FEBS Letters 588 (2014) 383–388

journal homepage: www.FEBSLetters.org

Hypothesis

Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion

Radha Krishnakumar [a,*], Jiqiang Ling [b]

[a] Synthetic Biology and Bioenergy, J. Craig Venter Institute, Rockville, MD 20850, USA
[b] Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, The University of Texas Graduate School of Biomedical Sciences, Houston, TX 77030, USA

ARTICLE INFO

Article history:
Received 24 October 2013
Revised 19 November 2013
Accepted 27 November 2013
Available online 12 December 2013

Edited by Michael Ibba

Keywords:
Sense codon recoding
Pyrrolysyl-tRNA synthetase
Genetic code expansion
Mycoplasma mycoides

ABSTRACT

The addition of new and versatile chemical and biological properties to proteins pursued through incorporation of non-canonical amino acids is at present primarily achieved by stop codon suppression. However, it is critical to find new "blank" codons to increase the variety and efficiency of such insertions, thereby taking 'sense codon recoding' to center stage in the field of genetic code expansion. Current thought optimistically suggests the use of the pyrrolysine system coupled with re-synthesis of genomic information towards achieving sense codon reassignment. Upon review of the serious experimental challenges reported in recent studies, we propose that success in this area will depend on the re-synthesis of genomes, but also on 'rewiring' the mechanism of protein synthesis and of its quality control.

1. Introduction

The primary objective of synthetic biology is the design and construction of biological molecules, metabolic and regulatory networks and even new organisms with functions and capabilities not found in nature. To realize these goals and reap benefit from this rapidly developing field, there is also an effort to incorporate new and versatile biochemical functionalities into proteins by expanding the amino acid repertoire beyond the 20 standard amino acids and six elements of the periodic table. The incorporation of non-canonical amino acids (ncAAs, also referred to as non-standard, non-genetically encoded or unnatural amino acids) into proteins has been of interest to chemists and biologists since the 1950s. Hundreds of amino acid analogs have already been developed by chemists and synthesis of an almost unlimited number of others is theoretically possible [1–4]. The advantages of adding these to proteins are numerous. Apart from use of an expanded amino acid repertoire as a microbial cell factory to generate proteins with novel structure–function relationship [5–12], genetic code rewriting is touted as a possible mechanism for intrinsic biocontainment of synthetic or engineered organisms by engineering dependency on non-canonical amino acids or creating genetic firewalls.

Present-day technology for the incorporation of new chemical moieties into proteins includes chemical synthesis, stop codon suppression, and supplementation incorporation [3], each of which is limited by the number of chemical modifications and/or scalability. The majority of genetic code expansion research relies on suppression of the three stop codons (amber UAG, opal UGA and ochre UAA) by introducing a new (and orthogonal) aminoacyl-tRNA synthetase (aaRS)/tRNA pair that is modified to insert ncAAs in response to the stop codon [13]. But we have yet to see a huge impact in the field of materials science and bio-pharmaceutical industry from this method developed 20 years ago, and benefits are mostly realized on a small scale laboratory level today. The many potential applications of using ncAAs would benefit greatly from increased efficiency and variety of the insertions [14]. Most *Escherichia coli* strains were shown to have background suppression of both amber (5–10%) and opal codons (10–15%) [15], which can complicate insertion of ncAA using currently available orthogonal translation systems (OTS). Typical efficiency of ncAA insertion in stop codon suppression is about 20–30% and using multiple stop codons for multiple site-specific insertion drops the efficiency of ncAA insertion and completion of protein translation by over 10-fold [16]. A more radical approach is the use quadruplet codons and evolved orthogonal ribosomes to accommodate 4-base "non-standard" tRNAs adapted from animal mitochondria or other

Abbreviations: ncAAs, non-canonical amino acids; aaRS, aminoacyl-tRNA synthetase; OTS, orthogonal translation system; Pyl, pyrrolysine

* Corresponding author.
E-mail address: rkrishna@jcvi.org (R. Krishnakumar).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2013.11.039

diverse sources for *in vitro* and *in vivo* incorporation of the new amino acid [17]. More recently orthogonal ribosomes were evolved to decode amber and quadruplet codons, which increases the efficiency of ncAA insertion to about 60% for single sites and about 20% for two sites [16,18]. One potential problem is that frameshifting may occur at the quadruplet codon due to the competition from endogenous tRNAs. An alternate to stop codon suppression is supplementation incorporation method, which essentially allows sense codon recoding without mutagenesis of the target DNA and has been used to label three different synthetic amino acids into a single target protein expressed in a polyauxotrophic strain of *E. coli* [19]. Despite these advances, expanded genetic codes primarily remain a valuable source of academic knowledge, and at present there are rare examples of scaled application [20] due to drawbacks with the above methods such as low protein yields, inefficient incorporation of the synthetic amino acid, and competition from endogenous translation termination machinery. A possible solution is to free up more codons in the genetic code to create new "blank" codons, and re-wire the protein synthesis machinery so that non-canonical amino acid insertion becomes part of the normal metabolism of the organism; an approach referred to as sense codon recoding (or reassignment). Although experimental information on this topic is rare [21,22], there remains the optimistic viewpoint that sense codon recoding is a trivial matter of simply changing the anticodon of a tRNA (such as $tRNA^{Pyl}$) and rewriting the genome of an organism to remove surplus sense codons. In this article, we review recent information regarding challenges of using existing orthogonal systems for genetic code expansion and suggest possible solutions to overcome them.

2. Evolving (reduced) genomes and sense codon recoding

The redundant nature of the genetic code and the appearance of slight deviations from the universal code reveal room for expansion through natural and non-natural recoding [23–25]. From our current knowledge of protein synthesis, codon-anticodon interactions and tRNA modifications, one can conclude that several sense codons could be dispensable and thereby designed to introduce new amino acids into proteins. Sense codon recoding requires an unused or rarely used sense codon and an orthogonal tRNA-synthetase pair where the tRNA anticodon reads the "blank" sense codon and inserts the 21st amino acid in response to the sense codon. However, unlike stop codons, every sense codon in the genetic code (although redundant) has a cognate tRNA and encodes for one of the 20 standard amino acids. Therefore, removing a sense codon does not necessarily create "blank" codons unless the cognate tRNA is missing as well. The codon usage of an organism is dictated by the number and variety of these tRNA isoacceptors, which varies greatly from organelle to organism and this leads to certain codons used more frequently than others. The observation that some microbial genomes lack the cognate tRNAs required to read some sense codons lead to the idea that these sense codons are open or blank, and perhaps could be prime targets for reassignment. *Micrococcus luteus* is one such example that was reported to have several rare codons and missing cognate tRNAs, and was used to demonstrate insertion of a standard amino acid into these unassigned sense codons *in vitro* ([26] and references within).

We chose mycoplasmas in our efforts to rewrite the genetic code through sense codon recoding. These organisms have been used in minimal cell studies because of their small genomes (0.5 to nearly 1.5 Mb) and metabolic simplicity, which generated information about essential and non-essential genes that could help towards genome-wide codon reassignment [27]. More specifically our studies focused on two closely related caprine pathogens. *Mycoplasma capricolum* and *Mycoplasma mycoides* JCVI syn 1.0, also known as the the first synthetic bacterial cell, which means it is driven by a fully synthetic chromosome. The entire genome was assembled from 1 kb synthetic DNA fragments by yeast genomic reengineering methods and transplanted into *M. capricolum* [28]. Another characteristic that proved useful for studying sense codon reassignment is the high AT richness of the *M. mycoides* and *M. capricolum* genomes, which results in very limited use of some GC codons. These rare codons could potentially be candidates for removal from the genome and reassigned for insertion of the 21st amino acid.

Mycoplasmas also have a non-standard genetic code and a reduced number of tRNAs (ranging from 29–35), where UGA is read as tryptophan. Many mycoplasma species are also missing several cognate tRNAs, the most obvious being some of $tRNA^{Arg}$ isotypes [29–31]. *M. capricolum and M. mycoides* have only two $tRNA^{Arg}$ isoacceptors to read the quadruplet and duplet arginine codon boxes ($tRNA^{Arg}_{ACG}$ and $tRNA^{Arg}_{UCU}$). The $tRNA^{Arg}_{ACG}$ is expected to undergo deamination to form the inosine containing $tRNA^{Arg}_{ICG}$, which would decode the codons ending in C, U and A. Wobble base-pairing allows for the same tRNA to read multiple sense codons, and relies on post-transcriptional modification of the tRNAs through specific RNA modification enzymes [32–34]. This leaves the codon CGG "open", due to lack of the cognate $tRNA_{CCG}$. The mycoplasma species that have only the two $tRNA^{Arg}$ species also use the CGG codons very sparingly [35]. Towards achieving the first experimental study of sense codon recoding, we attempted to reassign the CGG codon and evaluate feasibility of other rare sense codons in the mycoplasma genome. *M. capricolum* was selected as the organism for this study since it has well documented self-replicative plasmids that would enable testing of various combinations of tRNAs and aaRSs [36].

We initially attempted to provide an orthogonal system for insertion of arginine analogs as the 21st amino acid in *M. capricolum*. There are very few arginyl based orthogonal systems available at present, despite the need for such systems to understand the role of post-translational modifications of arginine residues and effect of arginine analogs in antimicrobial peptides. So far site specific incorporation of arginine analogs has been achieved only in a cell-free translation system using a 4-base anticodon tRNA and arginyl-tRNA synthetase (ArgRS) derived from yeast [37].

Finding an orthogonal pair to expand the amino acid repertoire of an organism requires understanding its tRNA identity elements and decoding strategy. Identity elements are known for a few organisms such as *E. coli* and *Saccharomyces cerevisiae,* and are often extrapolated to other organisms by sequence alignment and mutational studies. Considerable success has been achieved by importing a tRNA-aaRS pair from an evolutionarily distant organism (such as thermophilic archaea into *E. coli*) that behaves orthogonally with little or no re-engineering [5]. However in the case of mycoplasma, we noticed that the $tRNA^{Arg}$-ArgRS pair from the evolutionarily divergent bacterium, *E. coli,* showed cross-charging with the endogenous mycoplasma $tRNA^{Arg}$ and ArgRS. Mutation studies to decipher identity elements did not eliminate cross-charging, which only stressed the importance of understanding the tRNA identity elements and evolutionary biology of the organism being used as the recipient for an orthogonal pair. But in any arginyl based OTS, the arginine analogs will be outcompeted by the natural substrate (arginine), unless the ArgRS is engineered to accept only arginine analogs. Rather than engineering an OTS, we looked to other naturally occurring orthogonal systems such as the pyrrolysine system that could be manipulated to recognize rare arginine codons and insert the 21st amino acid.