JVI Accepts, published online ahead of print on 12 September 2007
J. Virol. doi:10.1128/JVI.01107-07

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 Claverie[1]

[1]Structural and Genomic Information Laboratory

CNRS-UPR2589, Institut de Biologie Structurale & Microbiologie IFR88,

Université de la Méditerranée,

163 avenue de Luminy, Case 934

13288 Marseille, Cedex 9, France

[2]Architecture et Réactivité de l'ARN

CNRS-UPR 9002, IBMC and Université Louis Pasteur

15 rue René Descartes, 67084 Strasbourg, Cedex, France

*__Correspondent footnote:__ Chantal Abergel, Structural and Genomic Information Laboratory CNRS-UPR2589, IBSM-IFR88, 163 avenue de Luminy, Case 934, 13288 Marseille Cedex 9, France. Tel : (+33) 491 825422 Fax : (+33) 491 825421. Chantal.Abergel@igs.cnrs-mrs.fr

Running Title: The Mimivirus aminoacyl tRNA-synthetases

Keywords: aminoacyl tRNA synthetase/3D-structure/tRNA specificity/amino acid specificity/ Mimivirus

Abstract words count: 234

Main text word count : 5134

ABSTRACT

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 of a 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, the TyrRS specificity for tyrosine and its conformity with the archea/eukarya identity rules for $tRNA^{Tyr}$. 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 functions as regular translation enzymes in infected amoebas. Their phylogenetic classification does not suggest that they have been recently acquired by horizontal gene transfer from a cellular host, and rather militate for an intricate evolutionary relationship between large DNA viruses and ancestral eukaryotes.

Introduction

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 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, Mimivirus genome encodes four aminoacyl-tRNA synthetases (aaRSs: 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 tRNA charging steps. The specificity of the tRNA recognition is ensured by stringent constraints referred to as "identity rules" (30). AaRSs 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 aaRSs share the sequence motifs HIGH and KMSKS, and have active sites based on a Rossmann-fold domain. Class II aaRSs 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 aaRSs attach amino acids to the 2'-hydroxyl of the terminal adenosine of the tRNA, whereas charging occurs on the 3'-hydroxyl for class II aaRSs (5). Tyrosyl-tRNA synthetases belong to class Ic of aaRS, together with TrpRS. Both are homodimers (18), a feature otherwise mostly shared 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 one, following many gene losses through a process of reductive evolution (such as observed for intracellular parasitic bacteria), or it is the results of multiple horizontal gene acquisitions conferring some selective advantages to the virus. The later is most often accomplished by diverting the acquired genes from their original functions. To address the above dilemma, we initiated a comprehensive study of the structure and activity of Mimivirus four class I aminoacyl-tRNA synthetases.

We complemented these experimental studies by searching evidence of horizontal gene transfers through a phylogenetic analysis of the four Mimivirus aminoacyl-tRNA synthetases.

This article reports a detailed functional study of two of the four viral class I aaRS, Mimivirus TyrRS ($TyrRS_{apm}$) and MetRS ($MetRS_{apm}$) and the structural study of the $TyrRS_{apm}$ in complex with tyrosinol. These enzymes were found to exhibit the specificity and function predicted from their sequences, but the three-dimensional structure of $TyrRS_{apm}$ exhibited significant differences from its cellular counterparts.

The phylogenetic analysis did not provide evidence for a recent acquisition of these genes from a cellular host close to modern acanthamoeba, but connects Mimivirus aminoacyl-tRNA synthetases to a variety of early diverging protozoan supergroups such as Amoebozoa and Excavata (3).

Results and Discussion

$TyrRS_{apm}$ and $MetRS_{apm}$ possess the predicted enzymatic activities

The presence of aminoacyl-tRNA synthetase 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 $MetRS_{apm}$ and $TyrRS_{apm}$. They were both tested for their 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 $TyrRS_{apm}$ and $MetRS_{apm}$ indeed respectively activate tyrosine and methionine, to the exclusion of any other amino acid. Furthermore, $TyrRS_{apm}$ specifically tyrosylates $tRNA^{Tyr}$, while $MetRS_{apm}$, as other methionyl-tRNA synthetases, aminoacylates both eukaryotic and bacterial $tRNA^{Met}$ (see below, Fig. 1C, Table I).

The structure of $TyrRS_{apm}$

The overall structure of $TyrRS_{apm}$ resembles that of other TyrRS.

TyrRSs are all similarly organized with an N-terminal catalytic domain, including the connective-polypeptide 1 (CP1) region responsible for dimerization, followed by a C-terminal domain (8). However TyrRSs do exhibit important differences in the sequence and architecture of their C-terminal domains. Archaeal 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, $TyrRS_{apm}$ exhibits the typical fold of the TyrRS core domain and is more similar to the archaeal type with a N-terminus

Rossmann-fold catalytic domain, an anticodon binding domain and no extra C-terminal domain. $TyrRS_{apm}$ shares 30% identity over 340 residues with the TyrRS of the hyper-thermophilic Euryarchaeota *Pyrococcus horikoshii*, its closest known structural homologue (2CYC, ref 38).

The protein was co-crystallized with ATP and the tyrosine analogue tyrosinol, and its structure 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 $TyrRS_{apm}$ active site with the *Methanococcus jannaschii* (1J1U, ref 37) and *Thermus thermophilus* (1H3E, ref 62) structures in complex with $tRNA^{Tyr}$ and L-tyrosine or tyrosinol, and the human TyrRS catalytic core structure (1Q11, ref 61) in complex with tyrosinol. The position of the residues in contact with tyrosinol in the $TyrRS_{apm}$ structure were found to be superimposable to their counterparts in the archaeal structures (Fig. 2) and correspond to conserved amino acids, except for two histidines replaced by F81 and N198, as in plants and protozoa sequences (Fig. 3).

The ATP binding site of $TyrRS_{apm}$ exhibits a modified version (a HIAQ motif as in protozoa and plants) of the catalytically important HIGH motif found in the archaeal and eukaryotic structures 1J1U and 1Q11 (Fig. 3). These motifs are superimposable in the structures. A perfect KMSKS motif is present in the $TyrRS_{apm}$ 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 $tRNA^{Tyr}$ (62). This loop, often disordered in ATP-free TyrRS structures, is also associated with large B-factor values in both the $TyrRS_{apm}$ and 1H3E structures.

Structural peculiarities of the TyrRS$_{apm}$ dimer.

In contrast to other tRNA/aaRS systems where the tRNA binds to a single subunit of the synthetase, the tRNA^{Tyr} recognition involves both TyrRS subunits (6, 62). The acceptor arm of tRNA^{Tyr} 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 others by a two-fold 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, with two tRNA molecules simultaneously interacting with one TyrRS dimer (62). However previous kinetic studies of tyrosine activation and tRNA^{Tyr} charging revealed an anti-cooperative behavior of the TyrRS dimer in solution (27). Other experiments suggested that each TyrRS dimer only binds and tyrosylates 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$_{apm}$ is a dimer in solution (data not shown), and crystallized as an homodimer. The superimposition of the TyrRS$_{apm}$ structure with all other available TyrRS dimeric structures nonetheless highlights a major difference. While the first monomer superimposes very well with the first monomer of other TyrRSs structures (RMSD <1.6 Å based on Cα superimposition), the second one is found at a near 90-degree angle relative to its position in other dimers (Fig. 4 and 5). Despite this dramatic conformational change, the CP1 domain (Fig. 3) is still central to the TyrRS$_{apm}$ dimer formation as in other TyrRS, but exhibits a significant alteration of a motif structurally conserved in all other TyrRS structures. This (α_6-turn-α_7) motif, a 14-residues long helix followed by a 90-degree turn and a 9-residue long helix, is now replaced by a 20-residue long

171 extended helix (α_7) in the $TyrRS_{apm}$. This helix is followed by a variable loop disordered in
172 the $TyrRS_{apm}$ as well as in the *M. jannaschii* apo TyrRS structure 1U7D. Interestingly, this
173 new and unique conformation of the $TyrRS_{apm}$ dimer is found in two crystal forms ($P2_1$, ref 2,
174 and $P2_12_12_1$, this work).

175 We investigated how this peculiar structure of the $TyrRS_{apm}$ dimer could be reconciled
176 with the previously proposed model of TyrRS dimers interaction with $tRNA^{Tyr}$. A first
177 possibility is that the conformation of the $TyrRS_{apm}$ dimer, although observed in two crystal
178 forms, is a crystallization artifact. The dimer interface should then appear more flexible than
179 the rest of the molecule, leading to higher B-factor values. This region of the $TyrRS_{apm}$
180 structure is actually more agitated than its surroundings. However, this is the case for all
181 available TyrRS structures, and this feature is thus not conclusive. We then compared the
182 buried surface area upon dimer formation across all TyrRS structures. This buried surface
183 value is 2360 $Å^2$ for $TyrRS_{apm}$ and ranges from 2970 to 3300 $Å^2$ for other TyrRSs. Forcing
184 the $TyrRS_{apm}$ dimer into the standard conformation would cost 1000 $Å^2$ of the buried surface
185 area and greatly reduces its stability. Conversely, forcing the other (canonical) TyrRS dimer
186 into the conformation observed for the $TyrRS_{apm}$ dimer, would also be energetically
187 unfavorable, again at a cost of about 1000 $Å^2$ (see Materials and Methods). Packing forces are
188 thus not likely to be responsible for the unique conformation of the $TyrRS_{apm}$ dimer at odds
189 with the "canonical" TyrRS structures and sheds new light on the previously observed
190 discrepancies between the enzyme properties in solution *versus* its crystallographic structure
191 (7,20,27).

192 We then examined the second possibility, namely that the tRNA could adopt a new
193 position to interact with the observed $TyrRS_{apm}$ dimer. Assuming an invariant structure for the
194 tRNA, we verified that a suitable model of the complex could be built through a rotation of
195 the tRNA relative to the $TyrRS_{apm}$ dimer as already observed for bacterial tRNA-TyrRS

196 complexes (37). In this crude model, the residues of the $TyrRS_{apm}$ interacting with the
197 acceptor arm of the tRNA are properly positioned, while the anticodon can still be recognized
198 by the anticodon binding site of the second $TyrRS_{apm}$ monomer.
199
200 **Functional idiosyncrasies of $TyrRS_{apm}$**
201 Wild-type $tRNA^{Tyr}$ from *E. coli* and yeast, in their native or transcribed versions, were
202 assayed for tyrosylation by $TyrRS_{apm}$. Both eukaryotic yeast $tRNA^{Tyr}$ (corresponding to TAC
203 and TAT codons) are tyrosylated up to 90% whereas only a weak activity (>2%) could be
204 detected with *E. coli* $tRNA^{Tyr}$ (Fig. 1C). Kinetic parameters for tyrosylation were slightly
205 altered when comparing the native molecule to the unmodified transcript resulting in a ~3-
206 fold reduction in aminoacylation efficiency (as defined by the k_{cat}/K_M ratio) (Table IA). These
207 results demonstrate that although the crystallographic dimer does not exhibit the canonical
208 conformation, it is active in solution. Upon interaction with the tRNA, it is thus likely that a
209 productive $tRNA/TyrRS_{apm}$ dimer complex is formed as discussed previously.
210 *Recognition of the $tRNA^{Tyr}$ acceptor stem: $TyrRS_{apm}$ obeys the archea/eukarya identity rules.*
211 All eukarya/archaea $tRNA^{Tyr}$ species possess a C_1–G_{72} base-pair located on the top of the
212 acceptor branch that is replaced by the reverse pair in prokaryotic and mitochondrial $tRNA^{Tyr}$
213 (55). The N_1–N_{72} base-pair, part of the tyrosine identity set, also determines the strong
214 phylogenic barrier preventing cross-tyrosylation between eukaryotic $tRNA^{Tyr}$ and bacterial
215 TyrRS (36,58).
216 $TyrRS_{apm}$ is active on yeast and inactive on bacterial or mitochondrial $tRNA^{Tyr}$ (data not
217 shown). To assess the conformity of the $TyrRS_{apm}$ to previously defined identity rules,
218 tyrosylation was assayed on yeast $tRNA^{Tyr}$ transcripts bearing mutations at the following
219 locations: the C_1-G_{72} base-pair, the anticodon triplet $G_{34}U_{35}A_{36}$, and the discriminator base
220 A_{73} (26). Replacement of the first base-pair by a G_1-C_{72} or A_1-U_{72} pair inactivates $tRNA^{Tyr}$

221 tyrosylation by $TyrRS_{apm}$. Similarly, no activity could be detected after mutation of A_{73} into
222 G_{73} (Table IA).

223 Four synthetase residues are known to be involved in the acceptor stem recognition in
224 archea: R132, R174, K175, and M178 (according to 1J1U numbering). The corresponding
225 amino acids are conserved in $TyrRS_{apm}$ (R149, R195, K196 and M199; Fig. 3). To analyze
226 their involvement in tRNA binding we superimposed a $TyrRS_{apm}$ monomer on the monomer
227 of the *M. jannaschii* TyrRS/$tRNA^{Tyr}$ complex at the acceptor site. Except for residue R149,
228 not visible in the $TyrRS_{apm}$ structure, the homologous residues are positioned as in the *M.*
229 *jannaschii* structure, with residues R195 and M199 contacting the C_1 base and residue K196
230 contacting the A_{73} discriminator base. In order to make contact with the G_{72} base (such as
231 seen with the homologous R132 residue in the *M. jannaschii* TyrRS/$tRNA^{Tyr}$ complex),
232 residue R149 in helix α_7 of the $TyrRS_{apm}$ structure would require a conformational change
233 breaking the helix into the canonical α_6-turn-α_7 motif. This is consistent with the high B-factor
234 values observed at these positions, and the results obtained with the C_1-G_{72}→G_1-C_{72} variant.
235
236 *Recognition of the $tRNA^{Tyr}$ anticodon.*

237 The other residues involved in tyrosine identity correspond to the anticodon nucleotides
238 $G_{34}U_{35}A_{36}$, the strength of which depends on the system studied (9). The two available crystal
239 structures of TyrRS/$tRNA^{Tyr}$ complexes identified the specific contacts established between
240 these identity elements and the amino acid residues on the synthetases (37, 42,62). By
241 comparison with all available TyrRS sequences and structures, the $TyrRS_{apm}$ loop involved in
242 anticodon recognition is much shorter (Fig. 3 and 6). To examine the functional consequence
243 of this unique feature, we studied the tyrosylation of variants derived from yeast $tRNA^{Tyr}$
244 transcript mutated at anticodon sites. In contrast with the yeast TyrRS, only weak effects were
245 observed upon mutation of the G_{34} anticodon nucleotide (Table I). The structural comparison

of the *M. jannaschii* complex with the $TyrRS_{apm}$ structure highlights that, in the archaeal complex, the G_{34} base is sandwiched between the F261 and H283 rings and is also recognized by D286 through hydrogen bonds. This part of the $TyrRS_{apm}$ structure exhibits the most significant differences with archaeal/eukaryotic TyrRS. The β-3_{10}-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_{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 IA) this suggests that the G_{34} base of the anticodon is not used as a discriminator for $tRNA^{Tyr}$ recognition by $TyrRS_{apm}$.

In contrast, mutations of the two other anticodon nucleotides (35 and 36) induce strong negative effects (from 87- to 696-fold) (Table IA) of the $TyrRS_{apm}$ 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 $MetRS_{apm}$

Wild-type native initiator $tRNA^{Met}$ from *E. coli* and *S. cerevisiae* were tested for methionylation by $MetRS_{apm}$. Both substrates were aminoacylated with kinetic parameters (K_M and k_{cat}) slightly decreased for the eukaryal tRNA that results in a 2-fold loss in methionylation efficiency as compared to the *E. coli* $tRNA^{Met}$ (Table IB). Unmodified yeast $tRNA^{Met}$ transcript behaves globally as its native counterpart with only a 2-fold loss in aminoacylation efficiency. Noticeable, however, are the low K_M and high k_{cat} values of yeast $tRNA^{Met}$ transcript towards $MetRS_{apm}$ as compared to the values determined with the yeast enzyme. Replacing individually the anticodon residues abolishes methionine acceptance (Table IB) as already demonstrated for *E. coli* and yeast MetRSs (52,53). Thus, as for all

271 MetRS, the $C_{34}U_{35}A_{36}$ anticodon residues govern specific recognition of $tRNA^{Met}$ by
272 $MetRS_{apm}$.
273
274 **The evolutionary origin of $aaRS_{apm}$**
275 Previous functional studies on TyrRSs revealed a phylogenetic barrier preventing cross-
276 tyrosylation of eubacterial $tRNA^{Tyr}$ by eukaryal TyrRS and *vice-versa* (36,58). Our work
277 demonstrates that $TyrRS_{apm}$ is a *bona fide* eukaryal enzyme in that respect. This corroborates
278 the sequence alignments revealing homologies of $TyrRS_{apm}$ with eukaryal synthetases (2).
279 Our functional assays of yeast $tRNA^{Tyr}$ transcript variants pointed out additional similarities
280 with the behavior of the yeast enzyme (26), with the C_1-G_{72} base-pair and the discriminator
281 nucleotide A_{73} being the strongest identity determinants. The anticodon residues also
282 contribute to the tyrosine identity but to a lesser extent. The largest difference was found for
283 the recognition of base G_{34}, the mutation of which has a strong effect in yeast and other
284 systems (9,26), but contributes weakly to tyrosylation by $TyrRS_{apm}$.
285 Two opposite scenarii could account for the presence of four aaRS in the genome of
286 Mimivirus. These enzymes, for the first time identified in a virus, could correspond to the
287 remains of an ancestral viral genome that encoded a complete and functional translation
288 apparatus, or they could have been acquired at once or sequentially from ancestral cellular
289 hosts.
290 The position of Mimivirus TyrRS was first analyzed by computing a phylogenetic tree of
291 TyrRS from all domains of life (Fig. 7). The tree robustly separates bacterial and eukaryal
292 TyrRSs on one side, from archaeal, Protozoa, and Plants TyrRSs on the other side. The
293 $TyrRS_{apm}$ sequence is most closely related to TyrRS found in Kingdom Protozoa and its
294 closest relative is the TyrRS from *Entamoeba histolytica* (54% identical residues over 330
295 residues). Although amoebas from genus Entamoeba are not host to Mimivirus (56), they

296 belong to the same Amoebozoa clade as Acanthamoeba (the virus's natural host). In the
297 absence of sequence data for Acanthamoeba TyrRS, we could speculate that the $TyrRS_{apm}$
298 was actually acquired from an ancestral amoebal host.

299 However, the TyrRS tree strongly disagrees with the accepted species tree for the
300 Protozoa (3). TyrRSs from reputed distant cellular organisms are found close from each
301 others (e.g. Excavata *Trichomonas vs*. Amoebozoa *Entamoeba*, Chromaveolata *Phytophthora*
302 *vs*. Excavata *Trypanosoma*, Chromaveolata *Plasmodium vs.* Amoebozoa *Dictyostelium*)
303 while reputed close organisms are found very distant (e.g. Amoebozoa *Dictyostelium* and
304 *Entamoeba*). The viral $TyrRS_{apm}$ sequence exhibits similar inconsistencies, its next closest
305 homologue belonging to *Trichomonas vaginalis* (52% identical residues), a member of clade
306 Excavata, in principle quite distant from the Amoebozoa supergroup (Fig. 7, ref 3, 46). At the
307 same time, the sequence of $TyrRS_{apm}$, presumably acquired from an amoebozoan ancestor
308 host, is only 43% identical to the TyrRS of *Dictyostelium discoideum* another member of the
309 Amoebozoa. Such inconsistencies suggest that TyrRS genes have been laterally transferred
310 between the ancestors of today's representatives of these various early diverging protozoan
311 clades, thus precluding the unambiguous identification of the source of Mimivirus TyrRS, or
312 even the direction of an eventual exchange with an amoebozoan ancestor.

313 Our experimental results also confirmed the intermediate (ancestral ?) status of
314 Mimivirus TyrRS: It was found significantly less active (Table IA) on *Plasmodium*
315 *falciparum* $tRNA^{Tyr}$ transcripts than on yeast $tRNA^{Tyr}$ transcripts, despite a greater sequence
316 similarity between $TyrRS_{apm}$ and *Plasmodium* TyrRS than with the yeast TyrRS. The
317 difference in activity might be due to the dissimilar D-loop organization between yeast
318 (member of the Opisthokonta) and *P. falciparum* (member of the Chromalveota) $tRNA^{Tyr}$, as
319 seen in other aminoacylation systems (47).

We extended our phylogenic study to the four Mimivirus enzymes using archeal and eukaryotic aaRSs sequences (including bacterial type mitochondrial eukaryotic aaRSs) to investigate their origin. The computed trees (Fig. 8) clearly do not militate in favor of a recent acquisition of any Mimivirus aaRS from an amoebal host. Both the 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 Giarda 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 aaRSs 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 $TyrRS_{apm}$, 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 non recognition of the third anticodon nucleotide by $TyrRS_{apm}$ might either be 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 aaRSs, might help answering these questions.

Conclusion

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 only possesses a handful of all the genes necessary to encode a functional protein translation apparatus, and the presence of only four aaRS (together with six tRNAs, five of which being substrates of other aaRSs) 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, ref 15), where some of them directly participate in the translation process (60). Sequenced chlorella virus genomes also exhibit an homologue of the fungal-specific translation elongation factor 3 (e.g. ORF A666L in PBCV1, ref. 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 aaRSs (including the mitochondrial TyrRS of *Neurospora crassa*) have been involved in splicing of group I introns (14, 41). Eukaryotic aaRSs 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 ref 28). More recently, the yeast AspRS has been shown to regulate its expression by interacting with its own mRNA (29). Noticeably, 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 $aaRS_{apm}$ did not reveal any unusual features that might suggest its involvement in non enzymatic, 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. The $MetRS_{apm}$ is also specific of methionine and behave as a regular eukaryotic MetRS. According to these results, the most parsimonious explanation remains that Mimivirus $aaRS_{apm}$ directly participates in the protein translation process in infected cells. Further experiments are now required to determine how essential these virus-encoded aaRS are to Mimivirus replication, and how they complement or interfere with the amoeba translation machinery.

Materials and Methods

Preparation of $TyrRS_{apm}$, $MetRS_{apm}$ and tRNAs

A. polyphaga Mimivirus TyrRS encoding gene was PCR amplified from genomic DNA and cloned into a Gateway system (Invitrogen, ref 33) as described earlier (1,2).

Oligomeric state of recombinant TyrRS was measured by gel filtration on an analytical S200 column. The purified TyrRS was loaded on the column at a 10 mg/ml 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 $tRNA^{Tyr}$ and initiator $tRNA^{Met}$ from *Escherichia coli* were purchased from Subriden and Sigma, respectively. Their equivalents in *Saccharomyces cerevisiae* were purified to homogeneity by counter current distribution (21) followed by appropriate column chromatographies. Transcripts of yeast $tRNA^{Tyr}$ and initiator $tRNA^{Met}$ as well as their variants were obtained as described elsewhere (19,26). Finally, transcripts of *P. falciparum* $tRNA^{Tyr}$ were obtained by *in vitro* transcription of synthetic genes using the "transzyme" method (25).

Crystallization and structure determination of $TyrRS_{apm}$

The selenomethionyl substituted protein was produced using a standard protocol (34). Crystallization was improved by introducing an anion exchange chromatography (Aktä explorer 10S, GE healthcare) on a Resource Q column (6 ml) using a NaCl gradient (0 to 1 M, 20 Column Volume). Three fractions were recovered (150 mM, 200 mM and 240 mM NaCl) and the best crystals were obtained with the second fraction desalted in 20 mM Tris (pH 7.4) and concentrated to 14 mg/ml. The crystals were grown at 298 K by vapor diffusion mixing 2 µl of TyrRS solution containing tyrosinol and ATP 1 mM with 0.5 µl of reservoir (500 µl) 0.1 M sodium citrate (pH 5.5) and 6 to 9 % PEG 4000 (w/v), 15 % MPD, 0.1 M KCl, 1 mM $MgCl_2$. Crystals belong to the orthorhombic space group $P2_12_12_1$ (a=63.25, b=107.19, c=148.67 Å) with one biological dimer per asymmetric unit. Both MAD and native datasets were collected at the ESRF synchrotron beamline ID29 and integrated using MOSFLM (40). Data were scaled and reduced using the CCP4i package (16,24) (Table II).

Even though selenomethionyl substituted protein crystals produced low resolution data and were fast decaying under X-ray exposure, a full 4 Å resolution MAD dataset was collected. AutoSHARP (10) was used to obtain initial phases and 23 selenium atoms were located, refined and solvent-flattening was performed to improved phases. The resulting electronic density map was used to superimpose in TURBO-FRODO (50) a model of the TyrRS structure produced using MODELLER (51) onto the refined selenium atoms positions. A round of manual building was performed to better fit the initial model in the electronic density map and to remove the portion of the model that were not in density.

The unrefined structure was then used to identify a molecular replacement solution (AMoRe, ref 43) with a 2.2 Å dataset of native TyrRS protein in complex with tyrosinol and ATP. The model was further refined by using rigid body refinement followed by several rounds of

positional refinement in CNS (11) with manual rebuilding using TURBO-FRODO (50). The quality of the structure was assessed using PROCHECK (39) (Table II).

Buried surface areas computations were performed using GRASP (44). The detailed calculation for the various TyrRS dimer conformations required the modeling of a 20-residue CP1 region next to the α_6-turn-α_7 motif, structurally conserved in all TyrRSs, but disordered in the $TyrRS_{apm}$ structure (Fig. 3). This region was not found to significantly contribute to the difference of buried surface areas between the two $TyrRS_{apm}$ dimer conformations. Structure graphical representations (Fig. 2 and Fig. 4-6) were produced using VMD (35).

The $TyrRS_{apm}$ structure coordinates have been deposited to the Protein Data Bank under accession code: 2J5B

Activity assays

ATP/PPi exchange reactions. Reaction media (200 µl) contained 100 mM Na-Hepes (pH 7.2), 10 mM MgCl2, 2 mM KF, 2 mM ATP, 2 mM [^{32}P]PPi (1-2 cpm/pmol) and 1 mM of a mixture of all amino acids, including or not, tyrosine or methionine. Reactions were initiated by addition of 1.5 µg of $TyrRS_{apm}$ or $MetRS_{apm}$. The [^{32}P]ATP formed after 5, 10, 15 and 20 min incubation at 37°C was determined as described elsewhere (12). Control experiments with either no aaRS or no amino acid were conducted in parallel.

$tRNA^{Tyr}$ 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 $tRNA^{Tyr}$ or $tRNA^{Tyr}$ transcripts (wild-type or mutated) was performed (50 µl) at 30°C in 50 mM Na-Hepes (pH 7.5), 25 mM KCl, 12 mM $MgCl_2$, 2.5 mM ATP, 0.2 mg/ml BSA, 1 mM spermine (54), 10 µM L-[^{14}C] tyrosine (adjusted to 750 cpm/pmol) and the required concentration of tRNA molecules and $TyrRS_{apm}$. Methionylation of initiator $tRNA^{Met}$ or

444 $tRNA^{Met}$ transcripts (wild-type or mutated) was performed (50 µl) at 30°C in 20 mM Na-
445 Hepes (pH 7.5), 10 mM $MgCl_2$, 1 mM dithioerythritol, 2 mM ATP (4), 10 µM L-[^{35}S]
446 methionine (adjusted to 400 cpm/pmol) and the required concentration of tRNA molecules
447 and $MetRS_{apm}$. At different incubation times, aliquots were spotted on 3MM Whatman paper
448 and 5% trichloroacetic acid precipitated. Incorporation of radioactive amino acid was
449 measured by liquid scintillation spectroscopy. Kinetic parameters (K_M and k_{cat}) were
450 determined from Lineweaver and Burk plots.

451 **Phylogenetic position of Mimivirus aaRS**

452 Homologous TyrRS protein sequences from all major phyla were aligned using the Muscle
453 software (22) and a maximun likelyhood tree was computed with PhyML (32) using the
454 default option on our phylogeny server at URL : phylogeny.fr. Bootstrap values are indicated
455 along the branches. Branches have been collapsed for bootstrap values smaller than 50%.

456 Additional phylogenetic analysis of each four Mimivirus aaRSs were performed using
457 archeal and eukaryotic aaRSs including mitochondrial sequences (bacterial type aaRSs). For
458 each aaRS, protein sequences were aligned using the T-coffee software (45). Maximun
459 likelyhood trees and Neighbour joining trees (Bootstrap 500) were computed as above. The
460 two methods produced similar trees. Only maximum likelihood trees are presented in figure 7.
461

462 **Acknowledgments**

463 We are grateful to D. Raoult for providing the Mimivirus genomic DNA and to Gérard Keith
464 and Franco Fasiolo (IBMC, Strasbourg) for the generous gift of yeast initiator $tRNA^{Met}$ and
465 $tRNA^{Met}$ clones, respectively. We also want to thank D. Nurizzo for data collection at ESRF
466 on ID29 beamline, Y. Ogata, S. Chenivesse and D. Byrne for technical assistance, A. Lartigue
467 for helpful discussions and P. Hingamp for reading the manuscript. We acknowledge the use
468 of the Marseille-Nice Génopole bioinformatic platform. This work was supported by Centre

469 National de la Recherche Scientifique (CNRS), Université Louis Pasteur, Strasbourg, and the
470 French Ministry for Research (ACI “BCMS”).

References:

1. **Abergel, C., Coutard, B., Byrne, D., Chenivesse, S., Claude, J. B., Deregnaucourt, C., Fricaux, T., Gianesini-Boutreux, C., Jeudy, S., Lebrun, R., Maza, C., Notredame, C., Poirot, O., Suhre, K., Varagnol, M., and Claverie, J-M.** 2003. Structural genomics of highly conserved microbial genes of unknown function in search of new antibacterial targets. J. Struct. Funct. Genomics. **4**:141–157.
2. **Abergel, C., Chenivesse, S., Byrne, D., Suhre, K., Arondel, V., and Claverie, J-M.** 2005. Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetase found in a virus. Acta Crystallograph. **F61**:212–215.
3. **Adl, S.M., Simpson, A.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser ,S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., Taylor, M.F.** 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot Microbiol. **52**:399-451.
4. **Aphasizhev, R., Théobald-Dietrich, A., Kostyuk, D., Kochetkov, S. N., Kisselev, L., Giegé, R., and Fasiolo, F.** 1997. Structure and aminoacylation capacities of tRNA transcripts containing deoxyribonucleotides. RNA. **3**:893-904.
5. **Arnez, J. G., and Moras, D.** 1997. Structural and functional considerations of the aminoacylation reaction. Trends Biochem. Sci. **22**:211–216.
6. **Bedouelle, H., and Winter, G.** 1986. A model of synthetase/transfer RNA interaction as deduced by protein engineering. Nature. **320**:371–373.
7. **Bedouelle, H.** 1990. Recognition of tRNAtyr by tyrosyl-tRNA synthetase. Biochimie. **72**:589–598.

8. **Bedouelle, H.** 2005. *In* Aminoacyl-tRNA synthetases, Ibba M, Francklyn C and Cusack S (eds.), Chapter 12, pp. 111–124. Landes Biosciences, Georgetown, TX.

9. **Bonnefond, L., Giegé, R., and Rudinger-Thirion, J.** 2005. Evolution of the $tRNA^{Tyr}$/TyrRS aminoacylation systems. Biochimie. **87**:873–883.

10. **Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., Paciorek, W.** 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallograph. **D59**:2023–2030.

11. **Brunger, 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., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L.** 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallograph. **D54**:905–921.

12. **Campanacci, V., Dubois, D. Y., Becker, H. D., Kern, D., Spinelli, S., Valencia, C., Pagot, F., Salomoni, A., Grisel, S., Vincentelli, R., Bignon, C., Lapointe, J., Giegé, R., and Cambillau, C.** 2004. The Escherichia coli YadB gene product reveals a novel aminoacyl-tRNA synthetase like activity. J. Mol. Biol. **337**:273–283.

13. **Carter, Jr. C. W.** 1993. Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Ann. Rev. Biochem. **62**:715–748.

14. **Cherniack, A. D., Garriga, G., Kittle, Jr. J. D., Akins, R. A., and Lambowitz, A. M.** 1990. Function of Neurospora mitochondrial tyrosyl-tRNA synthetase in RNA splicing requires an idiosyncratic domain not found in other synthetase. Cell. **62**:745–755.

15. **Claverie, J-M., Ogata, H., Audic, S., Abergel, C., Suhre, K., and Fournier, P-E.** 2006. Mimivirus and the emerging concept of giant virus. Virus Res. **117**:133–144.

16. **Collaborative Computational Project Number 4.** 1994 "The CCP4 suite: programs for protein crystallography". Acta Crystallograph. **D50**:760-763.

17. **Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., and Leberman, R.** 1990. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. Nature **347**:249–255.

18. **Delarue, M., and Moras, D.** 1993. The aminoacyl-tRNA synthetases family: modules at work. BioEssays. **15**:675–687.

19. **Despons, L., Senger, B., Fasiolo, F., and Walter, P.** 1992. The anticodon triplet is not sufficient to confer methionine acceptance to a transfer RNA. J. Mol. Biol. **225**:897-907.

20. **Dessen, P., Zaccaï, G. and Blanquet, S.** 1982. Neutron scattering studies of Escherichia coli tyrosyl-tRNA synthetase and of its interaction with tRNAtyr. J. Mol. Biol. **159**:651–664.

21. **Dirheimer, G., and Ebel, J-P.** 1967. Fractionation of Brewer's yeast t-RNA by countercurrent distribution. Bull. Soc. Chim. Biol. *(Paris)* **49**:1679–1687.

22. **Edgar, R. C.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. **5**:113.

23. **Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D.** 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature **347**:203–206.

24. **Evans, P. R.** 1993 "Data reduction", *In* Proceedings of CCP4 Study Weekend, on Data Collection and Processing, (Daresbury Laboratory, Warrington), pp. 114–122.

25. **Fechter, P., Rudinger, J., Giegé, R., and Théobald-Dietrich, A.** 1998. Ribozyme processed tRNA transcripts with unfriendly internal promoter for T7 RNA polymerase: production and activity. FEBS Lett**. 436**:99–103.

26. **Fechter, P., Rudinger-Thirion, J., Théobald-Dietrich, A., and Giegé, R.** 2000. Specific tyrosylation of the bulky tRNA-like structure of Brome Mosaic Virus RNA relies solely

on identity nucleotides present in its amino acid accepting domain. Biochemistry **39**:1725–1733.

27. **Fersht, A. R., and Jakes, R.** 1975. Tyrosyl-tRNA synthetase from Escherichia coli. Stoichiometry of ligand binding and half-of-the-sites reactivity in aminoacylation. Biochemistry. **14**:3344–3350.
28. **Francklyn, C., Perona, J. J., Puetz, J., and Hou, Y. M.** 2002. Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation. RNA. **8**:1363–1372.
29. **Frugier, M., Ryckelynck, M., and Giegé, R.** 2005. tRNA-balanced expression of a eukaryal aminoacyl-tRNA synthetase by an mRNA-mediated pathway. EMBO Rep. **6**:860–865.
30. **Giegé, R., Sissler, M., and Florentz, C.** 1998. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. **26**:5017–5035.
31. **Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F.** 1999. ESPript: multiple sequence alignments in PostScript. Bioinformatics. **15**:305–308.
32. **Guindon, S., and Gascuel, O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**:696-704.
33. **Hartley, J. L.., Temple, G. F., and Brasch, M. A.** 2000. DNA cloning using in vitro site-specific recombination. Genome Res. **10**:1788–1795.
34. **Hendrickson, W. A., Horton, J. R., and LeMaster, D. M.** 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. **9**:1665–1672.
35. **Humphrey, W., Dalke, A., and Schulten, K.** 1996 "VMD - Visual Molecular Dynamics". J. Mol. Graphics. **14**:33–38.

568 36. **Kleeman, T. A., Wei, D., Simpson, K. L., and First, E. A.** 1997. Human tyrosyl-tRNA
569 synthetase shares amino acid sequence homology with a putative cytokine. J. Biol. Chem.
570 **272**:14420–14425.
571 37. **Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S.,**
572 **Sakamoto, K., and Yokoyama, S.** 2003. Structural basis for orthogonal tRNA
573 specificities of tyrosyl-tRNA synthetases for genetic code expansion. Nat. Struct. Biol.
574 **10**:425–432.
575 38. **Kuratani, M., Sakai, H., Takahashi, M., Yanagisawa, T., Kobayashi, T., Murayama,**
576 **K., Chen, L., Liu, Z. J., Wang, B. C., Kuroishi, C., Kuramitsu, S., Terada, T., Bessho,**
577 **Y., Shirouzu, M., Sekine, S., and Yokoyama, S.** (2006). Crystal structures of tyrosyl-
578 tRNA synthetases from Archaea. J. Mol. Biol. **355**:395–408.
579 39. **Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M.** 1993.
580 PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl.
581 Cryst. **26**:283-291.
582 40. **Leslie, A.** 1993. "Data reduction", *In* Proceedings of CCP4 Study Weekend, on Data
583 Collection and Processing, (Daresbury Laboratory, Warrington), pp. 44–51.
584 41. **Myers, C. A., Kuhla, B., Cusack, S., and Lambowitz, A. M.** 2002. tRNA-like
585 recognition of group I introns by a tyrosyl-tRNA synthetase. Proc. Natl. Acad. Sci. *USA*.
586 **99**:2630–2635.
587 42. **Nair, S., Ribas de Pouplana, L., Houman, F., Avruch, A., Shen, X., and Schimmel, P.**
588 1997. Species-specific tRNA recognition in relation to tRNA synthetase contact residues.
589 J. Mol. Biol. **269**:1–9.
590 43. **Navaza J.** 2001. Implementation of molecular replacement in AMoRe. Acta
591 Crystallograph. **D57**:1367–1372.

44. **Nicholls, A., Sharp, K. A., and Honig, B.** 1991. Protein Folding and Association: Insights From the Interfacial and Thermodynamic Properties of Hydrocarbons. Proteins: Struct. Funct. and Genet. **11**:281–296.

45. **Notredame, C., Higgins, D., and Heringa, J.** 2000. T-Coffee: A novel method for multiple sequence alignments. J. Mol. Biol. **302**:205-217.

46. **Ogata, H., Abergel, C., Raoult, D., and Claverie, J-M.** 2004. Response to Comment on "The 1.2-Megabase Genome Sequence of Mimivirus". Science. **308**:1114.

47. **Perret, V., Florentz, C., Puglisi, J. D., and Giegé, R.** 1992. Effect of conformational features on the aminoacylation of tRNAs and consequences on the permutation of tRNA specificities. J. Mol. Biol. **226**:323–333.

48. **Poirot, O., Suhre, K., Abergel, C., O'Toole, E., and Notredame, C.** 2004. 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. Nucleic Acids Res. **32**:W37–40.

49. **Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J-M.** 2004. The 1.2-Mb Genome Sequence of Mimivirus. Science. **306**:1344–1350.

50. **Roussel, A., and Cambillau, C.** 1991. Turbo-Frodo *In* Silicon Graphics Geometry Partners Directory. Mountain View, CA, Silicon Graphics.

51. **Sali, A., and Blundell, T. L.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234**:779–815.

52. **Schulman, L. H. and Pelka, H.** 1988. Anticodon switching changes the identity of methionine and valine transfer RNAs. Science**. 242**:765-768.

53. **Senger, B., and Fasiolo, F.** 1996. Yeast tRNAMet recognition by methionyl-tRNA synthetase requires determinants from the primary, secondary and tertiary structure: a review. Biochimie. **78**:597-604.

54. **Sohm, B., Frugier, M., Brulé, H., Olszak, K., Przykorska, A., and Florentz, C.** 2003. Towards understanding human mitochondrial leucine aminoacylation identity. J. Mol. Biol. **328**:995–1010.

55. **Sprinzl, M., and Vassilenko, K. S.** 2005. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **33**:D139–D140.

56. **Suzan-Monti, M., La Scola, B., Raoult, D.** 2006. Genomic and evolutionary aspects of Mimivirus. Virus Res. **117**:145–155.

57. **Van Etten, J.L., Graves, M.V., Müller, D.G., Boland W., and Delaroque N.** 2002. *Phycodnaviridae* – large DNA algal viruses. Arch. Virol. **147**:1479-1516.

58. **Wakasugi, K., Quinn, C. L., Tao, N., and Schimmel, P.** 1998. Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant. EMBO J**.** **17**:297–305.

59. **Wakasugi, K., and Schimmel, P.** 1999. Two distinct cytokines released from a human aminoacyl-tRNA synthetase. Science. **284**:147–150.

60. **Yamada, T., Onimatsu, H., Van Etten, J. L.** 2006. Chlorella viruses. Adv. Virus Res. **66**:293-336.

61. **Yang, X. L., Otero, F. J,, Skene, R. J., McRee, D. E., Schimmel, P., and Ribas de Pouplana, L.** 2003. Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. Proc. Natl. Acad. Sci. *USA*. **100**:15376–15380.

62. **Yaremchuk, A., Kriklivyi, I., Tukalo, M., and Cusack, S.** 2002. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. EMBO J. **21**:3829–3840.

Figures legend:

Fig. 1. Functional assays on $TyrRS_{apm}$ and $MetRS_{apm}$. Amino acid activation of $TyrRS_{apm}$ and $MetRS_{apm}$ Reactions were conducted in the presence of a mix of all amino acids minus (□) or plus (■) tyrosine (A) or methionine (B). (C) Tyrosylation of *E. coli* (△) and yeast (▲) native $tRNA^{Tyr}$ by $TyrRS_{apm}$. Enzyme and tRNA concentrations were 20 µM and 1.3 µM, respectively.

Fig.2: Invariance of the tyrosinol binding site. (A) The amino acids closest to the tyrosinol molecule in the $TyrRS_{apm}$ and (**B**) in the *M. jannaschii* TyrRS structures (1J1U).

Fig. 3: Structure-based alignment of TyrRSs. $TyrRS_{apm}$ (2J5B) was aligned with eukaryal (1Q11: human, core structure) and archaeal structures (2CYC: *P. horikoshii*, 2CYA: *Aeropyrum pernix*, 2CYB: *Archaeoglobus fulgidus*, 1J1U and 1U7D: *M. jannaschii* complex and apo form). The closest $TyrRS_{apm}$ homologues from protozoa and plants are also included (EHISTO: *Entamoeba histolytica*, DISDIC: *Dictyostelium discoideum*, OSATI: *Oryza sativa*, PYOELI: *Plasmodium yoelii*, PFALC: *Plasmodium falciparum*, ATHAL: *Arabidopsis thaliana*, CHOMI: *Cryptosporidium hominis*, GLAMB: *Giardia lamblia*). The secondary structure elements of $TyrRS_{apm}$ 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 grey. This alignment was produced with 3D-coffee (http://www.igs.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi, ref 48) and the figure was produced with ESPript (31).

Fig. 4: Cartoon comparison of the dimer interface. The two α7 helices of $TyrRS_{apm}$ are in green, the conserved αs_6-turn-αs_7 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 near 90-degrees rotation of the $TyrRS_{apm}$ second monomer (colored according to secondary structure elements: red α-helices, blue β-strands, yellow coils and turns, pink η helices).

Fig. 5: Cartoon representation of the $TyrRS_{apm}$ dimer superimposed on the archaeal TyrRS 2CYC from *P. horikoshii*. The archaeal dimer is transparent and colored in silver excepted for the structural motif (αs_6-turn-αs_7 corresponding to the α_7 helix in $TyrRS_{apm}$) colored in opaque cyan. The $TyrRS_{apm}$ first monomer superimposed on 1CYC first monomer is colored in yellow. The second monomer secondary structure elements are colored as in Fig. 4. The arrow shows the rotation to be applied to the Mimivirus second monomer to superimpose its β_7-turn-β_8 onto the *P. horikoshii* βs_7-ηs_3-βs_8 motif.

668 **Fig. 6: Superimposition of the C-terminal anticodon binding domain of $TyrRS_{apm}$** (solid) on the *M.*

669 *jannaschii* TyrRS/tRNA complex (transparent). The anticodon appears solid in the transparent tRNA surface.

670 **Fig. 7: Phylogenetic position of Mimivirus TyrRS**

671 **Fig. 8: Phylogenetic position of Mimivirus aaRS:** Archeal sequences are colored in purple, eukaryotic in red,

672 mitochondrial (bacterial type) in cyan and Mimivirus in green.

Table I. Kinetic parameters for (A) tyrosylation of wild-type $tRNA^{Tyr}$ molecules or variants by $TyrRS_{apm}$ and (B) methionylation of wild-type $tRNA^{Met}$ molecules or variants by $MetRS_{apm}$.

tRNAs	K_M (μM)	k_{cat} ($10^{-3}sec^{-1}$)	k_{cat}/K_M	L^a (x-fold)
A- Tyrosine				
Wild-type molecules				
yeast (native)	0.5	126	252	0.3
yeast (transcript)	1.7	142	83.5	**1**
Pl. falciparum	4	18.3	4.6	18
Mutated molecules				
yeast C_1-G_{72}→A_1-U_{72}	nm	nm	nm	nm
yeast C_1-G_{72}→G_1-C_{72}	nm	nm	nm	nm
yeast C_1-G_{72}→A_{73}→G_{73}	nm	nm	nm	nm
yeast G_{34}→A_{34}	1.3	7.5	5.8	14.4
yeast G_{34}→C_{34}	1.4	9.3	6.6	12.6
yeast G_{34}→U_{34}	1.3	8.3	6.4	13
yeast U_{35}→A_{35}	16	2.5	0.16	522
yeast U_{35}→C_{35}	33.3	10	0.3	278
yeast U_{35}→G_{35}	9.3	1.1	0.12	696
yeast A_{36}→G_{36}	20	7.3	0.36	232
yeast A_{36}→U_{36}	15.4	14.9	0.96	87
B- Methionine				
Wild-type molecules				
yeast (native)	0.13	6.6	252	0.3
yeast (transcript)	0.74	17.4	83.5	**1**
E. coli (native)	0.47	47	252	0.3
Mutated molecules				
yeast C_{34}→G_{34}	nm	nm	nm	nm
yeast A_{35}→C_{35}	nm	nm	nm	nm
yeast U_{36}→C_{36}	nm	nm	nm	nm

nm : not measurable (loss > 10^5)

[a]L values correspond to losses of efficiency relative to yeast $tRNA^{Tyr}$ transcript in (**A**) and to yeast $tRNA^{Met}$ transcript in (**B**).

The value >1 correspond to a gain in efficiency. Experimental errors on k_{cat} and K_M varied at most by 20%. Experiments represent an average of at least two independent experiments.

715

Table II: X-ray data collection and refinement statistics

Data collection	
Beam line	ESRF/ID29
Wavelength (Å)	0.97925
Space group	$P2_12_12_1$
Unit cell dimensions(Å)	a=63.5 b= 107.3 c=148.9
Resolution range (Å)	87.7 to 2.2
(highest resolution shell)	(2.28 to 2.2)
Observations	238,357
Unique reflections	52537
Multiplicity[1]	4.5 (4.6)
Completeness[1] (%)	100 (100)
$\langle I / \sigma I \rangle$[1,2]	2.2 (2.3)
R_{sym} (%)[1,3]	9.1 (28.9)
Refinement	
Resolution range (Å)	29.8 to 2.2
R_{cryst} (%)[4]	21.6
R_{free} (%)	24.8
Δ_{bond} (Å)	0.006
Δ_{angle} (°)	1.1
N° Protein atoms	5169
N° water	330
Ligand tyrosinol	2
Average B factor (Å^2)	39
Protein main chain	37.1
Water	41.85
Ramachandran plot (%)	
Most favored	539
Allowed	40
Generously allowed	5
Disallowed regions	0

716

717 [1] values in parentheses are for the highest resolution shell.

718 [2] $\langle I / \sigma I \rangle$, is the mean signal to noise ratio, where I is the integrated intensity of a measured
719 reflection and σ is the estimated error in the measurement.

720 [3] $R_{sym} = \sum_h \sum_i \left| I_{h,i} - \langle I_h \rangle \right| / \sum_h \sum_i \left| I_{h,i} \right|$, where I is the integrated intensity of reflection h
721 having i observations and $\langle I_h \rangle$ is the mean recorded intensity of reflection h over multiple
722 recording.

723 [4] $R_{cryst} = \sum \left| |F_o| - |F_c| \right| / \sum |F_o|$, where F_o are observed and F_c calculated structure factor
724 amplitudes. R_{free} is calculated from a randomly chosen 9.9% of reflections.

725
726

727

Figure 1

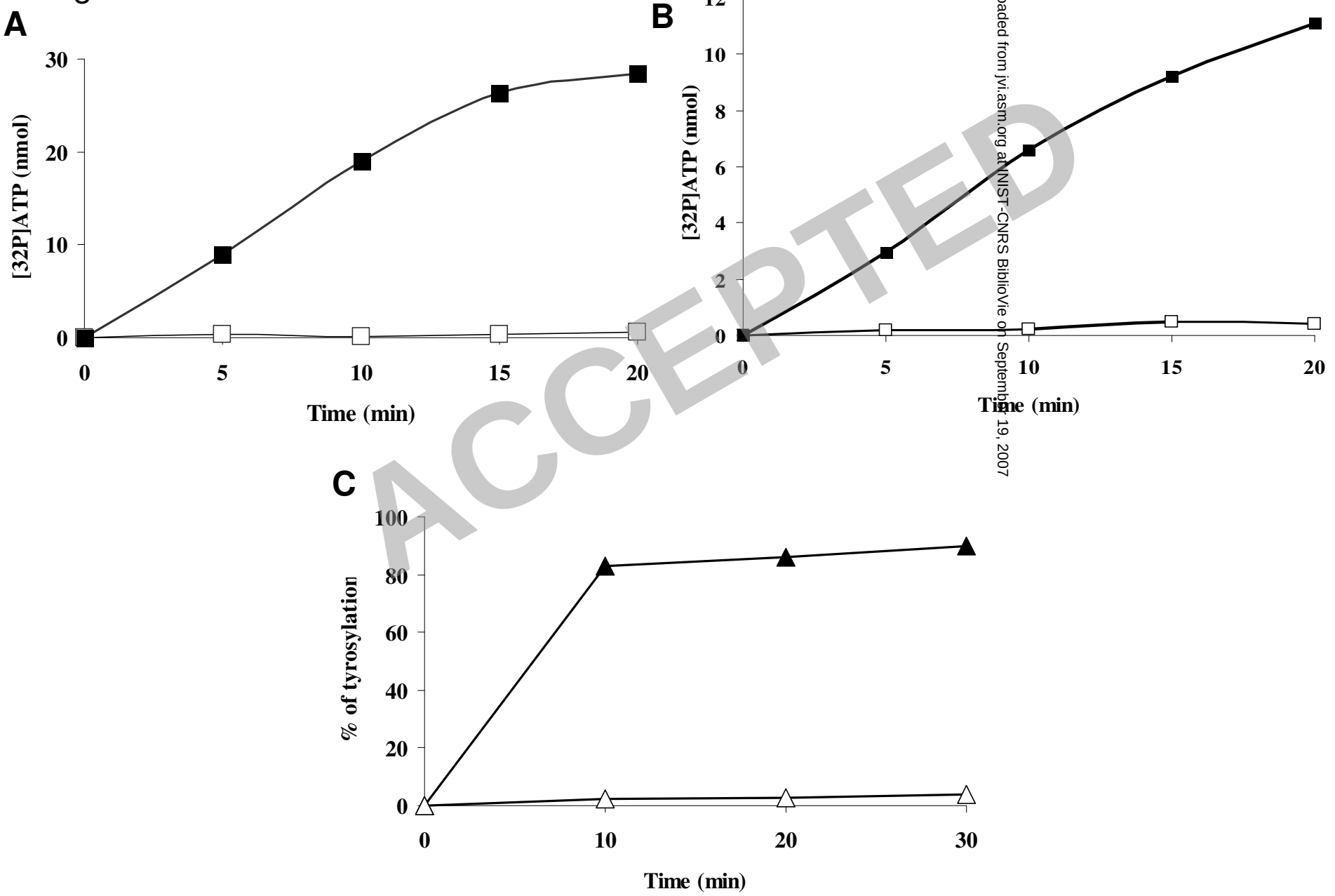

Figure 2

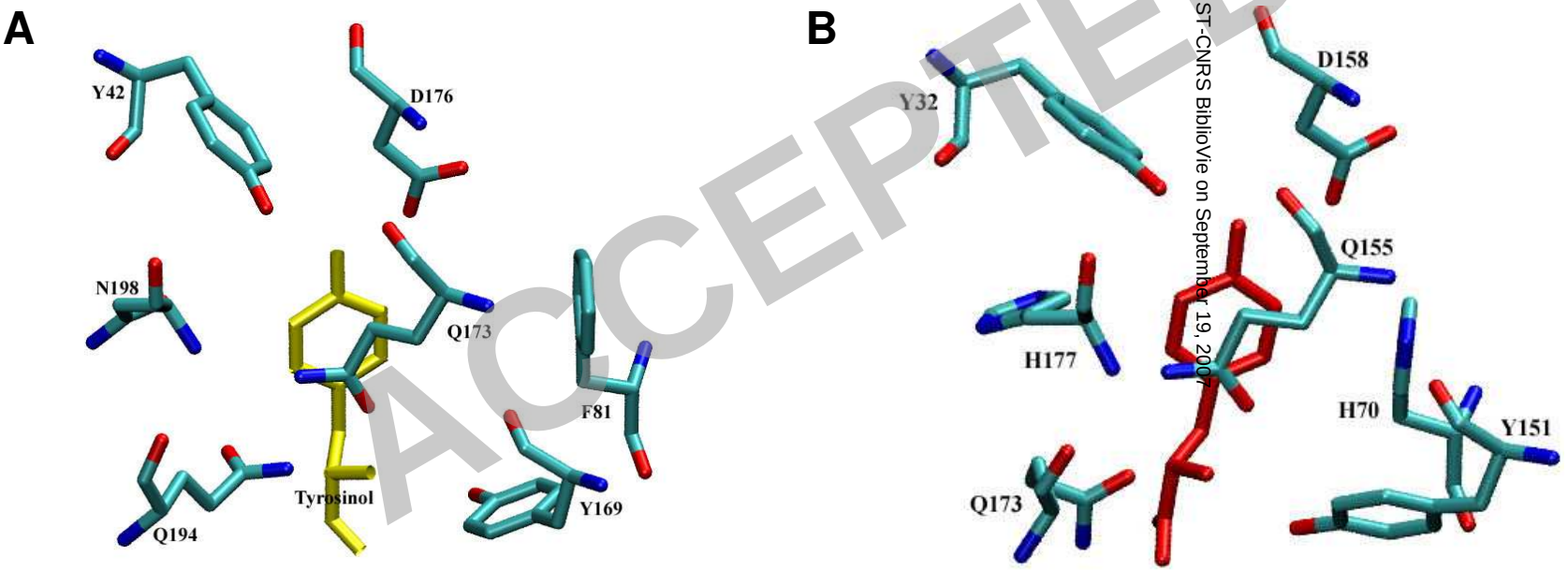

Figure 3

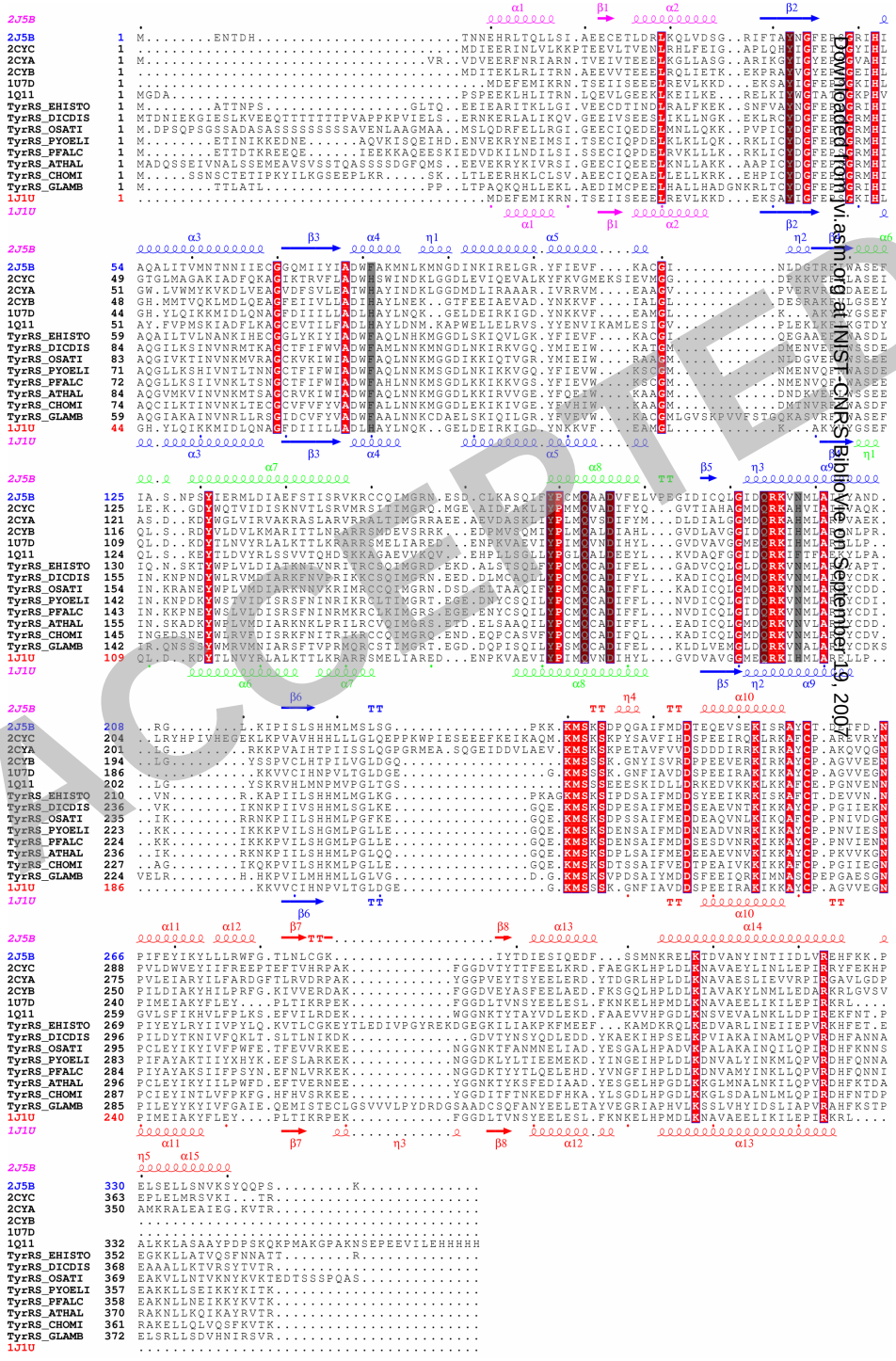

Figure 4

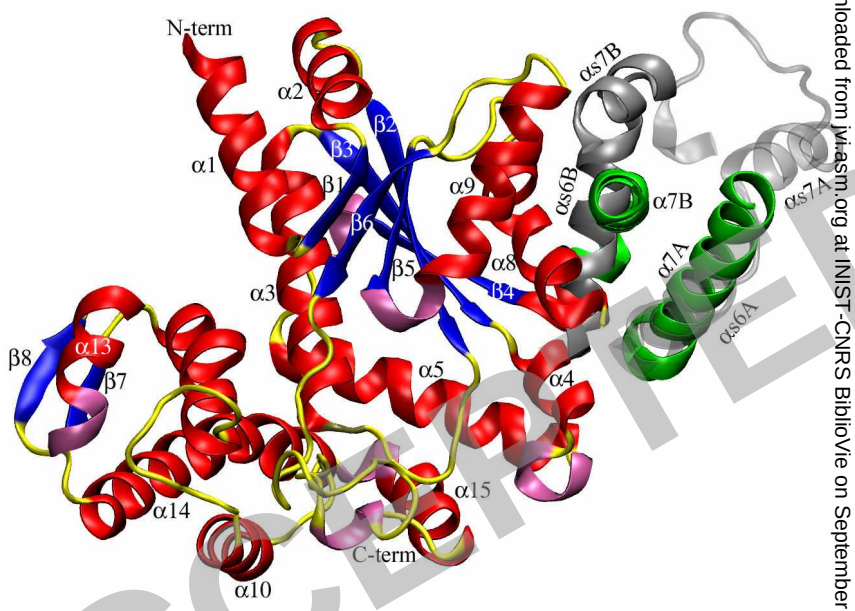

Figure 5

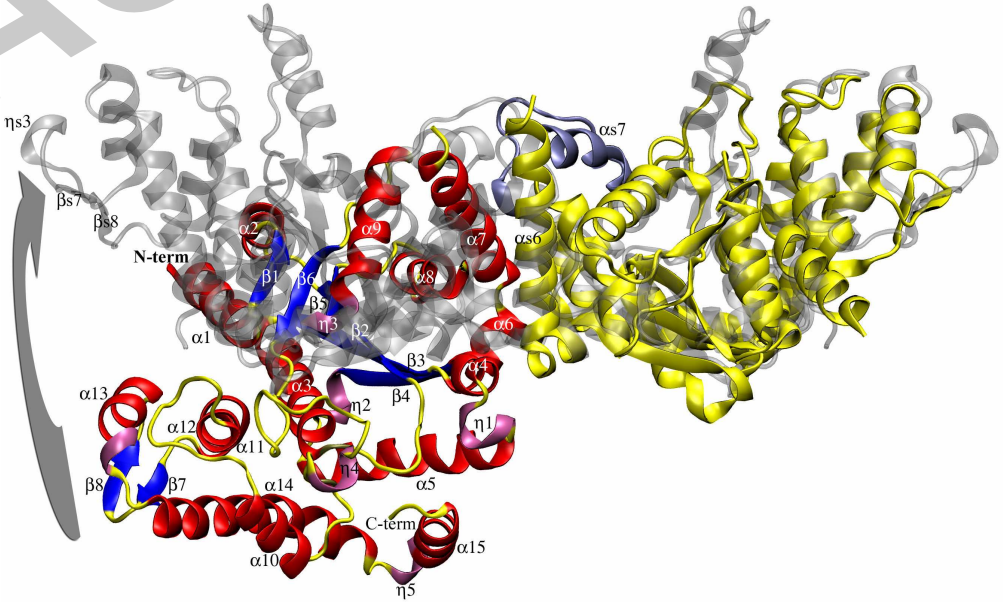

Figure 6

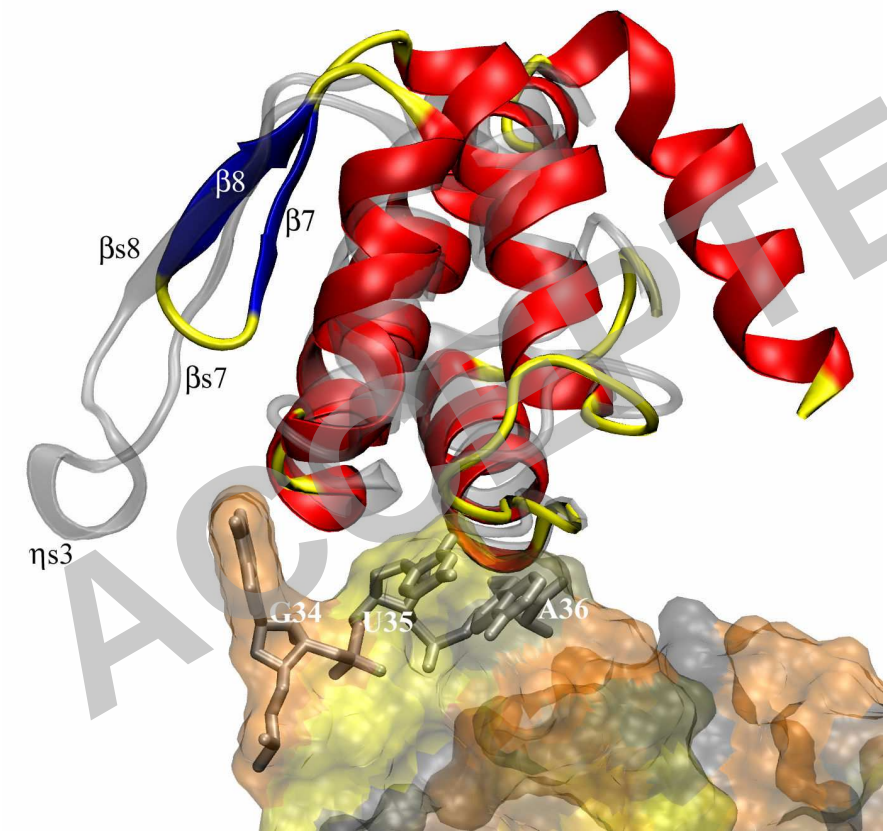

Figure 7

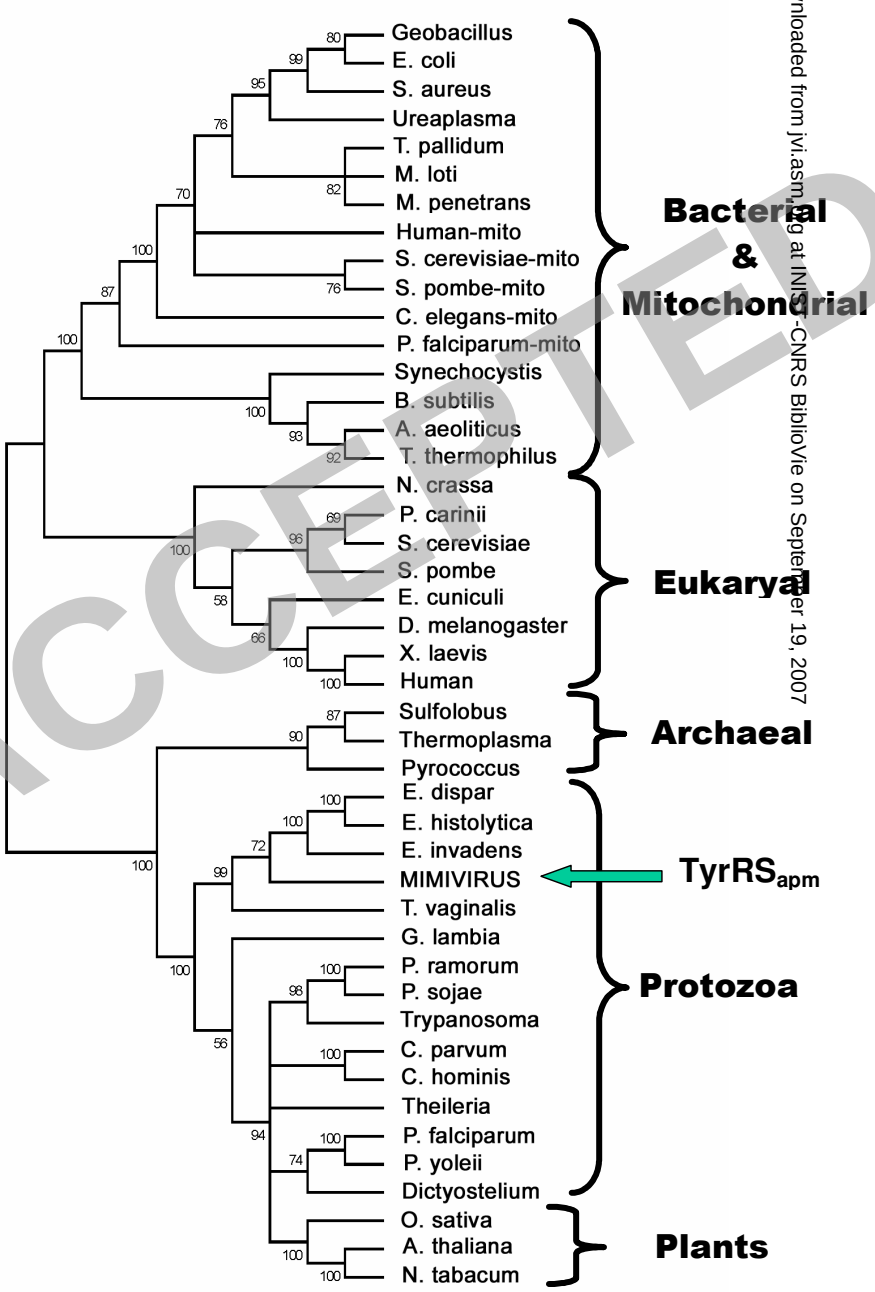

Figure 8