup&gt;) (protein/water/others)</td>
<td>33.6/35.3/26.0</td>
<td></td>
</tr>
<tr>
<td>RMSD bond length (Å)/angles (°)</td>
<td>0.006/1.3</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td>90.2</td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>9.6</td>
<td></td>
</tr>
<tr>
<td>Allowed</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Generous</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parentheses are for the outermost resolution shell.
<sup>b</sup>R<sub>merge</sub> = Σ<sub>j</sub>Σ<sub>i</sub>∥F<sub>o</sub><sup>i</sup>−F<sub>e</sub><sup>i</sup>∥/Σ<sub>j</sub>Σ<sub>i</sub>∥F<sub>e</sub><sup>i</sup>∥, where ∥F<sub>e</sub><sup>i</sup>∥ is the mean intensity of symmetry-equivalent reflections.
<sup>c</sup>R<sub>work</sub> = Σ<sub>j</sub>Σ<sub>i</sub>∥F<sub>o</sub><sup>i</sup>−F<sub>e</sub><sup>i</sup>∥/Σ<sub>j</sub>Σ<sub>i</sub>∥F<sub>o</sub><sup>i</sup>∥, where F<sub>o</sub><sup>i</sup> and F<sub>e</sub><sup>i</sup> are observed and calculated structure factor amplitudes.
<sup>d</sup>R<sub>free</sub> value was calculated for R factor, using only an unrefined subset of reflections data.

Determination of kinetic parameters with the <i>S. aureus</i> GatCAB enzymes was carried out as described for the <i>Helicobacter pylori</i> GatCAB (20). Determination of the kinetic parameters with the <i>M. thermototrophicus</i> GatCAB enzymes was carried out as described (10).

**RESULTS**

**Overall structure of a full-length GatCAB**

We crystallized <i>S. aureus</i> GatCAB by adding 3–10% (v/v) MPD to the previous crystallization condition (11), which dramatically improved the crystal quality (maximum resolution from 2.3 to 1.9 Å). Surprisingly, the high-resolution structure clearly shows the GatB C-terminal region encompassing amino acids 412–475 and a histidine-tag (Figure 1A and Supplementary Figure S1), whose electron density map was not visible in the previous crystal structures. Although the unit cell dimension and crystal packing of the full-length and the previous apo-form <i>S. aureus</i> GatCAB are very similar, in the current structure the C-terminal GatB region is sandwiched between two GatA molecules related by 2-fold screw axes. Successful resolution of the tail-domain may be due to MPD addition in the crystallization mixture making the protein domain less flexible.

The GatA apo-form is essentially the same as the previously deposited <i>S. aureus</i> GatA structures (11). Interestingly, the high-resolution structure shows an alternative conformation for R358, which recognizes the carboxyl group of the substrate Gln (Supplementary Figure S2). Furthermore, Y310, which made a hydrogen bond with the substrate Gln (Supplementary Figure S1), was disordered and could not be modeled.

GatB is comprised of three consecutive domains: a cradle domain (1–294), a helical domain (295–407) and a tail domain (412–475). The cradle and helical domains are identical to the <i>S. aureus</i> GatB apo-form (PDB ID: 2G5I), and the permanent Mg<sup>2+</sup> in the catalytic pocket was visible (11,18). The tail domain forms an anti-parallel helix bundle with three amphiphilic helices (113, 115 and 116) that construct a hydrophobic core (Figure 1B and Supplementary Figure S1). L472 participates in this hydrophobic core and is not exposed to the molecular surface, indicating L472 is important for maintaining the structure of the tail domain rather than directly recognizing the aa-tRNA substrate as previously suggested (11). The loop between 115 and 116 interacts with I412 and S413 by two main chain and one side chain hydrogen bonds. Therefore, the <i>S. aureus</i> GatB tail and helical domains are linked by a ~13 Å long inter-domain loop (408–411).

**Comparison of the GatB tail domain with YqeY**

The C-terminal end of GatB, comprised of the helical and tail domains (295–475), belongs to the same protein family.
as standalone YqeY polypeptides of unknown function present in many bacteria and in yeast (PFAM ID: PF09424) (15,16). The YqeY-like tail domain appended to the \textit{D. radiodurans} GlnRS enables the enzyme to productively bind to tRNA\textsubscript{Gln} (15). A similar role is proposed for this structure in GatB and GatE via recognition of the D-loop (11,12). The only YqeY structure so far resolved is from \textit{Bacillus subtilis} (PDB ID: 1NG6); it can now be compared with the YqeY region in the present full-length of \textit{S. aureus} GatCAB structure. The anti-parallel helix bundle of the tail domain superimposes well with the C-terminal domain of \textit{B. subtilis} YqeY protein (YqeY-C: 92-146) with an r.m.s. deviation of 1.97 Å for the 51 C\textsubscript{a} pairs compared (Figure 1C). Interestingly, the GatB tail domain has an additional \(\alpha\)-helix (\(\alpha\)14) and a short loop between \(\alpha\)14 and \(\alpha\)15 (439–442) not found in the YqeY-C. Based on the multiple sequence alignment with the tail domains and YqeY-C, this extended region contains a Lys rich motif (KXGKK) that is highly conserved in bacterial and archaeal GatB enzymes (Figure 1D and Supplementary Figure S3). The C-terminal extension of \textit{D. radiodurans} GlnRS also contains this extended region with a similar motif (RGGKTA). In contrast to GatB and \textit{D. radiodurans} GlnRS, the tail domain of GatE lacks \(\alpha\)14 and the Lys rich motif. Interestingly, the Lys rich motif of GatB is instead replaced in GatE with a GXXAXGX motif that has been implicated in GatDE distinguishing Glu-tRNA\textsubscript{Gln} from Asp-tRNA\textsubscript{Asn} (16).

Docking tRNA\textsubscript{Gln} into the \textit{S. aureus} GatCAB structure

The GatB C-terminal tail is essential for GatCAB binding tRNA\textsubscript{Gln} (11). In the co-crystal structure of the GatDE enzyme (PDB ID: 2D6F), the GatE tail domain was in the vicinity of the tRNA D-loop; however, its detailed structure could not be solved and the D-loop was fitted with the help of the \textit{B. subtilis} YqeY-C structure (12). We decided to create a docking model of full-length \textit{S. aureus} GatCAB with \textit{E. coli} tRNA\textsubscript{Gln} to understand how the enzyme uses the tail domain to bind the tRNA. The use of \textit{E. coli} tRNA\textsubscript{Gln} is reasonable as the sequences of the variable and D-loops of the \textit{S. aureus} and \textit{E. coli} tRNA\textsubscript{Gln} are identical (Figure 2A), and \textit{E. coli} Glu-tRNA\textsubscript{Gln} serves as substrate for GatCAB in vivo (39).

To construct the docking model we first superposed \textit{E. coli} tRNA\textsubscript{Gln} (PDB ID: 1QTQ) into the GatDE:tRNA\textsubscript{Gln} co-crystal structure by aligning all atoms of nucleotides U8, A14–G19, A21, C48 and U54–G57 of the tRNAs. The U8–A14–A21 base triple in the augmented D-stem is commonly found in tRNA\textsubscript{Gln} while nucleotide 15 in the D-loop and nucleotide 48 in the variable loop form the conserved Levitt pair (40,41).
The tertiary interaction between the D- and TΨC- loops of tRNA is also well conserved.

Next, we docked *S. aureus* GatCAB to tRNA\textsuperscript{Gln} by taking advantage of the homology between GatB and GatE (8,16), and the co-crystal structure of GatDE:tRNAGln (12), superposing GatB with *M. thermautotrophicus* GatE. The cradle domain of *S. aureus* GatB and *M. thermautotrophicus* GatE superpose well with an r.m.s. deviation of 1.7 Å for 244 pairs of C\textalpha atoms compared. However, initially there were severe clashes between the helical domain and tRNA\textsuperscript{Gln}. Therefore, we separated the helical and tail domains into three parts (293–363, 364–381 and 382–475) and then superposed into GatE independently. Such movements are predicted based on previous AdT structures (11,12,17). Finally, the cradle and helical domains of *S. aureus* GatB could be superposed into that of *M. thermautotrophicus* GatE with an r.m.s. deviation of 1.9 Å for 321 pairs of C\textalpha atoms compared (Supplementary Figure S4A).

The GatB tail domain possesses a highly conserved hydrophobic pocket comprised by residues (V449, M452, G457, A459 and P461) from the \( \alpha15, \alpha16 \) and the loop between them (Figure 3A), surrounded by positively charged residues (Figure 3B). In contrast, the opposite surface of the tail domain is composed of non-conserved, mostly negative residues (Figure 3A and B). These observations suggest that this conserved hydrophobic pocket may recognize the shape of the tRNA\textsuperscript{Gln} D-loop with the surrounding positive residues interacting with the tRNA phosphate backbone (Figure 3C).

Interaction of the tail domain with tRNA is expected, as deletion of the C-terminal portion of GatB gives rise to a GatCAB mutant enzyme unable to bind tRNA (11) and as mentioned the electron density map of the GatDE:tRNAGln structure places the YeqY-like tail domain in proximity of the tRNA D-loop (12). However, the initial *S. aureus* GatCAB:tRNAGln docking model placed the conserved hydrophobic pocket of the tail domain ~12 Å away from the D-loop of tRNAGln (Supplementary Figure S4B). It is likely that the tail domain can move to interact with tRNAGln due to the flexible loop connecting the tail and helical domains. This domain flexibility may explain why it has
been difficult to resolve the YqeY-like tail of previous AdT structures (11,12,17,18). Given the above, we manually fitted the pocket of the tail domain into the D-loop of the *E. coli* tRNAGln (Figure 3C). The distance between the N-terminal end of the tail domain (Q411) and the C-terminal end of the helical domain (N407) is 10 Å, a distance the inter-domain loop connecting the two domains can bridge (Figure 3C). Recognition of the D-loop is consistent with the fact bacterial GatCAB uses that tRNA element to distinguish transamidation substrates (Glu-tRNA Gln and Asp-tRNA Asn) from Glu-tRNAGlu and Asp-tRNA Asp (11,13).

In order to verify our final *S. aureus* GatB:tRNAGln docking model, we recorded small-angle X-ray diffusion spectra on the tRNA-free and -bound *S. aureus* GatCAB purified by a size exclusion chromatography (Supplementary Figure S5). The predicted curve calculated from the crystal structure of the full-length *S. aureus* GatCAB with the program CRYSOL (32) fits closely to the experimental curve of tRNA-free *S. aureus* GatCAB, as characterized by a discrepancy value $\chi$ of 0.050 (Figure 4A). Furthermore, the crystal structure fits well to the *ab initio* envelope calculated with the program DAMMIN (35), as reflected by a normalized spatial discrepancy (NSD) value of 1.21 (4A). The predicted scattering curve of the tRNA$^{Gln}$ docking model fits well the experimental curve of tRNA-bound *S. aureus* GatCAB ($\chi$ of 0.098) and the *ab initio* envelope (NSD of 0.97) (Figure 4A). However, the tRNA$^{Gln}$ docking model before fitting the tail domain shows a significantly higher discrepancy value with the experimental curve of tRNA-bound *S. aureus* GatCAB and the *ab initio* envelope ($\chi$ of 0.186 and NSD of 1.08, respectively) than using the tail domain fitting model. These results suggest that our docking model with the tail domain interacting with the tRNA is consistent with the solution structure of the *S. aureus* GatCAB:tRNA$^{Gln}$ complex.

In our docking model, the hydrophobic pocket of the GatB tail domain nicely accommodates the curve of the tRNA$^{Gln}$ D-loop (Figure 3C). The *S. aureus* enzyme distinguishes tRNA$^{Gln}$ from tRNA$^{Glu}$ based on the presence of an extra base (U20) in the D-loop of tRNA$^{Glu}$ (11). To gain a better understanding of how this is accomplished we superposed the structures of *E. coli* tRNAGln (PDB ID: 1QTQ) and *Thermus thermophilus* tRNAGlu (PDB ID: 1G59) in a similar manner as described above. The D-loop of *T. thermophilus* tRNAGlu like that of *S. aureus* tRNAGln possesses a U20 supernumerary base (Figure 2A). That extra base flips out from the tRNA D-loop and TWC-loop associating region (Figure 2B). In the docking model of the tail domain with tRNA, the flipped out U20 in the D-loop of tRNA$^{Glu}$ could not be accommodated; the extra base sterically clashes with the surface of the protein in the model (Figure 3C), suggesting that is the mechanism by which *S. aureus* GatCAB rejects tRNA$^{Glu}$.

In addition to recognition of the D-loop, the GatB tail domain may also bind to the variable loop of the tRNA. In our docking model, the GatB-specific Lys rich motif of...
the tail domain is situated in proximity to the variable loop of tRNA\textsuperscript{Gln}, in particular U46 which is pushed out from the tertiary core of the tRNA (Figures 2B and 3C). Consistent with this prediction, replacement of this Lys rich motif with the GXXAXGX motif from GatE results in a mutant \textit{S. aureus} GatCAB enzyme with reduced affinity for tRNA (Supplementary Figure S6).

The \textit{GatB} cradle domain recognizes the first base pair of the acceptor stem

Bacterial GatCAB enzymes recognize the U1–A72 base pair of the acceptor stem of Glu-tRNA\textsuperscript{Gln} and Asp-tRNA\textsuperscript{Asn} to discriminate them from Glu-tRNA\textsuperscript{Glu} and Asp-tRNA\textsuperscript{Asp} (11,13). In contrast, archaeal GatCAB does not use the first base pair of the tRNA acceptor stem to distinguish Asp-tRNA\textsuperscript{Asn} from Asp-tRNA\textsuperscript{Asp}, recognizing aa-tRNA species with either a U1–A72 or a G1–C72 base pair (10,13,14). However, the process of distinguishing a U1–A72 base pair from a G1–C72 base pair is not known.

In our GatCAB:tRNA\textsuperscript{Gln} docking model the tRNA\textsuperscript{Gln} acceptor stem U1–A72 base pair is located in the center of the cradle domain in a space constructed by a 3, two internal loops between β12–α3 and β13–β14, and a 3\textsubscript{10} turn between α2–β11. The 3\textsubscript{10} turn is adjacent to the U1–A72 base pair (Figure 4B). This 3\textsubscript{10} turn is constructed by a hydrogen bond between the main-chain carbonyl oxygen of K183 and amide of E186. Furthermore, M184 in the 3\textsubscript{10} turn forms a hydrophobic core with L167, L189, C191 and F217, suggesting the 3\textsubscript{10} turn is fixed with a 3 and β11 by a hydrophobic interaction. In the GatDE enzyme, GatE has a short loop in place of a 3\textsubscript{10} turn at the corresponding region (Figure 4B and C).

In an alignment of GatB and GatE sequences (Figure 4C, the region including the 3\textsubscript{10} turn (183–186 in \textit{S. aureus} GatB) is conserved in bacterial GatB sequences, and the corresponding region in archaeal GatB is 1–3 residues shorter, suggesting archaeal GatB has a loop in place of the 3\textsubscript{10} turn like GatE. The loop region of GatE is well conserved and is rich in positive residues, however the corresponding region in archaeal GatB is not well conserved. The docking model and alignment suggest these structural differences between bacterial GatB, archaeal GatB and GatE may enable the differences in first base pair recognition between the enzymes. In particular, they imply that the 3\textsubscript{10} turn enables bacterial GatB to distinguish a U1–A72 from a G1–C72 base pair.

Figure 4. (A) Experimental and computed SAXS scattering data. (Upper) the tRNA-bound GatCAB, (lower) tRNA-free GatCAB. The logarithm of the scattering intensity is plotted against the momentum transfer \( s = 4\pi\sin(\theta)/\lambda \), where \( 2\theta \) is the scattering angle and \( \lambda = 1.0 \text{Å} \) is the X-ray wavelength. The plots are displaced along the ordinate for better visualization. Red, cyan and pink curves indicated the computed scattering from tRNA-bound GatCAB model after the tail domain fitting, before the tail domain fitting, and tRNA-free GatCAB, respectively. (B) Comparison of the model of the \textit{S. aureus} GatB:tRNA\textsuperscript{Gln} complex (left) and the crystal structure of the \textit{M. thermautotrophicus} GatE:tRNA\textsuperscript{Gln} complex (right). \textit{S. aureus} GatB, \textit{M. thermautotrophicus} GatE and \textit{M. thermautotrophicus} tRNA\textsuperscript{Gln} are shown as green, magenta and grey ribbon diagram, respectively. The A1–U72 base pair of \textit{M. thermautotrophicus} tRNA\textsuperscript{Gln} is shown as a stick model. The 3\textsubscript{10} turn and the hydrophobic core of \textit{S. aureus} GatB and the short loop of \textit{M. thermautotrophicus} GatE are also shown. (C) Structural based sequence alignments of the cradle domain of bacterial GatB, archaeal GatB and GatE. The number at top is corresponding to \textit{S. aureus} GatB sequence. Secondary structure of \textit{S. aureus} GatB is indicated at top, and that of \textit{M. thermautotrophicus} GatE is indicated at bottom. The 3\textsubscript{10} turn of bacterial GatB and the corresponding region of archaeal GatB and GatE are shown as cyan box. The species aligned are as follows: CHLTR, \textit{Chlamydia trachomatis}; THETH, \textit{Thermus thermophilus HB8}; METJA, \textit{Methanococcus jannaschii}; AERPE, \textit{Aeropyrum pernix}; SULSO, \textit{Sulfolobus solfataricus}; ARCFU, \textit{Archaeoglobus fulgidus}.
The 3\textsubscript{10} turn of bacterial GatB is crucial for tRNA\textsubscript{Gln} acceptor helix U1–A72 base pair recognition

To evaluate the relation between the presence of the 3\textsubscript{10} turn in GatB and the specificity of first base pair recognition, we constructed a mutant \textit{S. aureus} GatCAB enzyme in which the 3\textsubscript{10} turn (K183, M184 and E185) is replaced by the short putative loop from \textit{M. thermautotrophicus} GatB. Based on the multiple sequence alignment (Figure 4C) we also constructed a mutant \textit{M. thermautotrophicus} GatCAB where the 3\textsubscript{10} turn (residues Lys, Met and Glu) is inserted into GatB between residues Gly169 and Glu170. We then tested the transamidase activities (Tables 2 and 3) by the [\textsuperscript{32}P]tRNA/nuclease P1 assay (38) of these mutant GatCAB enzymes with a variety of tRNA substrates.

As expected (10,13,14), wild-type \textit{S. aureus} GatCAB prefers tRNAs with a U1–A72 base pair over those with a G1–C72 base pair, while wild-type \textit{M. thermautotrophicus} GatCAB has no strong preference (Table 2). \textit{S. aureus} Glu-tRNA\textsubscript{Gln} was a poor substrate for wild-type \textit{M. thermautotrophicus} GatCAB (Tables 2 and 3), like \textit{B. subtilis} Glu-tRNA\textsubscript{Gln} for unknown reasons (10). However, the transamidase activity of wild-type \textit{M. thermautotrophicus} GatCAB with wild-type Glu-tRNA\textsubscript{Gln} containing a U1–A72 base pair was approximately the same as with the G1–C72 mutant Glu-tRNA\textsubscript{Gln} (0.0035 and 0.0014 s\textsuperscript{-1}, respectively; Table 2); this is in line with previous results archael GatCAB (10,13,14).

Deletion of the 3\textsubscript{10} turn from \textit{S. aureus} GatCAB resulted in a mutant enzyme that was 17-fold less efficient than wild-type mostly due to a difference in \(k_{\text{cat}}\) (Table 3), possibly due to the enzyme no longer recognizing the first base pair of the substrate aa-tRNA. However, consistent with our model, the removal of the 3\textsubscript{10} turn from the \textit{S. aureus} GatB resulted in a mutant \textit{S. aureus} GatCAB that no longer strongly preferred tRNA substrates with a U1–A72 base pair to those with a G1–C72 base pair (Tables 2 and 3). For example the mutant \textit{S. aureus} GatCAB could use the mutant \textit{S. aureus} Glu-tRNA\textsubscript{Gln} as a substrate about as well as wild-type \textit{S. aureus} Glu-tRNA\textsubscript{Gln} (0.03 and 0.05 s\textsuperscript{-1}, respectively, Table 2). In addition, the mutant \textit{S. aureus} GatCAB could use \textit{M. thermautotrophicus} Asp-tRNA\textsubscript{Asn} as a substrate about as well as wild-type \textit{S. aureus} Glu-tRNA\textsubscript{Gln} (0.03 and 0.05 s\textsuperscript{-1}, respectively, Table 2). In addition, the mutant \textit{S. aureus} GatCAB could use \textit{M. thermautotrophicus} Asp-tRNA\textsubscript{Asn} as a substrate about as well as \textit{M. thermautotrophicus} Asp-tRNA\textsubscript{Asn} (Table 3).

Consistent with our model, the insertion of the 3\textsubscript{10} turn into \textit{M. thermautotrophicus} GatCAB gave a mutant enzyme with preference for tRNA substrates with a U1–A72 base pair (Tables 2 and 3). The mutant \textit{M. thermautotrophicus} GatCAB was about as efficient as wild-type \textit{M. thermautotrophicus} GatCAB using \textit{S. aureus}

**Table 2.** Transamidase activity of the \textit{S. aureus} (\textit{Sa}) GatCAB and \textit{M. thermautotrophicus} (\textit{Mt}) GatCAB with different mischarged tRNA substrates

<table>
<thead>
<tr>
<th>aa-tRNA (^{a})</th>
<th>\textit{Sa} GatCAB(^{b})</th>
<th>\textit{Mt} GatCAB(^{b})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(k_{\text{cat}}) (s\textsuperscript{-1})</td>
<td>(k_{\text{cat}})/K(M) (s\textsuperscript{-1}/mM)</td>
</tr>
<tr>
<td>\textit{Sa} Glu-tRNA\textsubscript{Gln}</td>
<td>0.71 ± 0.03</td>
<td>1.7 ± 0.1</td>
</tr>
<tr>
<td>\textit{Sa} Glu-tRNA\textsubscript{Gln}</td>
<td>0.05 ± 0.01</td>
<td>0.03 ± 0.01</td>
</tr>
<tr>
<td>\textit{Mt} Asp-tRNA\textsubscript{Asn}</td>
<td>0.05 ± 0.03</td>
<td>0.04 ± 0.01</td>
</tr>
<tr>
<td>\textit{Ct} Asp-tRNA\textsubscript{Asn}</td>
<td>0.24 ± 0.01</td>
<td>0.02 ± 0.01</td>
</tr>
</tbody>
</table>

Measurements were from three separate experiments. Standard deviations are reported. Reactions were carried out at 37°C in the presence of ATP (4 mM), amide donor (4 mM) and aa-tRNA (1.25 μM) indicated.

\(^{a}\)The aa-tRNA substrates tested were the \textit{S. aureus} (\textit{Sa}) wild-type Glu-tRNA\textsubscript{Gln} and mutant Glu-tRNA\textsubscript{Gln}, the wild type of the \textit{M. thermautotrophicus} (\textit{Mt}) Asp-tRNA\textsubscript{Asn} and the \textit{C. trachomatis} (\textit{Ct}) Asp-tRNA\textsubscript{Asn}.

\(^{b}\)In the reactions, concentrations of the GatCAB indicated [\textit{Sa} wild-type (wt) GatCAB, \textit{Sa} 3\textsubscript{10} turn deleted mutant GatCAB (3\textsubscript{10}Δ), \textit{Mt} wt GatCAB, or \textit{Mt} 3\textsubscript{10} turn insertion mutant GatCAB (3\textsubscript{10} ins)] ranged from 20 nM to 1 μM.

**Table 3.** Kinetic data for the transamidase activity of the \textit{S. aureus} and \textit{M. thermautotrophicus} GatCAB mutants with different mischarged tRNA substrates

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>\textit{Sa} Glu-tRNA\textsubscript{Gln}</th>
<th>\textit{Mt} Asp-tRNA\textsubscript{Asn}</th>
<th>(L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(K_M) (mM)</td>
<td>(k_{\text{cat}}) (s\textsuperscript{-1})</td>
<td>(k_{\text{cat}}/K_M) (s\textsuperscript{-1}/mM)</td>
</tr>
<tr>
<td>\textit{Sa} wt</td>
<td>1.80 ± 0.22</td>
<td>1.7 ± 0.1</td>
<td>930 ± 120</td>
</tr>
<tr>
<td>\textit{Sa} 3\textsubscript{10}Δ</td>
<td>1.85 ± 0.59</td>
<td>0.1 ± 0.1</td>
<td>54 ± 18</td>
</tr>
<tr>
<td>\textit{Mt} wt</td>
<td>0.70 ± 0.26</td>
<td>0.006 ± 0.001</td>
<td>8 ± 3</td>
</tr>
<tr>
<td>\textit{Mt} 3\textsubscript{10} ins</td>
<td>0.97 ± 0.22</td>
<td>0.009 ± 0.001</td>
<td>9 ± 2</td>
</tr>
</tbody>
</table>

Measurements were from three separate experiments. Standard deviations are reported. Reactions were carried out at 37°C in the presence of excess ATP (4 mM) and Gln (4 mM). \(L\) is the relative catalytic efficiency \(\left[\frac{\text{\(k_{\text{cat}}/K_M\) of homologous substrate}}{\text{\(k_{\text{cat}}/K_M\) of non-homologous substrate}}\right]\). GatCAB enzymes labeled like in Table 2.
Glu-tRNA\textsubscript{Gln\textsubscript{UA}} as a substrate (Table 3). However, compared to wild-type \textit{M. thermautotrophicus} GatCAB, the mutant enzyme was 10-fold less efficient with the \textit{M. thermautotrophicus} Asp-tRNA\textsubscript{Gln\textsubscript{GC}} substrate (Table 3). In addition, the mutant \textit{M. thermautotrophicus} GatCAB preferred by 10-fold a substrate wild-type \textit{S. aureus} Glu-tRNA\textsubscript{Gln\textsubscript{UA}} substrate over the G1–C72 mutant Glu-tRNA\textsubscript{Gln\textsubscript{Gin}} (Table 2). Also, while wild-type \textit{M. thermautotrophicus} GatCAB was nearly as active with \textit{C. trachomatis} Asp-tRNA\textsubscript{Asn\textsubscript{UA}} as with \textit{M. thermautotrophicus} Asp-tRNA\textsubscript{Asn\textsubscript{GC}}, the insertion GatCAB mutant preferred \textit{C. trachomatis} Asp-tRNA\textsubscript{Asn\textsubscript{An}}, which has a U1–A72 base pair (Table 2). These results indicate that insertion of the 3\textsubscript{10} turn into GatB enables \textit{M. thermautotrophicus} GatCAB to distinguish aa-tRNA with a U1–A72 base pair from ones with a G1–C72 base pair. Taken with the mutant \textit{S. aureus} GatCAB data, these results strongly suggest that the 3\textsubscript{10} turn in \textit{S. aureus} GatCAB is crucial for distinguishing aa-tRNA with a U1–A72 from those with a G1–C72 base pair.

**DISCUSSION**

The bacterial GatCAB enzyme uses the first base pair of the acceptor stem and the tRNA D-loop for precise tRNA discrimination (11,13). Our results suggest that the U1–A72 base pair of tRNA\textsubscript{Asn} and tRNA\textsubscript{Gln} is recognized by a 3\textsubscript{10} turn in the cradle domain of GatB, while the tail domain of GatB binds the D-loop.

The transamidosome is a ternary complex of \textit{T. thermophilus} GatCAB, ND-AspRS and tRNA\textsubscript{Asn} stable during the overall catalytic process (42). Such a structure would enable Asn-tRNA\textsubscript{Asn} formation without the risk of free mischarged Asp-tRNA\textsubscript{Asn} being used in protein synthesis. The complex would also protect the Asn-tRNA\textsubscript{Asn} product from deacylation until it will be bound by EF-Tu and transported to the ribosome (42,43). Complexes between ND-GluRS, tRNA\textsubscript{Gln}, and either Glu-AdT (GatCAB or GatDE) are also predicted to exist (12,42). Formation of these complexes between ND-aaRS, AdT and tRNA may explain why the AdTs recognize the specific identity elements in their tRNA substrates.

The structural models of the transamidosomes predict the ND-aaRS enzyme binds to the acceptor stem and anticodon loop of the tRNA (tRNA\textsubscript{Gln}\textsubscript{Gin} or tRNA\textsubscript{Asn}\textsubscript{An}) (12,42) (Supplementary Figure S7) in the same fashion as they do in the absence of AdT (44,45). This places the tRNA 3'-end into the synthetase active site to be aminoacylated; initially the first base pair of the tRNA’s acceptor helix is not accessible to the 3\textsubscript{10} turn of bacterial GatB in its role to discriminate tRNA isoacceptors. However, the tRNA’s tertiary core (including the D-loop) is recognized (Supplementary Figure S7). Recognition of the D-loop by the AdT tail domain (11,12) may permit the amidotransferases to distinguish the ND-aaRS complexed with their tRNA transamidation substrates (tRNA\textsubscript{Gln}\textsubscript{Gin} or tRNA\textsubscript{Asn}\textsubscript{An}) from other tRNA isoacceptors (tRNA\textsubscript{Glu} or tRNA\textsubscript{Asp}) (12–14). For example, GatCAB recognition of the D-loop would enable the AdT to discriminate ND-AspRS bound to tRNA\textsubscript{Asn} from the aaRS bound to tRNA\textsubscript{Asp} despite the U1–A72 base pair being initially inaccessible to GatCAB. The variable loop of the tRNA in the transamidosome models is also accessible to the AdT, in particular the Lys rich motif in the tail domain of GatB, which may explain why archaeal GatCAB uses this element to discriminate tRNA\textsubscript{Asn} from tRNA\textsubscript{Asp} (13,14).

The transamidosome models (12,42) predict that after aminoacylation the tRNA’s 3’ CCA terminus flips from active site in the ND-aaRS to the kinase active site of the AdT (42), similar to the tRNA movements seen in certain aaRSs with editing domains (46,47). In the case of the transamidosome, this movement enables the aminoacyl-moiety of the mischarged tRNA to be amidated by the AdT in the complex. Once the 3’ end of the acceptor stem flips from the aaRS active site into the transamidase active site of the AdT, the U1–A72 base pair may become accessible to the 3\textsubscript{10} turn in bacterial GatB.

Recognition of the first base pair by bacterial GatCAB and the archaeal GatDE may be a proofreading step to ensure amidation of the mischarged tRNA substrate (Glu-tRNA\textsubscript{Gln} and/or Asp-tRNA\textsubscript{Asn}) and not the properly aminoacylated product of the ND-aaRS (Glu-tRNA\textsubscript{Gln} or Asp-tRNA\textsubscript{Asn}). Why this proofreading step is not required by the archaeal GatCAB is unclear, but presumably recognition of the tRNA tertiary core may be enough to ensure that GatCAB amidates Asp-tRNA\textsubscript{Asn} and not Asp-tRNA\textsubscript{Asp} (13,14).

It was speculated that the ancestor of GatB and GatE recognized the first base pair of its tRNA substrates (16). However, as the 3\textsubscript{10} turn in the cradle domain of the \textit{S. aureus} GatB for recognition of the U1–A72 base pair in tRNA\textsubscript{Gln} is conserved only in bacterial GatB and not in archaeal GatB and GatE, this may not be the case. Instead, the common ancestor of GatB and GatE may not have recognized the first base pair of its tRNA substrate, with bacterial GatCAB and GatDE independently evolving to recognize the first base pair of tRNA. Thus, bacterial GatCAB and GatDE both recognizing the first base pair may have been a case of convergent and not divergent evolution.

**ACCESSION NUMBERS**

Coordinate and structure factor have been deposited in PDB with accession number 3IP4.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

The authors are grateful to S. Chimnaronk, I. Heinemann, J. Ling and Y. Tanaka for many stimulating discussions, and to T. Tamura (National Institute of Advanced Industrial Science and Technology) for his gift of the \textit{R. erythropolis} expression system. They thank the staff of SPring-8 BL41XU, 40B2 (2009A1119) and 45XU
Conflict of interest statement. None declared.

REFERENCES


