Biochemistry of aminoacyl tRNA synthetase and tRNAs and their engineering for cell-free and synthetic cell applications

Ragunathan Bava Ganesh and Sebastian J. Maerkl*

School of Engineering, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

*** Correspondence:**
sebastian.maerkl@epfl.ch

Keywords: AARS, tRNA, genetic code, synthetic biology, cell-free systems, molecular engineering

Abstract

Cell-free biology is increasingly utilized for engineering biological systems, incorporating novel functionality, and circumventing many of the complications associated with cells. The central dogma describes the information flow in biology consisting of transcription and translation steps to decode genetic information. Aminoacyl-tRNA synthetases (AARSs) and tRNAs are key components involved in translation and thus protein synthesis. This review provides information on AARS and tRNA biochemistry, their role in the translation process, summarizes progress in cell-free engineering of tRNAs and AARSs, and discusses prospects and challenges lying ahead in cell-free engineering.

1 Cell-Free Synthetic Biology

Cell-free reactions were first used in 1897 when Buchner showed that yeast extract can be used for the fermentation process (Buchner, 1897). In the last century when the scientific community was exploring the molecular biology of the cell and identifying the key components involved in the central dogma of life, cell-free studies were used as a tool to understand their fundamental biochemical functions and served as an additional tool to validate or support their hypothesis. A classic example is the requirement of template RNA for amino acid incorporation during protein synthesis which was proved *in vitro* by Nirenberg in 1961 using a cell-free system extracted from *E. coli* (Nirenberg and Matthaei, 1961). Cell-free biology can be defined as, "the reproduction, study, and exploitation of complex biological processes without intact cells" (Swartz, 2006). Research that was not possible with intact cells or constrained by the limitations or complexities of the cell, was made possible by cell-free systems. For example, lysate-based systems have been successfully used for the study and implementation of synthetic gene regulatory networks (Swank, Laohakunakorn and Maerkl, 2019) and forward engineering of genetic oscillators (Niederholtmeyer *et al.*, 2015). Cell-free systems also provided a faster and more convenient way to synthesize proteins using linear rather than circular DNA templates. Cell-free systems opened up the possibility to work with a range of organisms from conventional model organisms like *E. coli* and yeast to more un-conventional systems such as *Bacillus megaterium* (Moore *et al.*, 2018), *Clostridium autoethanogenum* (Krüger *et al.*, 2020), and eukaryote derived systems like rabbit reticulocyte lysate (Gagoski *et al.*, 2016).

Initially, cell-free systems were prepared using cell lysates, where cells were lysed, chromosomal DNA and cell membrane debris removed, and the rest of the cellular contents used for studies. Lysate-based systems suffered from batch-to-batch variation, hampering the ability to obtain consistent results

(Hunter *et al.*, 2018; Dopp, Jo and Reuel, 2019). These systems also often contained inhibitory factors, nucleases, and proteases which lowered protein yield. Moreover, and in the context of molecular engineering of fundamental importance, lysates are complex and their composition unknown.

More recently, a recombinant system called "protein synthesis using recombinant elements" (PURE) has been generated where all individual components required for transcription, translation, and energy regeneration are expressed, purified, and then reconstituted to create a cell-free system. PURE consists of 36 reconstituted proteins capable of cell-free transcription and translation. Proteins were purified using His-tag based affinity chromatography (Shimizu *et al.*, 2001). A recombinant-based cell-free system was made possible as a result of an improved understanding of cellular biochemistry and the molecular machinery involved in transcription and translation. Various studies on the PURE system were performed to make PURE preparation easier and to decrease its cost (Shepherd *et al.*, 2017; Villarreal *et al.*, 2018; Lavickova and Maerkl, 2019). Productivity and functionality was increased by adjusting the various PURE components (Li *et al.*, 2014) and by supplementing additional factors to the system (Li, Zhang, *et al.*, 2017; Maddalena *et al.*, 2016). With the advent of cell-free systems and the PURE system becoming more accessible and affordable, it is beginning to be used in various applications. The PURE system has been explored as a platform for producing therapeutic proteins (Cai *et al.*, 2015; Dondapati *et al.*, 2020) and for molecular diagnostics (Pardee *et al.*, 2016). The modular nature of PURE has made this system also an appealing starting point for bottom-up synthetic cell approaches (Niederholtmeyer, Stepanova and Maerkl, 2013; Lavickova, Laohakunakorn and Maerkl, 2020). We have recently written a comprehensive review covering various aspects and applications of cell-free synthetic biology (Laohakunakorn *et al.*, 2020), and the focus of this review lies on aminoacyl tRNA synthetase and tRNA biochemistry and engineering in the context of cell-free systems.

2 The genetic code

2.1 Discovery of the genetic code

Genetic information in biology is decoded through transcription and translation steps to result in RNA and proteins, respectively. Scientific activities in the last century helped us reach our current understanding of the steps involved in processing genetic material. In the following, a brief history of the discovery of key steps and components involved in the central dogma is discussed. In the early 20th century, the scientific community believed that proteins should be the genetic material as they are structurally diverse being made from 20 different building blocks. Whereas nucleic acids have only 4 bases as their building blocks and were thus thought less likely to be the carrier of genetic information.

The nucleic acids (DNA and RNA) were first discovered in 1869 by Friedrich Miescher and were termed 'nuclein' since they were found in the nucleus (Miescher, 1869). Many years after his discovery, Levene in 1919 identified the components of nucleic acid and the sugar group of nucleotides (Levene, 1919). Experiments by Avery and his colleagues in 1944 provided definitive proof that DNA is the genetic information with its transforming ability in bacteria (Avery, Macleod and McCarty, 1944). On the other hand, proteins were conclusively proved not to be the genetic material by Hersley and Chase in 1952 (Hershey and Chase, 1952), and the work by Rosaling Franklin, Maurice Wilkins, James Watson and Francis Crick leading to the 1953 paper describing the three-dimensional structure of the genetic material DNA (Watson and Crick, 1953), and its double-helical structure with base-pairing according to Chargaff's rules (Chargaff, 1950).

The involvement of messenger RNA (mRNA) in the central dogma was floated since 1947 but the experimental discovery came in the year 1961 by Francois Gros and Francois Jacob (Jacob and Monod,

1961). For transfer RNA (tRNA), Crick hypothesized the existence of 'adaptor' molecules, which are unstable and help carry the amino acids to the ribosomes in the cytoplasm for protein synthesis (Crick, 1955). tRNA was discovered by Paul Zamecnik in 1958 as a soluble RNA intermediate in protein synthesis and was the first non-coding RNA to be discovered (Hoagland *et al.*, 1958). The ribosomal complex consists of ribosomal RNA (rRNA) and proteins, which are the site of protein synthesis located in the cytoplasm. It was discovered by George Palade in 1955 as small cytoplasmic bodies (Palade, 1955). Aminoacyl tRNA synthetases (AARS) were first identified as activating enzymes in 1958, responsible for activating amino acids in the presence of ATP. Only after undergoing this activation step were amino acids able to participate in protein synthesis (Hoagland, Keller and Zamecnik, 1956).

The genetic code linkage between nucleic acids and proteins was discovered by Marshall Nirenberg in 1961 (Nirenberg and Matthaei, 1961). Nirenberg's work showed that codon triplets of RNA gave rise to amino acid sequences during protein synthesis. This work laid the foundation for establishing the codon - amino acid relationship in protein synthesis, which is referred to as the second genetic code.

2.2 **Theories proposed on genetic code origin**

The two main theories were the stereochemical theory and the frozen accident theory. The stereochemical theory states that interactions between codon and amino acid gave rise to the genetic code. The theory was based on the notion that each codon is somehow stereo-chemically related to the particular amino acid. This theory was put forward by Gamow in 1954, where he speculated that the double helix structure of DNA gives rise to 'holes' along the axis of the chain and, each amino acid fits this 'hole' by some affinity to give rise to peptide and protein sequences (Gamow, 1954). The frozen accident theory on the other hand was proposed by Francis Crick in 1968 and it states that the genetic code was fixed or frozen because any change or mutation would be lethal to the organism. Mutations are only possible to occur provided the necessary corrections are made along the system to rectify the mutation. The theory presumed that all organisms evolved from a common ancestor since the genetic code is universal. The current relationship between codon and amino acid is by accident or by chance (Crick, 1968) and there is no special feature that governs this interaction. The theory also proposed why no new amino acids were newly incorporated in protein synthesis despite the availability of codons. Currently, our understanding of the genetic code has improved considerably. Decoding the genetic code is not purely by stereochemical interaction or by chance.

2.3 Current understanding of the genetic code

Many components are required to implement the genetic code and key components are DNA, DNA polymerase, RNA polymerase, mRNA, tRNA, ribosomal complex, amino acids, and AARS. Genetic code implementation is a result of specific interactions between these components. There is a nucleotide-based world (DNA, RNA) and an amino acid-based world (proteins). DNA replication duplicates and maintains the genetic code. The transcription step results in single-stranded RNA molecules using DNA as its template and the translation step uses RNA as its template to synthesize protein peptides and completes the decoding of genetic information. The translation process bridges the nucleotide world and the amino acid world. Specifically, AARS and tRNA connect the two worlds. Now we know that AARS are the enzymes responsible for charging cognate amino acid onto its cognate tRNA. The fidelity of the translation process is hugely dependent on the specificity of the AARS enzymes. These enzymes have a direct influence on the protein synthesis process, segregating proteogenic amino acids from non-proteogenic amino acids. The next section contains details on tRNAs and AARSs focusing on synthesis, biochemistry, function, and editing activity.

3 Biochemistry of tRNA and AARS

This section focuses on providing basic information on the structure, biochemistry, mode of action, classification, and the role of tRNA and AARS in the protein synthesis process. We are exclusively discussing *E. coli* AARS and tRNAs unless otherwise indicated.

3.1 tRNA

The tRNA molecule is a single-stranded, non-coding RNA molecule. The general structure of tRNA in 2-D and in 3-D is provided below in Figure 1. The tRNA structure consists of the following: the anticodon arm, D-arm, T-arm, acceptor stem, and the variable arm. Bases in the anticodon arm of tRNA molecules read the genetic information in mRNA codons and contain the corresponding amino acid present in the 3'-CCA sequence of the acceptor stem. tRNAs from the same species can exhibit a difference in sequence length, size of the variable arm, and length of acceptor stem. For example, $tRNA^{Sel}$, the tRNA for selenocysteine amino acid in *E. coli*, is 95 nucleotides in length. On average, tRNA sequence length ranges from 75 – 90 nucleotides (Shepherd and Ibba, 2015).

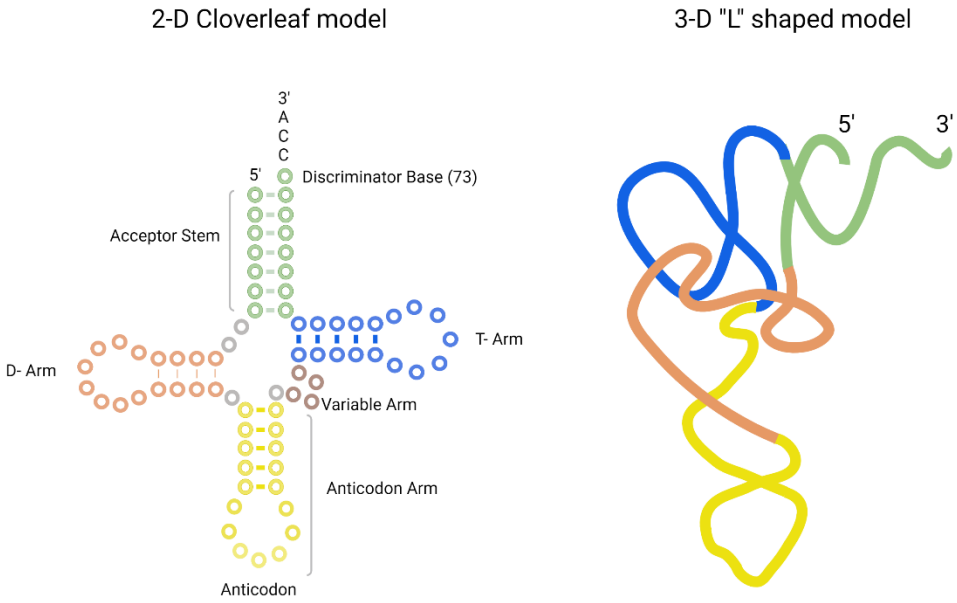

Figure 1. The structure of tRNA presented in 2-D cloverleaf model (A) and in 3-D "L" shaped model (right). The tRNA consists of acceptor stem (green), T-arm (blue), D-arm (orange), anticodon arm (yellow), and variable arm (purple).

Genes coding for tRNA are mostly arranged in groups in bacterial chromosomes with multiple copies present. Transcription is performed by RNA polymerase and results in tRNA precursor transcript having additional nucleotides on both 5' and 3' ends. Each tRNA transcript undergoes a maturation process where nucleotides are removed, specific nucleotide modifications occur, and structural integrity is gained, resulting in the cloverleaf shape. After maturation, a tRNA is available for amino acid charging at the 3'end by an AARS (Shepherd and Ibba, 2015).

The number of tRNAs present in an organism is dependent on codon usage. From a theoretical point of view, there are 64 (4^3) different codon sequences available. However, only 61 different tRNAs are used, each corresponding to a particular codon, and the remaining 3 codons (UAA, UAG, and UGA) are called stop codons and do not have a corresponding tRNA. These 61 tRNAs are shared amongst 20 amino acids. The number of tRNA acceptors for each amino acid is not the same and varies across amino acids. For example, there exists only one tRNA for the amino acid methionine with codon AUG. On the other hand, multiple tRNAs can carry the same amino acid at their 3' end and such tRNA groups are referred to as isoacceptors. For example, there are six isoacceptor tRNAs for the amino acid lysine with codons UUA, UUG, CUA, CUU, CUG, CUC.

3.2 AARS

AARSs are a family of enzymes responsible for adding an amino acid onto its cognate tRNA molecule. They are the enzyme implementing the genetic code. There are 21 AARS enzymes present, one for each amino acid except lysine, which has two AARS. In addition to these 21 AARS for proteogenic amino acids, there are AARS for non-proteogenic amino acids like pyrrolysyl tRNA synthetase and phosphoseryl tRNA synthetase. These additional AARS are found in archaea and bacteria. Each tRNA has a particular AARS for its activation. In general, tRNA charging by AARS with amino acid takes place in two steps (Figure 2).

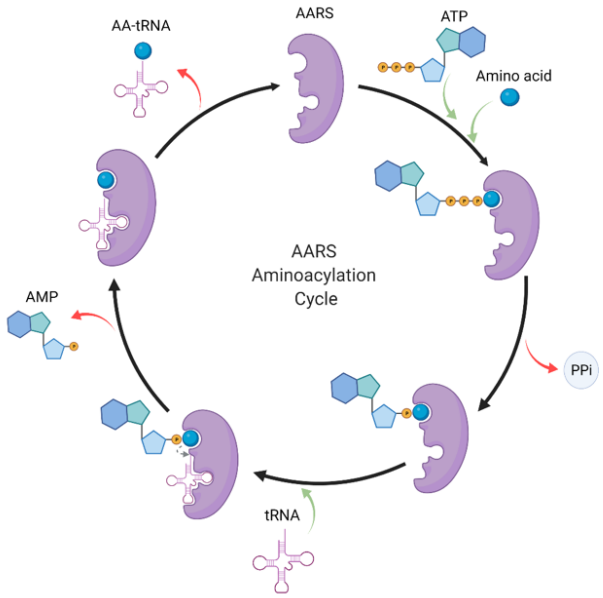

Figure 2. AARS aminoacylation process. In the first step, AARS binds to the ATP and amino acid to form an aminoacyl intermediate. In the last step, the amino acid is transferred onto tRNA resulting in activated tRNA being ready for the translation process. AARS becomes free for the next cycle of aminoacylation.

In the first step, amino acid activation takes place. ATP and amino acid bind to the AARS enzyme triggering a nucleophilic attack of the amino acid carboxyl oxygen to the α-phosphate group of ATP. This results in amino acid adenylate intermediate (AA-AMP) and release of pyrophosphate (PPi). In the second step, the hydroxyl group of adenosine at the 3'end of tRNA attacks the carboxyl carbon of AA-AMP intermediate resulting in the transfer of amino acid to tRNA. The amino acid and tRNA are linked by an ester bond. This step results in tRNA-AA, and AMP, which are released from the catalytic site of the enzyme and the AARS enzyme is free for the next cycle. In general, for the amino acid

activation step, tRNA is not required but some AARS like GlnRS, GluRS, ArgRS, and class I LysRS require tRNA as a prerequisite for amino acid activation. Activated tRNA-AA binds with the Elongation factor, EF-TU, and when reaching the ribosome participates in translation.

3.2.1 AARS classes

The 23 AARS enzymes are classified into two classes depending on the structure of the active site. In class I AARSs, the active site contains the Rossmann fold with five parallel β sheets connected by α helices. The Rossmann fold contains highly conserved motifs: HIGH and KMSKS, and both motifs are connected by sequence stretches called CP1. Class II AARSs active sites have several parallel β strands flanked by α helices. Class I enzymes are either monomeric or dimeric and class II enzymes are dimeric or tetrameric. There exist many differences between class I and class II AARS. Amino acid charging takes place at the 3'-OH group of tRNA in class I AARSs and PheRS, and 2'-OH group in class II AARSs. ATP binding during amino acid activation differs in both classes. ATP binds in an extended conformation in Class I AARSs and in a kink conformation in class II AARSs. AARSs differ in the way they bind tRNA with Class I AARSs binding the minor groove of tRNA and class II AARSs binding the major groove. In terms of reaction catalysis, the rate-limiting steps between both classes differ as well. In Class I AARSs the release of activated tRNA (tRNA-AA) is the rate-limiting step whereas in Class II AARSs it is the amino acid activation step. Each AARS class is further categorized into subclasses depending on the type of amino acid charged by the enzyme. Across both Class I and Class II, subclass A recognizes aliphatic and thiolated amino acids, subclass B recognizes charged polar amino acids, and subclass C recognizes aromatic amino acids (Gomez and Ibba, 2020).

3.2.2 Substrate recognition by AARSs

Amino acids are much smaller in size than tRNA, and so fewer chemical moieties are available for AARSs to distinguish cognate from non-cognate amino acids. Generally, amino acids are recognized based on their size, functional group, and ability to bind with metal ions present in the active site of the enzyme. In the enzyme PheRS, a conserved Alanine residue helps discriminate phenylalanine over tyrosine (Reynolds *et al.*, 2010). Glycine, the smallest amino acid, is recognized by the high negative charge in the binding pocket of its AARS. Five different conserved negative charges are used to identify glycine. Two threonine residues in the GlyRS binding pocket help to prevent other amino acids from activating. Crystallographic studies have shown that zinc metal ions in the active site of the enzyme help distinguish cognate from non-cognate amino acids (Valencia-Sánchez *et al.*, 2016). In the case of ThrRS (Sankaranarayanan *et al.*, 2000) and CysRS (Zhang *et al.*, 2003), the cognate amino acid is selected by its ability to interact with zinc ion present in the active site whereas non-cognate amino acids fail to do so.

Each AARS has a specific binding pocket for tRNA. Selecting the cognate tRNA is crucial for ensuring translation fidelity. Initial binding of tRNA to AARS is fast and non-specific and governed by electrostatic interaction. Upon initial binding, specific interactions between tRNA and AARS ensure recognition of the correct tRNA. Specific interactions are formed more slowly, accompanied by conformational changes in the active site of AARS. Specific interactions are mediated by identity elements such as modified nucleotides, conserved residues, base stacking, and different tRNA arm lengths. The identity elements include determinants and anti-determinants. Determinants favor binding of cognate tRNA with AARS while anti-determinants disfavor binding of non-cognate tRNA. The most commonly used identity elements are the anticodon bases 34, 35, and 36 in the anticodon arm and the discriminatory base 73 in the acceptor stem of the tRNA. In the case of AlaRS, tRNA is recognized exclusively based on the presence of G3-U70 base pair (McClain and Foss, 1988) . When non-cognate tRNAs were engineered *in vitro* containing the above base pair, AlaRS recognized those tRNAs and

charged them with alanine (Hou and Schimmel, 1988; McClain and Foss, 1988). For SerRS, the length of the variable arm is more crucial for its discrimination than sequence (Asahara *et al.*, 1993; Park and Schimmel, 1988). The complete list of identity elements and their location in the tRNA for each AARS from *E. coli* is provided below in Table 1. All AARS are classified into three groups based on the location of the tRNA identity elements. AARSs in group 1 have identity elements located in all regions of the tRNA, namely the acceptor stem, anticodon arm, other domains (T-arm, D-arm and the variable arm). AARSs in group 2 have identity elements located only in the acceptor stem and anticodon arm. AARSs in group 3 have identity elements located in the acceptor stem and other domains of the tRNA but not in anticodon arm. The list of AARSs in each group and the location of their identity elements on tRNA are provided in Figure 3.

Table 1: List of tRNA identity elements and their location on tRNA for aminoacylation by AARS from *E. coli*. Identity elements for fMet are provided in italics and bold.

AARS	Identity element location				References
	Acceptor Stem	Anticodon arm		Other domains (D-arm/T-arm/Variable arm)	
		Anticodon	other location		
Alanine	A73, G2:C71, G3:U70, G4:C69			G20	(Hou and Schimmel, 1988)(W H McClain and Foss, 1988)(Christopher, Jian-Ping and Paul, 1992)
Arginine	A/G73	C35, U/G36		A20	(W H McClain and Foss, 1988)(Schulman and Pelka, 1989)(McClain et al., 1990)(Tamura et al., 1992)
Asparagine	G73	G34, U35, U36			(Shimizu et al., 1992)(Li, Pelka and Schulman, 1993)
Aspartic acid	G73, G2:C71	G34, U35, C36,	C38	G10	(Hasegawa et al., 1989)(Nameki et al., 1992)
Cysteine	U73, G2:C71, C3:G70	G34, C35, A36		G15: G48, A13:A22	(Pallanck, Li and Schulman, 1992)(Shimizu et al., 1992)(Hou, Westhof and Giegé, 1993)(McClain, 1993)(Komatsoulis and Abelson, 1993)(Hamann and Hou, 1997)
Glutamine	G73, U1:A72, G2:C71, G3:C70	C/T34, U35, G36	A37, U38	G10	(Rogers and Soll, 1988)(Jahn, Rogers and Söll, 1991)(Hayase et al., 1992)(Ibba et al., 1996)

Glutamic acid	G1:C72, U2:A71	U34, U35,	A37	U11:A24, U13:G22-A46, 47	(Normanly et al., 1990)(Sylvers et al., 1993)(Gregory and Dahlberg, 1995)(Sekine et al., 1996)
Glycine	U73, G1:C72, C2:G71, G3:C70	C35, C36			(Shimizu et al., 1992)(Christopher, Jian-Ping and Paul, 1992)(McClain et al., 1991)
Histidine	C73, G1	Anticodon			(Shimizu et al., 1992)(Christopher, Jian-Ping and Paul, 1992)(Himeno et al., 1989)(Francklyn and Schimmel, 1990)(Yan and Francklyn, 1994)(Yan, Augustine and Francklyn, 1996)
Isoleucine	A73, C4:G69	G34, A35, U36	A37, A38	U12:A23, C29:G41	(Pallanck and Schulman, 1991)(Muramatsu et al., 1988)(Nureki et al., 1993)(Nureki et al., 1994)
Leucine	A73			U8:A14	(Normanly, Ollick and Abelson, 1992)(Asahara et al., 1993)
Lysine	A73	U34, U35, U36			(Normanly et al., 1990)(McClain et al., 1990)(Tamura et al., 1992)
Methionine ***fmet***	A73, U4:A69, A5:U68 ***G2:C71, C3:G70***	C34, A35, U36	***C32, U33, A37***		(Uemura et al., 1982)(H. and Heike, 1988)(Meinnel et al., 1993)(Lee et al., 1992)
Phenylalanine	A73	G34, A35, A36	G27:C43, G28:C42	U20, G44, U45, U59, U60	(Pallanck and Schulman, 1991)(William H McClain and Foss, 1988)(Peterson et al., 1993)(Peterson and Uhlenbeck, 1992)
Proline	A73, G72	G35, G36		G15:C48	(Shimizu et al., 1992)(McClain, Schneider and Gabriel, 1994)(Liu et al., 1995)

Serine	G73, C72, G2:C71, A3:U70, C11:G24, G/A4:T/C69			C11:G24 (Variable arm)	(Normanly, Ollick and Abelson, 1992)(Himeno et al., 1990)(Rogers and Soll, 1988)(Normanly et al., 1986)(Sampson and Saks, 1993)(Asahara et al., 1994)(Saks and Sampson, 1996)
Threonine	G1:C72, C2:G71	G34, G35, U36			(Schulman and Pelka, 1990)(Hasegawa et al., 1992)
Tryptophan	G73, A1:U72, G2:C71, G3:C70	C34, C35, A36			(Himeno et al., 1991)(Pak, Pallanck and Schulman, 1992)(Rogers et al., 1992)(Pak, Willis and Schulman, 1994)
Tyrosine	A73	U35			(Hou and Schimmel, 1989)(Bedouelle, 1990)(Himeno et al., 1990)(Sherman et al., 1992)
Valine	A73, G3:C70, U4:A69	A35, C36			(Chu and Horowitz, 1991)(Tamura et al., 1992)(Pallanck and Schulman, 1991)

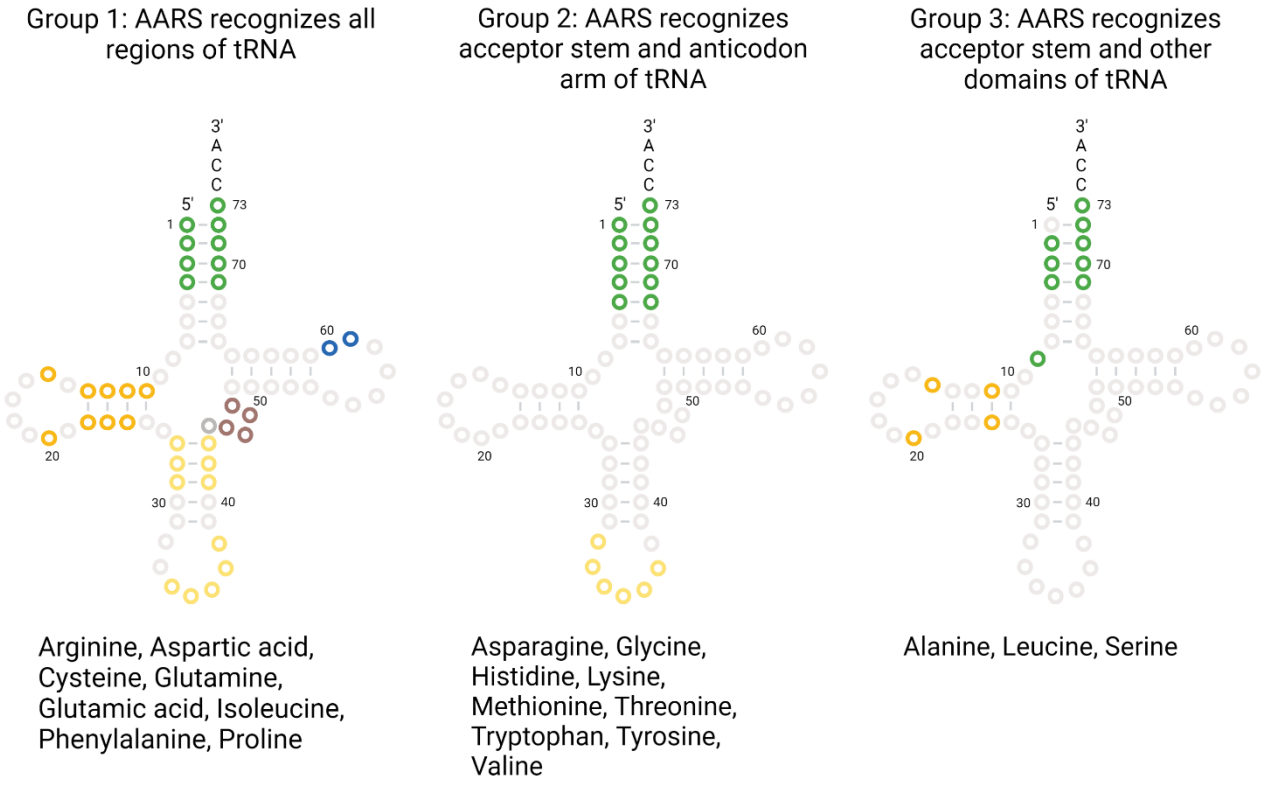

Figure 3: AARS identity elements on tRNA. AARSs are grouped based on the location of the identity elements present on tRNA regions. The list of AARS under each group and the tRNA bases utilized for recognition are highlighted.

Additional discrimination comes from the kinetic aspect of binding to discriminate cognate tRNAs. Aminoacylation with cognate tRNAs is more influenced by K_{cat} values than K_M values (Ebel *et al.*, 1973). Evolutionary conservation of the identity elements in tRNAs suggests their importance, even though these elements do not directly contribute to protein synthesis.

3.2.3 Proofreading mechanism by AARS

Pauling in 1958, theoretically predicted that amino acid misincorporation during translation should be about 1 in 200 (Pauling, 1958). However, *in vivo* experiments showed that this error rate is about 1 in 3000 (Loftfield, R. B., & Vanderjagt, 1972). Aminoacylation by AARS has an error rate of about 1 in 10000. This led to the suggestion of some editing mechanism being in place to account for these observations. The low error rate for AARSs is due to better recognition of cognate substrates and a proofreading/editing mechanism. This section briefly describes the editing/proofreading mechanism used by AARS to ensure faithful aminoacylation.

Fersht in 1977 proposed a “double sieve model” to explain the low error rate and presence of separate catalytic active and editing sites. According to this model, the active site of the enzyme acts as a first ‘coarse’ sieve to filter out amino acids that are larger than the cognate amino acid. Amino acids which are similar or smaller than cognate still will become activated in the active site. The second ‘fine’ sieve is the editing site capable of hydrolysis, which has a pocket size smaller than cognate amino acid. The

editing site serves to de-acylate any of the mischarged smaller amino acids which passed through the first sieve (Fersht, 1977). This way, de-acylation of cognate amino acid is prevented as it cannot enter the editing site. Evidence for the presence of a separate editing site is seen in 10 AARS from both Class I and II.

Editing activity can be divided into pre-transfer editing and post-transfer editing. In pre-transfer editing, the edit occurs before the amino acid is transferred to tRNA, and in post-transfer editing, editing occurs after amino acid is transferred to tRNA. Most AARSs use one of these editing mechanisms, but some AARS like LeuRS and ValRS use both mechanisms.

Pre-transfer editing occurs after the formation of amino acid adenylate (AA-AMP) but before transfer to tRNA. Pre-transfer editing is seen in both AARS classes. Pre-transfer editing can happen in two methods. In the first method, AA-AMP is released by the enzyme to the cytosol and the phosphoester bond is spontaneously hydrolyzed. In the second method, enzymatic hydrolysis of AA-AMP happens either in the active site or in a separate editing site. For example, thiolated non-proteogenic amino acids like homocysteine and ornithine are cleared by pre-transfer editing in the active site of the enzyme by MetRS and LysRS.

Post-transfer editing occurs after the transfer of amino acid to tRNA and happens in a separate editing site. This editing involves cleaving the ester bond between amino acid and tRNA. In general, once tRNA-AA is formed, amino acid triggers a conformational change in the 3' end of tRNA and results in tRNA translocation. Translocation results in an amino acid being in the editing site where it is hydrolyzed. For class II AARS, mischarged tRNA is rapidly released and in those cases, enzymes are capable of recapturing these mischarged tRNA for editing. In the case of PheRS, PheRS competes for Tyr-$tRNA^{Phe}$ with EF-TU to recapture and edit the tRNA.

The preference for the editing mechanism used is dependent on the rate of amino adenylate hydrolysis and transfer to tRNA. In case of a faster transfer rate to tRNA as in ValRS, post-transfer is preferred. For IleRS, both reaction rates are fairly equal and hence both editing mechanisms are used.

There also exists a separate group of editing proteins that act independently of AARS called trans-editing factors. These enzymes provide additional quality control in the editing process. The role of the trans-editing factors are to clear mischarged tRNAs before they reach the ribosome. D-aminoacyl-tRNA deacetylases are another class of trans-editing factors targeting in particular tRNAs charged with D-amino acids. The presence of such multiple editing mechanisms signifies the importance of the aminoacylation process and fidelity of the translation process. A schematic representation of various editing mechanism for aminoacylation is provided in Figure 4.

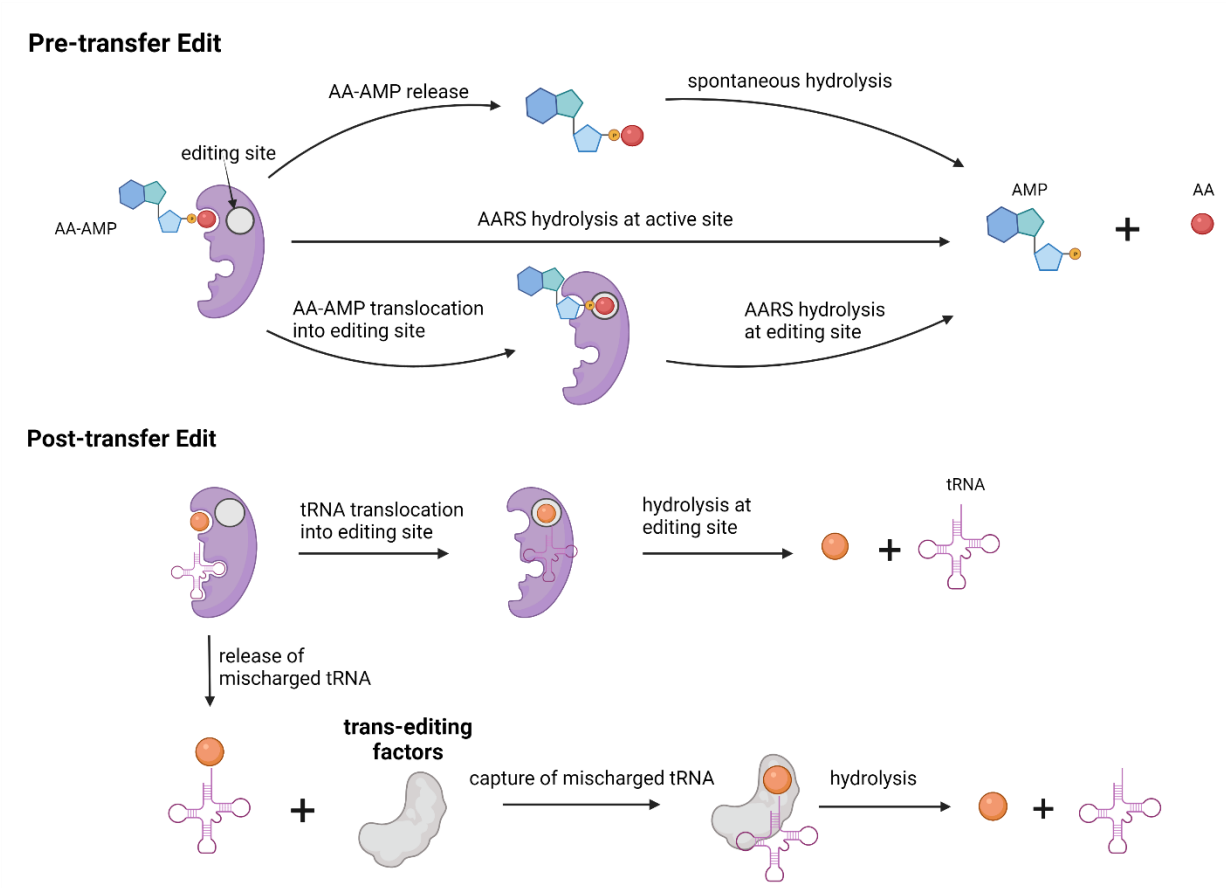

Figure 4: The editing mechanism for aminoacylation process. Pre-transfer edit occurs before amino acid gets charged onto tRNA. Post-transfer edit occurs after amino acid is charged onto tRNA. Trans editing factors are independent editing proteins involved in hydrolysis of mischarged tRNA.

3.2.4 Aminoacylation kinetics

The accuracy of protein synthesis relies on an AARS's ability to recognize cognate amino acid and tRNA. Kinetic analysis is useful to develop the mechanism of action in the steps involved in aminoacylation. This section provides information about different kinetic approaches used to study tRNA aminoacylation.

The parameters widely used to describe the kinetics of AARS are substrate affinity K_M, enzyme turnover K_{cat}, maximum velocity V_{max} and enzyme specificity K_{cat}/K_M. K_M refers to affinity of the enzyme to the substrate. K_{cat} is the catalytic constant for substrate to product conversion. K_{cat}/K_M is the specificity constant or catalytic efficiency of the enzyme. Steady-state kinetics is useful for initial characterization of the enzyme and to measure kinetic parameters. Since steady-state kinetics are used generally, parameters obtained can be compared across systems. For example, enzyme specificity for cognate and non-cognate amino acids can be measured and compared across AARS. Steady-state measurements are usually performed with substrate concentrations much higher than enzyme since the assay follows product formation. Minimal material requirements and fast readout make the steady-state approach suitable for initial characterization. The drawback of steady-state kinetics is that elementary reactions cannot be characterized. To determine the rate of the aminoacylation process, ATP pyrophosphate exchange assays and aminoacylation assays are performed under steady-state kinetics.

The ATP-PPi exchange assay is based on the amino acid activation step. ATP and amino acid form aminoacyl adenylate intermediate (AA-AMP) with the release of pyrophosphate (PPi). In one approach, [^{32}P]-PPi is used for the reaction. The radioactive group reacts with AA-AMP resulting in [^{32}P]-ATP. This assay measures the exchange of [^{32}P]-PPi into ATP to provide rate of the activation step. In another approach, radioactive [^{32}P]-ATP is used for amino acid activation and the rate of ATP consumption is measured using activated charcoal or thin layer chromatography (TLC) plates.

Aminoacylation assays are dependent on the second step of amino acid transfer to tRNA. Amino acids radiolabeled with [^{3}H] or [^{14}C] are used to measure the rate of product AA-tRNA^{AA} formation over time. One drawback of using radiolabeled amino acids is that attaining saturating conditions is difficult. This can be challenging while determining the K_M value of tRNA for AARS. The above limitation can be overcome by using a radiolabeled [^{32}P] group at 3' end of tRNA and unlabeled amino acids. In such a way, saturating amino acid concentration can be used in the assay. The table below contains the K_M and K_{cat} of amino acids for all AARS from *E. coli* unless mentioned otherwise.

Table 2: K_M and K_{cat} values of amino acids for AARS from *E. coli* unless mentioned otherwise. Unit for K_{cat} is s^{-1} unless mentioned otherwise.

AARS	Amino acid		References
	K_M (µm)	**K_{cat} (s^{-1})**	
AlaRS	240 ± 50	33 ± 7	(Hill and Schimmel, 1989)
ArgRS	12	2.2	(Lin, Wang and Wang, 1988)(Airas, 2006)
AsnRS	32	1.6	(Madern, Anselme and Härtlein, 1992)
AspRS	60	18	(Martin *et al.*, 1997)
CysRS	0.4 ± 0.1	680 ±60 (nmol min-1 mg protein-1)	(Komatsoulis and Abelson, 1993)
GlnRS	0.114 ± 0.012	157 ± 7 min-1	(Liu *et al.*, 1998)
GluRS	5	5.5 ± 1.0	(Lapointe and Söll, 1972)(Campanacci *et al.*, 2004)
GlyRS	0.03	0.31 ± 0.02 (*Homo sapiens*)	(Ostrem and Berg, 1974) *(Cader et al., 2007)*
HisRS	1.4 ± 0.6	2.4 ± 0.4	(Augustine and Francklyn, 1997)
IleRS	2.1 ± 0.2	3.1 ± 0.2	(Xu *et al.*, 1994)
LeuRS	15	3	(Chen *et al.*, 2000)
LysRS	230 ± 20	0.34 ± 0.009	(Wang *et al.*, 2006)
MetRS	1.2 ± 0.2	3.2 ± 0.2	(Ghosh *et al.*, 1991)
PheRS	1.8 ± 0.2	65 ± 3 min-1	(Moor, Klipcan and Safro, 2011)

ProRS	250 ± 35	70 ± 25	(Beuning and Musier-Forsyth, 2001)
SerRS	0.56 ± 0.15	2.6 ± 0.4	(Borel *et al.*, 1994)
ThrRS	12	0.3	(Hirsh, 1968)
TrpRS	0.53 ± 0.08	1.34 ± 0.26	(Chan and Koeppe, 1994)
TyrRS	3.3 ± 0.8	0.74± 0.06	(Hamano-Takaku *et al.*, 2000)
ValRS	4.3	13.9	(Tardif and Horowitz, 2004)

Pre-steady state kinetics is used to study elementary reaction steps. The pre-steady state kinetic approach is used to study fast reactions, in the order of a few milliseconds, present at an early stage of the interaction. This approach is best for understanding the mechanistic action of interaction. Parameters like individual rate constants of the reactants can be determined using pre-steady state kinetics. For example, substrate-binding order in the active site, formation, and consumption of intermediates can be studied by pre-steady state kinetics. Rapid chemical quench and stopped-flow fluorimetry are generally used to study AARS. Rapid kinetic approaches were used to mechanistically distinguish the two classes of AARS. As mentioned earlier, in class I AARSs product release of AA-tRNA^{aa} is the rate-limiting step, and in class II AARSs amino acid activation is the rate-limiting step.

Rapid chemical quench is a discontinuous assay providing a direct readout of the rate of the radiolabeled product formed. Stopped-flow fluorimetry is a continuous assay and provides an indirect readout of reaction progress. Progress is dependent on changes in intrinsic tryptophan fluorescence correlated to reaction chemistry.

The table below contains the half-lifes of activated tRNA-AA measured in *E. coli* (Hentzen, Mandel and Garel, 1972). The value represents the spontaneous hydrolysis rate of tRNA-AA at neutral or alkaline pH in a high ionic condition at 37°C. Under the same conditions, the stability of the ester bond depends purely on the amino acid attached to tRNA. Half-lives for all but tryptophan are presented and range from 2 to 65 min.

Table 3: Half-life values of tRNA-AA from *E. coli*. Values obtained based on ester bond hydrolysis under neutral or alkaline pH in a high ionic medium at 37 °C. (Hentzen, Mandel and Garel, 1972)

tRNA-aa	**$t_{1/2}$ (min)**
Ala	6
Arg	12
Asn	11
Asp	11
Cys	16
Gln	9

Glu	9
Gly	8
His	16
Ile	65
Leu	7
Lys	14
Met	12
Phe	16
Pro	2
Ser	17
Thr	38
Trp	-
Tyr	15
Val	60

4 Applications of tRNA and AARS in cell-free systems

4.1 tRNA and AARS *in vitro* synthesis

4.1.1 tRNA synthesis

The ability to synthesize tRNA *in vitro* was demonstrated in 1973 when T4 DNA was transcribed to yield T4 tRNAs when incubated with *E. coli* extract obtained after infecting with T4 bacteriophage (Nierlich *et al.*, 1973). Currently, a more sophisticated one-pot method for *in vitro* tRNA synthesis has been developed (Korencić, Söll and Ambrogelly, 2002). In this method, T7 RNAP is used for tRNA transcription from a ds/ssDNA hybrid template. Transcribed tRNAs were shown to be produced in full length and aminoacylated by AARS. This method did not have as high a yield as the control method using plasmid DNA as template, but is a useful approach to produce functional tRNAs *in vitro* (Korencić, Söll and Ambrogelly, 2002). *In vitro* synthesis of functional tRNA has made it easier to study interactions of tRNA with other components such as mRNA (Fauzi, Jack and Hines, 2005) and AARS (Wang *et al.*, 2015), to investigate tRNA stability *in vitro* (Serebrov *et al.*, 1998), and to determine mutational effects on tRNA function (Tamura *et al.*, 1992).

The ability of *in vitro* transcribed tRNA to decode codons on mRNA during translation was studied as well. In one study, native tRNAs were depleted from the cell lysate and replaced with an *in vitro* transcribed tRNA subset. The cell lysate containing replenished tRNA was found to produce proteins and was able to decode all 61 codons with 48 transcribed tRNAs (Cui *et al.*, 2015). This way, the tRNA pool directing the translation process can in principle be fully customized. In a similar study, efforts were taken to identify the minimal number of tRNAs required to support protein translation. The minimal set of tRNAs required for translation was explored by generating an entire set of 21 tRNAs *in vitro*, and they were shown to be able to synthesize proteins. A reduced number of tRNA was possible by using a single codon for each amino acid. The tRNAs were chosen such that they did not require

modifications after synthesis. Further, *in vitro* transcribed tRNAs provide flexibility in redesigning the genetic code and facilitate site-specific incorporation of amino acids. Protein yield obtained using this reduced set of tRNA was shown to be up to be 40% compared to the native system (Hibi *et al.*, 2020).

4.1.2 AARS synthesis

From a synthetic biology perspective, the ability to express and sustain proteins is crucial for building a self-replicating cell. To this end, AARS proteins were expressed individually in the PURE system, and it was demonstrated that all AARSs except PheRS were expressed as soluble proteins (Awai, Ichihashi and Yomo, 2015). These expressed proteins were functional with activity on par with their native counterparts purified from *E. coli*. The reason for the inactivity of PheRS was associated with insufficient formation of active dimers since dimer formation requires longer incubation at 4 °C and low salt concentrations *in vitro* (Awai, Ichihashi and Yomo, 2015). In a similar attempt, all 20 AARSs were expressed using a polycistronic plasmid encoding 32 proteins in total using the PURE system. The PURE system was able to synthesize all 32 proteins, including all 20 AARS as confirmed by mass spectrometric analysis but functionality was not assessed (Doerr *et al.*, 2021). Recently, in the bid to construct a self-replicating synthetic cell, a modified PURE system was used to demonstrate self-regeneration of up to 7 AARSs in a microfluidic reactor for more than 24 hours (Lavickova, Laohakunakorn and Maerkl, 2020).

4.2 tRNA and AARS engineering

4.2.1 tRNA engineering

Major work on tRNA engineering was achieved *in vivo* in regards to genetic code expansion (GCE). GCE involves increasing the genetic code alphabet by introducing new base nucleotides (Hoshika *et al.*, 2019), unnatural base pairs (UBP) (Mukba *et al.*, 2020), creating new codons and increasing the codon size to 4 (quadruplet codon) (Hohsaka *et al.*, 2001). The above approaches were used for incorporating modified or non-canonical amino acids (NC-AAs). Another approach is to reassign existing codons for NC-AA incorporation. Incorporating a NC-AA requires an orthogonal translation system (OTS) with a tRNA/AARS pair that does not cross-react with the endogenous tRNA/AARS present in the system. There should not be any interference of the endogenous AARS with the exogenous AARS in its ability to recognize the exogenous tRNA and NC-AA, and vise-versa. Orthogonality can be achieved by utilizing tRNA/AARS pairs from phylogenetically distinct species, this approach takes advantage of differences in codon usage. Another approach takes advantage of the difference in tRNA recognition by AARS (Doctor and Mudd, 1963). Orthogonal pairs TyrRS/tRNA-Tyr obtained from archaea *Methanococcus jannaschii* and pyrrolysyl-(Pyl)RS/tRNA-Pyl from *Methanosarcina barkeri* are most commonly used for incorporating NC-AAs. The open nature of Cell-Free Protein Synthesis (CFPS) provides a higher degree of freedom for structurally and functionally diverse NC-AA to be incorporated. Parameters like cellular toxicity, viability, and cross membrane transport are not constraints for CFPS allowing incorporation of many different NC-AAs.

Depending on the application, either site-specific or residue-specific incorporation is used. In site-specific incorporation a NC-AA is incorporated into a specific location whereas in residue specific incorporation a NC-AA is incorporated into all sites encoded by a specific codon. In site-specific incorporation, both sense codon and nonsense codons were utilized for NC-AA incorporation. Modifications are introduced at the identity elements present in the tRNA, mostly at anti-codon arm and in acceptor stem, to favor NC-AA incorporation. While using a nonsense codon, a stop codon (UAG, UGA, UAA) is reassigned to incorporate NC-AA instead. The amber codon (UAG) is usually used since this codon is least used as a termination signal in *E. coli*. In the input DNA sequence, all

instances of the amber codon are changed to either one of the other two stop codons which frees the amber codon for reassignment to a NC-AA. The stop codon UAG is then reassigned to the tRNA containing NC-AA and the tRNA anticodon arm is modified to CUA. With this approach, NC-AAs can be site-specifically incorporated at multiple places (Martin *et al.*, 2018; Hong *et al.*, 2014). Similarly tRNA suppressor targeted towards opal (UGA) and ochre (UAA) codons were developed and utilized for *in vitro* NC-AA incorporation (Gubbens *et al.*, 2010). As for sense codons, amino acids like leucine and arginine have up to 6 codons, therefore 6 tRNAs code for the same amino acid. One of the least commonly used isoacceptor tRNAs is generally used for codon reassignment. A report from 2016 demonstrated site-specific incorporation using sense codon reassignment using *in vitro* transcribed tRNAs charged with nc-aa in the PURE system. One of the codons for valine (GUG), arginine (CGC), and glycine (GGC) was reassigned to a different NC-AA (Iwane *et al.*, 2016). This approach suffered from low efficiency in NC-AA incorporation due to wobble decoding, the ability of the tRNA to recognize more than one codon. The commercially available FluoTech system has fluorescently modified lysine instead of lysine. The codon AAA for lysine is reassigned to BODIPY lysine. Here, tRNA pre-charged with BODIPY lysine is added to the *in vitro* protein expression system, incorporating fluorescent lysine into proteins. It should be noted that the native $tRNA_{AAA\text{-}lys}$ is still present in the system, leading to partial residue specific incorporation of the fluorescent lysine. This system is used for easy detection of *in vitro* protein expression. For residue-specific incorporation, all of the sense codons for a particular amino acid are reassigned to the NC-AA. The native amino acid is replaced by a NC-AA in the amino acid pool. All tRNAs now carry the NC-AA. Using an *E. coli* cell-free system canavanine amino acid, a toxic analog of arginine, was incorporated into model protein at all locations of arginine (Worst *et al.*, 2015).

As an alternative approach for NC-AA incorporation, quadruplet codons were used for tRNAs carrying the modified amino acids. The number of anticodons present in the anticodon arm is increased to 4 (UCCA) to match the corresponding 4 letter codon (AGGU). The modified amino acid was shown to be incorporated at multiple instances and up to two distinct NC-AA were incorporated with this approach. A modified nitrophenylalanine-tRNA was used to decode the 4 letter codons. Such a modified tRNA allowed incorporating NC-AA with better efficiency in an *E. coli in vitro* translation system (Hohsaka *et al.*, 2001). So far, tRNAs were engineered to incorporate NC-AA instead of a stop codon or amino acid (via sense codon). In an attempt to increase the amino acid diversity to more than 20, the codon table was split for amino acid arginine, glycine, and valine and, the free codons were assigned to three distinct NC-AA, thereby increasing the total number of amino acids to 23, in addition to the 20 native amino acids (Iwane *et al.*, 2016).

Charging of orthogonal tRNA with NC-AAs is a key step in the translation process and it is usually mediated by AARS enzymes. In addition to AARS enzymes, there are other methods such as chemoenzymatic methods, chemical methods, and ribozyme-based approaches for acylation to charge tRNA with NC-AAs. These synthetic methods are utilized *in vitro* to generate pre-charged tRNAs with NC-AAs and can be directly supplemented into the cell-free system. Such pre-charged tRNAs are useful to learn more about single turnover translation. The key advantage of chemical acylation methods being that structurally and chemically diverse groups can be added onto tRNAs without the need to re-engineer AARS.

4.2.2 AARS engineering

The central theme for AARS engineering is in genetic code expansion and incorporation of unnatural or NC-AAs during peptide synthesis. AARSs can be engineered by modifying the amino acid binding pocket, tRNA binding pocket, and editing domain for acylating NC-AA onto tRNA. In a relatively simple approach, native AARSs are used to incorporate unnatural amino acids. Here, the lack in

specificity in substrate recognition by native AARSs was exploited to charge tRNAs with unnatural amino acids to participate in protein synthesis. For example, research in non-ribosomal peptides utilized this approach to demonstrate simultaneously incorporation of 10 different amino acid analogs using the PURE based recombinant system. The amino acid analogs used were substrates for 12 native AARSs from *E. coli* (Josephson, Hartman and Szostak, 2005). Even though it is simple, this approach does suffer from low efficiency and therefore yield. Moreover, if cognate amino acids are present, they will compete with the same codon and lower the incorporation of NC-AA. In such cases, the relative ratio of NC-AA to cognate amoni acid should be tightly controlled to favor NC-AA incorporation.

The second approach uses engineered AARSs with altered specificity for amino acid or tRNA. In the report mentioned above, AARS substrate diversity was further expanded by a mutation in the binding pocket domain and editing domain to accept previously unaccepted amino acids. For example, in PheRS, a specific mutation at binding pocket domain Ala294Gly can accept the phenylalanine analog p-iodo-Phe. Similarly, inactivating the editing domain of LeuRS by Asp345Ala accepted allylglycine (Josephson, Hartman and Szostak, 2005).

Cell-free systems opened up the use of insoluble, non-canonical AARS for NC-AA incorporation. Pyrrolysyl synthetase (PylRS) is one such AARS which had limited use due to difficulties in purification. PylRS and its mutants are utilized in more than 100 unnatural amino acid incorporations (Wan, Tharp and Liu, 2014) and are highly useful due to their minimal cross-reactivity with the *E. coli* system. Expression of PylRS/tRNA pair from *Methanosarcina mazei* in *E. coli* helped obtain a cell-free extract containing the insoluble PylRS. This cell-free extract was used to demonstrate the site-specific incorporation of two analogs of pyrrolysine into a reporter gene (Chemla *et al.*, 2015). However site-specific incorporation requires a faster and more accurate aminoacylation reaction between NC-AA and tRNA, there is a need for a better strategy to identify mutants with better incorporation efficiency.

Identification of correct mutants of orthogonal AARS for NC-AA incorporation is largely benefited from methods like Multiplex Automated Genome Engineering (MAGE) and Phage Assisted Continuous Evolution (PAGE). These directed evolution-guided methods helped generate and screen a large library of AARS mutants in a short period, thereby saving time and effort. Despite the above methods being developed *in vivo*, orthogonal AARS can also be used in *in vitro* systems for high efficiency NC-AA incorporation (Amiram *et al.*, 2015; Bryson *et al.*, 2017). Similarly, computational techniques have been used to generate and screen AARS mutant libraries specifically targeted for ortho-nitrobenzyl tyrosine, an analog of tyrosine. The advent of such methods will pave the way for better and faster engineering of novel AARSs (Baumann *et al.*, 2019).

4.2.2.1 Aminoacylation methods

AARS enzymatic reactions are not the only way to charge tRNAs with a NC-AA. Other methods to charge tRNAs include the chemoenzymatic method, chemical method, and ribozyme method (Figure 4). In the chemoenzymatic method, the first step is the chemical acylation of hybrid dinucleotide 5'-phospho-2'-deoxyribocytidylylriboadenosine (pdCpA) using an activated amino acid donor with N-protected group. In the following step, the acylated dinucleotide is enzymatically ligated via T4 ligase to truncated tRNA lacking 3'-CA dinucleotide (Hecht *et al.*, 1978). Many functional groups can be ligated to tRNA with this approach, but it is cumbersome due to the laborious chemical process involved in the preparation of acylated nucleotides.

In the chemical method, the side chains of the amino acids are chemically modified after tRNAs were charged with canonical amino acids by the AARS. One can consider this method similar to post-

translational modification of proteins, instead, here it is the charged tRNA that is modified by a chemical reaction. Diverse functional groups like N-methyl amino acid (Merryman and Green, 2004), glycosyl amino acid (Fahmi *et al.*, 2007), and fluorescently-labeled amino acids (Iijima and Hohsaka, 2009) were generated with this approach. However, similar to the chemoenzymatic method, this method is laborious, technically demanding, and suffers from poor efficiency due to cyclic-tRNA by-product which inhibits protein synthesis, and also the short lifetime of aminoacylated tRNA (Yamanaka *et al.*, 2004).

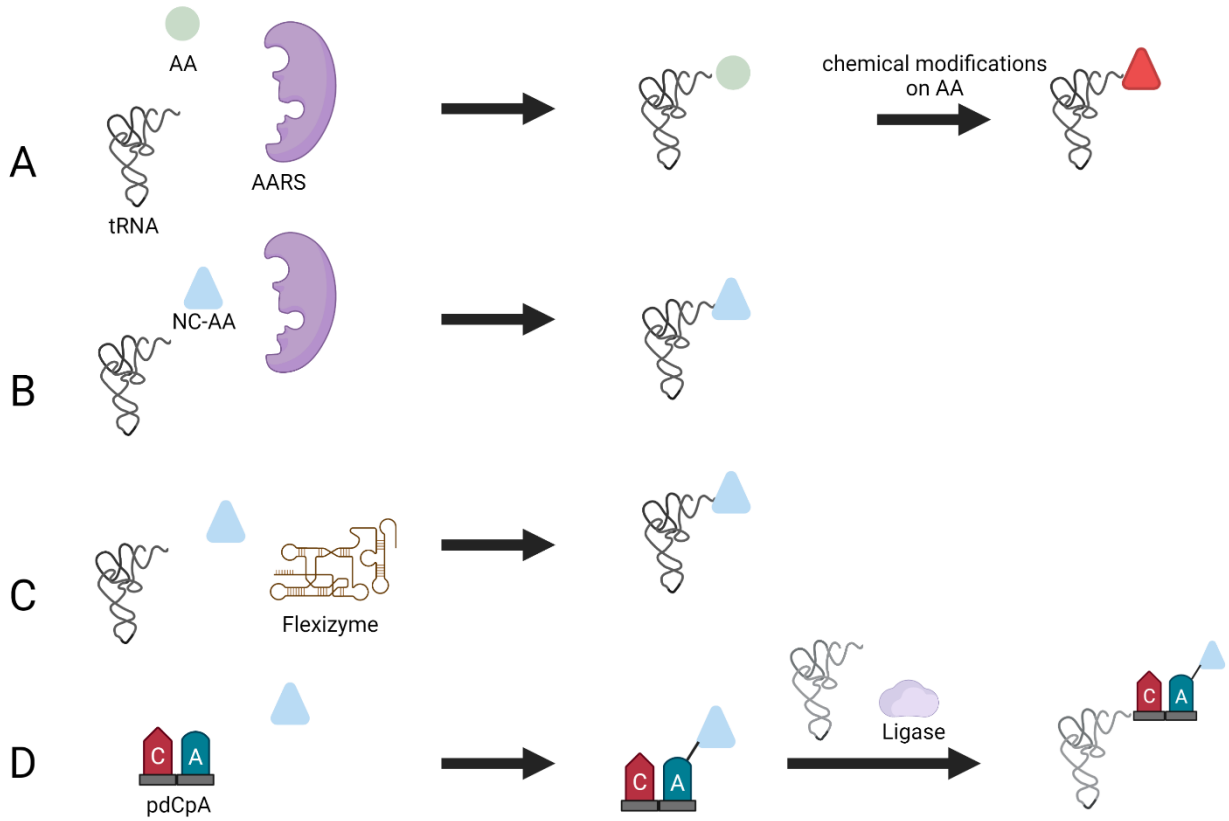

Figure 4: Different aminoacylation methods used for nc-aa incorporation. A) Chemical method, B) Enzymatic aminoacylation by AARS, C) Flexizyme method, D) Chemoenzymatic method.

Another approach for incorporating NC-AA that avoids the use of AARS altogether is performed by ribozymes called flexizymes. Flexizyme is a small artificial ribozyme (44 - 46 nucleotides) capable of generating NC-AA tRNAs (Murakami *et al.*, 2006). Flexizyme is derived from acyl-transferase ribozyme through directed evolution and sequence optimization. Flexizyme specifically recognize amino acids with an activated carboxyl group and charge the 3'-CCA end of tRNA irrespective of tRNA body and anticodon sequence (Xiao *et al.*, 2008). Flexizyme classes were expanded lately such that they can accept amino acids with a different active groups. The complete set of flexizymes includes dinitro flexizyme (dFx) recognizing amino acid with activated dinitrobenzyl ester, enhanced flexizyme (eFx) recognizing amino acid with chlorobenzyl ester, and amino flexizyme (aFx) recognizing amino acid with amino derivatized benzyl thioester group (Morimoto *et al.*, 2011). Flexizymes can accept almost all amino acids as acyl-donor substrates and expand the diversity of amino acids that can be incorporated. For instance, flexizyme mediated incorporation of NC-AA with D-α-amino acids (Katoh,

Tajima and Suga, 2017), β-amino acids (Katoh and Suga, 2018), γ-amino acids (Ohshiro *et al.*, 2011), N-alkyl-L- α-amino acids (Kawakami, Ishizawa and Murakami, 2013), benzoic acids (Kawakami *et al.*, 2016) and exotic peptides (Goto and Suga, 2009) to highlight a few. Aminoacylation is performed by incubating the desired tRNA and amino acid together to yield charged tRNA (Goto, Katoh and Suga, 2011). The only limitation that exists from the amino acid is the ability to activate the carboxyl group with a certain group and the chemical stability of amino acids during aminoacylation. This technique has opened up the possibility to charge any tRNA with nearly any amino acid and thereby reassigning any codon with NC-AAs. The *in vitro* translation system based on flexizyme and translation apparatus is called flexible *in vitro* translation system (FIT) and has been used for genetic code reprogramming (Goto, Katoh and Suga, 2011).

5 Challenges and Future Directions

Protein synthesis is a highly complex process with many players involved that insure fidelity. The process of translation has evolved over billions of years to attain the current state of efficiency and quality control. Attempts to modify, alter, or improve the efficiency of translation to suit a novel application require efforts at multiple levels *in vivo* to reach the desired function. CFPS systems eliminate some of the stringency associated with *in vivo* translation and simplified engineering of the protein synthesis process to some extent. For genetic code expansion, CFPS has opened new avenues for incorporating a diverse range of NC-AAs and helped achieve proteins with novel functional groups. The development of an orthogonal translation system requires a coordinated effect at every level of protein expression. Apart from tRNA and AARS engineering, engineering efforts on other translational elements such as elongation factors (Gan *et al.*, 2017; Fan, Ip and Söll, 2016), initiation factors (Goto and Suga, 2009; Goto *et al.*, 2008), ribosomes (Jewett *et al.*, 2013; Semrad and Green, 2002; Li, Haas, *et al.*, 2017), and termination factors (Korkmaz and Sanyal, 2017; Hong *et al.*, 2015) helped in the development of a better translational system. One key challenge lies in bringing together these individually engineered components. CFPS research will continue to increase our understanding and the capability to engineer improved translational system. Failures will reveal the gaps in our understanding and guide our scientific research. Despite ongoing effects, many areas require advancements. For instance, the efficiency of NC-AA is low due to the non-compatibility of structurally diverse NC-AA with ribosomes and elongation factors (Fujino *et al.*, 2013). This requires better engineered translation factors to successfully incorporate NC-AA with high efficiency. The advent of high-throughput methods and computational techniques like MAGE and PACE quickened the pace of library generation and screening, leading to faster protein engineering. GCE has increased amino acid repertoires with diverse functional and structural groups and has resulted in the creation of novel proteins that could not be synthesized before. The use of such novel proteins could be in creating better drugs, and functional biomaterials. Improvements and better characterization of translation elements, like AARS and tRNA, using the latest techniques will help to standardize translation elements and aid in creating a predictable biological system. From a synthetic biology perspective, engineering life from scratch remains a grand challenge in the field. Towards that goal, different aspects of living systems have been reconstituted *in vitro* including ATP synthesis (Berhanu, Ueda and Kuruma, 2019), DNA replication (Kurihara *et al.*, 2011), PURE protein component self-regeneration (Lavickova, Laohakunakorn and Maerkl, 2020), and ribosomal components (Jewett *et al.*, 2013). With respect to the PURE system, the efficiency is improved by addition of external components into the system such as EF-P, EF4, ArfA (Li, Zhang, *et al.*, 2017). Cell-free transcription and translation systems have enormous potential to overcome the limits of cell-based protein synthesis and could

become the next generation platform for protein engineering that can go well beyond the scope of what could be accomplished in a cellular environment.

6 Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7 Author Contributions

R.B.G. and S.J.M. wrote this review.

8 Funding

This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program Grant 723106 and a Swiss National Science Foundation grant (182019).

9 References

Airas, R. K. (2006) 'Analysis of the kinetic mechanism of arginyl-tRNA synthetase', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1764(2), pp. 307–319. doi: 10.1016/j.bbapap.2005.11.020.

Amiram, M. *et al.* (2015) 'Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids', *Nature biotechnology*. 2015/11/16, 33(12), pp. 1272–1279. doi: 10.1038/nbt.3372.

Asahara, H. *et al.* (1993) 'Recognition nucleotides of Escherichia coli tRNA(Leu) and its elements facilitating discrimination from tRNASer and tRNA(Tyr).', *Journal of molecular biology*, 231(2), pp. 219–229. doi: 10.1006/jmbi.1993.1277.

Augustine, J. and Francklyn, C. (1997) 'Design of an Active Fragment of a Class II Aminoacyl-tRNA Synthetase and Its Significance for Synthetase Evolution', *Biochemistry*, 36(12), pp. 3473–3482. doi: 10.1021/bi962395y.

Avery, O. T., Macleod, C. M. and McCarty, M. (1944) 'Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III', *The Journal of experimental medicine*, 79(2), pp. 137–158. doi: 10.1084/jem.79.2.137.

Awai, T., Ichihashi, N. and Yomo, T. (2015) 'Activities of 20 aminoacyl-tRNA synthetases expressed in a reconstituted translation system in Escherichia coli', *Biochemistry and Biophysics Reports*, 3, pp. 140–143. doi: 10.1016/j.bbrep.2015.08.006.

Baumann, T. *et al.* (2019) 'Computational Aminoacyl-tRNA Synthetase Library Design for Photocaged Tyrosine.', *International journal of molecular sciences*, 20(9). doi: 10.3390/ijms20092343.

Berhanu, S., Ueda, T. and Kuruma, Y. (2019) 'Artificial photosynthetic cell producing energy for protein synthesis', *Nature Communications*, 10(1). doi: 10.1038/s41467-019-09147-4.

610 Borel, F. *et al.* (1994) 'Seryl-tRNA synthetase from Escherichia coli: Implication of its N-terminal
611 domain in aminoacylation activity and specificity', *Nucleic Acids Research*, 22(15), pp. 2963–2969.
612 doi: 10.1093/nar/22.15.2963.

613 Bryson, D. I. *et al.* (2017) 'Continuous directed evolution of aminoacyl-tRNA synthetases.', *Nature*
614 *chemical biology*, 13(12), pp. 1253–1260. doi: 10.1038/nchembio.2474.

615 Buchner, E. (1897) 'Alkoholische Gärung ohne Hefezellen', *Berichte der deutschen chemischen*
616 *Gesellschaft*, 30(1), pp. 1110–1113. doi: 10.1002/cber.189703001215.

617 Cader, M. Z. *et al.* (2007) 'Crystal structure of human wildtype and S581L-mutant glycyl-tRNA
618 synthetase, an enzyme underlying distal spinal muscular atrophy', *FEBS Letters*, 581(16), pp. 2959–
619 2964. doi: 10.1016/j.febslet.2007.05.046.

620 Cai, Q. *et al.* (2015) 'A simplified and robust protocol for immunoglobulin expression in Escherichia
621 coli cell-free protein synthesis systems', *Biotechnology Progress*, 31(3), pp. 823–831. doi:
622 10.1002/btpr.2082.

623 Campanacci, V. *et al.* (2004) 'The Escherichia coli YadB Gene Product Reveals a Novel Aminoacyl-
624 tRNA Synthetase Like Activity', *Journal of Molecular Biology*, 337(2), pp. 273–283. doi:
625 10.1016/j.jmb.2004.01.027.

626 Chan, K. W. and Koeppe, R. E. (1994) 'Role of the TIGN sequence in E. coli tryptophanyl-tRNA
627 synthetase', *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 1205(2), pp.
628 223–229. doi: 10.1016/0167-4838(94)90237-2.

629 Chargaff, E. (1950) 'Chemical specificity of nucleic acids and mechanism of their enzymatic
630 degradation.', *Experientia*, 6(6), pp. 201–209. doi: 10.1007/BF02173653.

631 Chemla, Y. *et al.* (2015) 'Genetically expanded cell-free protein synthesis using endogenous
632 pyrrolysyl orthogonal translation system', *Biotechnology and Bioengineering*, 112(8), pp. 1663–
633 1672. doi: 10.1002/bit.25587.

634 Crick, F. (1955) 'On degenerate templates and the adaptor hypothesis.', *RNA Tie Club*. Available at:
635 https://wellcomecollection.org/works/j7qpsmdm.

636 Crick, F. H. C. (1968) 'The origin of the genetic code', *Journal of Molecular Biology*, 38(3), pp.
637 367–379. doi: 10.1016/0022-2836(68)90392-6.

638 Cui, Z. *et al.* (2015) 'Semisynthetic tRNA Complement Mediates in Vitro Protein Synthesis',
639 *Journal of the American Chemical Society*, 137(13), pp. 4404–4413. doi: 10.1021/ja5131963.

640 Doctor, B. P. and Mudd, J. A. (1963) 'Species Specificity of Amino Acid Acceptor Ribonucleic Acid
641 and Aminoacyl Soluble Ribonucleic Acid Synthetases', *Journal of Biological Chemistry*, 238(11),
642 pp. 3677–3681. doi: 10.1016/s0021-9258(19)75325-3.

643 Doerr, A. *et al.* (2021) 'In vitro synthesis of 32 translation-factor proteins from a single template
644 reveals impaired ribosomal processivity', *Scientific Reports*, 11(1), pp. 1–12. doi: 10.1038/s41598-
645 020-80827-8.

Dondapati, S. K. *et al.* (2020) 'Cell-Free Protein Synthesis: A Promising Option for Future Drug Development', *BioDrugs*, 34(3), pp. 327–348. doi: 10.1007/s40259-020-00417-y.

Dopp, J. L., Jo, Y. R. and Reuel, N. F. (2019) 'Methods to reduce variability in E. Coli-based cell-free protein expression experiments', *Synthetic and Systems Biotechnology*, 4(4), pp. 204–211. doi: 10.1016/j.synbio.2019.10.003.

Ebel, J. P. *et al.* (1973) 'Factors determining the specificity of the tRNA aminoacylation reaction: Non-absolute specificity of tRNA-aminoacyl-tRNA synthetase recognition and particular importance of the maximal velocity', *Biochimie*, 55(5), pp. 547–557. doi: 10.1016/S0300-9084(73)80415-8.

Fahmi, N. E. *et al.* (2007) 'Site-Specific Incorporation of Glycosylated Serine and Tyrosine Derivatives into Proteins', *Journal of the American Chemical Society*, 129(12), pp. 3586–3597. doi: 10.1021/ja067466n.

Fan, C., Ip, K. and Söll, D. (2016) 'Expanding the genetic code of Escherichia coli with phosphotyrosine.', *FEBS letters*, pp. 3040–3047. doi: 10.1002/1873-3468.12333.

Fauzi, H., Jack, K. D. and Hines, J. V (2005) 'In vitro selection to identify determinants in tRNA for Bacillus subtilis tyrS T box antiterminator mRNA binding ', *Nucleic Acids Research*, 33(8), pp. 2595–2602. doi: 10.1093/nar/gki546.

Fersht, A. R. (1977) 'Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase.', *Biochemistry*, 16(5), pp. 1025–1030. doi: 10.1021/bi00624a034.

Fujino, T. *et al.* (2013) 'Reevaluation of the d-Amino Acid Compatibility with the Elongation Event in Translation', *Journal of the American Chemical Society*, 135(5), pp. 1830–1837. doi: 10.1021/ja309570x.

Gagoski, D. *et al.* (2016) 'Performance benchmarking of four cell-free protein expression systems.', *Biotechnology and bioengineering*, 113(2), pp. 292–300. doi: 10.1002/bit.25814.

Gamow, G. (1954) 'Possible Relation between Deoxyribonucleic Acid and Protein Structures', *Nature*, 173(4398), p. 318. doi: 10.1038/173318a0.

Gan, R. *et al.* (2017) 'Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins.', *Biotechnology and bioengineering*, 114(5), pp. 1074–1086. doi: 10.1002/bit.26239.

Ghosh, G. *et al.* (1991) 'Activation of Methionine by Escherichia Coli Methionyl-tRNA Synthetase', *Biochemistry*, 30(40), pp. 9569–9575. doi: 10.1021/bi00104a002.

Gomez, M. A. R. and Ibba, M. (2020) 'Aminoacyl-tRNA synthetases', *Rna*, 26(8), pp. 910–936. doi: 10.1261/rna.071720.119.

Goto, Y. *et al.* (2008) 'Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides', *ACS Chemical Biology*, 3(2), pp. 120–129. doi: 10.1021/cb700233t.

Goto, Y., Katoh, T. and Suga, H. (2011) 'Flexizymes for genetic code reprogramming', *Nature Protocols*, 6(6), pp. 779–790. doi: 10.1038/nprot.2011.331.

Goto, Y. and Suga, H. (2009) 'Translation Initiation with Initiator tRNA Charged with Exotic Peptides', *Journal of the American Chemical Society*, 131(14), pp. 5040–5041. doi: 10.1021/ja900597d.

Gubbens, J. *et al.* (2010) 'In vitro incorporation of nonnatural amino acids into protein using tRNA(Cys)-derived opal, ochre, and amber suppressor tRNAs', *RNA (New York, N.Y.)*. 2010/06/25, 16(8), pp. 1660–1672. doi: 10.1261/rna.2024810.

Hamano-Takaku, F. *et al.* (2000) 'A mutant Escherichia coli tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine', *Journal of Biological Chemistry*, 275(51), pp. 40324–40328. doi: 10.1074/jbc.M003696200.

Hecht, S. M. *et al.* (1978) '"Chemical aminoacylation" of tRNA's.', *Journal of Biological Chemistry*, 253(13), pp. 4517–4520. doi: 10.1016/s0021-9258(17)30417-9.

Hentzen, D., Mandel, P. and Garel, J. P. (1972) 'Relation between aminoacyl-tRNA stability and the fixed amino acid.', *Biochimica et biophysica acta*, 281(2), pp. 228–232. doi: 10.1016/0005-2787(72)90174-8.

Hershey, A. D. and Chase, M. (1952) 'Independent functions of viral protein and nucleic acid in growth of bacteriophage', *The Journal of general physiology*, 36(1), pp. 39–56. doi: 10.1085/jgp.36.1.39.

Hibi, K. *et al.* (2020) 'Reconstituted cell-free protein synthesis using in vitro transcribed tRNAs', *Communications Biology*, 3(1), pp. 1–11. doi: 10.1038/s42003-020-1074-2.

Hoagland, M. B. *et al.* (1958) 'A soluble ribonucleic acid intermediate in protein synthesis.', *The Journal of biological chemistry*, 231(1), pp. 241–257. doi: 10.1016/s0021-9258(19)77302-5.

Hoagland, M. B., Keller, E. B. and Zamecnik, P. C. (1956) 'Enzymatic carboxyl activation of amino acids', *Journal of Biological Chemistry*, 218(1), pp. 345–358. doi: 10.1016/S0021-9258(18)65898-3.

Hohsaka, T. *et al.* (2001) 'Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system.', *Biochemistry*, 40(37), pp. 11060–11064. doi: 10.1021/bi0108204.

Hong, S. H. *et al.* (2014) 'Cell-free Protein Synthesis from a Release Factor 1 Deficient Escherichia coli Activates Efficient and Multiple Site-specific Nonstandard Amino Acid Incorporation', *ACS Synthetic Biology*, 3(6), pp. 398–409. doi: 10.1021/sb400140t.

Hong, S. H. *et al.* (2015) 'Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1.', *Chembiochem : a European journal of chemical biology*, 16(5), pp. 844–853. doi: 10.1002/cbic.201402708.

Hoshika, S. *et al.* (2019) 'Hachimoji DNA and RNA: A genetic system with eight building blocks', *Science (New York, N.Y.)*, 363(6429), pp. 884–887. doi: 10.1126/science.aat0971.

Hou, Y.-M. and Schimmel, P. (1988) 'A simple structural feature is a major determinant of the identity of a transfer RNA', *Nature*, 333(6169), pp. 140–145. doi: 10.1038/333140a0.

718 Hunter, D. J. B. *et al.* (2018) 'Unexpected instabilities explain batch-to-batch variability in cell-free
719 protein expression systems.', *Biotechnology and bioengineering*, 115(8), pp. 1904–1914. doi:
720 10.1002/bit.26604.

721 Iijima, I. and Hohsaka, T. (2009) 'Position-specific incorporation of fluorescent non-natural amino
722 acids into maltose-binding protein for detection of ligand binding by FRET and fluorescence
723 quenching.', *Chembiochem : a European journal of chemical biology*, 10(6), pp. 999–1006. doi:
724 10.1002/cbic.200800703.

725 Iwane, Y. *et al.* (2016) 'Expanding the amino acid repertoire of ribosomal polypeptide synthesis via
726 the artificial division of codon boxes', *Nature Chemistry*, 8(4), pp. 317–325. doi:
727 10.1038/nchem.2446.

728 Jacob, F. and Monod, J. (1961) 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal
729 of Molecular Biology*, 3(3), pp. 318–356. doi: 10.1016/S0022-2836(61)80072-7.

730 Jewett, M. C. *et al.* (2013) 'In vitro integration of ribosomal RNA synthesis, ribosome assembly, and
731 translation', *Molecular systems biology*, 9, p. 678. doi: 10.1038/msb.2013.31.

732 Josephson, K., Hartman, M. C. T. and Szostak, J. W. (2005) 'Ribosomal Synthesis of Unnatural
733 Peptides', *Journal of the American Chemical Society*, 127(33), pp. 11727–11735. doi:
734 10.1021/ja0515809.

735 Katoh, T. and Suga, H. (2018) 'Ribosomal Incorporation of Consecutive β-Amino Acids', *Journal of
736 the American Chemical Society*, 140(38), pp. 12159–12167. doi: 10.1021/jacs.8b07247.

737 Katoh, T., Tajima, K. and Suga, H. (2017) 'Consecutive Elongation of D-Amino Acids in
738 Translation.', *Cell chemical biology*, 24(1), pp. 46–54. doi: 10.1016/j.chembiol.2016.11.012.

739 Kawakami, T. *et al.* (2016) 'Directed Evolution of a Cyclized Peptoid-Peptide Chimera against a
740 Cell-Free Expressed Protein and Proteomic Profiling of the Interacting Proteins to Create a Protein-
741 Protein Interaction Inhibitor.', *ACS chemical biology*, 11(6), pp. 1569–1577. doi:
742 10.1021/acschembio.5b01014.

743 Kawakami, T., Ishizawa, T. and Murakami, H. (2013) 'Extensive Reprogramming of the Genetic
744 Code for Genetically Encoded Synthesis of Highly N-Alkylated Polycyclic Peptidomimetics',
745 *Journal of the American Chemical Society*, 135(33), pp. 12297–12304. doi: 10.1021/ja405044k.

746 Korencić, D., Söll, D. and Ambrogelly, A. (2002) 'A one-step method for in vitro production of
747 tRNA transcripts.', *Nucleic acids research*, 30(20), pp. 1–4. doi: 10.1093/nar/gnf104.

748 Korkmaz, G. and Sanyal, S. (2017) 'R213I mutation in release factor 2 (RF2) is one step forward for
749 engineering an omnipotent release factor in bacteria Escherichia coli.', *The Journal of biological
750 chemistry*, 292(36), pp. 15134–15142. doi: 10.1074/jbc.M117.785238.

751 Krüger, A. *et al.* (2020) 'Development of a clostridia-based cell-free system for prototyping genetic
752 parts and metabolic pathways', *Metabolic Engineering*, 62, pp. 95–105. doi:
753 10.1016/j.ymben.2020.06.004.

754 Kurihara, K. *et al.* (2011) 'Self-reproduction of supramolecular giant vesicles combined with the

amplification of encapsulated DNA.', *Nature chemistry*, 3(10), pp. 775–781. doi: 10.1038/nchem.1127.

Laohakunakorn, N. *et al.* (2020) 'Bottom-Up Construction of Complex Biomolecular Systems With Cell-Free Synthetic Biology', *Frontiers in Bioengineering and Biotechnology*. Available at: https://www.frontiersin.org/article/10.3389/fbioe.2020.00213.

Lavickova, B., Laohakunakorn, N. and Maerkl, S. J. (2020) 'A partially self-regenerating synthetic cell', *Nature Communications*, 11(1), pp. 1–11. doi: 10.1038/s41467-020-20180-6.

Lavickova, B. and Maerkl, S. J. (2019) 'A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-Free System', *ACS Synthetic Biology*, pp. 1–25. doi: 10.1021/acssynbio.8b00427.

Levene, P. A. (1919) 'The Structure of Yeast Nucleic Acid: IV. Ammonia hydrolysis', *Journal of Biological Chemistry*, 40(2), pp. 415–424. doi: 10.1016/S0021-9258(18)87254-4.

Li, J. *et al.* (2014) 'Improved cell-free RNA and protein synthesis system', *PLoS ONE*, 9(9). doi: 10.1371/journal.pone.0106232.

Li, J., Haas, W., *et al.* (2017) 'Cogenerating Synthetic Parts toward a Self-Replicating System', *ACS Synthetic Biology*, 6(7), pp. 1327–1336. doi: 10.1021/acssynbio.6b00342.

Li, J., Zhang, C., *et al.* (2017) 'Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system', *Translation*, 5(1), p. e1327006. doi: 10.1080/21690731.2017.1327006.

Loftfield, R. B., & Vanderjagt, D. (1972) 'The frequency of errors in protein biosynthesis.', *The Biochemical journal*, 128(5), pp. 1353–1356. doi: 10.1042/bj1281353.

Maddalena, L. L. de *et al.* (2016) 'GreA and GreB Enhance Expression of Escherichia coli RNA Polymerase Promoters in a Reconstituted Transcription–Translation System', *ACS Synthetic Biology*, 5(9), pp. 929–935. doi: 10.1021/acssynbio.6b00017.

Martin, R. W. *et al.* (2018) 'Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids', *Nature Communications*, 9(1), p. 1203. doi: 10.1038/s41467-018-03469-5.

McClain, W. H. and Foss, K. (1988) 'Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end.', *Science (New York, N.Y.)*, 240(4853), pp. 793–796. doi: 10.1126/science.2452483.

Merryman, C. and Green, R. (2004) 'Transformation of Aminoacyl tRNAs for the In Vitro Selection of "Drug-like" Molecules', *Chemistry & Biology*, 11(4), pp. 575–582. doi: 10.1016/j.chembiol.2004.03.009.

Miescher, F. (1869) 'Letter i; to wilhelm his; tübingen, february 26th, 1869', *Die histochemischen und physiologischen arbeiten von Friedrich Miescher-aus dem wissenschaftlichen Briefwechsel von F. Miescher*, 1, pp. 33–38.

Moore, S. J. *et al.* (2018) 'Rapid acquisition and model-based analysis of cell-free transcription–

translation reactions from nonmodel bacteria', *Proceedings of the National Academy of Sciences*, 115(19), p. E4340 LP-E4349. doi: 10.1073/pnas.1715806115.

Morimoto, J. *et al.* (2011) 'Flexizymes: their evolutionary history and the origin of catalytic function.', *Accounts of chemical research*, 44(12), pp. 1359–1368. doi: 10.1021/ar2000953.

Mukba, S. A. *et al.* (2020) '[Expanding the Genetic Code: Unnatural Base Pairs in Biological Systems].', *Molekuliarnaia biologiia*, 54(4), pp. 531–541. doi: 10.31857/S0026898420040126.

Murakami, H. *et al.* (2006) 'A highly flexible tRNA acylation method for non-natural polypeptide synthesis', *Nature Methods*, 3(5), pp. 357–359. doi: 10.1038/nmeth877.

Niederholtmeyer, H. *et al.* (2015) 'Rapid cell-free forward engineering of novel genetic ring oscillators.', *eLife*, 4, p. e09771. doi: 10.7554/eLife.09771.

Niederholtmeyer, H., Stepanova, V. and Maerkl, S. J. (2013) 'Implementation of cell-free biological networks at steady state', *Proceedings of the National Academy of Sciences*, 110(40), pp. 15985 LP – 15990. doi: 10.1073/pnas.1311166110.

Nierlich, D. P. *et al.* (1973) 'Transfer RNA synthesis in vitro', *Proceedings of the National Academy of Sciences of the United States of America*, 70(1), pp. 179–182. doi: 10.1073/pnas.70.1.179.

Nirenberg, M. W. and Matthaei, J. H. (1961) 'The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides', *Proceedings of the National Academy of Sciences*, 47(10), pp. 1588 LP – 1602. doi: 10.1073/pnas.47.10.1588.

Ohshiro, Y. *et al.* (2011) 'Ribosomal synthesis of backbone-macrocyclic peptides containing γ-amino acids.', *Chembiochem : a European journal of chemical biology*, 12(8), pp. 1183–1187. doi: 10.1002/cbic.201100104.

Palade, G. E. (1955) 'A small particulate component of the cytoplasm', *The Journal of biophysical and biochemical cytology*, 1(1), pp. 59–68. doi: 10.1083/jcb.1.1.59.

Pardee, K. *et al.* (2016) 'Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components', *Cell*, 165(5), pp. 1255–1266. doi: 10.1016/j.cell.2016.04.059.

Park, S. J. and Schimmel, P. (1988) 'Evidence for interaction of an aminoacyl transfer RNA synthetase with a region important for the identity of its cognate transfer RNA', *Journal of Biological Chemistry*, 263(32), pp. 16527–16530. doi: 10.1016/s0021-9258(18)37421-0.

Pauling, L. (1958) 'The probability of errors in the process of synthesis of protein molecules.', *Festschrift fur Prof. Dr. Arthur Stoll*, pp. 597–602.

Reynolds, N. M. *et al.* (2010) 'Cell-specific differences in the requirements for translation quality control.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(9), pp. 4063–4068. doi: 10.1073/pnas.0909640107.

Sankaranarayanan, R. *et al.* (2000) 'Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase.', *Nature structural biology*, 7(6), pp. 461–465. doi: 10.1038/75856.

827 Semrad, K. and Green, R. (2002) ‘Osmolytes stimulate the reconstitution of functional 50S
828 ribosomes from in vitro transcripts of Escherichia coli 23S rRNA’, *RNA (New York, N.Y.)*, 8(4), pp.
829 401–411. doi: 10.1017/s1355838202029722.

830 Serebrov, V. *et al.* (1998) ‘Mg2+ binding and structural stability of mature and in vitro synthesized
831 unmodified Escherichia coli tRNAPhe’, *Nucleic Acids Research*, 26(11), pp. 2723–2728. doi:
832 10.1093/nar/26.11.2723.

833 Shepherd, J. and Ibba, M. (2015) ‘Bacterial transfer RNAs’, *FEMS Microbiology Reviews*, 39(3), pp.
834 280–300. doi: 10.1093/femsre/fuv004.

835 Shepherd, T. R. *et al.* (2017) ‘De novo design and synthesis of a 30-cistron translation-factor
836 module’, *Nucleic acids research*, 45(18), pp. 10895–10905. doi: 10.1093/nar/gkx753.

837 Shimizu, Y. *et al.* (2001) ‘Cell-free translation reconstituted with purified components’, *Nature
838 Biotechnology*, 19(8), pp. 751–755. doi: 10.1038/90802.

839 Swank, Z., Laohakunakorn, N. and Maerkl, S. J. (2019) ‘Cell-free gene-regulatory network
840 engineering with synthetic transcription factors.’, *Proceedings of the National Academy of Sciences
841 of the United States of America*, 116(13), pp. 5892–5901. doi: 10.1073/pnas.1816591116.

842 Swartz, J. (2006) ‘Developing cell-free biology for industrial applications’, *Journal of Industrial
843 Microbiology and Biotechnology*, 33(7), pp. 476–485. doi: 10.1007/s10295-006-0127-y.

844 Tamura, K. *et al.* (1992) ‘In vitro study of E.coli tRNA(Arg) and tRNA(Lys) identity elements’,
845 *Nucleic acids research*, 20(9), pp. 2335–2339. doi: 10.1093/nar/20.9.2335.

846 Tardif, K. D. and Horowitz, J. (2004) ‘Functional group recognition at the aminoacylation and
847 editing sites of E. coli valyl-tRNA synthetase’, *Rna*, 10(3), pp. 493–503. doi: 10.1261/rna.5166704.

848 Valencia-Sánchez, M. I. *et al.* (2016) ‘Structural Insights into the Polyphyletic Origins of Glycyl
849 tRNA Synthetases.’, *The Journal of biological chemistry*, 291(28), pp. 14430–14446. doi:
850 10.1074/jbc.M116.730382.

851 Villarreal, F. *et al.* (2018) ‘Synthetic microbial consortia enable rapid assembly of pure translation
852 machinery’, *Nature Chemical Biology*, 14(1), pp. 29–35. doi: 10.1038/nchembio.2514.

853 Wan, W., Tharp, J. M. and Liu, W. R. (2014) ‘Pyrrolysyl-tRNA synthetase: an ordinary enzyme but
854 an outstanding genetic code expansion tool.’, *Biochimica et biophysica acta*, 1844(6), pp. 1059–
855 1070. doi: 10.1016/j.bbapap.2014.03.002.

856 Wang, N. *et al.* (2015) ‘Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for
857 efficient synthesis of proteins containing unnatural amino acids’, *ACS Synthetic Biology*, 4(3), pp.
858 207–212. doi: 10.1021/sb500195w.

859 Watson, J. D. and Crick, F. H. C. (1953) ‘Molecular Structure of Nucleic Acids: A Structure for
860 Deoxyribose Nucleic Acid’, *Nature*, 171(4356), pp. 737–738. doi: 10.1038/171737a0.

861 Worst, E. G. *et al.* (2015) ‘Cell-free expression with the toxic amino acid canavanine’, *Bioorganic &
862 Medicinal Chemistry Letters*, 25(17), pp. 3658–3660. doi: 10.1016/j.bmcl.2015.06.045.

Xiao, H. *et al.* (2008) 'Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme.', *Nature*, 454(7202), pp. 358–361. doi: 10.1038/nature07033.

Xu, B. *et al.* (1994) 'Probing the Metal Binding Sites of Escherichia coli Isoleucyl-tRNA Synthetase', *Biochemistry*, 33(2), pp. 398–402. doi: 10.1021/bi00168a002.

Yamanaka, K. *et al.* (2004) 'Efficient synthesis of nonnatural mutants in Escherichia coli S30 in vitro protein synthesizing system.', *Journal of bioscience and bioengineering*, 97(6), pp. 395–399. doi: 10.1016/S1389-1723(04)70225-X.

Zhang, C.-M. *et al.* (2003) 'Zinc-mediated amino acid discrimination in cysteinyl-tRNA synthetase.', *Journal of molecular biology*, 327(5), pp. 911–917. doi: 10.1016/s0022-2836(03)00241-9.