Emerging Topics in Life Sciences (2018)
https://doi.org/10.1042/ETLS20180094

Review Article

Pyrrolysine in archaea: a 22nd amino acid encoded through a genetic code expansion

Jean-François Brugère[1,2], John F. Atkins[3,4], Paul W. O'Toole[2] and Guillaume Borrel[5]

[1]Université Clermont Auvergne, Clermont-Ferrand F-63000, France; [2]School of Microbiology and APC Microbiome Institute, University College Cork, Cork, Ireland; [3]School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland; [4]Department of Human Genetics, University of Utah, Salt Lake City, UT, U.S.A.; [5]Evolutionary Biology of the Microbial Cell, Department of Microbiology, Institut Pasteur, Paris, France

Correspondence: Jean-François Brugère (j-francois.brugere@uca.fr)

The 22nd amino acid discovered to be directly encoded, pyrrolysine, is specified by UAG. Until recently, pyrrolysine was only known to be present in archaea from a methanogenic lineage (*Methanosarcinales*), where it is important in enzymes catalysing anoxic methylamines metabolism, and a few anaerobic bacteria. Relatively new discoveries have revealed wider presence in archaea, deepened functional understanding, shown remarkable carbon source-dependent expression of expanded decoding and extended exploitation of the pyrrolysine machinery for synthetic code expansion. At the same time, other studies have shown the presence of pyrrolysine-containing archaea in the human gut and this has prompted health considerations. The article reviews our knowledge of this fascinating exception to the 'standard' genetic code.

Introduction

It is now more than half a century ago since the genetic code was deciphered [1,2]. The understanding gained of how just four different nucleobases could specify the 20 universal amino acids to synthesise effective proteins reveals a fundamental feature of extant life. It is a remarkable feat of information transfer. Correct aminoacylation, supported by aminoacyl-tRNA synthetases (aaRS), and highly discriminatory selection of tRNA whose anticodon is complementary to the codon are essential for faithful decoding.

Deviations from the genetic code known in most organisms have been gathered under two different types termed 'reassignment' and 'recoding' [3]. In specialised niches, such as certain mitochondria [4,5], the meaning of particular codons is reassigned such that wherever that codon occurs it has the new meaning, and only the new meaning. Many but not all of these reassignments involve UAG or UGA encoding a specific one of the 20 universally encoded amino acids rather than signifying translation termination. This is distinct from context-dependent codon redefinition where the new meaning only applies at particular occurrences of the codon. Such redefinition is dynamic with specification of the new meaning being in competition with the standard meaning so that only a proportion of the product reflects the new meaning. It is a type of recoding. When dynamic redefinition occurs at UAG or UGA, which are by far the most frequent codons at which it occurs, one, or several different, of the 20 universal amino acids can be specified. Such occurrences are often termed stop codon readthrough as the identity of the amino acid specified is generally unimportant and functional significance derives from a proportion of the product having a C-terminal extension.

In addition to the 20 amino acids specified in all organisms, two additional amino acids, pyrrolysine and selenocysteine, are directly encoded by some organisms but not by others. In the organisms that encode selenocysteine, which includes several archaea [6], such specification is, with few though interesting exceptions [7], by UGA [8,9] and is mRNA context-dependent. With just one exception (the 3′-region of selenoprotein P mRNA), even when such specification does occur, it is in competition with the release factor-mediated termination, i.e. the redefinition is dynamic. In contrast, when the

Received: 18 July 2018
Revised: 24 September 2018
Accepted: 25 September 2018

Version of Record published:
22 November 2018

Emerging Topics in Life Sciences (2018)
https://doi.org/10.1042/ETLS20180094

22nd directly encoded amino acid, pyrrolysine, was discovered by Krzycki and colleagues [10,11] to be encoded by UAG in an archaeon, an early question was whether or not all ribosomes decoded all UAG codons in that organism as pyrrolysine, i.e. whether UAG was reassigned to specify pyrrolysine. One alternative considered was that all UAG codons present were sometimes decoded as either pyrrolysine or termination, and another was that a proportion of UAG codons were more prone than others to specify pyrrolysine because of a context feature of their encoding UAGs (a particular mRNA stem loop 3′ of such UAG codons). At issue is whether, unlike the selenocysteine specification, pyrrolysine is sometimes specified incidentally and the only selective advantage is a continuous synthesis (UAG reading) without importance to the identity of the amino acid incorporated. Pertinent to this issue is the frequency of utilised UAGs in mRNAs in organisms that encode pyrrolysine. While it was infrequent in the initially studied pyrrolysine-containing archaea, it was frequent in the anaerobic bacteria that also encode pyrrolysine. As discussed below, some of the dilemma posed led to an exciting regulatory discovery [12]. Related to the natural expansion of genetic decoding exemplified by pyrrolysine encoding was the excitement that arose with the realisation that the nature of pyrrolysine tRNA and of its aminoacylation could be exploited for synthetic code expansion. Another reason for high initial interest in pyrrolysine was the functional role it plays in methylamine-dependent methyltransferases. Recent discoveries of additional organisms, especially archaea that utilise pyrrolysine with distinctive traits, have substantially extended appreciation of its importance (e.g. climate change and biogenic methane production or potential biofuel production; human health and realisation of potential importance of pyrrolysine in gut microbes metabolism), making a review timely.

Pyrrolysine: a non-dispensable amino acid linked to anaerobic methylamine metabolism

From a biochemical perspective, Pyl ($C_{12}H_{21}N_3O_3$; N^6-{[(2*R*,3*R*)-3-methyl-3,4-dihydro-2*H*-pyrrol-2-yl] carbonyl}-L-lysine) is a typical L-lysine amino acid to which a pyrrole ring is branched on the lateral chain through an amide bond (biosynthesis in the section 'Synthesis of Pyl from a standard AA'). However, while L-lysine derivatives are already known in proteins, encompassing those from archaea like hypusine or methyllysine [13], Pyl does not originate from post-translational modifications of L-lysine, but is translationally incorporated, representing the 22nd proteinogenic amino acid. This presence has been mostly documented in three different kinds of proteins (MtmB, MtbB and MttB), which are methyltransferases involved in the metabolism of, respectively, mono-, di- and trimethylamine (abbreviated in MMA, DMA and TMA). Until recently, the only other known Pyl-containing proteins were certain transposases [14], and a $tRNA^{His}$-guanylyltransferase Thg1 [15] both present in a subset of *Methanosarcinales*. In methyltransferases, Pyl is found in the active site, thus arguing for a non-dispensable role for biological activity. It captures methylamines (through its imine electrophilic group) before transferring one methyl group to a Co(I)-corrinoid cofactor of an associated protein (MtmC/MtbC/MttC) [11]. Recently, this non-dispensable role of Pyl towards methylamines has been strengthened by the discovery of natural MttB analogues without Pyl [16]. These homologues of MttB have been identified in multiple bacteria as well as methanogenic and non-methanogenic archaea. One of these enzymes (named MtgB) has been characterised in *Desulfitobacterium hafniense* YF51 and catalyses the transfer of a methyl group from glycine betaine, but not trimethylamine, to a corrinoid protein. In consequence, Pyl in archaea seems tightly linked to methanogenesis from methylamines (MMA, DMA and TMA) and is non-dispensable for such a metabolism [17].

Components of the Pyl system

The first indication of unusual decoding of the *mtmB* gene of the archaeon *Methanosarcina barkeri* was that its mRNA contained an in-frame UAG codon that was not interpreted as a stop signal during translation [18]. This was also the case in other *Methanosarcinales*, certain anaerobic bacteria and even for genes encoding the different protein families, MtbB and MttB [19]. The crystal structure of *M. barkeri* MtmB1 revealed the presence of an unknown amino acid, which was named L-pyrrolysine [11]. It was determined that it relies on an UAG-decoding tRNA (PylT) charged by Pyl [10]. With the specification of a new non-universally encoded amino acid, this dramatically showed a natural expansion of the genetic code [20].

Pyl synthesis and insertion only rely on five genes, *pylTSBCD* (or six when $tRNA^{Pyl}$ synthetase is encoded by two genes, see below). In most cases, these genes are gathered in an operon-like structure as shown in Figure 1

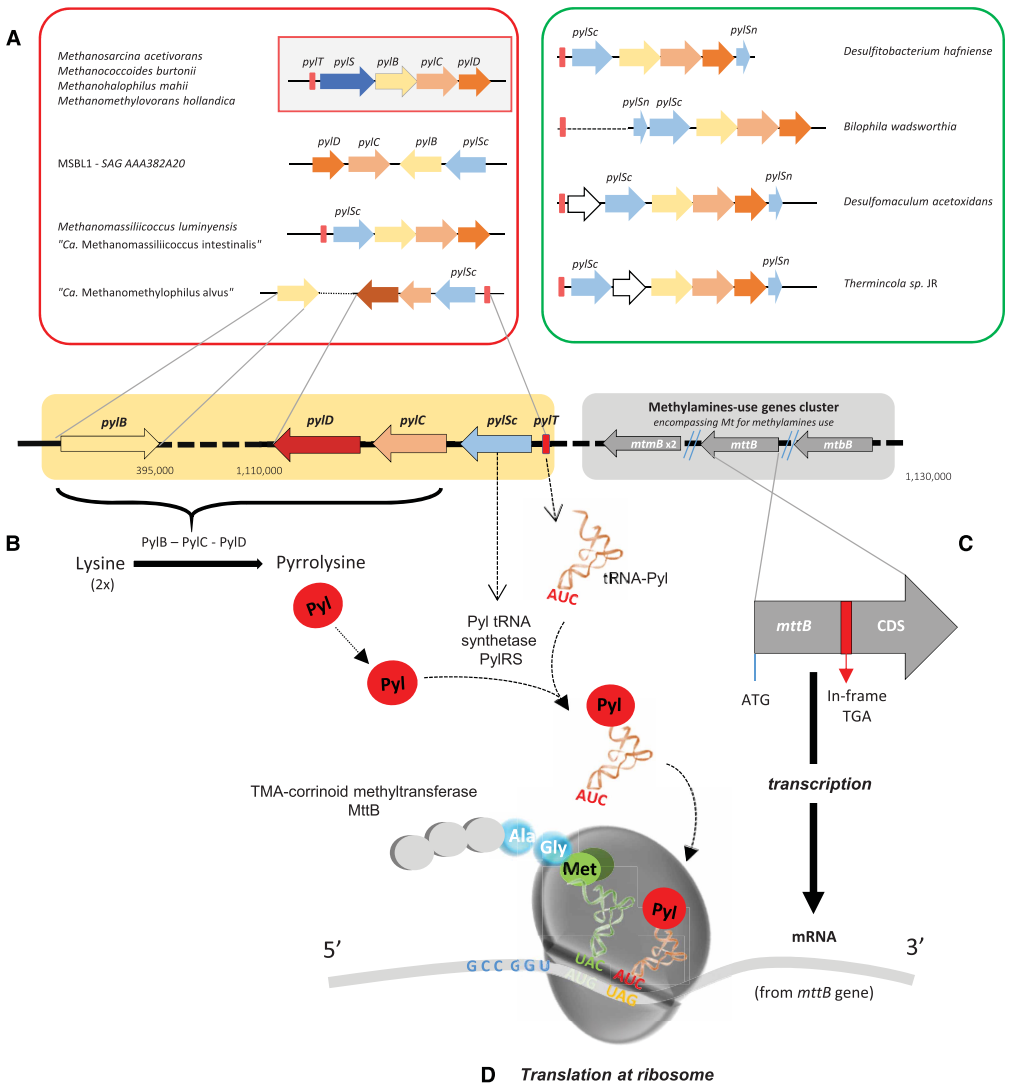

Figure 1. Overview of the components and functioning of the Pyl system.

(**A**) Examples of genomic organisation of genes involved in Pyl synthesis and encoding in archaea, in the red box on the left, and in bacteria, in the green box on the right (adapted from refs [21,23,53]). The 'canonical' organisation in archaea is given in the upper light grey box. (**B–D**) The functioning of the Pyl system is depicted here taken the example of Pyl insertion by readthrough of UAG codon contained in MttB mRNA of the *Methanomassiliicoccales* '*Candidatus* Methanomethylophilus alvus'. (**B**) It relies first on the biosynthesis of Pyl (depicted with more details in Figure 2). Pyl is then engrafted on a specific tRNA (encoded by *pylT*) whose anticodon recognises UAG codons. This implies the specific reaction catalysed by PylRS (in this example encoded by a truncated version of *pylS* corresponding to *PylSc*) (see section 'The case of Pyl in *Methanomassiliicoccales*' for more details). (**C**) All the genes encoding the (mono-, di- or tri-) methylamine-corrinoid methyltransferases (respectively, *mtmB*, *mtbB* and *mttB*) show an in-frame UAG codon, which should typically lead to a translational stop, i.e. a truncated protein. (**D**) Due to $tRNA^{Pyl}$ able to interact with UAG codon, translation is going on by readthrough this codon which leads to insertion of Pyl in the growing polypeptide. Functioning of such a system at the genomic and cellular level is discussed in section 'Implications and consequences of the natural presence of a UAG-decoding tRNA'.

(light grey box, upper left) [21,22]. However, some exceptions to such genomic organisation have emerged (see, some examples, the upper part of Figure 1) [21,23]. Evidence of this minimum set of genes was obtained from generating in *Escherichia coli*, an MtmB protein containing Pyl at the correct location by co-expressing *pylTSBCD* from *M. barkeri* [24]. Figure 1 gives a schematic overview of the Pyl system based on the decoding

Emerging Topics in Life Sciences (2018)
https://doi.org/10.1042/ETLS20180094

of UAG as Pyl in the MttB protein of '*Candidatus* Methanomethylophilus alvus', a *Methanomassiliicoccales* representative.

Synthesis of Pyl from a standard amino acid

The synthesis of pyrrolysine relies on three genes, *pylBCD* [22]. It was initially hypothesised that L-lysine was one of the precursors, with the pyrrole ring added on its acyl chain derived from L-isoleucine, L-proline, L-methionine, L-glutamate or D-ornithine. Indeed, supplying D-ornithine stimulates the readthrough of UAG codon of a reporter gene in *E. coli* expressing *pylTSBCD* [25]. D-ornithine supply leads, however, to a desmethylpyrrolysine insertion instead of Pyl due to the respective enzymatic role of *pylCD* products that were identified [26]. This suggested that PylB activity was mandatory to get Pyl and led Krzycki and colleagues to postulate that the reaction product catalysed by PylB was (*3R*)-3-methyl-D-ornithine. To obtain such a product, they demonstrated that PylB was an L-lysine mutase inverting the chiral centres in the pyrrole ring [27]. Therefore, Pyl has only one kind of precursor, which is surprisingly the common proteinogenic amino acid, L-lysine. It is used twice in a three-reaction pathway (Figure 2). This pathway encompasses first the L-lysine mutase PylB, a radical *S*-adenosylmethionine (SAM) protein with a TIM (triose phosphate isomerase) barrel [28]. The amide bond between a second L-lysine (at ε location) is then catalysed by PylC to form (*3R*)-3-methyl-D-ornithine. Finally, PylD initiates an (NAD-dependent) oxidation of this molecule to form a semi-aldehyde, intermediary compound that self-reacts to form the pyrrole ring by dehydration (Figure 2).

Linking Pyl to a specifically dedicated tRNA

The two other genes of the Pyl system, *pylT* and *pylS*, encode, respectively, a tRNA whose anticodon is complementary to the UAG codon and the subunit (PylS) of the $tRNA^{Pyl}$ synthetase. This tRNA synthetase/tRNA pair possesses a strict orthogonality towards the other 20 proteinogenic amino acids, leading specifically to link Pyl to this dedicated tRNA [10]. The structural analysis of near complete PylS (PylSc from the bacterium *D. hafniense*, see below) was deduced either from the crystallisation of the apoprotein alone or from its co-crystallisation with $tRNA^{Pyl}$. It revealed a pair highly distinct from the other ones, even the most closely related ones [29], with both $tRNA^{Pyl}$ synthetase and $tRNA^{Pyl}$ having original features compared with other known equivalents. Moreover, differences are also observable among Pyl-decoding bacteria and archaea, for their (unusually small) $tRNA^{Pyl}$ and for PylS (see below).

The active $tRNA^{Pyl}$ synthetase is a homodimer belonging to the class II family of aaRS. While being very different from other class II aaRS, its catalytic core shares common traits with $tRNA^{Phe}$ synthetase, suggesting that these two aaRS are evolutionarily related (section 'About the origin and evolution of Pyl'). This tRNA synthetase is encoded by a unique gene in *Methanosarcinales* (*pylS*) and two in bacteria, one encoding the Cter

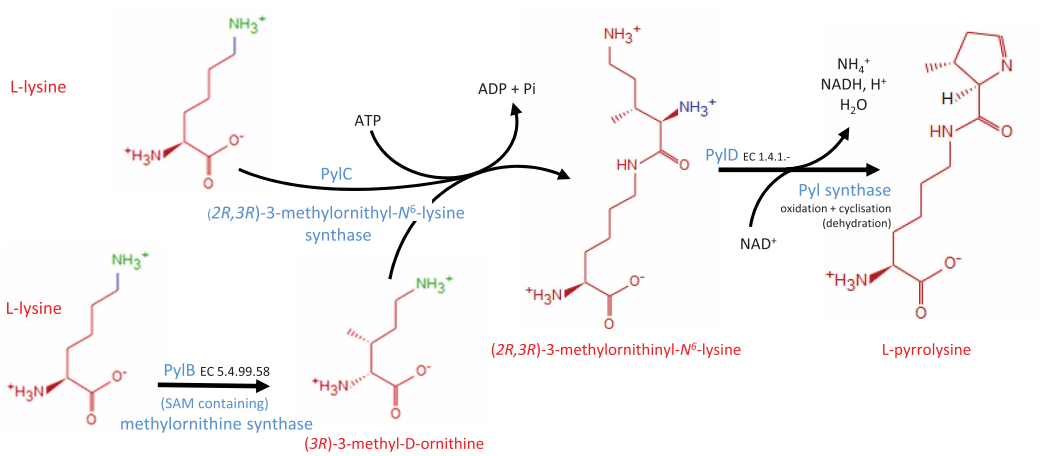

Figure 2. Biosynthesis of Pyl.

The complete biosynthesis pathway of L-pyrrolysine from two lysines catalysed by PylB, PylC and PylD is depicted. This corresponds, respectively, to reactions #13074, #13080 and #13084 adapted from MetaCyc [72], as determined from original data in refs [26,27].

portion of PylS (*pylSc*) and one encoding ~110 amino acids of PylS corresponding to the amino-terminal portion of *Methanosarcinales* PylS (*pylSn*). The amino-terminal part of *Methanosarcinales* PylS (homologue to PylSn) is dispensable for correct aminoacetylation *in vitro*, but is necessary for correct activity *in vivo* [30]. Due to its highly hydrophobic nature, likely preventing its *in vitro* crystallisation, there is no formal structural data on $tRNA^{Pyl}$ with complete PylS or PylSc + PylSn: initial work performed 10 years ago gave, however, interesting features with several structural data encompassing the catalytic domain structure (carboxy-terminal region 185–454) of the archaeal PylS from *Methanosarcina mazei* with Pyl [31] and the bacterial PylSc structure (from *D. hafniense* [29,32]). PylSn from bacteria has been shown to interact and recognise $tRNA^{Pyl}$ in a specific way, enhancing the affinity/specificity with $tRNA^{Pyl}$ [33]. Recently, the crystal structure of the amino-terminal domain of PylS (from the archaeon *M. mazei*) with $tRNA^{Pyl}$ gave precious new insights and notably, the determinants of its high specificity against canonical AAs (amino acids) not based on anticodon recognition [34]. These interactions are susceptible to nucleotide mutations of $tRNA^{Pyl}$ that are not involved in PylSc interactions with $tRNA^{Pyl}$ (for example, in the T-arm and variable loop). PylS carboxy-terminal part/PylSc are responsible at least for the catalytic aminoacylation and partly for the recognition of $tRNA^{Pyl}$. Nonetheless, anticodon UCA is not directly involved in its recognition by PylS [31], while only a few specific bases on anticodon stem and loop nucleotides of $tRNA^{Pyl}$ seem necessary, like, for example, the bases immediately adjacent to the anticodon in the archaeon *M. barkeri* [35]. However, recent genomic identification of the Pyl system in archaeal lineages other than *Methanosarcinales*, especially in the *Methanomassiliicoccales*, gives indications that these properties are not shared by all Pyl systems (see the section 'The case of Pyl in *Methanomassiliicoccales*').

Recent Pyl discoveries

As mentioned, work performed more than a decade ago showed that at least some Pyl gene cassettes from archaea or bacteria are fully functional when expressed in *E. coli* (i.e. UAG specification of Pyl) [24]. Moreover, desmethylpyrrolysine was incorporated when D-ornithine was provided to *E. coli* expressing a Pyl system, indicating a low susceptibility of the side chain for aminoacylation of $tRNA^{Pyl}$ by PylS. These properties were therefore used to divert Pyl system from its natural function, allowing non-canonical amino acids (ncAAs) to be included in proteins and, so in organisms, by expressing (wild-type or mutated) PylS/$tRNA^{Pyl}$ in the presence of (free) ncAAs in the growth medium. Achievements and strategies for using Pyl system components as a genetic code expansion tool have been recently reviewed (e.g. [36–39]). To date, more than 100 ncAAs have been successfully engineered and incorporated, in a broad phylogenetic spectrum of organisms ranging from bacteria and unicellular eukaryotes (e.g. *Saccharomyces cerevisiae* [40]) to metazoans, like *Caenorhabditis elegans* [41], *Drosophila melanogaster* [42], the Zebrafish [43] or the mammal *Mus musculus* [44].

Pyl encoding, a widespread feature in the archaeal world?

For a long time, all known methanogens were part of six orders gathered into two large classes (Class I: *Methanobacteriales*, *Methanococcales*, *Methanopyrales*; Class II: *Methanosarcinales*, *Methanomicrobiales*, *Methanocellales*) [45]. Recent data based on culture and metagenomic approaches have largely modified our vision of the diversity of the methanogens. Indeed, methanogenesis was predicted, and sometimes experimentally confirmed, in several recently discovered lineages that are only distantly related to the Class I and Class II methanogens: the *Methanomassiliicoccales*[46,47], the *Methanofastidiosa* [48], possibly the *Bathyarchaeota* [49], the *Verstraetearchaeota* [50] and the *Methanonatronarchaeia* [51] (Figure 3). A striking characteristic of the members of all these lineages is their dependence on methyl-compounds for methanogenesis, a metabolism present in a minority of previously known methanogens. Along with the methyl-dependent hydrogenotrophic methanogenesis, the Pyl system was also found to be more widespread, being present in some of these new methanogen lineages and for the first time in non-methanogenic archaea (Figure 3). Among these lineages, the Pyl system was first described in the *Methanomassiliicoccales* where it exhibits several unusual characteristics compared with previously characterised Pyl systems (see the section 'The case of Pyl in *Methanomassiliicoccales*'). In addition, one of the five currently known genomes of the *Verstraetearchaeota* harbours the genes for the synthesis and encoding of Pyl (Table 4 in supplementary data of [50]). In this genome, PylS is mentioned as being encoded by two genes, *pylSn* and *pylSc*. This is different from Methanosarcinaceae members that have a single gene coding for both PylSn and PylSc domains, but similar to what was observed in bacteria so far. Recently, the Pyl system was also reported from the *Methanonatronarchaeia*, a class composed of extreme halophiles [51]. In all these novel lineages, the occurrence of the Pyl system was coupled with the presence of one or several Pyl-containing methyltransferases

Emerging Topics in Life Sciences (2018)
https://doi.org/10.1042/ETLS20180094

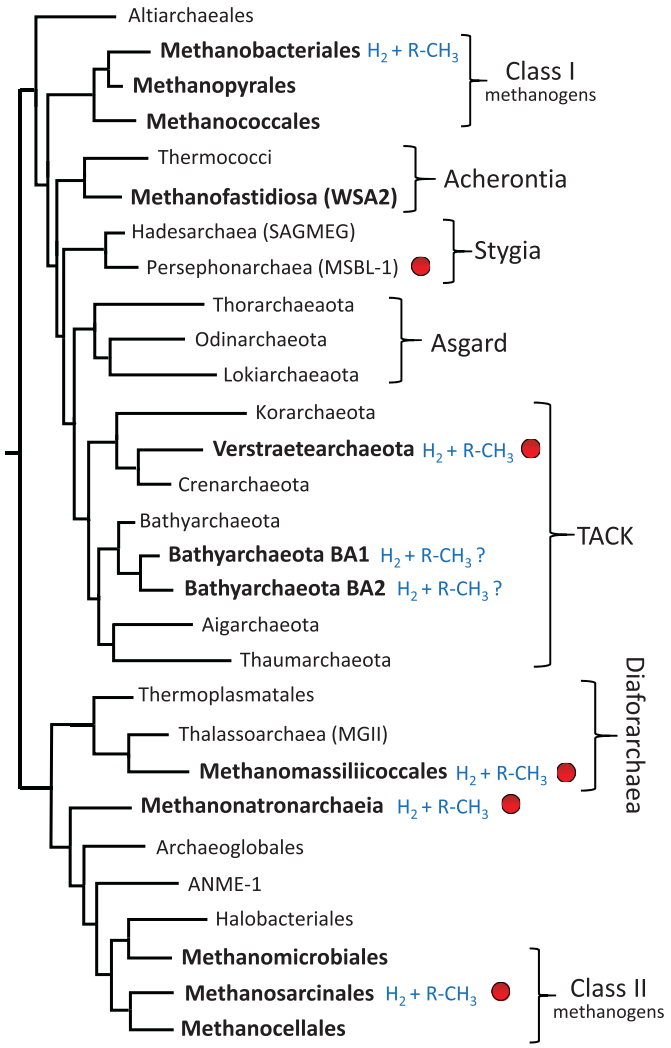

Figure 3. Distribution of the Pyl-encoding microorganisms in the domain Archaea, as indicated by red dots.
Lineages in bold contain predicted (based on genome analysis) or experimentally characterised methanogens. 'H_2+R-CH_3' indicates that some members (e.g. *Methanosphaera* in *Methanobacteriales*, *Methanimicrococcus* in *Methanosarcinales*) or all members of the lineage can gain energy through methyl-dependent hydrogenotrophic methanogenesis. Depending on the methanogen species, one or several methylated compounds (R-CH_3, e.g. methanol, TMA, DMA, TMA, methanethiol, glycine betaine) can be used. Some of the names and the topology of the tree are derived from ref. [56].

(i.e. MtmB, MtbB and MttB) needed for methylamine utilisation. This strengthens previous observations that the Pyl system is dedicated to the incorporation of Pyl in these methyltransferases and thus associated with methylamine utilisation. Other methanogens missing the Pyl-containing methyltransferases (e.g. *Methanosphaera* spp. [52], *Methanofastidiosa* spp. [48]) also lack the Pyl system, which argues that this system is more linked to methylamine methyltransferases than to archaea relying on a methanogenesis based on methyl-compounds. In addition, several members of the candidate division MSBL-1 (or *Persephonarchaea*) harbour a complete set of genes for Pyl synthesis and encoding, as well as *mtmB*, *mtbB* and *mttB* genes, as determined from single-cell genomics [53]. Interestingly, genomic analyses of these uncultured archaea revealed that they are likely not methanogens, but rather sugar-fermenters. Moreover, the components of the Pyl system in MSBL-1 are phylogenetically related to the bacterial ones and more specifically with the Pyl system found in the bacteria *Acetohalobium arabaticum* [53]. This possibly indicates a horizontal gene transfer (HGT) between these organisms living in the same kinds of the environment (hypersaline) and gives new missing pieces to the puzzle of the unknown history, evolution and dissemination of Pyl. The main conclusions to be drawn from

these results are that (i) archaea with a Pyl-encoding capacity do not only belong to the *Methanosarcinales*, but have a much wider distribution in the archaeal phylogeny, probably not limited to the recently discovered lineages, and that (ii) Pyl is not necessarily associated with methanogenesis in archaea.

The case of Pyl in *Methanomassiliicoccales*

The Pyl system of *Methanomassiliicoccales* was the first archaeal one found outside of the Methanosarcinaceae [54] and it revealed several originalities compared with Methanosarcinaceae or to bacteria [23]. It is not shared by all members of this methanogen order, even among closely related species living in an identical environment, the human gut [55]. This patchy distribution of the Pyl system in the *Methanomassiliicoccales* is most likely due to its complete loss in some representatives. In this order, all *pyl* genes are not always grouped together in an operon-like structure. This is, for example, the case in '*Candidatus* Methanomethylophilus alvus', in which *pylB* is distant (more than 600 kb) from *pylTSCD* (Figure 1). It is notable that important genes for methanogenesis are present within a few kilobases from either *pylB* or *pylTSCD*, those of the methyl-coenzyme M reductase catalysing methane formation and those of the methyltransferases dedicated to methylamines, respectively [23,56]. Also, *Methanomassiliicoccus luminyensis* is the only known organism to possess more than one copy of all *pyl* genes [23]. The most amazing feature of the *Methanomassiliicoccales* Pyl system concerns the nature of both $tRNA^{Pyl}$ and PylS. Indeed, the (UAG-recognizing) $tRNA^{Pyl}$ shows a low number of universally conserved bases among $tRNA^{Pyl}$ [21,23]. It has the usual cloverleaf secondary structure of tRNAs, although in a condensed manner (Figure 4), and a longer anticodon stem [23]. In *Methanomassiliicoccales*, $tRNA^{Pyl}$ shows the surprising feature of having one or two mismatches in the anticodon 'stem' [23] (Figure 4). This is a special feature among all natural tRNAs. *Methanomassiliicoccales* PylS corresponds to the carboxy-terminal part of Methanosarcinaceae PylS or to the PylSc of bacteria. No homologue of the amino-terminal part of PylS/PylSn has been identified in all complete genomes of *Methanomassiliicoccales* analysed so far. The structure of PylS in *Methanomassiliicoccales* is thus likely different from those of Methanosarcinaceae members and bacteria. This case seems also present in other archaea (e.g. *Persephonarchaea* (MSBL-1)). Phylogenetic analyses revealed that the *Methanomassiliicoccales* Pyl system has an intermediary position between the *Methanosarcinales* and the bacteria/MSBL-1 [57]. In view of these very atypical features, the functionality of the Pyl system in *Methanomassiliicoccales* was therefore questionable. However, the utilisation of methylamines

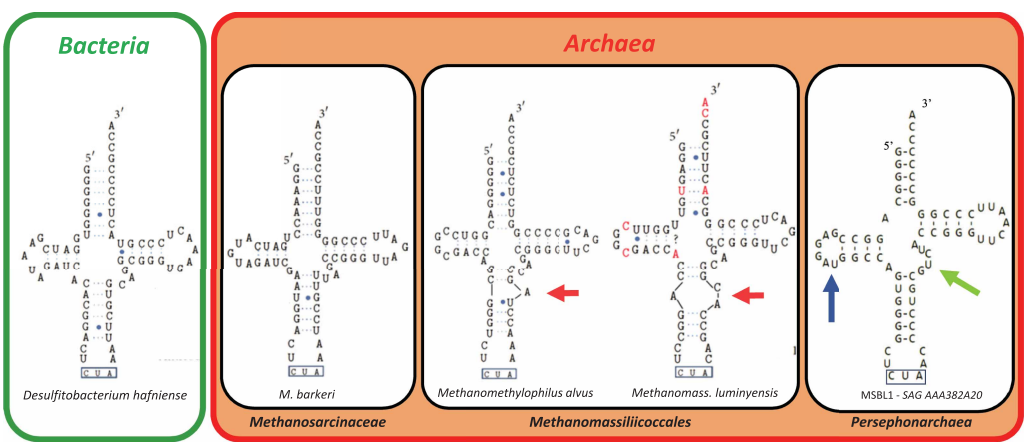

Figure 4. Sequence diversity among $tRNA^{Pyl}$.

The sequence and structure of $tRNA^{Pyl}$ are depicted here (with a cloverleaf representation) using selected examples of known sequences, encompassing a bacterial one (green box) and four archaeal ones (red box). Compared with other tRNAs, $tRNA^{Pyl}$ is smaller and has marked differences among Pyl lineages, especially between Methanosarcinaceae and other recently discovered Pyl-encoding archaeal lineages. The *Methanomassiliicoccales* $tRNA^{Pyl}$ shown here have mismatches/a loop in their anticodon "arm" (red arrows), which is a unique characteristic. The MSBL-1 $tRNA^{Pyl}$ shows a larger variable loop (green arrow) and D-loop (blue arrow) than the *Methanomassiliicoccales* $tRNA^{Pyl}$. This is especially evident in *M. luminyensis* and '*Ca.* M. intestinalis' [23]. This figure is an adaptation of the data available in refs [23,53]. Only one of the two $tRNA^{Pyl}$ recovered from *M. luminyensis* genome is indicated ($tRNA^{Pyl}$ #1 according to ref. [23]).

Emerging Topics in Life Sciences (2018)
https://doi.org/10.1042/ETLS20180094

as a methanogenesis substrate by several members of the *Methanomassiliicoccales* strongly suggests that they possess a functional Pyl system. TMA, DMA and MMA utilisation has been reported for a pure culture of *M. luminyensis* [58], as well as MMA utilisation by an enrichment culture of '*Ca.* Methanoplasma termitum' [59] and TMA utilisation by '*Ca.* Methanomethylophilus alvus' [55]. The remaining question is, therefore, how these organisms can accommodate such a different $tRNA^{Pyl}$ and the absence of amino-terminal part of PylS/PylSn. One possibility is that in this lineage, PylSc-like/shortened PylS efficiently binds and links Pyl to its dedicated $tRNA^{Pyl}$ in the absence of the PylSn domain. Another one is that it requires an unknown factor that would functionally replace PylSn and/or enhance the effectiveness of this restricted system. In any case, this indicates that other mechanisms than those described so far are possible.

Finally, it seems that the presence of the Pyl system in *Methanomassiliicoccales* has led to different genomic adaptations among clusters of *Methanomassiliicoccales* [56], some of which are mentioned in the section 'Implications and consequences of the natural presence of a UAG-decoding tRNA'.

About the origin and evolution of Pyl

Recent data have revealed that both methanogenesis based on methylamines (methylotrophic one and methyl-dependent hydrogenotrophic one) [60] and Pyl-encoding behaviour are more widespread than stated just a few years ago (Figure 3). The hypothesis that methanogenesis is a very ancient metabolism that arose before the last common ancestor of *Archaea* has been strengthened by these data [61] and in the same way, a similar hypothesis can be made for the origin of the 22nd amino acid.

It is generally accepted that LUCA (Last Unique Common Ancestor) did not have Pyl and that the genetic code based on 20 proteinogenic amino acid catalogue was already established. Among the hypotheses of the presence of Pyl in various and extremely distant lineages, Gogarten and colleagues postulated that Pyl could have been encoded by cells living at the time of LUCA, whose descendants are, however, either extinct or not yet discovered [62]. In this hypothesis, Pyl and associated coding behaviour persisted until present time because, before their extinction, some of the descendants of this hypothetical lineage would have spread the Pyl system by HGT to some LUCA descendants, an archaeon (leading to Pyl-encoding in the lineage *Methanosarcinales*), and a bacterium. Molecular and structurally based phylogenetic studies seem to indicate that PylS arose from a duplication of PheRS-coding gene anterior to LUCA [31]. The extra copy of this gene could have been the matrix for Pyl invention, once in one lineage, or possibly independently in several lineages, later during the evolution [31]. The evidence of Pyl in *Methanomassiliicoccales* and the nature of its encoding may indicate a unique origin of the Pyl system (in either archaea or bacteria) and its later spread to the other domain of life by HGT(s). This would make its origin subsequent to LUCA [23]. The recent discovery and the nature of the Pyl system in *Persephonarchaea* (MSBL-1) also support this view. Irrespective of such considerations, current data strongly suggest that Pyl-encoding is a very ancient trait, which arose likely just once as a post-LUCA invention possibly linked to methanogenesis. It could have then evolved and been retained in some organisms for which methylamine metabolism was key to survival and could have been further transferred across the bacterial and archaeal domains. While our understanding is likely to deepen further with the probable discovery of new Pyl-containing organisms, interesting questions can now be asked about how organisms have adapted to this genetic code expansion. One type of informative analysis has come from studying the consequences of deleting a functional Pyl system [63].

Implications and consequences of the natural presence of a UAG-decoding tRNA

While genetic studies with mutants of tRNA (*amber* suppressors) that permit decoding of UAG in competition with release factors have provided decoding insights, the organisms studied have not adapted over long evolutionary time to the presence of the UAG-decoding tRNA. In contrast, such adaptation has occurred in Pyl-containing organisms. Though the situation in such organisms is not as extreme as the ciliate *Condolysoma magnum* where UAG, UAA and UAG have been reassigned to be sense codons, with termination being specified by a recoding event [64,65], since whole organisms are involved, the issues are nevertheless more substantial than in the case of decoding small genomes such as mitochondria.

The mechanism used for UGA specification of selenocysteine was first considered as a potent paradigm from which Pyl insertion could be extrapolated. UGA specification of selenocysteine is dependent on special features of the mRNA 3′ of the UGA codon. UGA codons that do not have the recoding signals involved function in

the standard manner of specifying termination [66,67]. A counterpart Pyl insertion sequence (PYLIS) element was first postulated [25]. However, consistent with a detailed bioinformatics analysis [14], such a *cis*-acting element was found to be not necessary for insertion of Pyl in engineered *E. coli* [68,25]. The resulting hypothesis is that Pyl insertion relies 'simply' on a competition between $tRNA^{Pyl}$ and aRF1 on UAG codon during mRNA translation at the ribosome. This is likely promoted by a peculiar aRF1, as some distinct patterns are observed in Pyl-coding Methanosarcinaceae compared with other archaeal aRF1 [14]. In such a case, one can imagine that UAG codon that does not need to be reassigned to a Pyl would be counter-selected in favour of the other two non-sense codons. In agreement with this hypothesis, it is observed that the UAG codon usage in Pyl-coding archaea is low [14,69] with Methanosarcinaceae having only ∼4–6% CDSs with an in-frame UAG codon [12]. Recent data indicate that UAG usage is even lower in some *Methanomassiliicoccales* (in-frame UAG predicted in 2.8% of CDSs in '*Ca.* M. alvus'). However, the UAG usage varies among *Methanomassiliicoccales*, independently of GC% of considered genomes [56]. More surprisingly, members of *Methanomassiliicoccales* are predicted, through bioinformatics analysis, to use Pyl in a larger repertoire of proteins than the usual methyltransferases described above [56]. This encompasses 16 non-MT proteins in '*Ca.* M. alvus' [56] and more than 40 non-MT proteins in another representative (methanogenic archaeon ISO4-H5 [70]). This questions a possible 'genetic code domestication' of Pyl in *Methanomassiliicoccales* corroborating the hypothesis of an on-going irreversible conversion of UAG sense from a stop sense to a Pyl-encoding one proposed in *Methanosarcinales* [69]. Indeed, the peculiar properties observed in some *Methanomassiliicoccales* may reflect the on-going establishment of a genetic code having 21 amino acids encoded by 62 codons and only two codons instead of three being non-sense. This 21 proteinogenic AA repertoire would be currently invading the proteome of some of these organisms by neutral selection on most proteins, while it cannot be excluded that some cases are providing/will provide benefits for the cell.

The frequency of UAG codons is much higher in Pyl-coding bacteria than in archaea, with ∼20–28% noted in one study [14,12]. In the absence of evidence for mRNA context dependency for the natural Pyl specification that occurs in some bacteria, how, if it is, can specification of Pyl by the numerous UAG codons be avoided? For *A. arabaticum*, a member of the phylum *Clostridia*, a fascinating resolution to the dilemma has been discovered. While UAG specifies Pyl when the cells are grown in trimethylamine, when instead the cells are grown on pyruvate as a carbon source, UAG only specifies termination [12]. This presumably means that when the cells are grown on trimethylamine, many proteins for which pyrrolysine has no specific function have C-terminal extensions with several amino acids encoded after pyrrolysine. Indeed, some such proteins have been identified [12]. Doubtless, selection has had ample time to adjust, where significant, further downstream stop codon position. However, for most genes such selective pressure may be weak since where analysed there is an abundance of proteases that act on incorrectly folded peptides, and the efficiency of newly isolated genetically selected *E. coli* amber suppressors with little or no discernible growth rate. Nevertheless, a decoding-wide strategy has evolved for when trimethylamine is absent.

Conclusion

Recent data about Pyl in archaea have renewed our comprehension of this original and still enigmatic exception in the 'universal' genetic code. It was recently uncovered that this system is not confined to the *Methanosarcinales* in the domain Archaea, but present in at least four other orders, and not only methanogens. In members of those novel lineages, the divergences from canonical Pyl systems offer new opportunities to understand its functioning. These advances will likely facilitate improved utilisation of the Pyl system as a biotechnological tool for engineering of proteins/organisms by incorporating ncAAs in their proteins. Also, Pyl is mandatory for methylamines utilisation under anoxic conditions which may explain its presence and maintenance in diverse and phylogenetically distant organisms. One of these methylamines, TMA, was recently shown to be a by-product of human intestinal bacteria, and the unique precursor of plasma TMAO, an important atherosclerogenic factor [71]. Thus, autochthonous archaea able to metabolise TMA in the gut (some *Methanomassiliicoccales* members), or engineered microbes with a Pyl-encoding behaviour to 'trap' TMA in the gut may be valuable auxiliaries for preventing cardiovascular diseases in humans [58]. Besides these biotechnological aspects, its study will likely facilitate understanding of the genetic code evolution among lineages and the intimate mechanisms of such important biological functions as translation. All of these bode well for increasing interest in Pyl and its host organisms in the coming years.

Emerging Topics in Life Sciences (2018)
https://doi.org/10.1042/ETLS20180094

Summary

- Pyrrolysine (Pyl) is the 22nd proteinogenic amino acid whose origin remains enigmatic.
- It corresponds to a genetic code expansion using UAG codon readthrough.
- It is encoded in a few bacteria and in distantly related archaeal lineages, but its phylogenetic distribution is likely underestimated.
- Modalities of the functioning of such a genetic code exception are different among organisms, which has likely different consequences on evolution and biology of these organisms.

Abbreviations

AA, amino acid; aaRS, aminoacyl tRNA synthetase; DMA, Dimethylamine; HGT, horizontal gene transfer; LUCA, Last Unique Common Ancestor; MMA, monomethylamine; ncAA, non-canonical AA; Pyl, pyrrolysine; TMA, trimethylamine.

Author Contribution

All authors contributed to manuscript writing.

Acknowledgements

J.-F.B. thanks the French national commission '64^{e} CNU' for agreeing with a sabbatical in P.W.O.T.'s laboratory (School of Microbiology and APC Microbiome Institute, University College Cork).

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1 Crick, F.H., Barnett, L., Brenner, S. and Watts-Tobin, R.J. (1961) General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 https://doi.org/10.1038/1921227a0
2 'The Genetic Code'. (1966) Collective work of proceedings of symposium yeld in Cold Spring Harbor (USA) in 1966. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 3–9 https://doi.org/10.1101/SQB.1966.031.01.007
3 Atkins, J.F. and Baranov, P.V. (2010) The distinction between recoding and codon reassignment. *Genetics* **185**, 1535–1536 https://doi.org/10.1534/genetics.110.119016
4 Barrell, B.G., Bankier, A.T. and Drouin, J. (1979) A different genetic code in human mitochondria. *Nature* **282**, 189 https://doi.org/10.1038/282189a0
5 Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.* **56**, 229–264 PMID:1579111
6 Rother, M. and Quitzke, V. (2018) Selenoprotein synthesis and regulation in *Archaea*. *Biochim. Biophys. Acta* **1862**, 2451–2462 https://doi.org/10.1016/j.bbagen.2018.04.008
7 Mukai, T., Englert, M., Tripp, H.J., Miller, C., Ivanova, N.N., Rubin, E.M. et al. (2016) Facile recoding of selenocysteine in nature. *Angew. Chem. Int. Ed. Engl.* **55**, 5337–5341 https://doi.org/10.1002/anie.201511657
8 Zinoni, F., Heider, J. and Böck, A. (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl Acad. Sci. U.S.A.* **87**, 4660–4664 https://doi.org/10.1073/pnas.87.12.4660
9 Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W. and Harrison, P.R. (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, TGA. *EMBO J.* **5**, 1221–1227 https://doi.org/10.1002/j.1460-2075.1986.tb04350.x
10 Srinivasan, G., James, C.M. and Krzycki, J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 https://doi.org/10.1126/science.1069588
11 Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A. and Chan, M.K. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**, 1462–1466 https://doi.org/10.1126/science.1069556
12 Prat, L., Heinemann, I.U., Aerni, H.R., Rinehart, J., O'Donoghue, P. and Söll, D. (2012) Carbon source-dependent expansion of the genetic code in bacteria. *Proc. Natl Acad. Sci. U.S.A.* **109**, 21070–21075 https://doi.org/10.1073/pnas.1218613110
13 Eichler, J. and Adams, M.W. (2005) Posttranslational protein modification in *Archaea*. *Microbiol. Mol. Biol. Rev.* **69**, 393–425 https://doi.org/10.1128/MMBR.69.3.393-425.2005

14 Zhang, Y., Baranov, P.V., Atkins, J.F. and Gladyshev, V.N. (2005) Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* **280**, 20740–20751 https://doi.org/10.1074/jbc.M501458200
15 Heinemann, I.U., O'Donoghue, P., Madinger, C., Benner, J., Randau, L., Noren, C.J. et al. (2009) The appearance of pyrrolysine in tRNAHis guanylyltransferase by neutral evolution. *Proc. Natl Acad. Sci. U.S.A.* **106**, 21103–21108 https://doi.org/10.1073/pnas.0912072106
16 Ticak, T., Kountz, D.J., Girosky, K.E., Krzycki, J.A. and Ferguson, D.J. (2014) A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. *Proc. Natl Acad. Sci. U.S.A.* **111**, E4668–E4676 https://doi.org/10.1073/pnas.1409642111
17 Krzycki, J.A. (2004) Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* **8**, 484–491 https://doi.org/10.1016/j.cbpa.2004.08.012
18 Burke, S.A., Lo, S.L. and Krzycki, J.A. (1998) Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. *J. Bacteriol.* **180**, 3432–3440 PMID:9642198
19 Paul, L., Ferguson, D.J. and Krzycki, J.A. (2000) The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *J. Bacteriol.* **182**, 2520–2529 https://doi.org/10.1128/JB.182.9.2520-2529.2000
20 Atkins, J.F. and Gesteland, R. (2002) The 22nd amino acid. *Science* **296**, 1409–1411 https://doi.org/10.1126/science.1073339
21 Gaston, M.A., Jiang, R. and Krzycki, J.A. (2011) Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **14**, 342–349 https://doi.org/10.1016/j.mib.2011.04.001
22 Krzycki, J.A. (2013) The path of lysine to pyrrolysine. *Curr. Opin. Chem. Biol.* **17**, 619–625 https://doi.org/10.1016/j.cbpa.2013.06.023
23 Borrel, G., Gaci, N., Peyret, P., O'Toole, P.W., Gribaldo, S. and Brugère, J.F. (2014) Unique characteristics of the pyrrolysine system in the 7th order of methanogens: implications for the evolution of a genetic code expansion cassette. *Archaea* **2014**, 1–11 https://doi.org/10.1155/2014/374146
24 Longstaff, D.G., Larue, R.C., Faust, J.E., Mahapatra, A., Zhang, L., Green-Church, K.B. et al. (2007) A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl Acad. Sci. U.S.A.* **104**, 1021–1026 https://doi.org/10.1073/pnas.0610294104
25 Namy, O., Zhou, Y., Gundllapalli, S., Polycarpo, C.R., Denise, A., Rousset, J.P. et al. (2007) Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Lett.* **581**, 5282–5288 https://doi.org/10.1016/j.febslet.2007.10.022
26 Cellitti, S.E., Ou, W., Chiu, H.P., Grünewald, J., Jones, D.H., Hao, X. et al. (2011) D-Ornithine coopts pyrrolysine biosynthesis to make and insert pyrroline-carboxy-lysine. *Nat. Chem. Biol.* **7**, 528 https://doi.org/10.1038/nchembio.586
27 Gaston, M.A., Zhang, L., Green-Church, K.B. and Krzycki, J.A. (2011) The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **471**, 647 https://doi.org/10.1038/nature09918
28 Quitterer, F., List, A., Eisenreich, W., Bacher, A. and Groll, M. (2012) Crystal structure of methylornithine synthase (PylB): insights into the pyrrolysine biosynthesis. *Angew. Chem. Int. Ed.* **51**, 1339–1342 https://doi.org/10.1002/anie.201106765
29 Nozawa, K., O'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T. et al. (2009) Pyrrolysyl-tRNA synthetase–tRNA Pyl structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163 https://doi.org/10.1038/nature07611
30 Herring, S., Ambrogelly, A., Gundllapalli, S., O'Donoghue, P., Polycarpo, C.R. and Söll, D. (2007) The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett.* **581**, 3197–3203 https://doi.org/10.1016/j.febslet.2007.06.004
31 Kavran, J.M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D. and Steitz, T.A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl Acad. Sci. U.S.A.* **104**, 11268–11273 https://doi.org/10.1073/pnas.0704769104
32 Lee, M.M., Jiang, R., Jain, R., Larue, R.C., Krzycki, J. and Chan, M.K. (2008) Structure of *Desulfitobacterium hafniense* PylSc, a pyrrolysyl-tRNA synthetase. *Biochem. Biophys. Res. Commun.* **374**, 470–474 https://doi.org/10.1016/j.bbrc.2008.07.074
33 Jiang, R. and Krzycki, J.A. (2012) Pylsn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. *J. Biol. Chem.* **287**, 32738–32746 https://doi.org/10.1074/jbc.M112.396754
34 Suzuki, T., Miller, C., Guo, L.T., Ho, J.M., Bryson, D.I., Wang, Y.S. et al. (2017) Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **13**, 1261–1266 https://doi.org/10.1038/nchembio.2497
35 Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C. and Söll, D. (2007) Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl Acad. Sci. U.S.A.* **104**, 3141–3146 https://doi.org/10.1073/pnas.0611634104
36 Wan, W., Tharp, J.M. and Liu, W.R. (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **1844**, 1059–1070 https://doi.org/10.1016/j.bbapap.2014.03.002
37 Brabham, R. and Fascione, M.A. (2017) Pyrrolysine amber stop-codon suppression: development and applications. *Chembiochem* **18**, 1973–1983 https://doi.org/10.1002/cbic.201700148
38 Mukai, T., Lajoie, M.J., Englert, M. and Söll, D. (2017) Rewriting the genetic code. *Annu. Rev. Microbiol.* **71**, 557–577 https://doi.org/10.1146/annurev-micro-090816-093247
39 Tharp, J.M., Ehnbom, A. and Liu, W.R. (2017) tRNAPyl: structure, function, and applications. *RNA Biol.* **15**, 441–452 https://doi.org/10.1080/15476286.2017.1356561
40 Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z. and Schultz, P.G. (2003) An expanded eukaryotic genetic code. *Science* **301**, 964–967 https://doi.org/10.1126/science.1084772
41 Greiss, S. and Chin, J.W. (2011) Expanding the genetic code of an animal. *J. Am. Chem. Soc.* **133**, 14196–14199 https://doi.org/10.1021/ja2054034
42 Bianco, A., Townsley, F.M., Greiss, S., Lang, K. and Chin, J.W. (2012) Expanding the genetic code of *Drosophila melanogaster*. *Nat. Chem. Biol.* **8**, 748 https://doi.org/10.1038/nchembio.1043
43 Liu, J., Hemphill, J., Samanta, S., Tsang, M. and Deiters, A. (2017) Genetic code expansion in zebrafish embryos and its application to optical control of cell signaling. *J. Am. Chem. Soc.* **139**, 9100–9103 https://doi.org/10.1021/jacs.7b02145
44 Han, S., Yang, A., Lee, S., Lee, H.W., Park, C.B. and Park, H.S. (2017) Expanding the genetic code of *Mus musculus*. *Nat. Commun.* **8**, 14568 https://doi.org/10.1038/ncomms14568
45 Bapteste, É., Brochier, C. and Boucher, Y. (2005) Higher-level classification of the *Archaea*: evolution of methanogenesis and methanogens. *Archaea* **1**, 353–363 https://doi.org/10.1155/2005/859728

46 Borrel, G., O'Toole, P.W., Harris, H.M., Peyret, P., Brugere, J.F. and Gribaldo, S. (2013) Phylogenomic data support a seventh order of methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol. Evol.* **5**, 1769–1780 https://doi.org/10.1093/gbe/evt128
47 Petitjean, C., Deschamps, P., López-García, P., Moreira, D. and Brochier-Armanet, C. (2015) Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. *Mol. Biol. Evol.* **32**, 1242–1254 https://doi.org/10.1093/molbev/msv015
48 Nobu, M.K., Narihiro, T., Kuroda, K., Mei, R. and Liu, W.T. (2016) Chasing the elusive Euryarchaeota class WSA2: genomes reveal a uniquely fastidious methyl-reducing methanogen. *ISME J.* **10**, 2478–2487 https://doi.org/10.1038/ismej.2016.33
49 Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D. et al. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434–438 https://doi.org/10.1126/science.aac7745
50 Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P. et al. (2016) Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 https://doi.org/10.1038/nmicrobiol.2016.170
51 Sorokin, D.Y., Makarova, K.S., Abbas, B., Ferrer, M., Golyshin, P.N., Galinski, E.A. et al. (2017) Discovery of extremely halophilic, methyl-reducing Euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat. Microbiol.* **2**, 17081 https://doi.org/10.1038/nmicrobiol.2017.81
52 Fricke, W.F., Seedorf, H., Henne, A., Krüer, M., Liesegang, H., Hedderich, R. et al. (2006) The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H2 for methane formation and ATP synthesis. *J. Bacteriol.* **188**, 642–658 https://doi.org/10.1128/JB.188.2.642-658.2006
53 Guan, Y., Haroon, M.F., Alam, I., Ferry, J.G. and Stingl, U. (2017) Single-cell genomics reveals pyrrolysine-encoding potential in members of uncultivated archaeal candidate division MSBL-1. *Environ. Microbiol. Rep.* **9**, 404–410 https://doi.org/10.1111/1758-2229.12545
54 Borrel, G., Harris, H.M., Tottey, W., Mihajlovski, A., Parisot, N., Peyretaillade, E. et al. (2012) Genome sequence of '*Candidatus* Methanomethylophilus alvus' Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J. Bacteriol.* **194**, 6944–6945 https://doi.org/10.1128/JB.01867-12
55 Borrel, G., McCann, A., Deane, J., Neto, M.C., Lynch, D.B., Brugère, J.F. et al. (2017) Genomics and metagenomics of trimethylamine-utilizing Archaea in the human gut microbiome. *ISME J.* **11**, 2059 https://doi.org/10.1038/ismej.2017.72
56 Borrel, G., Parisot, N., Harris, H.M., Peyretaillade, E., Gaci, N., Tottey, W. et al. (2014) Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **15**, 679 https://doi.org/10.1186/1471-2164-15-679
57 Mayumi, D., Mochimaru, H., Tamaki, H., Yamamoto, K., Yoshioka, H., Suzuki, Y. et al. (2016) Methane production from coal by a single methanogen. *Science* **354**, 222–225 https://doi.org/10.1126/science.aaf8821
58 Brugère, J.F., Borrel, G., Gaci, N., Tottey, W., O'toole, P.W. and Malpuech-Brugère, C. (2014) Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes* **5**, 5–10 https://doi.org/10.4161/gmic.26749
59 Lang, K., Schuldes, J., Klingl, A., Poehlein, A., Daniel, R. and Brune, A. (2015) New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of 'Candidatus Methanoplasma termitum'. *Appl. Environ. Microbiol.* **81**, 1338–1352 https://doi.org/10.1128/AEM.03389-14
60 Adam, P.S., Borrel, G., Brochier-Armanet, C. and Gribaldo, S. (2017) The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* **11**, 2407 https://doi.org/10.1038/ismej.2017.122
61 Borrel, G., Adam, P.S. and Gribaldo, S. (2016) Methanogenesis and the Wood–Ljungdahl pathway: an ancient, versatile, and fragile association. *Genome Biol. Evol.* **8**, 1706–1711 https://doi.org/10.1093/gbe/evw114
62 Fournier, G.P., Huang, J. and Gogarten, J.P. (2009) Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2229–2239 https://doi.org/10.1098/rstb.2009.0033
63 O'Donoghue, P., Prat, L., Kucklick, M., Schäfer, J.G., Riedel, K., Rinehart, J. et al. (2014) Reducing the genetic code induces massive rearrangement of the proteome. *Proc. Natl Acad. Sci. U.S.A.* **111**, 17206–17211 https://doi.org/10.1073/pnas.1420193111
64 Swart, E.C., Serra, V., Petroni, G. and Nowacki, M. (2016) Genetic codes with no dedicated stop codon: context-dependent translation termination. *Cell* **166**, 691–702 https://doi.org/10.1016/j.cell.2016.06.020
65 Heaphy, S.M., Mariotti, M., Gladyshev, V.N., Atkins, J.F. and Baranov, P.V. (2016) Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in *Condylostoma magnum*. *Mol. Biol. Evol.* **33**, 2885–2889 https://doi.org/10.1093/molbev/msw166
66 Heider, J., Baron, C. and Böck, A. (1992) Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *EMBO J.* **11**, 3759–3766 https://doi.org/10.1002/j.1460-2075.1992.tb05461.x
67 Berry, M.J., Banu, L., Harney, J.W. and Larsen, P.R. (1993) Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J.* **12**, 3315–3322 https://doi.org/10.1002/j.1460-2075.1993.tb06001.x
68 Longstaff, D.G., Blight, S.K., Zhang, L., Green-Church, K.B. and Krzycki, J.A. (2007) In vivo contextual requirements for UAG translation as pyrrolysine. *Mol. Microbiol.* **63**, 229–241 https://doi.org/10.1111/j.1365-2958.2006.05500.x
69 Alkalaeva, E., Eliseev, B., Ambrogelly, A., Vlasov, P., Kondrashov, F.A., Gundllapalli, S. et al. (2009) Translation termination in pyrrolysine-utilizing archaea. *FEBS Lett.* **583**, 3455–3460 https://doi.org/10.1016/j.febslet.2009.09.044
70 Li, Y., Leahy, S.C., Jeyanathan, J., Henderson, G., Cox, F., Altermann, E. et al. (2016) The complete genome sequence of the methanogenic archaeon ISO4-H5 provides insights into the methylotrophic lifestyle of a ruminal representative of the *Methanomassiliicoccales*. *Stand. Genomic Sci.* **11**, 59 https://doi.org/10.1186/s40793-016-0183-5
71 Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., DuGar, B. et al. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57 https://doi.org/10.1038/nature09922
72 Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M. et al. (2015) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **44**, D471–D480 https://doi.org/10.1093/nar/gkv1164