The renaissance of aminoacyl-tRNA synthesis

Michael Ibba* & Dieter Söll†,‡,§,∥

Center for Biomolecular Recognition, Department of Medical Biochemistry and Genetics, Laboratory B, The Panum Institute, Blegdamsvej 3c, DK-2200, Copenhagen N, Denmark, †Departments of Molecular Biophysics and Biochemistry, ‡Chemistry and §Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520-8114, USA

Received February 6, 2001; revised March 13, 2001; accepted March 21, 2001

The role of tRNA as the adaptor in protein synthesis has held an enduring fascination for molecular biologists. Over four decades of study, taking in numerous milestones in molecular biology, led to what was widely held to be a fairly complete picture of how tRNAs and amino acids are paired prior to protein synthesis. However, recent developments in genomics and structural biology have revealed an unexpected array of new enzymes, pathways and mechanisms involved in aminoacyl-tRNA synthesis. As a more complete picture of aminoacyl-tRNA synthesis now begins to emerge, the high degree of evolutionary diversity in this universal and essential process is becoming clearer.

Introduction

Aminoacyl-tRNA synthesis is the ubiquitous cellular process that provides substrates for ribosomal translation of messenger RNA during protein synthesis. Much of the long-lived interest in aminoacyl-tRNA and how it is made, derives from its pivotal role in providing the interface between nucleic acid and amino acid sequences in translation. Consequently, the specificity of aminoacyl-tRNA synthesis in pairing the appropriate tRNAs and amino acids is a key determinant of the faithful transmission of genetic information (Figure 1). The synthesis of any given protein may require the co-translational insertion of up to 20 canonical amino acids plus selenocysteine, whose unusual formation was elucidated more recently (see below). In turn, this necessitates the use of at least 21 distinct pathways of aminoacyl-tRNA formation. Recent years have seen dramatic advances in our knowledge of all aspects of aminoacyl-tRNA synthesis, which have emphasized not only the conserved features, but also the mechanistic complexity and diversity of this essential process.

Aminoacyl-tRNA synthesis: back to diversity

The direct attachment of the correct amino acid to a specific tRNA is carried out by an aminoacyl-tRNA synthetase (AARS) in a two-step process involving ATP-dependent amino acid activation followed by transfer of the amino acid onto tRNA (reviewed in Ibba and Söll, 2000). The discovery of 20 distinct AARSs was a milestone in the study of these enzymes, as it confirmed one of the central tenets of Crick’s adaptor hypothesis (Crick, 1958). The subsequent finding that these 20 enzymes could be partitioned into two highly conserved structural classes of 10 members, each with unrelated active site topologies, provided a satisfying symmetry which seemed to support the notion that aminoacyl-tRNA synthesis varies little among contemporary organisms (see for example Arnez and Moras, 1997, for discussion of the AARS classes). Numerous studies endorsed the notion that aminoacyl-tRNA synthesis had been highly conserved during evolution, and as recently as 1997 a detailed and fairly complete picture of the various processes was seemingly in place. Answers to some long-standing questions were also starting to emerge. For example, the means by which tRNA1le induces the editing (i.e. removal) of misactivated valine by isoleucyl-tRNA synthetase [IleRS; first observed in 1966 (Baldwin and Berg, 1966)] and the alternative route to Asn-tRNA synthesis [suspected since 1972 (White and Bayley, 1972)] were both becoming more transparent (Curnow et al., 1996; Hale et al., 1997). As has subsequently become clear, however, rather than tidying up loose ends, these and many other studies have marked the starting point for the remarkable recent progress in our understanding of many diverse aspects of aminoacyl-tRNA synthesis. Various mechanisms of this process have proven more complex and dissimilar than anticipated, prompting re-assessments of entrenched views on their origin and evolution. Similarly, AARSs and AARS-like proteins

*Corresponding authors. D.S. at: Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, 266 Whitney Avenue, New Haven, CT 06520-8114, USA. Tel: +1 203 432 6200; Fax: +1 203 432 6202; E-mail: soll@trna.chem.yale.edu
M.I. at: Tel: +45 35327771; Fax: +45 35396042; E-mail: mibba@imbg.ku.dk
have been found to play a variety of unexpected roles outside of translation. These range from group I intron splicing (e.g. Chen et al., 2000; Rho and Martinis, 2000) to apoptosis (reviewed with other examples in Schimmel and Ribas de Pouplana, 2000). In all these respects, the title of a keynote address given at a 1997 Keystone Symposium by Dino Moras, ‘Aminoacyl-tRNA synthetases: back to diversity’, has proven to be remarkably prophetic.

Quality control during translation

Correctly aminoacylated tRNAs are essential for faithful translation of the genetic code: any mistakes in the matching of tRNAs and amino acids may be perpetrated as errors in protein synthesis. The range of unique tRNA structures available from the different combinations of nucleotides, both unmodified and modified, ensures that the correct nucleic acid substrate can be specifically selected from the cellular pool by the appropriate AARS. By comparison, the structural simplicity of the amino acid substrates presents a more challenging problem. AARSs achieve the amino acid substrate specificity necessary to keep errors to an acceptable level in essentially two ways, by preferential binding of the correct (cognate) amino acid and by selective editing of incorrect (non-cognate) amino acids. While many structural and biochemical principles underlying amino acid selection and rejection have long been understood (e.g. Jakubowski and Goldman, 1992; Fersht, 1999), recent advances have provided data at the molecular level and revealed unanticipated editing reactions. For example, the crystal structures of IleRS complexed with the non-cognate amino acid valine (Nureki et al., 1998) and with tRNAVal (Silvian et al., 1999) confirmed the existence of distinct catalytic and editing sites in the enzyme, as had long been suspected based on biochemical studies. Subsequent investigations have now revealed the mechanism by which tRNAVal translocates misactivated non-cognate amino acids such as valine, threonine, α-aminobutyrate and cysteine between these catalytic and editing sites, which are over 25 Å apart (Nomanbhoy and Schimmel, 2000). Interestingly, many of the principles of editing described for IleRS have recently been observed in other AARSs. For example valyl-tRNA synthetase (ValRS), which is required to discriminate valine from isoleucine and threonine, displays a high degree of structural homology to IleRS in terms both of its catalytic and editing sites and the role of its cognate tRNA in facilitating hydrolysis of non-cognate amino acids (Fukai et al., 2000). The parallels between IleRS and ValRS are not entirely unexpected, given that the two enzymes are closely related members of the class I AARSs. Indeed leucyl-tRNA synthetase, which is also closely related to IleRS and ValRS, appears to contain a homologous editing domain (Cusack et al., 2000) possibly required for discrimination against the metabolite homocysteine (Jakubowski, 1995). More surprisingly, structural studies have also shown that editing of misactivated serine by threonyl-tRNA synthetase (ThrRS), a class II AARS, is also achieved by tRNA-dependent translocation of the non-cognate amino acid between catalytic and editing sites (Sankaranarayanan et al., 2000). Structure-based comparisons between ThrRS and the class I enzymes IleRS and ValRS indicate that tRNA-dependent editing in both AARS classes requires conformational switching of the tRNA’s 3′-terminal CCA end between a hairpin and a helical conformation; the two classes’ mechanisms are effectively mirror images of each other (Dock-Bregenon et al., 2000). This surprising finding reinforces previous observations that class I and II AARSs appear to be mirror images of each other with respect to tRNA and ATP binding (Ruff et al., 1991), and it substantiates proposals that this property reflects the origin of the two different AARS classes. Significant insight into this last point has come from the finding that certain class I and II AARSs form sterically compatible pairs, whose respective amino acid substrates are themselves sterically related (Ribas de Pouplana and Schimmel, 2001; see also below).

In addition to the editing mechanisms which target mischarged tRNAs as a means of quality control, the aminoacylation of tRNAs may also provide an important check-point for translational fidelity. In much the same way that tRNA maturation and transport are dependent on the recognition of specific conformations (e.g. McClain and Seidman, 1975; Lipowsky et al., 1999) a number of tRNAs are only recognized and exported from eukaryotic nuclei after aminoacylation (Lund and Dahlberg, 1998; Sarkar et al., 1999). Recent studies have shown that many of the AARSs present in mammalian nuclei are part of a large multi-enzyme complex (Nathanson and Deutscher, 2000), which might itself also facilitate tRNA export through association with elongation factor 1 (Grosshans et al., 2000). The existence of a nuclear multi-synthetase complex is not in itself surprising given that comparable cytoplasmic complexes have been well documented (see for example Quevillon et al., 1999 and refer-
ences therein). In addition to AARSs, the cytoplasmic complexes also contain accessory proteins (e.g. Quevillon et al., 1999), elongation factors (e.g. Kim et al., 2000) and molecular chaperones (Kang et al., 2000). Accessory proteins functional in tRNA-binding have also been described in bacteria, suggesting that higher order aminoacylation complexes may also exist outside the eukaryal kingdom (Swairjo et al., 2000). While the exact functional significance of such higher order complexes is not known, it is possible that they may provide a means of optimizing aminoacyl-tRNA synthesis and its subsequent channeling to the ribosome (discussed in Grosshans et al., 2000; Nathanson and Deutscher, 2000).

The biochemical scope of aminoacyl-tRNA synthesis

Recent years have seen dramatic advances in our understanding of canonical aminoacyl-tRNA synthesis, i.e. the direct attachment of cognate tRNAs to amino acids by the 20 canonical AARSs. At the same time it has become clear that the formation of a minority of aminoacyl-tRNAs does not follow the canonical pattern, either in terms of the enzymes and/or the pathways employed for synthesis (Ibba et al., 2000). Pathways whereby aminoacylated tRNAs are further modified prior to their use in translation are known for Asn-tRNA, Gln-tRNA, selenocysteinyl-tRNA and formylmethionyl-tRNA formation (reviewed in Blanquet et al., 2000; Ibba et al., 2000), and AARSs with novel enzymatic activities have been identified for Cys-tRNA and Lys-tRNA formation. Non-canonical aminoacyl-tRNA synthesis is most prevalent in microbes; for example, some archaea have been shown to use such enzymes to synthesize up to five of the 21 aminoacyl-tRNAs they require for protein synthesis (Figure 2). Recent advances in our knowledge of non-canonical aminoacyl-tRNA synthesis are summarized below.

The transamidation pathways for Gln-tRNA and Asn-tRNA synthesis. Many bacteria and archaea lack genes encoding the asparaginyl- (AsnRS) or glutaminyl-tRNA synthetases (GlnRS) or both, as shown in Figure 2. In these organisms, Asn-tRNA and Gln-tRNA are instead synthesized by aminoacylation-transamidation (equations 1–4).

\[
\begin{align*}
\text{Glu} + \text{tRNAGln} &+ \text{ATP} \equiv \text{Glu-tRNAGln} + \text{AMP} + \text{PPi} \quad (1) \\
\text{Glu-tRNAGln} + \text{Gln} + \text{ATP} &\equiv \text{Gln-tRNAGln} + \text{Glu} + \text{ADP} + \text{Pi} \quad (2) \\
\text{Asp} + \text{tRNAAsn} &+ \text{ATP} \equiv \text{Asp-tRNAAsn} + \text{AMP} + \text{PPi} \quad (3) \\
\text{Asp-tRNAAsn} + \text{Asn} + \text{ATP} &\equiv \text{Asn-tRNAAsn} + \text{Asp} + \text{ADP} + \text{Pi} \quad (4)
\end{align*}
\]

Glu-tRNAGln is first misaminoacylated with glutamate (equation 1) by a non-discriminating glutamyl-tRNA synthetase (GluRS), which is able to synthesize both Glu-tRNAGlu and Glu-tRNAGln (Sekine et al., 2001). The resulting mischarged tRNA is then converted by glutamyl-tRNA Gln amidotransferase (GluAdT) into Gln-tRNAGln (equation 2). Similarly, Asn-tRNAAsn is formed via a non-discriminating aspartyl-tRNA synthetase (AspRS, equation 3) and an aspartyl-tRNAAsn amidotransferase (AspAdT, equation 4). In bacteria and archaea, transamidation (equations 2 and 4) is catalyzed by two related multimeric enzymes, GatCAB and GatDE (Tumbula et al., 2000). To date GatDE has only been found in archaea, where it functions solely as a GluAdT. In contrast GatCAB can function as both a GluAdT and as AspAdT, and has been shown to perform one or both reactions in a genome-context specific fashion (Table I; see also Curnow et al., 1998).

Selenocysteinyl-tRNA synthesis. The mechanism by which selenocysteine (Sec) is formed from serine and then co-translationally...
tRNA synthesis is an alternative pathway for Cys-tRNA formation. Most unexpected of all the non-canonical routes of aminoacyl-tRNA synthetases, the dual function prolyl/cysteinyl-tRNA synthetase providing no obvious clues as to why two unrelated LysRS enzymes in most archaea and a few bacteria (see for example Söll et al., 2000) are otherwise members of class II. Functionally, the class II forms of LysRS behave as archetypal members of class I and class II lysyl-tRNA synthetases but, with few exceptions (see below), do not reveal how AARSs but, with few exceptions (see below), do not reveal how the two cognate amino acid substrates are achieved by a specific requirement for binding tRNA<sup>Sec</sup> prior to cysteine activation, but not for tRNA<sup>Pro</sup> prior to proline activation (Stathopoulos et al., 2001).

**The origins of aminoacyl-tRNA synthesis**

The intimate connection between aminoacyl-tRNA synthesis and the genetic code has made this a fertile area for speculation on the origins of coded protein synthesis. While a consensus has arisen that aminoacyl-tRNA synthesis per se may have originated as an RNA-catalyzed process (e.g. Illangasekare and Yarus, 1999; Lee et al., 2000), the evolution of extant AARSs remains more contentious. Phylogenetic comparisons (Nagel and Doolittle, 1995) and structural sub-groupings (Ribas de Pouplana and Schimmel, 2001) indicate clear relationships between AARSs but, with few exceptions (see below), do not reveal how they may have evolved from a presumably smaller number of ancestral proteins. The only AARSs for which clear evolutionary histories have been established are GlnRS and AsnRS. Phylogenetic and structural considerations leave little doubt that GlnRS and AsnRS were the last AARSs to emerge, resulting from duplication and diversification of ancestral GluRS and AspRS, respectively (reviewed in Woese et al., 2000). The comparative evolutionary history of GlnRS and AsnRS also seems to have been relatively recent (reviewed in Woese et al., 2000). Whether the comparatively recent emergence of GlnRS and AsnRS also reflects a late recruitment of glutamine and asparagine to coded protein synthesis is more open to question, as it seems likely that the origin of GlnRS and AsnRS is much more recent (reviewed in Doolittle, 1998). Functional studies with one such archaeon, *Methanococcus jannaschii*, revealed that both Cys-tRNA<sup>Cys</sup> and Pro-tRNA<sup>Pro</sup> can be synthesized by prolyl-tRNA synthetase (ProRS; Stathopoulos et al., 2000). Subsequent investigations have shown that, while the ability to synthesize Cys-tRNA<sup>Cys</sup> is not a feature of all ProRS proteins, it is nevertheless observed in some organisms (e.g. *Giardia, Methanococcus maripaludis* and *Thermus thermophilus*) that also encode a conventional CysRS (Bunjun et al., 2000; Stathopoulos et al., 2000; C. Stathopoulos and I. Ahel, unpublished results). The mechanism that allows certain ProRS enzymes to specifically synthesize both Cys-tRNA<sup>Cys</sup> and Pro-tRNA<sup>Pro</sup> is not apparent from the high-resolution structure of the protein (Yaremchuk et al., 2000). Despite some initial dispute (Lipman et al., 2000), recent biochemical studies indicate that discrimination between the two cognate amino acid substrates is achieved by a specific requirement for binding tRNA<sup>Cys</sup> prior to cysteine activation, but not for tRNA<sup>Pro</sup> prior to proline activation (Stathopoulos et al., 2001).

### Table I. Known tRNA-dependent amidotransferase activities

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function&lt;sup&gt;a&lt;/sup&gt;</th>
<th>AspAdt</th>
<th>AspAdt and GluAdt</th>
</tr>
</thead>
<tbody>
<tr>
<td>gatCAB</td>
<td>many bacteria, chloroplasts</td>
<td>few bacteria and archaea</td>
<td>some bacteria</td>
</tr>
<tr>
<td>gatDE</td>
<td>archaea</td>
<td>none known</td>
<td>none known</td>
</tr>
</tbody>
</table>

<sup>a</sup>Function is defined as the proven ability of the given gene products to catalyze the specified reaction *in vivo* and/or *in vitro*, and is dependent upon the presence of the corresponding non-discriminating aminoacyl-tRNA synthetases. The absence of known mitochondrial GlnRS encoding genes suggests that GluAdt might also exist in some examples of this organelle.

The possibility that Cys-tRNA<sup>Cys</sup> can be formed independently of cysteine-tRNA synthetase (CysRS) was first raised when it became clear that some archaeal genomes do not encode CysRS (reviewed in Doolittle, 1998). Functional studies with one such archaeon, *Methanococcus jannaschii*, revealed that both Cys-tRNA<sup>Cys</sup> and Pro-tRNA<sup>Pro</sup> can be synthesized by prolyl-tRNA synthetase (ProRS; Stathopoulos et al., 2000). Subsequent investigations have shown that, while the ability to synthesize Cys-tRNA<sup>Cys</sup> is not a feature of all ProRS proteins, it is nevertheless observed in some organisms (e.g. *Giardia, Methanococcus maripaludis* and *Thermus thermophilus*) that also encode a conventional CysRS (Bunjun et al., 2000; Stathopoulos et al., 2000; C. Stathopoulos and I. Ahel, unpublished results). The mechanism that allows certain ProRS enzymes to specifically synthesize both Cys-tRNA<sup>Cys</sup> and Pro-tRNA<sup>Pro</sup> is not apparent from the high-resolution structure of the protein (Yaremchuk et al., 2000). Despite some initial dispute (Lipman et al., 2000), recent biochemical studies indicate that discrimination between the two cognate amino acid substrates is achieved by a specific requirement for binding tRNA<sup>Cys</sup> prior to cysteine activation, but not for tRNA<sup>Pro</sup> prior to proline activation (Stathopoulos et al., 2001).

### Table I. Known tRNA-dependent amidotransferase activities

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function&lt;sup&gt;a&lt;/sup&gt;</th>
<th>AspAdt</th>
<th>AspAdt and GluAdt</th>
</tr>
</thead>
<tbody>
<tr>
<td>gatCAB</td>
<td>many bacteria, chloroplasts</td>
<td>few bacteria and archaea</td>
<td>some bacteria</td>
</tr>
<tr>
<td>gatDE</td>
<td>archaea</td>
<td>none known</td>
<td>none known</td>
</tr>
</tbody>
</table>

<sup>a</sup>Function is defined as the proven ability of the given gene products to catalyze the specified reaction *in vivo* and/or *in vitro*, and is dependent upon the presence of the corresponding non-discriminating aminoacyl-tRNA synthetases. The absence of known mitochondrial GlnRS encoding genes suggests that GluAdt might also exist in some examples of this organelle.

### The origins of aminoacyl-tRNA synthesis

The intimate connection between aminoacyl-tRNA synthesis and the genetic code has made this a fertile area for speculation on the origins of coded protein synthesis. While a consensus has arisen that aminoacyl-tRNA synthesis per se may have originated as an RNA-catalyzed process (e.g. Illangasekare and Yarus, 1999; Lee et al., 2000), the evolution of extant AARSs remains more contentious. Phylogenetic comparisons (Nagel and Doolittle, 1995) and structural sub-groupings (Ribas de Pouplana and Schimmel, 2001) indicate clear relationships between AARSs but, with few exceptions (see below), do not reveal how they may have evolved from a presumably smaller number of ancestral proteins. The only AARSs for which clear evolutionary histories have been established are GlnRS and AsnRS. Phylogenetic and structural considerations leave little doubt that GlnRS and AsnRS were the last AARSs to emerge, resulting from duplication and diversification of ancestral GluRS and AspRS, respectively (reviewed in Woese et al., 2000). Whether the comparatively recent emergence of GlnRS and AsnRS also reflects a late recruitment of glutamine and asparagine to coded protein synthesis is more open to question, as it seems likely that Gln-tRNA<sup>Glu</sup> and Asn-tRNA<sup>Asn</sup> were originally synthesized by tRNA-dependent transamidation reactions. Nevertheless, given the widespread use of the AspAdT-catalyzed reaction as the sole
means of asparagine synthesis in many bacteria (e.g. Becker and Kern, 1998; Curnow et al., 1998; B. Min and D. Söll, unpublished results), it is tempting to speculate that asparagine may have been one of the last amino acids recruited to protein synthesis.

**Outlook**

Advances in structural biology and functional genomics have already provided an unprecedented new impetus to the study of aminoacyl-tRNA synthesis, and will undoubtedly continue to do so. For example, genome-wide array analyses offer a means to see how aminoacyl-tRNA synthesis fits into the broader process of translation, and together with gene knockouts and expression arrays may also suggest functions for some of the numerous AARS-like proteins (Schimmel and Ribas de Pouplana, 2000). While AARS interactions with synthetases or other proteins are well known (see above) in higher eukaryotes, little is known about the existence or potential relevance of such interactions in other organisms. Genomic tools provide an obvious means to address this issue. For example, the protein–protein interaction map of *Helicobacter pylori* (Rain et al., 2001) indicates a variety of potential interactions of interest that include not only AARSs and a subunit of the GluAtD, but also other proteins of unknown function. The correlation of function with new protein domains (obtained by structural genomic and computational analyses) may lead to in silico approaches to tease out of the amino acid sequences of unknown open reading frames their putative enzymatic activities. This may be a way of searching for ‘lost’ progenitors of components of the current aminoacyl-tRNA synthesis machinery. Despite the obvious advantages such new methodologies offer, the full realization of their potential still requires careful checks by more classical approaches. For example, transposon insertion studies of the minimal genome of *Mycoplasma genitalium* initially had suggested, quite surprisingly, that both the isoleucyl- and tyrosyl-tRNA synthetases might be dispensable for cellular viability (Hutchison et al., 1999). However, subsequent genetic and biochemical experiments indicated that transposon mutagenesis had removed only non-essential portions of these AARS, and thus not disrupted aminoacyl-tRNA synthesis (B. Ruan and D. Söll, unpublished results). Finally, it is worth noting that current technologies offer no obvious means to rapidly investigate the true nature of the tRNAs themselves in terms of the large number of modified nucleotides they are known to contain (see for example Björk et al., 2001 and references therein). Given that key roles in the recognition of AARSs and regulation of translation have already been ascribed to modified RNAs, such modifications may contribute to a further unanticipated level of complexity in the process of aminoacyl-tRNA synthesis itself.

**Acknowledgements**

The authors would like to thank Lluis Ribas de Pouplana and Paul Schimmel for providing a preprint of a manuscript and Fabrizio Arigoni for data prior to publication. M.I. is the recipient of an Investigator Fellowship from the Alfred Benzon Foundation. Work from the authors’ laboratories was supported by grants from the European Commission (QLG-CT-99-00660, M.I.), the National Institute of General Medical Sciences, the National Aeronautics and Space Administration (D.S.) and the Office of Basic Energy Sciences of the Department of Energy (D.S.).

**References**


Renaissance of aminoacyl-tRNA synthesis


Michael Ibba and Dieter Söll

DOI: 10.1093/embo-reports/kve095