Virus-Encoded Aminoacyl-tRNA Synthetases: Structural and Functional Characterization of Mimivirus TyrRS and MetRS

Chantal Abergel,1* Joëlle Rudinger-Thirion,2 Richard Giegé,2 and Jean-Michel Claverie1

Structural and Genomic Information Laboratory, CNRS-UPR2589, Institut de Biologie Structurale & Microbiologie IFR888, Université de la Méditerranée, 163 avenue de Luminy, Case 934, 13288 Marseille, Cedex 9, France,1 and Architecture et Réactivité de l’ARN, CNRS-UPR 9002, IBMC, and Université Louis Pasteur, 15 rue René Descartes, 67084 Strasbourg, France2

Received 22 May 2007/Accepted 4 September 2007

Aminoacyl-tRNA synthetases are pivotal in determining how the genetic code is translated in amino acids and in providing the substrate for protein synthesis. As such, they fulfill a key role in a process universally conserved in all cellular organisms from their most complex to their most reduced parasitic forms. In contrast, even complex viruses were not found to encode much translation machinery, with the exception of isolated components such as tRNAs. In this context, the discovery of four aminoacyl-tRNA synthetases encoded in the genome of mimivirus together with a full set of translation initiation, elongation, and termination factors appeared to blur what was once a clear frontier between the cellular and viral world. Functional studies of two mimivirus tRNA synthetases confirmed the MetRS specificity for methionine and the TyrRS specificity for tyrosine and conformity with the identity rules for tRNA^tyr for archea/eukarya. The atomic structure of the mimivirus tyrosyl-tRNA synthetase in complex with tyrosinol exhibits the typical fold and active-site organization of archaeal-type TyrRS. However, the viral enzyme presents a unique dimeric conformation and significant differences in its anticodon binding site. The present work suggests that mimivirus aminoacyl-tRNA synthetases function as regular translation enzymes in infected amoebas. Their phylogenetic classification does not suggest that they have been acquired recently by horizontal gene transfer from a cellular host but rather militates in favor of an intricate evolutionary relationship between large DNA viruses and ancestral eukaryotes.

Acanthamoeba polyphaga mimivirus is the largest known DNA virus. Its particle size (750 nm), genome length (1.2 million bp), and large gene repertoire (>910 protein-coding genes) blur the established boundaries between viruses and parasitic cellular organisms (49). On the one hand, mimivirus exhibits the standard features of nucleocytoplasmic large DNA viruses (capsid structure, life cycle, and core gene set), within which it now constitutes the prototype of the Mimiviridae family besides the previously defined Poxviridae, Asfarviridae, Iridoviridae, and Phycodnaviridae families (49). On the other hand, the mimivirus genome exhibits numerous genes never encountered before in any other virus. Among the most intriguing are genes corresponding to central components of the protein translation machinery, a biochemical process widely thought to be an exclusive signature of cellular organisms. For instance, the mimivirus genome encodes four aminoacyl-tRNA synthetases (aaRS): ArgRS, CysRS, MetRS, and TyrRS. These key enzymes link the genetic code with the proper 20 amino acids and provide the basic substrates for the translation process. In cellular organisms, they catalyze the esterification of a given amino acid to the 3’ ends of their cognate tRNAs in a two-step reaction comprising the activation of the amino acid as an aminoacyl-adenylate followed by its transfer onto the 3’-terminal ribose of the cognate tRNA (13). The correct interpretation of the genetic code thus requires a perfect specificity at both the amino acid activation and the tRNA-charging step. The specificity of the tRNA recognition is ensured by stringent constraints referred to as “identity rules” (30). aaRS are structurally diverse enzymes with modular architectures, traditionally partitioned into two classes based on signature sequences and common features of their catalytic sites (17, 23). Class I aaRS share the sequence motifs HIGH and KMSKS and have active sites based on a Rossmann-fold domain. Class II aaRS share three other signature motifs, and their active sites are built on an antiparallel β-sheet surrounded by α-helices. There are also functional differences between the two classes: class I aaRS attach amino acids to the 2’-hydroxyl end of the terminal adenosine of the tRNA, whereas charging occurs on the 3’-hydroxyl end for class II aaRS (5). Tyrosyl-tRNA synthetases belong to class Ic of aaRS, together with TrpRS. Both are homodimers (18), a feature otherwise shared mostly by class II synthetases.

Two main opposite hypotheses can be proposed to account for the presence of the largely incomplete translation machinery exhibited by mimivirus. Either it is the remnant of a complete translation machinery, following many gene losses through a process of reductive evolution (such as observed for intracellular parasitic bacteria), or it is the result of multiple horizontal gene acquisitions conferring some selective advantages to the virus. The latter is most often accomplished by diverting the acquired genes from their original functions. To address the above-mentioned dilemma, we initiated a comprehensive study of the structures and activities of the four mimivirus class I aaRS. We complemented these experimental stud-
ies by searching for evidence of horizontal gene transfers through a phylogenetic analysis of the four mimivirus aaRSs.

This article reports a detailed functional study of two of the four viral class I aaRS, *A. polyphaga* mimivirus TyrRS (TyrRS<sub>apm</sub>) and *A. polyphaga* mimivirus MetRS (MetRS<sub>apm</sub>), and the structural study of TyrRS<sub>apm</sub> in complex with tyrosinol. These enzymes were found to exhibit the specificity and function predicted from their sequences, but the three-dimensional structure of TyrRS<sub>apm</sub> exhibited significant differences from its cellular counterparts.

Phylogenetic analysis did not provide evidence for a recent acquisition of these genes from a cellular host close to modern *Acanthamoeba* but connected mimivirus aaRS to a variety of early diverging protozoan supergroups, such as Amoebozoa and Excavata (3).

**MATERIALS AND METHODS**

**Preparation of TyrRS<sub>apm</sub>, MetRS<sub>apm</sub> and tRNAs.** The *A. polyphaga* mimivirus TyrRS-encoding gene was PCR amplified from genomic DNA and cloned into a Gateway system (Invitrogen) (33) as described earlier (1, 2).

The oligomeric state of recombinant TyrRS was measured by gel filtration on an analytical Superose 6 column. Purified TyrRS was loaded on the column from a 10-μM concentration in 10 mM Tris buffer (pH 7.4), 100 mM KCl. Five molecular weight standards were initially run onto the column, using the same buffer to calibrate the column.

Native tRNA<sup>Tyr</sup> and initiator tRNA<sup>Met</sup><sup>Met</sup> from *Escherichia coli* were purchased from Subrden and Sigma, respectively. Their equivalents in *Saccharomyces cerevisiae* were purified by homogenization by countercurrent distribution (21) followed by appropriate column chromatographies. Transcripts of yeast tRNA<sup>Tyr</sup> and initiator tRNA<sup>Met</sup><sup>Met</sup> as well as their variants were obtained as described elsewhere (19, 26). Finally, transcripts of *Saccharomyces cerevisiae* tRNA<sup>Tyr</sup> were obtained by in vitro transcription of synthetic genes using the *transzyme* method (25).

Cryocrystals of a 20-residue connective polypeptide 1 (CP1) region next to the α<sub>ε</sub>-turn-α<sub>ε</sub> motif, structurally conserved in all TyrRSs but disordered in the TyrRS<sub>apm</sub> structure (see Fig. 3). This region was not found to contribute significantly to the difference of buried surface areas between the two TyrRS<sub>apm</sub> dimer conformations. Structure graphical representations (see Fig. 2 and 4 to 6) were produced using VMD (35). Activity assays. (i) ATP/PP<sub>i</sub> exchange reactions. Reaction media (200 μl) contained 100 mM Na-HEPES (pH 7.2), 10 mM MgCl<sub>2</sub>, 2 mM KF, 2 mM ATP, 2 mM [γ-<sup>35</sup>S]PP<sub>i</sub> (1 to 0.2 cpm/pmol), and 1 mM of a mixture of all amino acids, with or without tyrosine or methionine. Reactions were initiated by addition of 1.5 μg of TyrRS<sub>apm</sub> or MetRS<sub>apm</sub>. The levels of [γ-<sup>35</sup>S]ATP formed after 5, 10, 15, and 20 min of incubation at 37°C were determined as described elsewhere (12). Control experiments with either no aaRS or no amino acid were conducted in parallel. (ii) tRNA<sup>Tyr</sup> aminoacylation reactions. Before aminoacylation, the eluted transcripts were heated at 65°C for 2 min and cooled for 10 min to allow native conformation. Tyrosylation of native tRNA<sup>Tyr</sup> or initiator tRNA<sup>Met</sup><sup>Met</sup> transcripts (wild type or mutated) was performed (50 μl) at 30°C in 50 mM Na-HEPES (pH 7.5), 2 mM MgCl<sub>2</sub>, 2.5 mM ATP, 0.2 mg/ml bovine serum albumin, 1 μM spermine (54), 10 mM L-[<sup>35</sup>S]tyrosine (adjusted to 750 cpm/pmol), and the required concentration of tRNA molecules and MetRS<sub>apm</sub>. Methionylation of initiator tRNA<sup>Met</sup> or tRNA<sup>Met</sup><sup>Met</sup> transcripts (wild type or mutated) was performed (50 μl) at 30°C in 20 mM Na-HEPES (pH 7.5), 10 mM MgCl<sub>2</sub>, 1 mM dithioerythritol, 2 mM ATP (4), 10 μM L-[<sup>35</sup>S]methionine (adjusted to 400 cpm/pmol), and the required concentration of tRNA molecules and MetRS<sub>apm</sub>. At different incubation times, aliquots were spotted on 35% Whatman paper and 5% trichloroacetic acid precipitated. Incorporation of radioactive amino acid was measured by liquid scintillation spectroscopy. Kinetic parameters (K<sub>m</sub> and V<sub>max</sub>) were determined from Lineweaver and Burk plots.

Phylogenetic position of mimivirus aaRSs. Homologous TyrRS protein sequences from all major phyla were aligned using MUSCLE software (22), and a maximum-likelihood tree was computed with PhyML (32) using the default

---

**TABLE 1. X-ray data collection and refinement statistics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value or type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td></td>
</tr>
<tr>
<td>Beam line</td>
<td>ESRF/ID29</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97925</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>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td>a = 63.5, b = 107.3, c = 148.9</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>8.77 to 2.2 (2.28 to 2.2)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>238,357</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>52,537</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.5 (4.6)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>21.6</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>24.8</td>
</tr>
<tr>
<td>Δρ&lt;sub&gt;sig&lt;/sub&gt; (%)</td>
<td>0.006</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>5,169</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>330</td>
</tr>
<tr>
<td>Ligand tyrosinol (Å)</td>
<td>2</td>
</tr>
<tr>
<td>Average B factor (Å)</td>
<td>39</td>
</tr>
<tr>
<td>Protein main chain</td>
<td>37.1</td>
</tr>
<tr>
<td>Water</td>
<td>41.85</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td>Most favored regions... 539</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>40</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>5</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0</td>
</tr>
</tbody>
</table>

---

<sup>a</sup> Values in parentheses are for the highest-resolution shell.

<sup>b</sup> <i>]<i>factor</i></i> is the mean signal-to-noise ratio, where <i>l</i> is the integrated intensity of a measured reflection and σ<i>l</i> is the estimated error in the measurement.

<sup>c</sup> R<sub>free</sub> = <i>S</i>(<i>l</i>/<i>S</i><i>c</i>)<sub>obs</sub> - <i>S</i><i>c</i><sub><i>free</i></sub>, where <i>F</i><i>c</i> and <i>F</i><i><i><i>obs</i></i></i> are observed and calculated structure factor amplitudes, respectively.

<sup>d</sup> <i>R</i><sub>cryst</sub> was calculated from a randomly chosen 5% of reflections.
option on our phylogeny server at the URL http://www.phylogeny.fr/. Bootstrap values are indicated along the branches. Branches have been collapsed for bootstrap values smaller than 50%.

Additional phylogenetic analyses of all four mimivirus aaRS were performed using archaean and eukaryotic aaRS, including mitochondrial sequences (bacterial-type aaRS). For each aaRS, protein sequences were aligned using T-Coffee software (45). Maximum-likelihood trees and neighbor-joining trees (Bootstrap 500) were computed as described above. The two methods produced similar trees. Only maximum-likelihood trees are presented in Fig. 7.

**Protein structure accession number.** The TyrRS apm structure coordinates have been deposited in the Protein Data Bank under accession code 2J5B.

**RESULTS AND DISCUSSION**

TyrRSapm and MetRSapm possess the predicted enzymatic activities. The presence of aaRS genes in a viral genome was unexpected and immediately raised the question of the activity of the corresponding protein. We expressed and purified two of the four mimivirus aaRS (MetRS, TyrRS, ArgRS, and CysRS), namely, MetRSapm and TyrRSapm. They were both tested for amino acid activation activity in the absence of tRNA (2, 49). Exchange reactions were assayed in the presence of TyrRS and all amino acids with or without tyrosine (Fig. 1A) and in the presence of MetRS and all amino acids with or without methionine (Fig. 1B). These experiments clearly demonstrate that TyrRSapm and MetRSapm indeed activate tyrosine and methionine, respectively, to the exclusion of any other amino acid. Furthermore, TyrRSapm specifically tyrosylates tRNA^Tyr^, while MetRSapm, as other methionyl-tRNA synthetases, aminocylates both eukaryotic and bacterial tRNA^Met^ (Fig. 1C; Table 2).

**Structure of TyrRSapm.** (i) The overall structure of TyrRSapm resembles that of other TyrRSs. TyrRSs are all organized similarly, with an N-terminal catalytic domain, including the CP1 region responsible for dimerization, followed by a C-terminal domain (8). However, TyrRSs do exhibit important differences in sequence and architecture of their C-terminal domains. Archaean TyrRSs are encoded by the shortest polypeptides, whereas vertebrate TyrRSs exhibit a large extra C-terminal domain. For the human cytoplasmic TyrRS and *Neurospora crassa* mitochondrial TyrRS, these differences have been associated with functions unrelated to the translation process (14, 59).

As expected from its sequence similarity to other TyrRSs, TyrRSapm exhibits the typical fold of the TyrRS core domain and is more similar to the archaean type, with an N-terminal Rossmann-fold catalytic domain, an anticodon binding domain, and no extra C-terminal domain. TyrRSapm shares 30% identity over 340 residues with the TyrRS of the hyperthermophilic Euryarchaeota *Pyrococcus horikoshii*, its closest known structural homologue (2CYC [38]).

The protein was co-crystallized with ATP and the tyrosine analogue tyrosinol, and its structure was solved at 2.2-Å resolution. While tyrosinol was clearly located in the electronic density map, only residual density appears at the binding site of ATP, probably due to its instability at the crystallization pH (5.5). We compared the TyrRSapm active site with the *Methanococcus jannaschii* (1J1U [37]) and *Thermus thermophilus* (1H3E [62]) structures in complex with tRNA^Tyr^ and L-tyrosine or tyrosinol and the human TyrRS catalytic core struc-
ture (1Q11 [61]) in complex with tyrosinol. The positions of the residues in contact with tyrosinol in the TyrRS<sub>apm</sub> structure were found to be superimposable to their counterparts in the archaean structures (Fig. 2) and to correspond to conserved amino acids, except for two histidines replaced by F81 and N198, as in plant and protozoan sequences (Fig. 3).

The ATP binding site of TyrRS<sub>apm</sub> exhibits a modified version (a HIAQ motif, as in protozoa and plants) of the catalytically important HIGH motif found in the archaean and eukaryotic structures 1J1U and 1Q11 (Fig. 3). These motifs are superimposable in the structures. A perfect KMSKS motif is present in the TyrRS<sub>apm</sub> sequence but in an open-state conformation compared to the one observed in 1H3E, the structure of a bacterial TyrRS in complex with tyrosinol and tRNA<sub>Tyr</sub> (62). This loop, often disordered in ATP-free TyrRS structures, is also associated with large B-factor values in both the TyrRS<sub>apm</sub> and 1H3E structures.

(ii) Structural peculiarities of the TyrRS<sub>apm</sub> dimer. In contrast to other tRNA/aaRS systems, where the tRNA binds to a single subunit of the synthetase, the tRNA<sub>Tyr</sub> recognition involves both TyrRS subunits (6, 62). The acceptor arm of tRNA<sub>Tyr</sub> interacts with the catalytic domain of one monomer, whereas the anticodon arm is sandwiched between the α-helical and C-terminal domains of the other monomer. In all TyrRS structures, the monomers are related to each other by a twofold rotational axis and the dimers are superimposable, with very small variations in the orientations of the monomers. All available crystal structures of tRNA/TyrRS complexes are also planar, with a symmetrical conformation of the two monomers in the dimer and with two tRNA molecules simultaneously interacting with one TyrRS dimer (62). However, previous kinetic studies of tyrosine activation and tRNA<sub>Tyr</sub> charging revealed an anticooperative behavior of the TyrRS dimer in solution (27). Other experiments suggested that each TyrRS dimer binds and tyrosylates only one tRNA molecule at a time (7, 20), again putting into question the fact that the symmetrical conformation observed in all crystal structures to date corresponds to the active conformation in solution (62).

Similarly to the other members of the family, TyrRS<sub>apm</sub> is a dimer in solution (data not shown), and it crystallized as a homodimer. The superimposition of the TyrRS<sub>apm</sub> structure with all other available TyrRS dimeric structures nonetheless highlights a major difference. While the first monomer super-

### Table 2. Kinetic Parameters for Tyrosylation of Wild-Type tRNA<sub>Tyr</sub> Molecules or Variants by TyrRS<sub>apm</sub> and Methionylation of Wild-Type tRNA<sub>Met</sub> Molecules or Variants by MetRS<sub>apm</sub>

<table>
<thead>
<tr>
<th>tRNA</th>
<th>Km (µM)</th>
<th>k&lt;sub&gt;cat&lt;/sub&gt; (10&lt;sup&gt;-3&lt;/sup&gt; s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>k&lt;sub&gt;cat&lt;/sub&gt;/Km (n-fold)</th>
<th>L value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyrosine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast (native)</td>
<td>0.5</td>
<td>126</td>
<td>252</td>
<td>0.3</td>
</tr>
<tr>
<td>Yeast (transcript)</td>
<td>1.7</td>
<td>142</td>
<td>83.5</td>
<td>1</td>
</tr>
<tr>
<td>P. falciparum</td>
<td>4</td>
<td>18.3</td>
<td>4.6</td>
<td>18</td>
</tr>
<tr>
<td>Mutated molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast C&lt;sub&gt;T&lt;/sub&gt;G&lt;sub&gt;72&lt;/sub&gt;→A&lt;sub&gt;T&lt;/sub&gt;U&lt;sub&gt;72&lt;/sub&gt;</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
</tr>
<tr>
<td>Yeast C&lt;sub&gt;T&lt;/sub&gt;G&lt;sub&gt;72&lt;/sub&gt;→G&lt;sub&gt;T&lt;/sub&gt;C&lt;sub&gt;72&lt;/sub&gt;</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
</tr>
<tr>
<td>Yeast C&lt;sub&gt;T&lt;/sub&gt;G&lt;sub&gt;72&lt;/sub&gt;→A&lt;sub&gt;T&lt;/sub&gt;U&lt;sub&gt;73&lt;/sub&gt;→G&lt;sub&gt;T&lt;/sub&gt;C&lt;sub&gt;73&lt;/sub&gt;</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
</tr>
<tr>
<td>Yeast G&lt;sub&gt;34&lt;/sub&gt;A&lt;sub&gt;34&lt;/sub&gt;</td>
<td>1.3</td>
<td>7.5</td>
<td>5.8</td>
<td>14.4</td>
</tr>
<tr>
<td>Yeast G&lt;sub&gt;34&lt;/sub&gt;C&lt;sub&gt;34&lt;/sub&gt;</td>
<td>1.4</td>
<td>9.4</td>
<td>6.6</td>
<td>12.6</td>
</tr>
<tr>
<td>Yeast U&lt;sub&gt;35&lt;/sub&gt;C&lt;sub&gt;35&lt;/sub&gt;</td>
<td>9.3</td>
<td>1.1</td>
<td>0.12</td>
<td>696</td>
</tr>
<tr>
<td>Yeast A&lt;sub&gt;56&lt;/sub&gt;G&lt;sub&gt;56&lt;/sub&gt;</td>
<td>20</td>
<td>7.3</td>
<td>0.36</td>
<td>232</td>
</tr>
<tr>
<td>Yeast A&lt;sub&gt;56&lt;/sub&gt;C&lt;sub&gt;56&lt;/sub&gt;</td>
<td>15.4</td>
<td>14.9</td>
<td>0.96</td>
<td>87</td>
</tr>
<tr>
<td>Methionine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast (native)</td>
<td>0.13</td>
<td>6.6</td>
<td>252</td>
<td>0.3</td>
</tr>
<tr>
<td>Yeast (transcript)</td>
<td>0.74</td>
<td>17.4</td>
<td>83.5</td>
<td>1</td>
</tr>
<tr>
<td>E. coli (native)</td>
<td>0.47</td>
<td>47</td>
<td>252</td>
<td>0.3</td>
</tr>
<tr>
<td>Mutated molecules</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast C&lt;sub&gt;34&lt;/sub&gt;G&lt;sub&gt;34&lt;/sub&gt;</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
</tr>
<tr>
<td>Yeast A&lt;sub&gt;35&lt;/sub&gt;C&lt;sub&gt;35&lt;/sub&gt;</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
</tr>
<tr>
<td>Yeast U&lt;sub&gt;36&lt;/sub&gt;C&lt;sub&gt;36&lt;/sub&gt;</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
<td>NM</td>
</tr>
</tbody>
</table>

<sup>a</sup> L values correspond to losses of efficiency relative to yeast tRNA<sub>Tyr</sub> transcript or yeast tRNA<sub>Met</sub> transcript. Values of >1 correspond to gains in efficiency. Experimental errors for k<sub>cat</sub> and Km varied at most by 20%. Results represent averages of at least two independent experiments.

<sup>b</sup> NM, not measurable (loss of 10<sup>3</sup>).
FIG. 4. Cartoon comparison of the dimer interface. The two α7 helices of TyrRS<sub>apm</sub> are in green, and the conserved α-turn-α<sub>z</sub>, structural motif (found in other TyrRS dimers) is colored in silver, transparent for the first monomer and opaque for the second one. This figure illustrates the nearly 90° rotation of the TyrRS<sub>apm</sub> second monomer (colored according to secondary-structure elements: red, α-helices; blue, β-strands; yellow, coils and turns; pink, η helices). N-term, N terminus; C-term, C terminus.

FIG. 3. Structure-based alignment of TyrRSs. TyrRS<sub>apm</sub> (2J5B) was aligned with eukaryal (1Q11, human, core structure) and archaeal (2CYC, P. horikoshii; 2CYA, Aeropyrum pernix; 2CYB, Archaeoglobus fulgidus; 1J1U and 1U7D, M. jannaschii complex and apo form) structures. The closest TyrRS<sub>apm</sub> homologues from protozoa and plants are also included (EHISTO, Entamoeba histolytica; DICDIS, Dicyostelium discoideum; OSATI, Oryza sativa; PYOELL, Plasmodium yoelii; PFALC, Plasmodium falciparum; ATHAL, Arabidopsis thaliana; CHOML, Cryptosporidium hominis; GLAMB, Giardia lamblia). The secondary-structure elements of TyrRS<sub>apm</sub> and M. jannaschii are, respectively, indicated above and below the multiple alignment. The N-terminal, Rossmann fold, CP1, and C-terminal domains are colored in pink, blue, green, and red, respectively. Strictly conserved residues are boxed in red. Residues involved in tyrosine binding (Fig. 2) are highlighted in gray. This alignment was produced with ESPript (31).
tyrosylation was assayed on yeast tRNA^{Tyr} transcripts bearing mutations at the following locations: the C→G72 base pair, the anticodon triplet G34U35A36, and the discriminator base A73 (26). Replacement of the first base pair by a G1-C72 or an A1-U72 pair inactivates tRNA{\text{TYR}}{\text{A}}{\text{PM}} tyrosylation by TyrRS apm. Similarly, no activity could be detected after mutation of A73 into G73 (Table 2).

Four synthetase residues are known to be involved in the acceptor stem recognition in archaea: R132, R174, K175, and M178 (according to 1J1U numbering). The corresponding amino acids are conserved in TyrRS{\text{A}}{\text{PM}} (R149, R195, K196, and M199) (Fig. 3). To analyze their involvement in tRNA binding, we superimposed a TyrRS{\text{A}}{\text{PM}} monomer on the monomer of the M. jannaschii TyrRS/tRNA{\text{TYR}} complex at the acceptor site. Except for residue R149, not visible in the TyrRS{\text{A}}{\text{PM}} structure, the homologous residues are positioned as in the M. jannaschii structure, with residues R195 and M199 contacting the C1 base and residue K196 contacting the A73 discriminator base. In order to make contact with the G72 base (as seen with the homologous R132 residue in the M. jannaschii TyrRS/tRNA{\text{TYR}} complex), residue R149 in helix α7 of the TyrRS{\text{A}}{\text{PM}} structure would require a conformational change, breaking the helix into the canonical α7-turn-α8 motif. This is consistent with the high B-factor values observed at these positions and the results obtained with the C1-G72→G1-C72 variant.

Recognition of the tRNA^{Tyr} anticodon. The other residues involved in tyrosine identity correspond to the anticodon nucleotides G34U35A36, the strength of which depends on the system studied (9). The two available crystal structures of TyrRS/tRNA{\text{TYR}} complexes identified the specific contacts established between these identity elements and the amino acid residues on the synthetases (37, 42, 62). By comparison with all available TyrRS sequences and structures, the TyrRS{\text{A}}{\text{PM}} loop involved in anticodon recognition is quite short (Fig. 3 and 6). To examine the functional consequence of this unique feature, we studied the tyrosylation of variants derived from yeast tRNA^{Tyr} transcript mutated at anticodon sites. In contrast with the yeast TyrRS, only weak effects were observed upon mutation of the G34 anticodon nucleotide (Table 2). The structural comparison of the M. jannaschii complex with the TyrRS{\text{A}}{\text{PM}} structure highlights that, in the archaeal complex, the G34 base is sandwiched between the F261 and H283 rings and is also

![FIG. 5. Cartoon representation of the TyrRS{\text{A}}{\text{PM}} dimer superimposed on the archaeal TyrRS 2CYC from P. horikoshii. The archaeal dimer is transparent and colored in silver, except for the structural motif (α7-turn-α8, corresponding to the α7 helix in TyrRS{\text{A}}{\text{PM}}), colored in opaque cyan. The TyrRS{\text{A}}{\text{PM}} first monomer superimposed on the 1CYC first monomer is colored in yellow. The second monomer secondary-structure elements are colored as outlined in the legend for Fig. 4. The arrow shows the rotation to be applied to the mimivirus second monomer to superimpose its β7-turn-β8 onto the P. horikoshii β7-β8-β9 motif. N-term, N terminus; C-term, C terminus.](http://jvi.asm.org/)

![FIG. 6. Superimposition of the C-terminal anticodon binding domain of TyrRS{\text{A}}{\text{PM}} (solid) on the M. jannaschii TyrRS/tRNA complex (transparent). The anticodon appears solid in the transparent tRNA surface.](http://jvi.asm.org/)
recognized by D286 through hydrogen bonds. This part of the TyrRS\textsubscript{apm} structure exhibits the most significant differences from archaeal/eukaryotic TyrRS. The β-3\textsubscript{ter}-β motif encompassing the F261 position has no counterpart in the mimivirus structure, where it is replaced by a shorter β-turn-β motif, too far away to possibly interact with base G\textsubscript{34} on the other side of the anticodon. The H283 and D286 positions are uniquely replaced in the viral sequence by N303 and E306, respectively. Together with our experimental results (Table 2), this suggests that the G\textsubscript{34} base of the anticodon is not used as a discriminator for tRNA\textsuperscript{Tyr} recognition by TyrRS\textsubscript{apm}.

In contrast, mutations of the two other anticodon nucleotides (nucleotides 35 and 36) induce strong negative effects (from 87- to 696-fold) on the TyrRS\textsubscript{apm} activity on the corresponding tRNA, which is not observed with the M. jannaschii TyrRS, suggesting the viral enzyme specifically recognizes these nucleotides.

**Methionylation properties of MetRS\textsubscript{apm}**. Wild-type native initiator tRNA\textsuperscript{Met}\textsuperscript{E. coli} from E. coli and S. cerevisiae were tested for methionylation by MetRS\textsubscript{apm}. Both substrates were amionocyated with kinetic parameters (K\textsubscript{m} and k\textsubscript{cat}) slightly decreased for the eukaryal tRNA\textsuperscript{Met}, which resulted in a twofold loss in methionylation efficiency compared to that of the E. coli tRNA\textsuperscript{Met} (Table 2). Unmodified yeast tRNA\textsuperscript{Met} transcript behaves globally as its native counterpart, with only a twofold loss in aminocacylation efficiency. Noticeable, however, are the low K\textsubscript{m} and high k\textsubscript{cat} values of yeast tRNA\textsuperscript{Met} transcript towards MetRS\textsubscript{apm} compared to the values determined with the yeast enzyme. Replacing individually the anticodon residues abolishes methionine acceptance (Table 2), as has already been demonstrated for E. coli and yeast MetRS\textsubscript{s} (52, 53). Thus, as for all MetRS\textsubscript{s}, the C\textsubscript{34}U\textsubscript{35}A\textsubscript{36} anticodon residues govern specific recognition of tRNA\textsuperscript{Met} by MetRS\textsubscript{apm}.

**Evolutionary origin of aaRS\textsubscript{apm}**. Previous functional studies of TyrRS\textsubscript{s} revealed a phylogenetic barrier preventing cross-tyrosylation of eu bacterial tRNA\textsuperscript{Tyr} by eukaryal TyrRS and vice versa (36, 58). Our work demonstrates that TyrRS\textsubscript{apm} is a bona fide eukaryal enzyme in that respect. This corroborates the sequence alignments revealing homologies of TyrRS\textsubscript{apm} with eukaryal synthetases (2). Our functional assays of yeast tRNA\textsuperscript{Tyr} transcript variants pointed out additional similarities with the behavior of the yeast enzyme (26), with the C\textsubscript{1},G\textsubscript{72} base pair and the discriminator nucleotide A\textsubscript{73} being the strongest identity determinants. The anticodon residues also contribute to the tyrosine identity but to a lesser extent. The largest difference was found for the recognition of base G\textsubscript{34}, the mutation of which has a strong effect in yeast and other systems (9, 26) but contributes weakly to tyrosylation by TyrRS\textsubscript{apm}.

Two opposite scenarios could account for the presence of four aaRS in the genome of mimivirus. These enzymes, identified in a virus for the first time, could correspond to the remains of an ancestral viral genome that encoded a complete and functional translation apparatus, or they could have been acquired at once or sequentially from ancestral cellular hosts.

The position of mimivirus TyrRS was first analyzed by computing a phylogenetic tree of TyrRS from all domains of life (Fig. 7). The tree robustly separates bacterial and eukaryal TyrRS\textsubscript{s} on one side from archaeal, protozoan, and plant TyrRS\textsubscript{s} on the other side. The TyrRS\textsubscript{apm} sequence is most closely related to TyrRS found in the kingdom Protozoa, and its closest relative is the TyrRS from Entamoeba histolytica (54% identical residues over 330 residues). Although amoebas from the genus Entamoeba are not hosts to mimivirus (56), they belong to the same Amoebozoa clade as Acanthamoeba (the virus’s natural host). In the absence of sequence data for Acanthamoeba TyrRS, we could speculate that the TyrRS\textsubscript{apm} was actually acquired from an ancestral amoebal host.

However, the TyrRS tree strongly disagrees with the accepted species tree for the Protozoa (3). TyrRS\textsubscript{s} from reputed distant cellular organisms are found close to each other (e.g., Excavata Trichomonas versus Amoebozoa Entamoeba, Chromalveolata Phytophthora versus Excavata Trypanosoma, and Chromalveolata Plasmodium versus Amoebozoa Dictyostelium), while reputed close organisms are found very distant (e.g., Amoebozoa Dictyostelium and Entamoeba). The viral TyrRS\textsubscript{apm} sequence exhibits similar inconsistencies, its next closest homologue belonging to Trichomonas vaginalis (52% identical residues), a member of clade Excavata, in principle quite distant from the Amoebozoa supergroup (Fig. 7) (3, 46). At the same time, the sequence of TyrRS\textsubscript{apm}, presumably acquired from an Amoebozoa ancestor host, is only 43% identical to the TyrRS of Dictyostelium discoideum, another member of the Amoebozoa. Such inconsistencies suggest that TyrRS\textsubscript{s} genes have been transferred laterally between the ancestors of today’s representatives of these various early diverging protozoan clades, thus precluding the unambiguous identification of the source of mimivirus TyrRS or even the direction of an eventual exchange with an Amoebozoa ancestor.

Our experimental results also confirmed the intermediate (ancestral?) status of mimivirus TyrRS. It was found significantly less active (Table 2) on Plasmodium falciparum tRNA\textsuperscript{Tyr} transcripts than on yeast tRNA\textsuperscript{Tyr} transcripts, despite a greater sequence similarity between TyrRS\textsubscript{apm} and Plasmodium TyrRS than between TyrRS\textsubscript{apm} and yeast TyrRS. The difference in activity might be due to the dissimilar D-loop organizations between yeast (member of the Opisthokonta) and P. falciparum (member of the Chromalveota) tRNA\textsuperscript{Tyr}, as seen in other aminocacylation systems (47).

We extended our phylogenetic study to the four mimivirus enzymes by using archael and eukaryotic aaRS sequences (including bacterial-type mitochondrial eukaryotic aaRS) to investigate their origin. The computed trees (Fig. 8) clearly do not mitigate in favor of a recent acquisition of any mimivirus aaRS from an amoebal host. Both ArgRS and MetRS exhibit a basal branching, predating the radiation of the eukaryotic kingdom and of most protozoan supergroups (3). In contrast, mimivirus CysRS exhibits a strong affinity with the Excavata member Giardia lamblia. However, like the previous TyrRS tree (Fig. 7 and 8) these trees exhibit serious inconsistencies with the accepted protozoan species tree, again suggesting an active lateral exchange of these genes among early diverging protozoan ancestors. Altogether, these results argue against a recent acquisition of these aaRS by the mimivirus lineage and are not incompatible with their presence in a mimivirus ancestor genome prior to the divergence of the various protozoan supergroups.

If the evolutionary origin of TyrRS\textsubscript{apm} remains ambiguous,
the specificity of its activity, its structure, and its phylogenetic position classify it as a typical archaeal/eukaryotic TyrRS in a protozoan lineage closer to the archaeal lineage than to the one encompassing the cytoplasmic metazoan and yeast TyrRSs (Fig. 8). Its short sequence and its good specific activity are reminiscent of the concept of “optimized” viral enzymes already proposed in the context of other large DNA viruses (60). The lack of recognition of the third anticodon nucleotide by

FIG. 7. Phylogenetic position of mimivirus TyrRS. S. aureus, Staphylococcus aureus; T. pallidum, Treponema pallidum; M. loti, Mesorhizobium loti; M. penetrans, Mycoplasma penetrans; S. pombe, Schizosaccharomyces pombe; C. elegans, Caenorhabditis elegans; B. subtilis, Bacillus subtilis; A. aeolicus, Aquifex aeolicus; P. carinii, Pneumocystis carinii; E. cuniculi, Encephalitozoon cuniculi; D. melanogaster, Drosophila melanogaster; X. laevis, Xenopus laevis; E. dispar, Entamoeba dispar; P. ramorum, Phytophthora ramorum; P. sojae, Phytophthora sojae; C. parvum, Cryptosporidium parvum; C. hominis, Cryptosporidium hominis; P. yoelii, Plasmodium yoelii; O. sativa, Oryza sativa; A. thaliana, Arabidopsis thaliana; N. tabacum, Nicotiana tabacum. All other organisms are defined in the text. mito, mitochondrial sequences.
FIG. 8. Phylogenetic position of mimivirus aaRS. Archeal sequences are colored in purple, eukaryotic in red, mitochondrial (Mito) (bacterial type) in cyan, and mimivirus in green. *A. fulgidus*, *Archaeoglobus fulgidus*; *P. abyssi*, *Pyrococcus abyssi*; *A. pernix*, *Aeropyrum pernix*; *T. volcanium*, *Thermoplasma volcanium*; *A. gambiae*, *Anopheles gambiae*; *G. gallus*, *Gallus gallus*; *B. taurus*, *Bos taurus*; *H. sapiens*, *Homo sapiens*; *D. rerio*, *Danio rerio*; *X. tropicalis*, *Xenopus tropicalis*. All other organisms are defined in the text or the legend for Fig. 7.
TyrRS$_{apm}$ might be either a simplification made possible by the absence of tRNA for TAG and TAA codons or, more boldly, a reminiscence of an ancestral two-letter genetic code. Future biochemical and structural studies on the other mimivirus aaRS might help to answer these questions.

Conclusions. The discovery of four aaRS in the genome of mimivirus, together with other proteins central to the translation apparatus (initiation factor, elongation factor, peptide release factors, and tRNA modification enzyme) (49), violated the established view that protein translation is a process uniquely encoded by cellular organisms. Still, mimivirus possesses only a few of all of the genes necessary to encode a functional protein translation apparatus, and the presence of only four aaRS (together with six tRNAs, five of which are substrates of other aaRS) does not make immediate biochemical sense. Yet, the presence of tRNAs in DNA virus genomes is not uncommon, in particular in large phycodnaviruses (to date, the closest lineage to Mimiviridae [15]), where some of them participate directly in the translation process (60). Sequenced chlorella virus genomes also exhibit a homologue of the fungal-specific translation elongation factor 3 (e.g., open reading frame A666L in Paramecium bursaria [57]), most similar to a chlorella homologue, strongly suggesting (in contrast with mimivirus aaRS) a recent horizontal transfer with an ancestor of today’s chlorovirus hosts (data not shown).

In the absence of an obvious reason for cellular-specific functions found to be encoded by viruses, the traditional interpretation is that they correspond to a regulatory function helping the virus hijack the cell metabolism. A number of accessory roles have been described for synthetases. For instance, two aaRS (including the mitochondrial TyrRS of Neurospora crassa) have been involved in splicing of group I introns (14, 41). Eukaryotic aaRS have also been involved in cell cycle control (36, 59), the direct biosynthesis of pyrrolysine and selenocysteine, and the regulation of the pool of cytoplasmic/nuclear tRNA (reviewed in reference 28). Moreover, these accessory functions all require the presence of special features or domains not found in mimivirus aaRS. None of these functions have yet been described in the context of virus infections.

Rather unexpectedly, our structural and functional analysis of the two viral aam$_{apm}$ did not reveal any unusual features that might suggest involvement in nonenzymatic, accessory processes. On the contrary, the structure of TyrRS$_{apm}$ corresponds to the minimal archaeal-like core catalytic domain, is indeed specific of tyrosine, and obeys the expected rules of interaction with eukaryal cognate tRNAs. MetRS$_{apm}$ is also specific for methionine and behaves as a regular eukaryotic MetRS. According to these results, the most parsimonious explanation remains that mimivirus aaRS$_{apm}$ directly participate in the protein translation process in infected cells. Further experiments are required to determine how essential these virus-encoded aaRS are to mimivirus replication and how they complement or interfere with the amoeba translation machinery.

ACKNOWLEDGMENTS

We are grateful to D. Raoult for providing the mimivirus genomic DNA and to Gérard Keith and Franco Fasiolo (IBMC, Strasbourg, France) for the generous gifts of yeast initiator tRNA$m^{60}$ and tRNA$^{65}$ clones. We also thank D. Nurizzo for data collection at ESRF on ID29 beamline; Y. Ogata, S. Chenisvise, and D. Byrne for technical assistance; A. Lartigue for helpful discussions; and P. Hingamp for reading the manuscript. We acknowledge the use of the Marseille-Nîmes Génopôle bioinformatic platform.

This work was supported by Centre National de la Recherche Scientifique (CNRS), Université Louis Pasteur, Strasbourg, and the French Ministry for Research (ACI BCMS).

REFERENCES


Escherichia coli tyrosyl-tRNA synthetase and of its interaction with tRNA.


