Published in final edited form as:
FEBS Lett. 2010 January 21; 584(2): 342–349. doi:10.1016/j.febslet.2009.11.005.

Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems

Jing Yuan[a,1,*], **Patrick O'Donoghue**[a,1,*], **Alex Ambrogelly**[c], **Sarath Gundllapalli**[c], **R. Lynn Sherrer**[a], **Sotiria Palioura**[a], **Miljan Simonović**[d], and **Dieter Söll**[a,b]

[a]Departments of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA

[b]Chemistry, Yale University, New Haven, CT 06520-8114, USA

[c]Schering-Plough Corporation, Kenilworth, NJ 07033-0530, USA

[d]Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA

Abstract

Selenocysteine and pyrrolysine, known as the 21st and 22nd amino acids, are directly inserted into growing polypeptides during translation. Selenocysteine is synthesized via a tRNA-dependent pathway and decodes UGA (opal) codons. The incorporation of selenocysteine requires the concerted action of specific RNA and protein elements. In contrast, pyrrolysine is ligated directly to $tRNA^{Pyl}$ and inserted into proteins in responses to UAG (amber) codons without the need for complex recoding machinery. Here we review the latest updates on the structure and mechanisms of molecules involved in Sec-$tRNA^{Sec}$ and Pyl-$tRNA^{Pyl}$ formation as well as the distribution of the Pyl-decoding trait.

Selenocysteine biogenesis

Selenocysteine (Sec) is the major biological form of the element selenium, which in trace amounts is essential for human health. Sec is incorporated into polypeptides to form selenoproteins during translation. The 21st amino acid is typically found in catalytic centers of selenoproteins where it plays a functionally essential role. Unlike most amino acids, Sec is universally synthesized on its cognate tRNA [1-4]. During translation, selenocysteinyl-$tRNA^{Sec}$ (Sec-$tRNA^{Sec}$) is delivered to the ribosome by a specific translation factor that requires a characteristic stem-loop structure in the mRNA to actively recode an inframe UGA from stop codon to Sec sense codon. The human genome encodes only 25 selenoproteins [5], yet variations in these Sec-containing proteins or their synthetic machinery is linked to a range of human disorders including cancer and numerous diseases affecting the nervous, immune, and endocrine systems [6].

*Corresponding authors. jing.yuan@yale.edu (J. Yuan), patrick.odonoghue@yale.edu (P. O'Donoghue).
[1]Both authors contributed equally.

The machineries to synthesize Sec and incorporate it into selenoproteins are divergent in bacteria compared to archaea and eukaryotes. In bacteria, serine (Ser) as the precursor of Sec is initially attached to $tRNA^{Sec}$ by seryl-tRNA synthetase (SerRS). The resulting Ser-$tRNA^{Sec}$ is then converted to Sec-$tRNA^{Sec}$ by selenocysteine synthase (SelA) in the presence of the selenium donor selenophosphate. This pathway has been well characterized in *Escherichia coli* [7] and extensively reviewed before [8,9].

The pathway for Sec biosynthesis in archaea and eukaryotes (Fig. 1) was revealed only in the last few years. The missing component was an archaeal/eukaryotic analog of SelA, since no clear sequence-based homolog could be found. An additional enzymatic step is involved in Sec biosynthesis in archaea and eukaryotes. *O*-phosphoseryl-tRNA kinase (PSTK) [10] catalyzes the phosphorylation of Ser-$tRNA^{Sec}$ to form *O*-phosphoseryl-$tRNA^{Sec}$ (Sep-$tRNA^{Sec}$). The Sep-tRNA:Sec-tRNA synthase (SepSecS), an independently evolved protein that is distantly related to SelA [11], then forms the final product Sec-$tRNA^{Sec}$ from selenophosphate and Sep-$tRNA^{Sec}$ [2-4]. Phylogenetic analysis indicated that PSTK and SepSecS co-evolved and are restricted to the archaeal and eukaryotic domains [2,12]. An interesting similarity, as reviewed previously [13], exists between the archaeal and eukaryotic Sec biosynthetic pathway and archaeal tRNA-dependent cysteine biosynthesis, which also proceeds via a Sep-tRNA intermediate [14]. In the first section of this paper, we focus on recent biochemical and structural work that further elucidated the mechanism and specificity of the tRNA and enzymes involved in Sec biosynthesis in archaea and eukaryotes.

$tRNA^{Sec}$ has a distinct structure

$tRNA^{Sec}$ was identified more than two decades ago [1]. It is the longest tRNA with an extended acceptor stem resulting from an abnormal RNase P cleavage specificity [15]. $tRNA^{Sec}$ has an 8-bp acceptor stem and 5-bp T-stem (a 8/5 secondary structure) in bacteria and a 9/4 arrangement in archaea and eukaryotes, and both can told into a 13-bp long acceptor-TΨC helix (Fig. 2A). In contrast, canonical tRNAs typically have a 7-bp acceptor stem and a 5-bp T-stem forming a 12-bp acceptor-TΨC helix. Several other features of $tRNA^{Sec}$, including an elongated D-stem (6-bp instead of 4-bp), a smaller D loop (4-bp instead of 8-bp), a long variable arm, and the absence of the highly conserved U8 residue, make $tRNA^{Sec}$ distinct from canonical tRNAs.

The tertiary structures of *E. coli* and eukaryotic $tRNA^{Sec}$ were investigated by chemical and enzymatic probing during the 1990s [16,17]. Very recently, the first $tRNA^{Sec}$ crystal structure was solved. The unacylated human $tRNA^{Sec}$ transcript was crystallized in complex with human SepSecS [18] and in an unbound state [19]. The tertiary structure confirms the predicted distinct features of the acceptor, T and D-stems. Compared to canonical tRNAs, the overall structure of $tRNA^{Sec}$ is less compact due to the lack of tertiary interactions from the D-arm and the variable arm, i.e., absence of the G15:C48 base pair, which creates a hole in the tertiary core [18] [19]. The D-stem does not stack perfectly along the axis formed by the anticodon stem and the T-loop and shifts to the minor groove side of the tRNA. The long variable arm protrudes from the tertiary core on the side of $tRNA^{Sec}$ opposite from the D-arm [18] [19]. Remarkably, despite different secondary and tertiary interactions, the relative position of the T stem and the variable arm in the free $tRNA^{Sec}$ is almost indistinguishable to that in $tRNA^{Sec}$. The variable arm is one of the major identity elements for the recognition of $tRNA^{Ser}$ and $tRNA^{Sec}$ by seryl-tRNA synthetase (SerRS), and these observations in part explain the dual specificity of SerRS for $tRNA^{Sec}$ and $tRNA^{Sec}$ [18]. Another model of the SerRS-RNA^{Sec} complex suggests that SerRS may also recognize both the G19:C56 base pair and the discriminator base of $tRNA^{Sec}$ [19]. Thus, all the major indentity elements for tRNA serylation are present and can be recognized in $tRNA^{Sec}$.

Uridine at position 8 is highly conserved and forms a tertiary base pair with A14 in all other mature tRNAs. This conserved U8:A4 base pair defines the elbow region of the tRNA [20] and U8 acts as a sensor for the response mechanism to UV exposure [21]. The $tRNA^{Sec}$ has an adenosine at position 8 instead. An A8:A14:U21 base triple was suggested from the chemical and enzymatic probing data [17], yet A8 does not appear to interact with other bases in the two recently reported $tRNA^{Sec}$ crystal structures [18,19]. A tertiary base interaction change is observed when comparing the SepSecB-bound and the free $tRNA^{Sec}$ crystal structures. In the unbound $tRNA^{Sec}$, the base of U20 stabilizes interactions between the D and TΨC loops by forming a base triple with the G19:C56 Watson-Crick base pair [19]. In the SepSecS-bound state the base of U20 does not participate in the base triple. Its electron density is weak, suggesting increased mobility in this part of the D-loop. Thus, the base triple stabilizes an unbound conformation of $tRNA^{Sec}$, whereas on binding to SepSecS the base triple breaks apart. Besides canonical tertiary base interactions (i.e. G18:U55 and G19:C56), unique tertiary interactions are present in $tRNA^{Sec}$ such as the U16:U59 base pair, which stacks on top of A15:A20a and forms a 13-bp stacking helix together with the D and anticodon stems.

Further analysis, based on superimposing the two solved structures of $tRNA^{Sec}$, reveals that while the individual arms of $tRNA^{Sec}$ adopt the same structure, their relative orientations differ in the SepSecS-bound state and unbound state. Indeed, the acceptor, T and variable arms undergo a conformational change upon binding to SepSecS (Fig. 2B): (*i*) the variable arm makes a 33° rotation around the axis that is parallel to the anticodon arm, (*ii*) the T arm moves in the same direction, but to a somewhat lesser extent (17°), and (*iii*) the acceptor arm makes a 6° rotation around the anticodon stem as well as a 24° rotation around the axis that is parallel to the variable arm (Fig. 2C). The conformational change facilitates the interaction between $tRNA^{Sec}$ and SepSecS and presumably orients the CCA end toward the active site. It will be interesting to see if $tRNA^{Sec}$ undergoes a similar conformational change as it binds to other components of the Sec biosynthetic apparatus and the translating ribosome.

$tRNA^{Sec}$ plays a pivotal role in the Sec biosynthesis pathway. It interacts with Sec specific proteins including PSTK and SepSecS in archaea and eukaryotes. The distinct structural features of $tRNA^{Sec}$ are proven to be major recognition elements throughout Sec biosynthesis as detailed below. IT is unknown, however, it these features of $tRNA^{Sec}$ are also important for Sec incorporation, which requires the participation of multiple protein factors in Sec decoding during translation.

PSTK is a $tRNA^{Sec}$-dependent kinase

In archaea and eukaryotes, PSTK catalyzes the second step of Sec biosynthesis (Fig. 1, center), the phosphorylation of Ser-$tRNA^{Sec}$ to Sep-$tRNA^{Sec}$ in a reaction requiring ATP and magnesium. Such kinase activity was first observed in rat and rooster liver lysate almost 40 years ago [22,23]. In 2004, the identity of this elusive kinase as PSTK was finally revealed in mouse via a comparative genomic approach [10]. Subsequently the archaeal homolog was shown to have the same activity [24], and the biochemical and structural properties of the *Methanocaldococcus jannaschii* PSTK (MjPSTK) have been extensively studied [12,24-26].

MjPSTK is a homodimeric enzyme with each monomer consisting of a N-terminal kinase domain and a C-terminal domain that is putatively involved in tRNA binding. PSTK is a member of the P-loop kinase family with the conserved Walker A and B motifs and the RX_3R motif that are responsible for the recognition of the Mg^{2+}-ATP complex [12,26]. Mutations to either the Walker A or B motifs ablate PSTK activity both *in vitro* and *in vito* [26]. Mutagenesis also revealed that the distal Arg residue in the RX_3R is critical for catalysis both *in vivo* and *in vitro*. Interestingly, the kinase activity is not highly specific for the phosphate donor, since all four rNTPs and even dATP are competent substrates [12]. Structural analysis

of the PSTK:AMPPNP complex supported this observation since the adenosine ring makes few specific contacts, with the exception of a cation-π interaction involving the RX_3R motif [26].

The conserved Asp residue (Asp41) in the Walker B motif plays an essential role in MjPSTK catalysis. Mutation of Asp41 to Asn leads to total loss of activity [26]. The structure shows that Asp41 stabilizes the ATP bound Mg^{2+} ion via a water mediated hydrogen bond, while modeling of the Ser substrate suggests direct contact between Asp41 and the hydroxyl group of Ser [26]. Based on these reports, we can propose a plausible reaction mechanism for PSTK. The reaction begins when both substrates, Ser-$tRNA^{Sec}$ and Mg^{2+}-ATP, are bound to the enzyme. Asp41 then attracts a proton from the hydroxyl group of the Ser in the substrate. The deprotonated Ser nucleophilically attacks the γ-phosphate of the ATP yielding a pentavalent transition-state intermediate. Finally, the phosphodiester bond between the β and γ phosphate in the ATP is broken to complete the transphosphorylation reaction, and the product Sep-$tRNA^{Sec}$ is formed. Mg^{2+} aids in the departure of the leaving group ADP.

The enzyme activity of PSTK is strictly $tRNA^{Sec}$-dependent. PSTK does not hydrolyze ATP in the absence of tRNA nor in the presence of Ser-$tRNA^{Ser}$. The binding of $tRNA^{Ser}$, however, promotes ATP hydrolysis [12]. This suggests that $tRNA^{Sec}$ might play an essential role in positioning the Ser moiety for initiating phosphoryl transfer. Compared to aminoacyl-tRNA synthetases, PSTK has approximately 20-fold higher affinity toward its substrate, Ser-$tRNA^{Sec}$ (K_m = 40 nM) [12], which may compensate for the low abundance of $tRNA^{Sec}$ *in vivo*. The concentration of $tRNA^{Sec}$ *in vivo* is at least 10-fold lower than $tRNA^{Ser}$ in $tRNA^{Sec}$-rich tissues such as liver, kidney and tests in rat [27].

The crystal structure of PSTK reveals an unusual large central groove that is formed in the MjPSTK dimer interface, a feature not observed in other P-loop kinases previously. The central groove exposes the monomer active sites at either end, and contains many positively charged residues. Computational docking suggested that the central groove provides a complementary surface for the two $tRNA^{Sec}$ substrates. The model places archaeal PSTK identity elements (G2:C71 and the C3:G70 [25]) within contact of the protein dimer interface. Interestingly, the second base pair in the acceptor stem is highly conserved as C2:G71 in eukaryotic $tRNA^{Sec}$, and mutation of G2:C71 to C2:G71 in archaeal $tRNA^{Sec}$ resulted in a Ser-$tRNA^{Sec}$ variant that is phosphorylated inefficiently [25]. Moreover, the eukaryotic PSTK has been reported to recognize the unusual D-arm of $tRNA^{Sec}$ as the major identity element for phosphorylation [28], and phylogeny indicates a deep divide between the archaeal and eukaryotic PSTK proteins [12]. Given these observations, it is quite likely that the eukaryotic PSTK/tRNA complex may involve interactions that are distinct from its archaeal counterpart. Future co-crystal structures of $tRNA^{Sec}$ with the archaeal and eukaryotic PSTKs will provide a definitive picture of these differences.

SepSecS requires Sep-$tRNA^{Sec}$ as the precursor for Sec biosynthesis

The conversion of phosphoseryl-$tRNA^{Sec}$ (Sep-$tRNA^{Sec}$) to selenocysteinyl-$tRNA^{Sec}$ (Sec-$tRNA^{Sec}$) is the last step of Sec biosynthesis in both archaea and eukaryotes, and it is catalyzed by *O*-phosphoseryl-tRNA:selenocysteinyl-tRNA synthase (SepSecS). SepSecS forms its own branch in the phylogenetic tree of the fold type I family of the pyridoxal phosphate (PLP)-dependent enzymes and displays a distinct homotetrameric ($[\alpha_2]_2$) quaternary structure [11]. Two active sites are formed at each homodimer interface and each active site contains a PLP-binding pocket. Thus, the tetrameric SepSecS has four active sites and four PLP molecules bound. The PLP cofactor forms a reversible Schiff-base linkage (internal aldimine) with a highly conserved Lys284. The PLP-dependent enzymes form a functionally diverse group, which is responsible for more than 140 distinct activities [29]. Remarkably, almost all the PLP-

dependent enzymes utilize a catalytic mechanism that proceeds through a carbanion intermediate, which is, in turn, stabilized by PLP [30]. Based on both *in vitro* and *in vivo* activity assays, a catalytic mechanism of SepSecS that is dependent of PLP was recently proposed [11,18,31].

The tetrameric SepSecS is distinct from all other members of its family including *O*-phosphoseryl-tRNA:cysteinyl-tRNA synthase (SepCysS). SepCysS acts on a tRNA-based substrate (Sep-tRNA^{Cys}), but it does so as a dimer [11]. Moreover, SepSecS is a highly specific enzyme that acts on Sep-tRNA^{Sec}, and not on Sep-tRNA^{Cys}, Ser-tRNA^{Sec}, or free Sep. Crystal structures of both the archaeal and murine apo-SepSecS [11,31] and that of the human SepSecS-tRNA^{Sec} complex [18] have provided insights into the substrate specificity and the catalytic mechanism of SepSecS.

SepSecS binds both the unacylated tRNA^{Sec} and Sep-tRNA^{Sec} with comparable affinities *in vitro* [4], which presumably allowed crystallization of the human complex between SepSecS and unacylated tRNA^{Sec}. The SepSecS/tRNA^{Sec} complex structure revealed that one SepSecS homodimer interacts with the sugar-phosphate backbone of both the acceptor-TΨC and the variable arms of tRNA^{Sec}, while the other homodimer interacts specifically with the tip of the acceptor arm through interaction between the conserved Arg398 and the discriminator base G73 of tRNA^{Sec} [18]. Thus, one homodimer measures both the length of the acceptor-TΨC arm and the distance between the variable arm and the CCA end, whereas the other homodimer ensures that the correct tRNA is bound to the enzyme. The latter homodimer also provides the catalytic site that would act on the Sep moiety. Finally, structural modeling revealed that canonical tRNAs, including tRNA^{Ser} (which is the closest structural homolog of tRNA^{Sec}), cannot bind to SepSecS due to the shorter acceptor-TΨC arm. These observations explain why only the SepSecS tetramer can bind tRNA^{Sec} and why SepSecS cannot bind canonical tRNAs.

Unlike its bacterial counterpart SelA, SepSecS cannot directly convert Ser-tRNA^{Sec} to Sec-tRNA^{Sec} and has a significantly reduced affinity for Ser-tRNA^{Sec} *in vitro* [4]. This argues that the phosphate group of Sep plays an important role in substrate binding. Indeed, the crystal structure of SepSecS/tRNA^{Sec} complex in presence of free Sep shows extensive hydrogen bonding between the active site and the phosphate group of the free Sep [18]. This suggests that the phosphate group may serve as an anchor to bring Sep in close proximity of the cofactor PLP. However, the α-amino group of the free Sep appears quite flexible in the structure and it is not oriented properly for the nucleophilic attack on the Schiff base [18]. Thus, the covalent attachment of Sep to tRNA^{Sec} may restrain the conformation of the Sep moiety as well as facilitate the proper orientation of the α-amino group for efficient PLP-dependent catalysis to occur. This would explain why SepSecS could not act on either free Sep or Ser-tRNA^{Sec}, as both molecules bear only one anchoring moiety.

Sec biosynthesis in all domains of life is achieved by RNA-dependent amino acid modification. As such it is reminiscent of other pathways of tRNA-dependent amino acid synthesis that form Asn-tRNA, Gln-tRNA and Cys-tRNA in many organisms [32]. The individual reactions in this stepwise process may actually be carried out in a large tRNA:protein complex. One example is the transamidosome [33], a complex of tRNA^{Asn}, aspartyl-tRNA synthetase, and Asp-tRNA^{Asn} amidotransferase, that provides an efficient route to bacterial Asn-tRNA^{Asn} formation. The complex architecture should also aid protein quality control, as it may prevent release of mis-acylated aminoacyl-tRNA that would be used in protein synthesis. A similar complex possibly involving tRNA^{Sec}, SerRS, PSTK and SepSecS may exist to facilitate efficient selenocysteine formation.

Pyrrolysine, the uncommon 22nd amino acid

Evidence for the genetic encoding of pyrrolysine (Pyl) is only present in 1% of all sequenced genomes thus far. The rarity of Pyl is matched by its unique chemical structure, a lysine in N^{ε}-linkage ot a pyrroline ring. Pyl was first observed in a crystal structure at the active site of the monomethylamine methyltransferase (MtmB1) from the methanogenic archaeon *Methanosarcina barkeri* [34]. Pyl has now been characterized in three distinct methylamine methyltransferases from the *Methanosarcinaceae* that allow these methanogens to utilize the unique growth substrates mono-, di- and trimethylamines [35]. Although each of the genes encoding these methyltransferases contains a in-frame amber (TAG) codon (a stop condon in most organisms) the methylamine methyltransferase genes are read-through to the next opal or ochre stop codon and Pyl is inserted in-response to the in-frame amber codon. Due to its proximity to the active site, Pyl was hypothesized to be a key catalytic residue in these Pyl-proteins [36].

While some work has attempted to address the mechanism of stop codon suppression in Pyl-decoding organisms [37-39], it remains unclear whether all amber codons in these organisms encode Pyl or whether some are translational stop signals. In recent work [40], Pyl was also found in the $tRNA^{His}$ guanylyltransferase (Thg1) from *Methanosarcina acetivorans.* Thg1 is responsible for addition of the key identity element for His-tRNA formation, a G-1 residue ligated to the 5′ end of immature $tRNA^{His}$. Interestingly, Pyl does not play a role in the catalytic action of Thg1 and represents the first example of a dispensable Pyl residue. Efficient read-through of the in-frame amber codon in Thg1, without the apparent presence of re-coding signals, indicates that Pyl insertion may be more similar to natural stop codon suppression [41] than to the more elaborate re-coding of UGA codons that is known to occur for selenocysteine [9].

Previous reviews focused on how Pyl enters the genetic code [42,43], namely by the catalytic action of a pyrrolysyl-tRNA synthetase (PylRS) that ligates pyrrolysine to the amber codon decoding $tRNA^{Pyl}$ [44,45]. Since that time the most dramatic advances have been the structural characterization of PylRS and $tRNA^{Pyl}$. This work, which is reviewed in detail below, revealed the unique nature of the Pyl binding pocket and provided atomic detail for the PylRS/$tRNA^{Pyl}$ interaction which underlies the specific and exclusive association between the protein and tRNA. This section of the review concludes with an overview of an expanded number of possible Pyl-decoding organisms uncovered from recently sequenced genomes and metagenomic surveys.

Recognition of Pyl by PylRS

While PylRS was first crystallized in 2006 [46], the first structure was solved independently the following year [47]. This structure of PylRS was determined in complex with Pyl and the Pyl analogue N-*ε*-[(cyclopentyloxy)carbonyl]-L-lysine (Cyc). As for all PylRS structures reported to date the highly insoluble N-terminal domain was not included (annotated as pylSn below and in Fig. 3). This and other structures reported subsequently describe the unique organization of the binding pocket for the large Pyl side chain, thus elucidating the principles governing Pyl recognition by PylRS [47,48]. The core catalytic domain contains the typical features of the class II aaRS family, and structural phylogeny suggested the divergence of PylRS from an ancestral version of PheRS early in evolution, i.e., during the time of the last common ancestor of all life on earth [47].

The structure also revealed that hydrophobic interactions account for most of the contact surface between PylRS and the Pyl or Cyc substrates, but specific recognition is derived from hydrogen bonds formed between PylRS and the amino acid substrate (Fig. 3A). The interactions of two residues (Asn346 and Arg330 in *Methanosarcina mazei* numbering) with

the primary and secondary carbonyl of the substrate were noticed in complexes with Pyl or Cyc. Although the cyclic components of both substrates were bound in a roughly equivalent position in the hydrophobic pocket of PylRS, subtle differences in the interactions are likely responsible for a significantly greater charging efficiency of $tRNA^{Pyl}$ with Pyl as opposed to Cyc [45,49]. The residue Tyr384 in PylRS is believed to play an important role in orienting Pyl into the binding pocket through its hydrogen bond to the pyrrole ring nitogen. Tyr384 is located on a mobile loop that only appears ordered in the presence of the cognate substrate Pyl [47,48]. A report of the first bacterial PylRS crystal structure revealed a somewhat more sterically constrained active site that can nevertheless accommodate Pyl [50]. The large active site of PylRS has been shown to accommodate several Pyl analogs [51-53], including some that can be chemically altered post-translationaly using “click chemistry” to specifically attach fluorescent probes [54,55].

Molecular basis for the PylRS:$tRNA^{Pyl}$ orthogonality

As is typical for class II aaRSs, PylRS is a dimer of two identical subunits, each of which bind one tRNA molecule from the major groove side. Each $tRNA^{Pyl}$ interacts with one subunit and makes few specific contacts with the second protomer. Since PylRS and $tRNA^{Pyl}$ display no significant cross reactivity with other aaRS/tRNA pairs in native or heterologous contexts [38,56,57], they are said to be orthogonal to other aminoacylation systems. Several unique structural features in both the protein and its substrate RNA have emerged in the course of evolution that explain the specificity of PylRS for its tRNA (Fig. 3B).

In both archaea and bacteria, the amber suppressor $tRNA^{Pyl}$ has several unique structural features when compared to canonical tRNAs. These include an anticodon stem of 6 base pairs instead of 5, a short variable loop of only 3 bases, a single base separating the acceptor and D-steams, a small D-loop with only 5 bases, the absence of the almost universally conserved G18G19 sequence in the D-loop and the T54Ψ55C56 sequence in the T-Loop, and a small number of post-transcriptional base modifications. *M. barkeri* $tRNA^{Pyl}$ was found to have only a 4-thiouridine at position 8 and a 1-methyl-pseudouridine at position 50 [45].

Crystallization of the *Desulfitobacterium hafniense* PylRS (DhPylRS) in complex with its cognate tRNA provided the first view of a naturally evoloved orthogonal aaRS/tRNA pair. The structure supports what was predicted from sequence, that PylRS/$tRNA^{Pyl}$, unlike engineered orthogonal pairs, is highly distinct from even the most closely related aaRS/tRNA pair [57] pair [57]. The struture showed that while the unusual features of $tRNA^{Pyl}$ do not prevent the formation of the expected L-shape tertiary structure fold, they result in a unique core more compact than those seen in canonical tRNAs. To match the structural peculiarity of its tRNA substrate, PylRS has evolved unique protein domains from which residues emerge and engage in specific interactions with $tRNA^{Pyl}$ core nucleotides. The tRNA core comes in contact with one side of the protein hence termed core binding surface. This surface groups the tRNA binding domain 1 (N-terminus of pylSc domain), the C terminal tail and helix α6 on the second protomer (Fig. 3B, upper).

Structural data identified that tRNA core nucleotides involved in specific interactions with the protein. Nucleotide G9 and D-stem nucleotide pairs G10:C25 and A11:U24 are engaged in a series of hydrogen bonding interaction with PylRS amino acid side chains. Earlier biochemical evidence had indicated the importance of these particular nucleotides for $tRNA^{Pyl}$ aminoacylation [49,58]. Their mutation resulted in dramatic decrease in aminoacylation efficiency both *in vitro* and *in vivo.* It was, however, unclear whether the deficient aminoacylation resultedfrom destabilization of the tRNA structure or the direct disruption of specific contacts with PylRS. Structural data now unambiguously identify these nucleotides as key elements contributing to the orthogonality of PylRS/$tRNA^{Pyl}$.

The tRNA binding domain 1, C-terminal tail and bulge domain of the opposite subunit form a U shaped concave structure complementary to the acceptor helix of the tRNA that directs its 3′ terminus toward the PylRS catalytic site (Fig. 3B, lower). The structural data also show how PylRS selectively recognizes, through a network of specific hydrogen bonding interactions, the discriminator base and the G1:C72 base pair at the top of the acceptor stem. Compelling biochemical evidence demonstrating the critical importance of the same nucleotides for the aminoacylation of $tRNA^{Pyl}$ with pyrrolysine *in vitro* and *in vivo* were also obtained [49,58]. The discriminator base G73 and the G1:C72 base pair along with those of the core domain define the complete $tRNA^{Pyl}$ identity set. Transfer of these nucleotides is able to convert a catalytically non-competent tRNA substrate into a pyrrolysine accepting tRNA [49]. Although biochemical studies have shown that $tRNA^{Pyl}$ anticodon nucleotides are dispensable for the formation of PylRS/$tRNA^{Pyl}$ complex to tRNA aminoacylation, the two bases adjacent to the anticodon (U33 and A37) were shown to be identity elements for the *M. mazei* PyLRS [49]. These two bases are possibly recognized in a sequence-specific manner by residues from the N-terminal domain, a domain lacking in the DhPylRS and all other PylRS crystal structures.

The strict orthogonality of the PylRS/$tRNA^{Pyl}$ system, which has been demonstrated [57], and available crystal structures have already been exploited for the genetic encoding, and thus site-specific incorporation of typically post-translationally modified amino acids directly into proteins. PylRS variants were selected to generate a mutant that specifically charges $tRNA^{Pyl}$ with N-acetyl-lysine [59]. Wild type PylRS was employed to aminoacylate $tRNA^{Pyl}$ with a reactive Pyl analog which was chemically converted to the common histone modification N-methyl-lysine [60].

Expanding the Pyl-decoding biosphere

Data from newly sequenced genomes and metagenomic projects has doubled the potential number of Pyl-decoding organisms. The Pyl-decoding trait is associated with the presence of the Pyl operon, which includes genes for the $tRNA^{Pyl}$ (*pylT*) and the PylRS (*pylS*) followed by genes that encode the putative Pyl biosynthetic machinery (*PylBCD*). The Pyl operon is sufficient to encode the Pyl-decoding trait as shown by transformation of the operon into a heterlogous *E. coli* context that then supported translational read-through of amber codons [38,56]. Earlier reports [37] found the Pyl-decoding trait confined to 5 archael organisms from the *Methanosarcinaceae* family and one bacterial organism (*D. hafniense*), while later studies [47,58,61] identified components of the Pyl operon in a δ-proteobacterial endosymbiont of the marine worm *Olavius algarvensis*, based on sequence data from a metagenomic survey [62].

There are now four additional examples of the bacterial version of the Pyl-trait, and one additional *Methanosarcinaceae* (*Methanohalophilus mahii*) is shown to have the complete Pyl operon (Fig. 3C). For reasons unknown, only *M. barkeri* among the archaea appears to deviate from the otherwise strict operonal organization and includes a small (100 codon) ORF that is a mutated duplication of the 3′ end of *pylC*. The bacterial examples show greater diversity in Pyl operon structure. In keeping with this observation, an updated phylogeny of the PylRS sequences continues to show a clear divide between the bacterial and archaeal type PylRSs, with a statistically significant root placed between the bacterial and archaeal clades according to an alignment with other subclass IIC aaRSs. While the archaeal Pyl-decoding trait is still confined to the *Methanosarcinaceae*, bacteria examples appear in only a few members of the phyla Clostridia and δ-proteobacteria.

A key difference between the archaeal and bacterial PylRS is that the bacteria encoded the N-terminal domain (about 100 amino acids) of PylRS as a separate gene, often at the 3′ end of the operon, leading to the organization (*pylT*, *pylSc*, *pylBCD*, *pylSn*). While the evolutionary scenario for the *pylS* gene split and operonal re-arrangement was previously unclear, the newly

sequenced and deepest branching bacterial Pyl-decoding representative *Acetohalobium arabaticum*, appears to display a transitional form. In this organism, *pylSn* terminates 11 bases upstream of and is thus out-of-frame with *pylSc*. The phylogeny indicates that the split *pylS* evolved after the divergence of the archaeal and bacterial *pylS* genes and that the re-arrangement (placement of pylSn at the 3′ end of the operon) occurred later in bacterial evolution.

Only two organisms interrupt the Pyl operon with other genes. *A. arabaticum* encodes the Pyl-containing trimethylamine methyltransferase (MttB) and cognate corrinoid protein (MttC) in the operon as well as a gene related to N-methylhydantoinase (hyuA) and 5-oxoprolinase between *pylB* and *pylC*. The gene has close homologs in most Pyl-decoding organisms and various bacteria. Proteins from this superfamily (hyuA) are typically involved in the breakdown of cyclic amines (similar to the pyrroline ring of Pyl), and this gene could be the first example of a catabolic Pyl enzyme. *Desfulotomaculum acetoxidans* encodes an unknown iron-sulfur protein between *pylT* and *pylSc* and an exact copy of the 3′ 30 bases of $tRNA^{Pyl}$ downstream of the operon, which is an indication that the Pyl operon is on a mobile genetic element [63]. Finally, the complete genome of *Desulfobacterium autotrophicum* contains the first example of a degraded Pyl operon, only *pylSc* and *pylB* are present. Gene loss has been assumed to explain in part the sparsity of the Pyl-operon, and this example provides the first evidence of a transitional form towards loss of the Pyl-decoding trait.

The 21st and 22nd amino acids promise exciting new directions

The enzymes involved in the archaeal and eukaryotic Sec biosynthesis pathways achieve specificity for $tRNA^{Sec}$ by recognizing distinct structural features of the tRNA substrate. While a similar theme is observed in the Pyl-decoding system, Sec and Pyl followed different evolutionary and metabolic routes to enter the genetic code. Sec relied on a specialized tRNA that is recognized by a canonical aaRS and on the evolution of enzymes that specifically modify the aminoacyl-tRNA to ultimately form Sec-$tRNA^{Sec}$. In contrast, Pyl entered the code as a result of the establishment of a tRNA-independent biosynthetic pathway and co-evolution of a aaRS/tRNA pair that is quite different from its closest known relatives. Despite their differences, both systems will be important for future biotechnological applications. Just as detailed biochemical and structural investigation of the PylRS:$tRNA^{Pyl}$ system facilitated the engineering of this orthogonal pair to incorporate other non-canonical amino acids, the archaeal/eukaryotic Sec-$tRNA^{Sec}$ synthesis machinery, specifically the intermediate Sep-$tRNA^{Sec}$, provides another possible means to incorporate phosphoserine or possibly other phosphorylated amino acids directly into a growing polypeptide chain.

Acknowledgments

We thank Ilka Heinemann for advice and Lennart Randau for suggestions and help locating bacterial *pylT* genes. Work in the authors' laboratory was supported by grants (to D.S.) from the Department of Energy, the National Science Foundation, and the National Institute for General Medical Sciences.

References

[1]. Leinfelder W, Zehelein E, Mandrand-Berthelot MA, Böck A. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature 1988;331:723–725. [PubMed: 2963963]

[2]. Yuan J, Palioura S, Salazar JC, Su D, O'Donoghue P, Hohn MJ, Cardoso AM, Whitman WB, Söll D. RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. Proc Natl Acad Sci USA 2006;103:18923–18927. [PubMed: 17142313]

[3]. Abe K, Mihara H, Tobe R, Esaki N. Characterization of human selenocysteine synthase involved in selenoprotein biosynthesis. Biomed Res Trace Elements 2007;19:80–83.

[4]. Xu XM, Carlson BA, Mix H, Zhang Y, Saira K, Glass RS, Berry MJ, Gladyshev VN, Hatfield DL. Biosynthesis of selenocysteine on its tRNA in eukaryotes. PLoS Biol 2007;5:e4. [PubMed: 17194211]

[5]. Kryukov GV, Castellano S, Novoselov SV, Lobanov AV, Zehtab O, Guigo R, Gladyshev VN. Characterization of mammalian selenoproteomes. Science 2003;300:1439–1443. [PubMed: 12775843]

[6]. Bellinger FP, Raman AV, Reeves MA, Berry MJ. Regulation and function of selenoproteins in human disease. Biochem J 2009;422:11–22. [PubMed: 19627257]

[7]. Ehrenreich A, Forchhammer K, Tormay P, Veprek B, Böck A. Selenoprotein synthesis in *E. coli*. Purification and characterisation of the enzyme catalysing selenium activation. Eur J Biochem 1992;206:767–773. [PubMed: 1606960]

[8]. Böck, A.; Thanbichler, M.; Rother, M.; Resch, A. Selenocysteine. In: Ibba, M.; Francklyn, CS.; Cusack, S., editors. The Aminoacyl-tRNA Synthetases. Landes Bioscience; Georgetown, TX: 2005. p. 320-327.

[9]. Yoshizawa S, Böck A. The many levels of control on bacterial selenoprotein synthesis. Biochim Biophys Acta 2009;1790:1404–1414. [PubMed: 19328835]

[10]. Carlson BA, Xu XM, Kryukov GV, Rao M, Berry MJ, Gladyshev VN, Hatfield DL. Identification and characterization of phosphoseryl-tRNA$^{[Ser]Sec}$ kinase. Proc Natl Acad Sci USA 2004;101:12848–12853. [PubMed: 15317934]

[11]. Araiso Y, Palioura S, Ishitani R, Sherrer RL, O'Donoghue P, Yuan J, Oshikane H, Domae N, Defranco J, Söll D, Nureki O. Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. Nucleic Acids Res 2008;36:1187–1199. [PubMed: 18158303]

[12]. Sherrer RL, O'Donoghue P, Söll D. Characterization and evolutionary history of an archael kinase involved in selenocysteinyl-tRNA formation. Nucleic Acids Res 2008;36:1247–1259. [PubMed: 18174226]

[13]. Su D, Hohn MJ, Palioura S, Sherrer RL, Yuan J, Söll D, O'Donoghue P. How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. IUBMB Life 2009;61:35–39. [PubMed: 18798524]

[14]. Sauerwald A, Zhu W, Major TA, Roy H, Palioura S, Jahn D, Whitman WB, Yates JR 3rd, Ibba M, Söll D. RNA-dependent cysteine biosynthesis in archaea. Science 2005;307:1969–1972. [PubMed: 15790858]

[15]. Burkard U, Söll D. The unusually long amino acid acceptor stem of *Escherichia coli* selenocysteine tRNA results from abnormal clevage by RNase P. Nucleic Acids Res 1988;16:11617–11624. [PubMed: 3062578]

[16]. Baron C, Westhof E, Böck A, Giege R. Solution structure of selenocysteine-inserting tRNA^{Sec} from *Escherichia coli*. Comparsion with canonical tRNA^{Ser}. J Mol Biol 1993;231:274–292. [PubMed: 8510147]

[17]. Sturchler C, Westhof E, Carbon P, Krol A. Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA^{Sec}. Nucleic Acids Res 1993;21:1073–1079. [PubMed: 8464694]

[18]. Palioura S, Sherrer RL, Steitz TA, Söll D, Simonović M. The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. Science 2009;325:321–325. [PubMed: 19608919]

[19]. Itoh Y, Chiba S, Sekine SI, Yokoyama S. Crystal structure of human selenocysteine tRNA. Nucleic Acids Res 2009;37:6259–6268. [PubMed: 19692584]

[20]. Westhof E, Dumas P, Moras D. Crystallographic refinement of yeast aspartic acid transfer RNA. J Mol Biol 1985;184:119–145. [PubMed: 3897553]

[21]. Caldeira de Araujo A, Favre A. Near ultraviolet DNA damage induces the SOS responses in *Escherichia coli*. EMBO J 1986;5:175–179. [PubMed: 3514208]

[22]. Mäenpää PH, Bernfield MR. A specific hepatic transfer RNA for phosphoserine. Proc Natl Acad Sci USA 1970;67:688–695. [PubMed: 4943179]

[23]. Sharp SJ, Stewart TS. The characterization of phosphoseryl tRNA from lactating bovine mammary gland. Nucleic Acids Res 1977;4:2123–2136. [PubMed: 242796]

[24]. Kaiser JT, Gromadski K, Rother M, Engelhardt H, Rodnina MV, Wahl MC. Structural and functional investigation of a putative archaeal selenocysteine synthase. Biochemistry 2005;44:13315–13327. [PubMed: 16201757]

[25]. Sherrer RL, Ho JML, Söll D. Divergence of selenocysteine tRNA recognition by archaeal and eukaryotic *O*-phosphoseryl-tRNA^{Sec} kinase. Nucleic Acids Res 2008;36:1871–1880. [PubMed: 18267971]

[26]. Araiso Y, Sherrer RL, Ishitani R, Ho JML, Söll D, Nureki O. Structure of a tRNA-dependent kinase essential for selenocysteine decoding. Proc Natl Acad Sci USA 2009;106:13.

[27]. Diamond AM, Choi IS, Crain PF, Hashizume T, Pomerantz SC, Cruz R, Steer CJ, Hill KE, Burk RF, McCloskey JA, Hatfield DL. Dietary selenium affects mehtylation of the wobble nucleoside in the anticodon of selenocysteine tRNA$^{[Ser]Sec}$. J Biol Chem 1993;268:14215–14223. [PubMed: 8314785]

[28]. Wu XQ, Gross HJ. The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. EMBO J 1994;13:241–248. [PubMed: 8306966]

[29]. Percudani R, Peracchi A. A genomic overview of pyridoxal-phosphate-dependent enzymes. EMBO Rep 2003;4:850–854. [PubMed: 12949584]

[30]. Eliot AC, Kirsch JF. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. Annu Rev Biochem 2004;73:383–415. [PubMed: 15189147]

[31]. Ganichkin OM, Xu XM, Carlson BA, Mix H, Hatfield DL, Gladyshev VN, Wahl MC. Structure and catalytic mechanism of eukaryotic selenocysteine synthase. J Biol Chem 2007;283:5849–5865. [PubMed: 18093968]

[32]. Yuan J, Sheppard K, Söll D. Amino acid modifications on tRNA. Acta Biochim Biophys Sin (Shanghai) 2008;40:539–553. [PubMed: 18604446]

[33]. Bailly M, Blaise M, Lorber B, Becker HD, Kern D. The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. Mol Cell 2007;28:228–239. [PubMed: 17964262]

[34]. Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science 2002;296:1462–1466. [PubMed: 12029132]

[35]. Soares JA, Zhang L, Pitsch RL, Kleinholz NM, Jones RB, Wolff JJ, Amster J, Green-Church KB, Krzycki JA. The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. J Biol Chem 2005;280:36962–36969. [PubMed: 16096277]

[36]. Krzycki JA. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. Curr Opin Chem Biol 2004;8:484–491. [PubMed: 15450490]

[37]. Zhang Y, Baranov PV, Atkins JF, Gladyshev VN. Pyrrolysine and selenocysteine use dissimilar decoding strategies. J Biol Chem 2005;280:20740–20751. [PubMed: 15788401]

[38]. Namy O, Zhou Y, Gundllapalli S, Polycarpo CR, Denise A, Rousset JP, Söll D, Ambrogelly A. Adding pyrrolysine to the *Escherichia coli* genetic code. FEBS Lett 2007;581:5282–5288. [PubMed: 17967457]

[39]. Longstaff DG, Blight SK, Zhang L, Green-Church KB, Krzycki JA. *In vivo* contextual requirements for UAG translation as pyrrolysine. Mol Microbiol 2007;63:229–241. [PubMed: 17140411]

[40]. Heinemann IU, O'Donoghue P, Madinger C, Benner J, Randau L, Noren CJ, Söll D. The appearance of pyrrolysine in tRNA^{His} guanylyltransferase by neutral evolution. Proc Natl Acad Sci USA. 2009 In press.

[41]. Beier H, Barciszewska M, Krupp G, Mitnacht R, Gross HJ. UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs with suppressor activity from tobacco plants. EMBO J 1984;3:351–356. [PubMed: 16453503]

[42]. Ambrogelly A, Palioura S, Söll D. Natural expansion of the genetic code. Nat Chem Biol 2007;3:29–35. [PubMed: 17173027]

[43]. Krzycki JA. The direct genetic encoding of pyrrolysine. Curr Opin Microbiol 2005;8:706–712. [PubMed: 16256420]

[44]. Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G, Kang PT, Green-Church KB, Chan MK, Krzycki JA. Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vitro* and *in vivo*. Nature 2004;431:333–335. [PubMed: 15329732]

[45]. Polycarpo C, Ambrogelly A, Berube A, Winbush SM, McCloskey JA, Crain PF, Wood JL, Söll D. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proc Natl Acad Sci USA 2004;101:12450–12454. [PubMed: 15314242]

[46]. Yanagisawa T, Ishii R, Fukunaga R, Nureki O, Yokoyama S. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*. Acta Crystallogr Sect F Struct Biol Cryst Commun 2006;62:1031–1033.

[47]. Kavran JM, Gundllapalli S, O'Donoghue P, Englert M, Söll D, Steitz TA. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc Natl Acad Sci USA 2007;104:11268–11273. [PubMed: 17592110]

[48]. Yanagisawa T, Ishii R, Fukunaga R, Kobayashi T, Sakamoto K, Yokoyama S. Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. J Mol Biol 2008;378:634–652. [PubMed: 18387634]

[49]. Ambrogelly A, Gundllapalli S, Herring S, Polycarpo C, Frauer C, Söll D. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci U S A 2007;104:3141–3146. [PubMed: 17360621]

[50]. Lee MM, Jiang R, Jain R, Larue RC, Krzycki J, Chan MK. Structure of *Desulfitobacterium hafniense* PylSc, a pyrrolysyl-tRNA synthetase. Biochem Biophys Res Commun 2008;374:470–474. [PubMed: 18656445]

[51]. Kobayashi T, Yanagisawa T, Sakamoto K, Yokoyama S. Recognition of non-alpha-amino substrates by pyrrolysyl-tRNA synthetase. J Mol Biol 2009;385:1352–1360. [PubMed: 19100747]

[52]. Li WT, Mahapatra A, Longstaff DG, Bechtel J, Zhao G, Kang PT, Chan MK, Krzycki JA. Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. J Mol Biol 2009;385:1156–1164. [PubMed: 19063902]

[53]. Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K, Yokoyama S. Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun 2008;371:818–822. [PubMed: 18471995]

[54]. Fekner T, Li X, Lee MM, Chan MK. A pyrrolysine analogue for protein click chemistry. Angew Chem Int Ed Engl 2009;48:1633–1635. [PubMed: 19156778]

[55]. Nguyen DP, Lusic H, Neumann H, Kapadnis PB, Deiters A, Chin JW. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/$tRNA_{CUA}$ pair and click chemistry. J Am Chem Soc 2009;131:8720–8721. [PubMed: 19514718]

[56]. Longstaff DG, Larue RC, Faust JE, Mahapatra A, Zhang L, Green-Church KB, Krzycki JA. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. Proc Natl Acad Sci USA 2007;104:1021–1026. [PubMed: 17204561]

[57]. Nozawa K, O'Donoghue P, Gundllapalli S, Araiso Y, Ishitani R, Umehara T, Söll D, Nureki O. Pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}$ structure reveals the molecular basis of orthogonality. Nature 2009;457:1163–1167. [PubMed: 19118381]

[58]. Herring S, Ambrogelly A, Polycarpo CR, Söll D. Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. Nucleic Acids Res 2007;35:1270–1278. [PubMed: 17267409]

[59]. Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N^{ε}-acetyllysine in recombinant proteins. Nat Chem Biol 2008;4:232–234. [PubMed: 18278036]

[60]. Nguyen DP, Alai MM, Kapadnis PB, Neumann H, Chin JW. Genetically encoding N^{ε}-methyl-L-lysine in recombinant histones. J Am Chem Soc 2009;131:14194–14195. [PubMed: 19772323]

[61]. Zhang Y, Gladyshev VN. High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*. Nucleic Acids Res 2007;35:4952–4963. [PubMed: 17626042]

[62]. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland

C, Rubin EM, Dubilier N. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature 2006;443:950–955. [PubMed: 16980956]

[63]. Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol 2000;54:641–679. [PubMed: 11018140]

[64]. Itoh Y, Sekine S, Kuroishi C, Terada T, Shirouzu M, Kuramitsu S, Yokoyama S. Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon *Pyrococcus horikoshii*. RNA Biol 2008;5:169–177. [PubMed: 18818520]

[65]. Markowitz VM, Kyrpides NC. Comparative genome analysis in the integrated microbial genomes (IMG) system. Methods Mol Biol 2007;395:35–56. [PubMed: 17993666]

[66]. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52:696–704. [PubMed: 14530136]

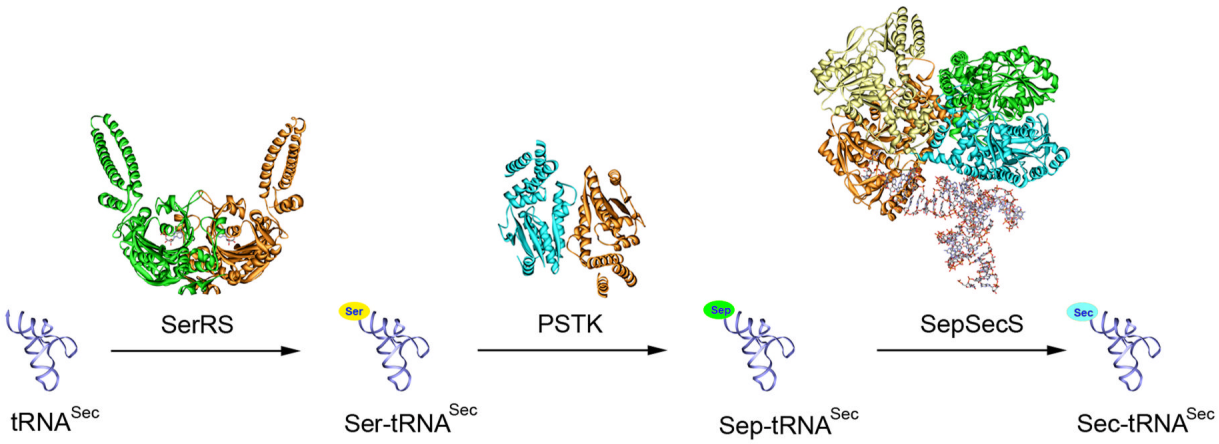

Figure 1.

The Sec biosynthesis pathway in archaea and eukaryotes. Sec in synthesized on $tRNA^{Sec}$ in three steps. 1. The unacylated $tRNA^{Sec}$ is serylated by SerRS; 2. the resulting Ser-$tRNA^{Sec}$ is phosphorylated by PSTK forming Sep-$tRNA^{Sec}$; 3. the phosphorylated intermediate is converted to the final product Sec-$tRNA^{Sec}$ by SepSecS. The crystal structures of tRNA and enzymes in this pathway are presented: $tRNA^{Sec}$ (ribbon) from *Homo sapiens* [19], SerRS (ribbon) with AMP (stick) from *Pyrococcus horikoshii* [64], PSTK (ribbon) from *M. jannaschii* [26], and SepSecS (ribbon) with $tRNA^{Sec}$ (stick, only one tRNA is shown) from *Homo sapiens* [18].

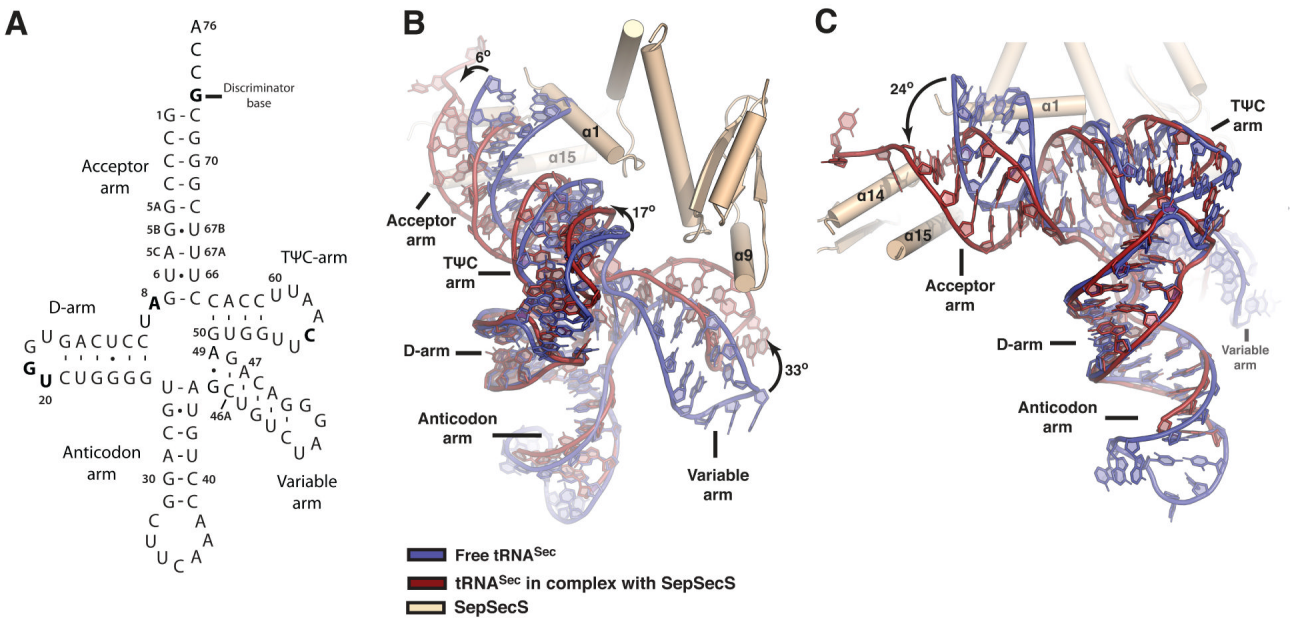

Figure 2.

Binding to SepSecS promotes a conformational change in $tRNA^{Sec}$. (A) The secondary structure of human $tRNA^{Sec}$ with bases mentioned in the text highlighted in bold. (B) Superposition of the sugar-phosphate backbone of both the D and anticodon arms of free $tRNA^{Sec}$ (blue) on the corresponding atoms of $tRNA^{Sec}$ complexed with SepSecS (red) reveals a conformational change in the tRNA molecule on binding to the enzyme. The variable arm of $tRNA^{Sec}$ rotates by 33°, the T arm swings 17° and the acceptor arm rotates by 6° around the axis that runs parallel to both the D and anticodon arms. The free $tRNA^{Sec}$ conformer cannot bind to SepSecS because the tip of its acceptor arm would clash with the helix α1 (arrow) and the variable arm would be positioned away from the helix α9. (C) The acceptor arm also slides down towards the anticodon arm on binding to SepSecS through a 24° rotation around the axis that is parallel to the variable arm. This movement positions G73 to interact with Arg398 in the helix α14 and orients the CCA end toward the active site. Free $tRNA^{Sec}$ is blue, $tRNA^{Sec}$ complexed with SepSecS is red and SepSecS is beige. Only the secondary structure elements of SepSecS that interact with tRNA are shown. The view is first rotated ~30° clockwise around the horizontal axis, and then another 30° anticlockwise around the vertical axis relative to panel B.

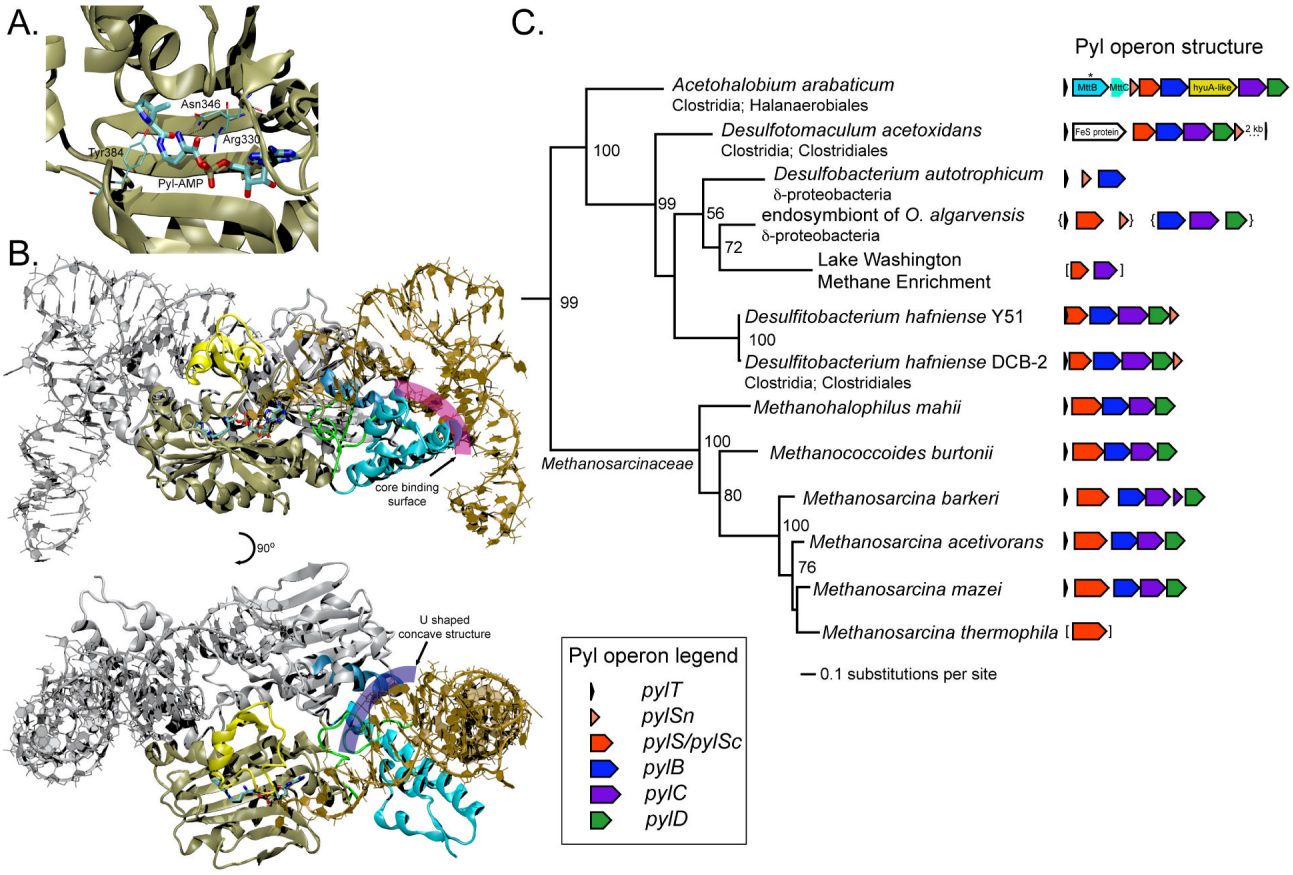

Figure 3.

Structure and evolution of PylRS and the Pyl operon. (A) Key hydrogen-bond interactions in the active site of *M. mazei* PylRS with pyrrolysyl-adenylate (Pyl-AMP) bound. (B) Two views of the *D. hafniense* PylRS/tRNA^{Pyl} complex showing the tRNA binding domain 1 (cyan), the catalytic core domain (tan), the bulge domain (yellow), the C-terminal tail domain (green), and α-helix 6 (blue) emerging from the other protomer (gray). The tRNA core binding surface (magenta highlight) and U shaped concave surface (purple highlight) are also illustrated. Pyl-AMP from the *M. mazei* PylRS structure was superimposed onto the DhPylRS structure for reference. (C) The PylRS maximum likelihood phylogenetic tree includes all known putative Pyl-decoding organisms to date, and the organization of the Pyl operon in each genome sequence is shown adjacent to the tree. Square brackets indicate partial sequence data from a short sequence read, and curly brackets indicate that the Pyl operon is split between different contigs from a metagenomic survey. * indicates an in-frame TAG codon. Sequence data were downloaded from the Integrated Microbial Genomes database [65]. The tree was calculated with PHYML [66] using a BioNJ starting tree and SPR tree search followed by NNI branch swapping to optimize the tree. All alignment positions were considered, likelihood parameters were estimated from the alignment, and the JTT+Γ model with 8 rate categories was applied. Bootstrap values (only those > 50 are shown) were computed according to the Shimodaira-Hasegawa re-estimation of log-likelihood test implemented in PHYML.