1

2

3

4

5

6

Reprogramming the genetic code in vitro

7

8

Toby Passioura¹ and Hiroaki Suga^{1,2}

9

10 ¹Department of Chemistry, Graduate School of Science, The University of Tokyo, 7-
11 3-1 Bunkyo-ku, Tokyo 113-0033, Japan.

12 ²Japan Science and Technology Agency (JST), Core Research for Evolutionary
13 Science and Technology (CREST), Saitama 332-0012, Japan.

14

15

16 Corresponding author: Hiroaki Suga (hsuga@chem.s.u-tokyo.ac.jp)

17

18 Keywords: genetic code reprogramming, translation, aminoacyl-tRNA synthetase,
19 flexizyme.

20 **Abstract**

21 The site-specific introduction of non-canonical amino acids into polypeptides through
22 genetic code reprogramming has become a powerful tool for biochemical studies and
23 bioorganic synthesis. Whilst a variety of such techniques have been developed, all
24 are based on the “mis-acylation” of tRNA molecules with non-canonical amino acids.
25 Multiple strategies have been devised to synthesize such non-canonical aminoacyl-
26 tRNAs; those based on protein or ribozyme aminoacyl-tRNA synthetase enzymes are
27 particularly useful. Such techniques have enabled the incorporation of hundreds of
28 different non-canonical amino acids into polypeptides in vitro. This review discusses
29 the development and application of in vitro genetic code reprogramming techniques,
30 especially enzymatic mis-acylation, and examines recent efforts to engineer the
31 translational machinery to increase the range of translatable non-canonical amino
32 acids.

33 **Beyond the twenty canonical amino acids**

34 The site-specific incorporation of non-canonical amino acids (Glossary) into
35 polypeptides through genetic code reprogramming has developed into a powerful
36 technique for biochemical investigation of proteins, and an extraordinary tool for
37 biologically mediated organic chemical synthesis of peptides. For example, the
38 ability to replace a key amino acid in a protein of interest with a non-canonical
39 analogue allows finely-tuned structure-function analyses [1], and the lessons learnt
40 from engineering the translation machinery to accept non-canonical amino acids have
41 provided valuable insights into the mechanism of translation itself [2, 3]. Further, the
42 combination of genetic code reprogramming with selection-based techniques allows
43 the parallel screening of trillions of peptide-like molecules for pharmacological
44 activity or for co-crystallization studies [4-12]. In these cases, the ability to introduce
45 non-canonical residues into peptides allows for the synthesis of libraries of peptide-
46 like molecules with much greater structural diversity (and therefore much greater
47 likelihood of discovering a “hit” during screening) than would be possible using the
48 20 canonical amino acids alone. Moreover, peptide-like molecules containing
49 unusual structural features such as backbone cyclization, *N*-methylation, *N*-alkylation,
50 D-stereochemistry or β -residues exhibit more “drug-like” qualities than canonical
51 peptides (e.g. increased stability in biological environments, greater target affinity,
52 etc), and are structurally similar to naturally-occurring bioactive peptides (e.g.
53 cyclosporin and actinomycin D)[12].

54 The genetic reprogramming techniques developed to date can be broadly classified
55 into those techniques which involve genetic reprogramming in a living cell (Box 1)
56 and those that utilize in vitro translation techniques. Each of these approaches exhibit
57 relative advantages and disadvantages in comparison with each other, and have

58 differing applications; in general terms, intracellular reprogramming is less flexible
59 with respect to the chemistries available but allows for the biochemical investigation
60 of reprogrammed peptides or proteins in the relevant biological environment.
61 Comparisons of reprogramming efficiency and total peptide yield are complicated by
62 the diversity of conditions employed in different studies (choice of amino acid, codon,
63 mRNA template, etc), however in vitro systems may be considered to generally
64 produce lower yields (because they are not readily scaled-up) but greater
65 reprogramming efficiency (because of reduced competition from endogenous
66 elements – discussed later) than intracellular systems.

67 Whilst there is some overlap in the approaches used to engineer the translation
68 machinery for intracellular and in vitro genetic code reprogramming, the different
69 challenges faced in each case have generally necessitated different experimental
70 strategies. The approaches used for intracellular genetic code reprogramming have
71 recently been reviewed elsewhere [13], and so the present review focuses upon
72 strategies for engineering the translation machinery for in vitro genetic code
73 reprogramming.

74

75

76 **Early studies of genetic code expansion**

77 The vast majority of in vitro genetic code reprogramming studies used to date have
78 utilized the *Escherichia coli* translation machinery (either from crude cell lysate or
79 through a purified recombinant translation system; discussed later) for polypeptide
80 synthesis (Figure 1). In such systems, the sequence of the polypeptide synthesized is,
81 of course, defined by the sequence of its cognate mRNA. However, translational
82 fidelity is equally dependent upon the coupling of each amino acid to its respective

83 tRNA, and thus, through translation of “mis-acylated” tRNAs bearing unusual amino
84 acids, it is possible to introduce these residues into the translated polypeptide.

85 Following such a strategy, multiple techniques have been developed for the synthesis
86 and translation of these “mis-acylated” tRNAs. The earliest such reports involved the
87 use of so-called semi-enzymatic synthesis, in which an aminoacyl-dinucleotide was
88 synthesized by standard organic techniques and then enzymatically acylated onto a
89 tRNA body through the activity of T4 RNA ligase [14, 15]. An *E. coli* cell lysate-
90 based system was then used to incorporate the relevant aminoacyl moiety at a specific
91 codon in a specific mRNA. In these examples, incorporation of the non-canonical
92 amino acid was directed to the UAG stop codon (amber stop codon) in order to avoid
93 competition with canonical tRNAs present in the cell lysate system. Such approaches
94 are known as stop codon suppression techniques, and their application has
95 demonstrated that a variety of non-canonical amino acids may be incorporated into
96 polypeptides site-specifically, including several substituted analogues of
97 phenylalanine and *N*-alkylated or α,α -disubstituted analogues of alanine [14, 15]. As
98 an alternative to stop codon suppression, four-base codons (based on so-called “rare”
99 codons) have also been used to direct incorporation of non-canonical amino acids [16-
100 20].

101 Stop codon suppression and four-base codon based methods result in the synthesis of
102 polypeptides containing all 20 canonical amino acids in addition to one (or sometimes
103 two or three) extra non-canonical amino acid(s). As such, they may be considered to
104 be genetic code expansion techniques, rather than genetic code reprogramming
105 techniques *per se*. Nonetheless, they have proven to be extremely useful tools for the
106 synthesis of proteins with non-canonical functional groups for biochemical studies,

107 and form the basis of more recent intracellular genetic code reprogramming
108 techniques [21-24].

109 However, these genetic code expansion approaches have two significant limitations,
110 with respect to in vitro applications. First, reliance on stop or four-base codons
111 restricts the number of different non-canonical amino acids that may be incorporated
112 into a single polypeptide, with the vast majority of published examples limited to
113 incorporation of a single non-canonical residue. The second limitation is that the
114 efficiency of incorporation of any non-canonical amino acids is substantially
115 compromised by competition from release factors or endogenous elongator tRNAs.
116 In practical terms, this means that only the most efficiently incorporated non-
117 canonical amino acids may be employed because, in many cases, incorporation of
118 less-efficient substrates for translation (e.g. *N*-alkyl- or D-amino acids) is very poor or
119 undetectable [14, 15, 25, 26].

120

121

122 **Genetic code reprogramming using recombinant translation systems**

123 An elegant solution to the two major limitations of stop codon suppression or four-
124 base codon based genetic expansion systems has been the use of fully reconstituted
125 translation systems. Such systems are completely composed of purified, defined
126 components (ribosomes, aminoacyl-tRNA synthetases (AARSs), release factors,
127 initiation factors, elongation factors, tRNAs, amino acids, etc). In consequence,
128 competition by release factors or endogenous acylated tRNAs at a given codon can be
129 largely obviated by simply omitting the relevant release factor or amino acid (and/or
130 its cognate AARS) from the reaction.

131 In an early report of this approach, Forster and co-workers used a highly simplified
132 translation system composed principally of purified initiation factors, elongation
133 factors, ribosomes, mRNA and pre-acylated tRNAs which had been synthesized by
134 the semi-enzymatic method described earlier or through the enzymatic activity of
135 purified amino acid-tRNA transferases. They used this system to synthesize short
136 peptides composed of combinations of the canonical amino acids formyl methionine
137 (the initiator), valine and glutamic acid, and the non-canonical amino acids *O*-methyl-
138 serine, 2-aminopent-4-enoic acid and 2-aminopent-4-ynoic acid [27]. This study
139 demonstrated that the use of a purified recombinant translation system not only
140 improves the incorporation efficiency of non-canonical amino acids by reducing
141 competition, but also allows for the radical reassignment of amino acid encoding
142 codons (as compared to stop or four-base codons) to other amino acids (either
143 canonical or non-canonical). Subsequent studies, particularly facilitated by the
144 flexizyme technology (*vide infra*), have refined and extended this approach [2, 4, 28-
145 40]; for example, the addition of release factors to the translation reaction allows for
146 multiple turnover peptide synthesis, and the inclusion of a limited set of amino acids
147 and their cognate AARSs enables the synthesis of peptides comprising a mixture of
148 canonical and non-canonical peptides in which only the non-canonical amino acids
149 need to be included as pre-acylated tRNAs.

150

151

152 **AARS-mediated acylation**

153 The use of fully reconstituted translation systems, as described in the previous
154 section, now enables the routine in vitro synthesis of peptides (and peptide-like
155 molecules) comprising diverse non-canonical residues. However, the relatively

laborious synthesis of acylated tRNAs through semi-enzymatic synthesis has now been largely replaced with more practical enzymatic techniques utilizing either protein AARS enzymes or aminoacylating ribozymes (flexizymes).

Although naturally occurring AARSs appear to exhibit high selectivity for both their amino acid and tRNA substrates in vivo, many are sufficiently promiscuous to allow acylation of tRNAs with close analogues of canonical amino acids. For example, lysine and isoleucine AARSs will accept thioether-containing analogues of their respective canonical substrates, and phenylalanine and tyrosine AARSs will accept fluorinated analogues of phenylalanine and tyrosine, respectively (Figure 2) [41]. Indeed, extensive studies of AARS promiscuity have identified around 100 such non-canonical amino acids that are efficiently acylated onto tRNAs by *E. coli* AARSs [41, 42]. Importantly, many of these are also efficiently incorporated into polypeptides during translation, enabling the synthesis of peptides containing numerous consecutive non-canonical amino acid residues through the reassignment of up to 47 of the 64 (73%) three base codons [10, 41, 43-45].

AARS-mediated acylation is substantially less labor-intensive than semi-enzymatic synthesis. Indeed, this ease-of-use is its major practical advantage; because many non-canonical amino acids are available from commercial suppliers, they can be used for genetic code reprogramming simply by substituting them for their canonical analogues in a translation reaction in a process that requires no synthetic organic chemical techniques. Incorporation of non-canonical amino acids using this technique can be very efficient, and polypeptides comprising up to 10 consecutive non-canonical amino acids have been reported [41]. This efficiency may reflect that genetic code reprogramming via AARS-mediated acylation involves multiple turnover acylation (*i.e.* re-acylation of tRNAs continues during the translation

181 process, in contrast to approaches in which pre-synthesized aminoacyl-tRNAs are
182 added to the reaction at the outset) of endogenous tRNAs with analogues of their
183 cognate canonical amino acids (as compared to the use of pre-synthesized synthetic
184 tRNAs acylated with diverse amino acid structures) and thus closely mirrors natural
185 translation.

186 However, compared to both semi-enzymatic synthesis and acylation using flexizymes
187 (discussed later), AARS-mediated acylation sacrifices flexibility because only close
188 analogues of canonical amino acids may be translated and then only at their original
189 codons. As a result, more exotic amino acid structures (*i.e.* those that differ
190 substantially from their canonical counterparts) cannot be used, and radical codon
191 reassignment (e.g. translation of a tyrosine analogue at a glycine codon) would only
192 be possible through the use of engineered tRNAs and AARSs designed specifically
193 for each amino acid substrate [46], a relatively laborious approach which has, to the
194 best of our knowledge, not been applied to reconstituted *in vitro* reprogramming
195 systems. Further, AARS-mediated approaches are generally most effective when
196 using non-canonical amino acids with unusual side chain structures [43]. However,
197 synthesis of backbone-modified peptides using this technique is somewhat
198 problematic, and, to the best of our knowledge, incorporation of only two *N*-
199 methylated (aspartate and histidine) and two α,α -disubstituted amino acids has been
200 reported, whilst incorporation of D-amino acids does not appear to be feasible. In the
201 case of *N*-methylated amino acids, these limitations can be circumvented to some
202 degree through post-acylation chemical *N*-methylation, which has been shown to be
203 effective for synthesis of polypeptides containing *N*-methylated leucine, threonine and
204 valine [35, 47, 48]. However, such an approach partially negates the advantages of

205 AARS-mediated acylation, namely ease-of-use and in-reaction multiple turnover
206 acylation.

207

208

209 **Flexizyme-mediated acylation**

210 An alternative to either semi-enzymatic synthesis or AARS-mediated acylation are
211 acylating ribozymes, known as flexizymes, which have been used to great effect.

212 Flexizymes are a family of small (45–46 nucleotide) tRNA acylating ribozymes
213 originally identified through in vitro selection experiments [40, 49-52]. Because they
214 exhibit marked promiscuity towards both their tRNA and amino acid substrates,
215 flexizymes may be used to acylate diverse amino acids onto tRNAs bearing any anti-
216 codon; the only substrate requirement in the tRNA is the presence of a 3'-CCA motif.

217 Three flexizymes are currently in general use, with each having slightly different
218 amino acid substrate requirements, although all of them require activation of the
219 carboxylic acid through an appropriate ester leaving group and recognize aromatic
220 moieties either in the amino acid side chain or the leaving group. Enhanced
221 flexizyme (eFx) acts upon amino acids with aromatic side chains and a cyanomethyl
222 ester leaving group, or on amino acids with diverse side chains activated with a
223 chlorobenzylthioester leaving group. Dinitro-flexizyme (dFx) acts upon amino acids
224 with diverse side chains activated with a 3,5-dinitrobenzyl ester leaving group.

225 Amino-flexizyme (aFx) acts upon amino acids with diverse side chains activated with
226 a (2-aminoethyl)amidocarboxybenzyl thioester leaving group. Compared to the other
227 two flexizymes, aFX substrates are more soluble in water. Thus, by using an
228 appropriately matched flexizyme and activated amino acid it is possible to acylate
229 essentially any amino acid onto a tRNA bearing any anticodon, which can then be

230 incorporated into a peptide using a recombinant translation system in a process that
231 has been termed a flexible in vitro translation (FIT) reaction.

232 Like the AARS-mediated approach described earlier, flexizyme-mediated acylation is
233 substantially less labor-intensive than semi-enzymatic synthesis, and synthesis of the
234 activated amino acid substrates requires only simple organic synthetic techniques.
235 Compared with AARS-mediated acylation, flexizyme-mediated acylation is slightly
236 more labor-intensive, because the activated amino acid substrates must be synthesized
237 (although a single such synthesis typically produces enough substrate for thousands of
238 translation reactions) and the tRNAs must be pre-acylated before translation.
239 However, the extreme promiscuity of flexizymes enables tRNA acylation using a
240 much wider range of amino acid substrates than AARS-mediated acylation, including
241 D-, β -, γ -, *N*-alkyl, and *N*-acyl-amino acids as well as amino acids with diverse side
242 chains, α -hydroxy acids and even short peptides (Figure 3) [4, 30, 33, 34, 37, 53, 54].
243 Using this approach, more than three hundred non-canonical amino acids have now
244 been used to produce exotic peptides, polyesters and peptoids (*N*-alkyl polyglycines)
245 [4-6, 30, 31, 33, 34, 36-39, 55-61]. Further, as with the semi-enzymatic approach of
246 Forster et al. described earlier, FIT reactions also allow for the radical reassignment of
247 amino acids to desired codons. For example, *N*-methylated serine and glycine may be
248 translated at leucine and isoleucine codons, respectively [4]. With respect to the
249 relative efficiencies and yields of AARS-mediated and flexizyme-mediated
250 reprogramming, we are not aware of any instances in which direct comparisons have
251 been made. Yields may be expected to be similar in both cases, since both strategies
252 use reconstituted translation systems, although the multiple turnover acylation
253 kinetics of the AARS-mediated approach (discussed above) may lead to higher yields
254 in some cases. Comparisons of reprogramming efficiency are likely to be highly

255 confounded by differences in the choice of amino acid, codon and mRNA template,
256 and the data in the literature to date are insufficient for any conclusions to be drawn.
257 More recently, the FIT system has been applied to the reinvestigation of D-amino acid
258 incorporation into peptide chains [54]. In previous reports using amber suppression,
259 although there were some contradicting data, D-amino acids were generally found to
260 be very poor elongation substrates (<10 % compared with their natural counterpart).
261 However, because the FIT system allows for almost complete knock-down of
262 competing background incorporation by canonical amino acids, this enables
263 evaluation of the intrinsic ability of the ribosome to efficiently incorporate D-amino
264 acids. This reinvestigation has revealed that 8 D-amino acids (out of 19 tested,
265 because glycine has no chiral center) are incorporated with higher efficiency than
266 expected (>40% compared with the canonical L-stereoisomer), 4 with modest
267 efficiency (10–40%), and only 7 with less than 10% efficiency. This strongly suggests
268 that the previous observations using amber suppression were largely misinterpreted as
269 a result of termination competition, because they did not represent the intrinsic
270 compatibility of the ribosome. Further, this work also revealed that both consecutive
271 incorporation of D-amino acids and D-L-D sequential incorporation are completely
272 prohibited, suggesting that the ribosome peptidyl-transfer center cannot accommodate
273 such reactions.

274

275

276 **Molecular engineering of translation reaction components**

277 As outlined in the preceding sections, the combined use of an acylation technology
278 (semi-enzymatic synthesis, AARS-mediated acylation or flexizyme-mediated
279 acylation) with a recombinant, purified translation system allows the synthesis of

280 diverse peptide-like molecules in vitro. We consider that these technologies are in
281 many ways mature (for example, the development of new flexizymes seems
282 unwarranted because judicious use of dFx, eFx and aFx allows acylation of tRNAs
283 with virtually any amino acid). However, the diversity of peptidic chemical structures
284 accessible through such systems remains limited in some respects. Most notably,
285 some *N*-methylated, D-, α,α -disubstituted, and β -amino acids, as well as some amino
286 acids with very bulky non-canonical side chains, have been identified that can be
287 efficiently acylated onto tRNAs but which are not efficient substrates for translation
288 [34, 43, 54, 62-65]. Thus, some amino acids appear to be intrinsically
289 “untranslatable” using the canonical translation machinery (Box 2).

290 However, ongoing efforts to engineer the translation machinery at a molecular level in
291 order to broaden the range of acceptable substrates have met with some success. The
292 most obvious strategy for such approaches is engineering of the ribosome itself.
293 Indeed, Hecht and co-workers have shown that targeted mutations in the ribosomal
294 23S RNA lead to modest increases in the efficiency of amber stop codon suppression
295 using D-methionine and D-phenylalanine [66, 67], and these studies have been
296 extended to the directed evolution of ribosomes for inclusion of β -amino acids [63,
297 68]. These studies appear promising; however, the reported stop codon suppression
298 efficiency using non-canonical amino acids was generally 20% or less, even for the
299 most efficient mutant ribosomes, and further studies are warranted to assess
300 translation efficiency using a broader range of D- and β -amino acid substrates and a
301 more diverse set of target codons.

302 Whilst engineering of the ribosome appears to improve the translation efficiency of
303 some non-canonical amino acids, other components of the translation machinery also
304 play critical roles in determining the range of translatable substrates. Elongation

factor-Tu (EF-Tu) is of particular importance in this respect, because it binds to the acylated tRNA molecules and delivers them to the A site of the ribosome (Figure 4). Uhlenbeck and co-workers have shown that the binding affinities of EF-Tu to different canonical acylated tRNAs fall within a narrow range of values, and that disruption of this relationship (either through increasing or decreasing the strength of the EF-Tu and acyl-tRNA interaction) significantly impairs translation efficiency [69, 70]. This relatively uniform affinity of EF-Tu for acylated tRNAs is observed even though different aminoacyl moieties exhibit different affinities for the amino acid binding pocket of EF-Tu. However, in each case, compensatory alterations in the binding affinity of each tRNA for EF-Tu (mediated primarily by three non-conserved nucleotides in the T-stem) modulate the overall binding affinity, so that amino acids with high affinities for EF-Tu are coupled with tRNAs exhibiting low affinities for EF-Tu and vice versa [70, 71]. Thus, EF-Tu proof-reads the acylated tRNAs, ensuring that each tRNA is acylated with the appropriate amino acid, the substitution of which with a non-canonical amino acid can greatly impair translation efficiency [72].

Molecular engineering of EF-Tu to relax this proof-reading function has been demonstrated in at least two instances. In the first, the amino acid binding pocket of EF-Tu was enlarged through alanine mutagenesis of an aspartate residue in order to accommodate amino acids with particularly bulky sidechains (poly-aromatic analogues of phenylalanine), and it was demonstrated that the resultant mutant EF-Tu afforded the translation of such amino acids in a purified recombinant translation system, whereas wild-type EF-Tu did not [73]. In the second instance, Park and co-workers used the slightly more sophisticated approach of directed evolution to identify an EF-Tu mutant with 6 mutations in the amino acid binding pocket that facilitated translation of phosphoserine in an *E. coli* intracellular system [74]. These

330 studies have provided proof-of-principle demonstrations that targeted engineering of
331 EF-Tu can improve the translation efficiency of some non-canonical amino acids,
332 however the general applicability of such approaches remains uncertain because it
333 remains unclear whether these mutant EF-Tu proteins are generally more permissive
334 or are simply selective for the non-canonical amino acid--tRNA for which they were
335 designed. In the latter case, a specific mutant EF-Tu, necessitating substantial protein
336 engineering, may be required for each non-canonical amino acid--tRNA pair. This
337 could potentially lead to complex competitive interactions if multiple different mutant
338 EF-Tu proteins are used to incorporate multiple non-canonical amino acids.

339 EF-Tu recognizes elements of both the amino acid and tRNA, therefore an alternative
340 to engineering EF-Tu for optimal binding affinity to the acylated tRNA would be to
341 engineer the tRNA so as to compensate for altered properties of the non-canonical
342 amino acid. Such studies have been described in intracellular reprogramming systems
343 [75], however it remains unclear whether tRNAs engineered for improved
344 intracellular reprogramming efficiency would be efficacious if used in in vitro
345 reactions. Nonetheless, it is notable that the use of different synthetic versions of
346 endogenous tRNAs has been shown to alter the efficiency of in vitro reprogramming
347 using non-canonical amino acids. For example, synthetic tRNAs based on *E. coli*
348 phenylalanine, asparagine, alanine and proline tRNAs have been shown to mediate
349 translation with different efficiencies when charged with *N*-methylated variants of
350 phenylalanine, glycine and alanine [29, 76, 77], demonstrating that the choice of
351 tRNA used for a specific non-canonical amino acid can be an important determinant
352 of translation efficiency. Thus, it appears that matching the non-canonical amino acid
353 of choice to an appropriate tRNA (either a synthetic variant of a naturally-occurring
354 tRNA or one specifically engineered for that purpose) might be a useful strategy for

355 improving the translation efficiency of at least some poorly translated non-canonical
356 amino acids.

357

358

359 **Concluding remarks**

360 As described in this Review, in vitro genetic code reprogramming techniques allow
361 the synthesis of polypeptides comprising combinations of diverse non-canonical
362 amino acid residues. These techniques have developed from early studies using semi-
363 enzymatically synthesized aminoacyl-tRNAs for stop codon suppression to the
364 current state of the art, in which any one of hundreds of non-canonical amino acids
365 can be incorporated at a codon of choice. Such approaches have found utility in
366 biochemical experiments designed to investigate the mechanics of translation or to
367 test the effect of subtle sequence changes on protein structure, and have proven to be
368 particularly useful (when combined with selection based technologies) for the
369 identification of small peptide-like molecules with drug-like properties [4-12].
370 Looking ahead, we predict that the range of chemistries available through in vitro
371 genetic reprogramming will continue to grow, as the limits of the available
372 technologies are fully defined, and molecular engineering of individual components
373 of the translation systems appears likely to widen this range further. This in turn will
374 broaden the scope for the application of genetic code reprogramming to biochemical
375 investigation and drug discovery studies, further improving the utility of techniques
376 that already exhibit substantial utility.

377

378

379 **Acknowledgements**

380 This work was supported by a Grant-in-Aid for JSPS Specially Promoted Research
381 (21000005), JST CREST, and Platform for Drug Discovery, Informatics, and
382 Structural Life Science from MEXT, Japan to H.S..

383 **References**

- 384 1. Ojemalm, K., et al. (2011) Apolar surface area determines the efficiency of
385 translocon-mediated membrane-protein integration into the endoplasmic
386 reticulum. *Proc. Natl. Acad. Sci. U. S. A.* 108, E359-364.
- 387 2. Pavlov, M.Y., et al. (2009) Slow peptide bond formation by proline and other
388 N-alkylamino acids in translation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 50-54.
- 389 3. Terasaka, N., et al. (2014) An orthogonal ribosome-tRNA pair via engineering
390 of the peptidyl transferase center. *Nat. Chem. Biol.*
- 391 4. Yamagishi, Y., et al. (2011) Natural product-like macrocyclic N-methyl-
392 peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-
393 expressed de novo library. *Chem. Biol.* 18, 1562-1570.
- 394 5. Hayashi, Y., et al. (2012) In vitro selection of anti-Akt2 thioether-macrocyclic
395 peptides leading to isoform-selective inhibitors. *ACS Chem. Biol.* 7, 607-613.
- 396 6. Morimoto, J., et al. (2012) Discovery of Macrocyclic Peptides Armed with a
397 Mechanism-Based Warhead: Isoform-Selective Inhibition of Human
398 Deacetylase SIRT2. *Angew. Chem. Int. Ed. Engl.* 51, 3423-3427.
- 399 7. Kawakami, T., et al. (2013) In Vitro Selection of Multiple Libraries Created
400 by Genetic Code Reprogramming To Discover Macrocyclic Peptides That
401 Antagonize VEGFR2 Activity in Living Cells. *ACS Chem. Biol.* In press.
- 402 8. Tanaka, Y., et al. (2013) Structural basis for the drug extrusion mechanism by
403 a MATE multidrug transporter. *Nature* 496, 247-251.
- 404 9. Hofmann, F.T., et al. (2012) In vitro selection of functional lantipeptides. *J.*
405 *Am. Chem. Soc.* 134, 8038-8041.

- 406 10. Schlippe, Y.V., et al. (2012) In vitro selection of highly modified cyclic
407 peptides that act as tight binding inhibitors. *J. Am. Chem. Soc.* 134, 10469-
408 10477.
- 409 11. Hipolito, C.J., et al. (2013) A macrocyclic peptide that serves as a
410 cocrystallization ligand and inhibits the function of a MATE family
411 transporter. *Molecules* 18, 10514-10530.
- 412 12. Passioura, T., et al. (2014) Selection-Based Discovery of Druglike
413 Macrocyclic Peptides. *Annu. Rev. Biochem.*
- 414 13. Chin, J.W. (2014) Expanding and Reprogramming the Genetic Code of Cells
415 and Animals. *Annu. Rev. Biochem.*
- 416 14. Ellman, J.A., et al. (1992) Site-specific incorporation of novel backbone
417 structures into proteins. *Science* 255, 197-200.
- 418 15. Noren, C.J., et al. (1989) A general method for site-specific incorporation of
419 unnatural amino acids into proteins. *Science* 244, 182-188.
- 420 16. Hohsaka, T., et al. (2002) Nonnatural mutagenesis in *E. coli* and rabbit
421 reticulocyte lysates by using four-base codons. *Nucleic Acids Res Suppl*, 201-
422 202.
- 423 17. Hohsaka, T., et al. (2004) Position-specific incorporation of dansylated non-
424 natural amino acids into streptavidin by using a four-base codon. *FEBS Lett.*
425 560, 173-177.
- 426 18. Neumann, H., et al. (2010) Encoding multiple unnatural amino acids via
427 evolution of a quadruplet-decoding ribosome. *Nature* 464, 441-444.
- 428 19. Taira, H., et al. (2005) Four-base codon-mediated incorporation of non-natural
429 amino acids into proteins in a eukaryotic cell-free translation system. *J Biosci*
430 *Bioeng* 99, 473-476.

- 431 20. Hayashi, G., et al. (2010) Ribosome evolution for two artificial amino acids in
432 E. coli. *Chem. Biol.* 17, 320-321.
- 433 21. Davis, L. and Chin, J.W. (2012) Designer proteins: applications of genetic
434 code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* 13, 168-182.
- 435 22. Hoesl, M.G. and Budisa, N. (2012) Recent advances in genetic code
436 engineering in *Escherichia coli*. *Curr. Opin. Biotechnol.* 23, 751-757.
- 437 23. Liu, C.C. and Schultz, P.G. (2010) Adding new chemistries to the genetic
438 code. *Annu. Rev. Biochem.* 79, 413-444.
- 439 24. Neumann, H. (2012) Rewiring translation - Genetic code expansion and its
440 applications. *FEBS Lett.* 586, 2057-2064.
- 441 25. Bain, J.D., et al. (1991) Site-specific incorporation of nonnatural residues
442 during in vitro protein biosynthesis with semisynthetic aminoacyl-tRNAs.
443 *Biochemistry (Mosc.)* 30, 5411-5421.
- 444 26. Bain, J.D., et al. (1991) Site-specific incorporation of non-natural residues into
445 peptides: Effect of residue structure on suppression and translation
446 efficiencies. *Tetrahedron* 47, 2389-2400.
- 447 27. Forster, A.C., et al. (2003) Programming peptidomimetic syntheses by
448 translating genetic codes designed de novo. *Proc. Natl. Acad. Sci. U. S. A.*
449 100, 6353-6357.
- 450 28. Murakami, H., et al. (2006) A highly flexible tRNA acylation method for non-
451 natural polypeptide synthesis. *Nat. Methods* 3, 357-359.
- 452 29. Zhang, B., et al. (2007) Specificity of translation for N-alkyl amino acids. *J.*
453 *Am. Chem. Soc.* 129, 11316-11317.
- 454 30. Goto, Y., et al. (2008) Initiating translation with D-amino acids. *RNA* 14,
455 1390-1398.

- 456 31. Goto, Y., et al. (2008) Reprogramming the translation initiation for the
457 synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* 3, 120-
458 129.
- 459 32. Kang, T.J. and Suga, H. (2008) Ribosomal synthesis of nonstandard peptides.
460 *Biochem. Cell Biol.* 86, 92-99.
- 461 33. Kawakami, T., et al. (2008) Ribosomal synthesis of polypeptoids and peptoid-
462 peptide hybrids. *J. Am. Chem. Soc.* 130, 16861-16863.
- 463 34. Kawakami, T., et al. (2008) Messenger RNA-programmed incorporation of
464 multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.*
465 15, 32-42.
- 466 35. Subtelny, A.O., et al. (2008) Ribosomal synthesis of N-methyl peptides. *J.*
467 *Am. Chem. Soc.* 130, 6131-6136.
- 468 36. Goto, Y., et al. (2009) Ribosomal synthesis of dehydrobutyrine- and
469 methyllanthionine-containing peptides. *Chem Commun (Camb)*, 3419-3421.
- 470 37. Goto, Y. and Suga, H. (2009) Translation initiation with initiator tRNA
471 charged with exotic peptides. *J. Am. Chem. Soc.* 131, 5040-5041.
- 472 38. Kawakami, T., et al. (2009) Diverse backbone-cyclized peptides via codon
473 reprogramming. *Nat. Chem. Biol.* 5, 888-890.
- 474 39. Yamagishi, Y., et al. (2009) Ribosomal synthesis of cyclic peptides with a
475 fluorogenic oxidative coupling reaction. *Chembiochem* 10, 1469-1472.
- 476 40. Goto, Y., et al. (2011) Flexizymes for genetic code reprogramming. *Nat.*
477 *Protoc.* 6, 779-790.
- 478 41. Josephson, K., et al. (2005) Ribosomal synthesis of unnatural peptides. *J. Am.*
479 *Chem. Soc.* 127, 11727-11735.

480 42. Hartman, M.C., et al. (2006) Enzymatic aminoacylation of tRNA with
481 unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4356-4361.

482 43. Hartman, M.C., et al. (2007) An expanded set of amino acid analogs for the
483 ribosomal translation of unnatural peptides. *PLoS ONE* 2, e972.

484 44. Seebeck, F.P., et al. (2011) Artificial lantipeptides from in vitro translations.
485 *Chem Commun (Camb)* 47, 6141-6143.

486 45. Seebeck, F.P. and Szostak, J.W. (2006) Ribosomal synthesis of
487 dehydroalanine-containing peptides. *J. Am. Chem. Soc.* 128, 7150-7151.

488 46. Neumann, H., et al. (2010) De novo generation of mutually orthogonal
489 aminoacyl-tRNA synthetase/tRNA pairs. *J. Am. Chem. Soc.* 132, 2142-2144.

490 47. Merryman, C. and Green, R. (2004) Transformation of aminoacyl tRNAs for
491 the in vitro selection of "drug-like" molecules. *Chem. Biol.* 11, 575-582.

492 48. Subtelny, A.O., et al. (2011) Optimal codon choice can improve the efficiency
493 and fidelity of N-methyl amino acid incorporation into peptides by in-vitro
494 translation. *Angew. Chem. Int. Ed. Engl.* 50, 3164-3167.

495 49. Murakami, H., et al. (2003) A versatile tRNA aminoacylation catalyst based
496 on RNA. *Chem. Biol.* 10, 655-662.

497 50. Morimoto, J., et al. (2011) Flexizymes: their evolutionary history and the
498 origin of catalytic function. *Acc Chem Res* 44, 1359-1368.

499 51. Passioura, T. and Suga, H. (2013) Flexizyme-Mediated Genetic
500 Reprogramming As a Tool for Noncanonical Peptide Synthesis and Drug
501 Discovery. *Chemistry*.

502 52. Goto, Y. and Suga, H. (2012) Flexizymes as a tRNA acylation tool facilitating
503 genetic code reprogramming. *Methods Mol. Biol.* 848, 465-478.

- 504 53. Goto, Y., et al. (2006) Translation initiation by using various N-acylaminoacyl
505 tRNAs. *Nucleic Acids Symp Ser (Oxf)*, 293-294.
- 506 54. Fujino, T., et al. (2013) Reevaluation of the D-amino acid compatibility with
507 the elongation event in translation. *J. Am. Chem. Soc.* 135, 1830-1837.
- 508 55. Murakami, H., et al. (2006) Flexizyme as a versatile tRNA acylation catalyst
509 and the application for translation. *Nucleic Acids Symp Ser (Oxf)*, 35-36.
- 510 56. Ohta, A., et al. (2007) Synthesis of polyester by means of genetic code
511 reprogramming. *Chem. Biol.* 14, 1315-1322.
- 512 57. Ohta, A., et al. (2008) Polymerization of alpha-hydroxy acids by ribosomes.
513 *Chembiochem* 9, 2773-2778.
- 514 58. Sako, Y., et al. (2008) Ribosomal synthesis of peptidase-resistant peptides
515 closed by a nonreducible inter-side-chain bond. *ACS Chem. Biol.* 3, 241-249.
- 516 59. Sako, Y., et al. (2008) Ribosomal synthesis of bicyclic peptides via two
517 orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* 130, 7232-7234.
- 518 60. Nakajima, E., et al. (2009) Ribosomal synthesis of peptides with C-terminal
519 lactams, thiolactones, and alkylamides. *Chembiochem* 10, 1186-1192.
- 520 61. Ohshiro, Y., et al. (2011) Ribosomal synthesis of backbone-macrocyclic
521 peptides containing gamma-amino acids. *Chembiochem* 12, 1183-1187.
- 522 62. Starck, S.R., et al. (2003) The puromycin route to assess stereo- and
523 regiochemical constraints on peptide bond formation in eukaryotic ribosomes.
524 *J. Am. Chem. Soc.* 125, 8090-8091.
- 525 63. Dedkova, L.M., et al. (2012) beta-Puromycin selection of modified ribosomes
526 for in vitro incorporation of beta-amino acids. *Biochemistry (Mosc.)* 51, 401-
527 415.

- 528 64. Hohsaka, T., et al. (1993) Adaptability of nonnatural aromatic amino acids to
529 the active center of the E. coli ribosomal A site. FEBS Lett. 335, 47-50.
- 530 65. Murakami, H., et al. (2003) Using a solid-phase ribozyme aminoacylation
531 system to reprogram the genetic code. Chem. Biol. 10, 1077-1084.
- 532 66. Dedkova, L.M., et al. (2003) Enhanced D-amino acid incorporation into
533 protein by modified ribosomes. J. Am. Chem. Soc. 125, 6616-6617.
- 534 67. Dedkova, L.M., et al. (2006) Construction of modified ribosomes for
535 incorporation of D-amino acids into proteins. Biochemistry (Mosc.) 45,
536 15541-15551.
- 537 68. Maini, R., et al. (2013) Incorporation of beta-amino acids into dihydrofolate
538 reductase by ribosomes having modifications in the peptidyltransferase center.
539 Bioorg. Med. Chem. 21, 1088-1096.
- 540 69. LaRiviere, F.J., et al. (2001) Uniform binding of aminoacyl-tRNAs to
541 elongation factor Tu by thermodynamic compensation. Science 294, 165-168.
- 542 70. Schrader, J.M., et al. (2011) Tuning the affinity of aminoacyl-tRNA to
543 elongation factor Tu for optimal decoding. Proc. Natl. Acad. Sci. U. S. A. 108,
544 5215-5220.
- 545 71. Nissen, P., et al. (1995) Crystal structure of the ternary complex of Phe-
546 tRNAPhe, EF-Tu, and a GTP analog. Science 270, 1464-1472.
- 547 72. Nakata, H., et al. (2006) Binding efficiency of elongation factor Tu to tRNAs
548 charged with nonnatural fluorescent amino acids. Anal. Biochem. 348, 321-
549 323.
- 550 73. Doi, Y., et al. (2007) Elongation factor Tu mutants expand amino acid
551 tolerance of protein biosynthesis system. J. Am. Chem. Soc. 129, 14458-
552 14462.

553 74. Park, H.S., et al. (2011) Expanding the genetic code of *Escherichia coli* with
554 phosphoserine. *Science* 333, 1151-1154.

555 75. Guo, J., et al. (2009) Evolution of amber suppressor tRNAs for efficient
556 bacterial production of proteins containing nonnatural amino acids. *Angew.*
557 *Chem. Int. Ed. Engl.* 48, 9148-9151.

558 76. Jeong, K.W., et al. (2014) A tRNA body with high affinity for EF-Tu hastens
559 ribosomal incorporation of unnatural amino acids. *RNA* 20, 632-643.

560 77. Wang, J., et al. (2014) Peptide Formation by N-Methyl Amino Acids in
561 Translation Is Hastened by Higher pH and tRNA. *ACS Chem. Biol.*

562 78. Rackham, O. and Chin, J.W. (2005) A network of orthogonal ribosome x
563 mRNA pairs. *Nat. Chem. Biol.* 1, 159-166.

564 79. Liu, D.R., et al. (1997) Engineering a tRNA and aminoacyl-tRNA synthetase
565 for the site-specific incorporation of unnatural amino acids into proteins in
566 vivo. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10092-10097.

567 80. Chin, J.W., et al. (2003) An expanded eukaryotic genetic code. *Science* 301,
568 964-967.

569 81. Xie, J. and Schultz, P.G. (2005) An expanding genetic code. *Methods* 36, 227-
570 238.

571 82. Davis, L. and Chin, J.W. (2012) Designer proteins: applications of genetic
572 code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* 13, 168-182.

573

574

Box 1. Intracellular genetic code reprogramming

In contrast to in vitro genetic code reprogramming (which generally uses a purified recombinant translation system), intracellular genetic code reprogramming exploits the translation machinery in living bacterial or eukaryotic cells to synthesize polypeptide(s) containing one or more non-canonical amino acids [13]. Thus, each approach faces significantly different challenges with respect to achieving efficient reprogramming. For intracellular genetic code reprogramming, the greatest challenges arise from undesired interactions with the natural translation processes of the host cell. As a consequence, many advances in this area have involved the engineering of orthogonality into the reprogrammed translation machinery, such that it is able to function in parallel to the normal metabolism of the cell [78]. For example, the identification of AARS/tRNA pairs that do not cross react with the endogenous AARSs and tRNAs is fundamental to the field [46, 79-81]. This requirement for orthogonality decreases the flexibility of intracellular genetic code reprogramming (as compared to in vitro reprogramming), and most studies involve the introduction of only a single non-canonical amino acid at a single site of a given polypeptide. However, the great strength of such techniques is that the synthesized proteins can be studied in their normal biochemical environment. Applications include the use of photo-crosslinkable residues for studying weak protein interactions, and studies of signaling that use genetic code reprogramming to introduce moieties that would normally occur through post-translational modification [82]. More recent expansion of intracellular genetic code reprogramming techniques to multi-cellular organisms will doubtless provide further opportunities for such in situ biochemical studies.

600 **Box 2. Outstanding questions**

- 601 1. What are the limitations of peptide synthesis using the natural translational
602 machinery? That is, of the almost infinite number of non-canonical amino
603 acids, which can be efficiently incorporated into polypeptides?
- 604 2. How can the translation machinery be engineered to allow the practical and
605 efficient incorporation of amino acids that are not translatable using current
606 techniques?
- 607

608 **Glossary**

609

610 **Aminoacyl-tRNA synthetase (AARS):** An enzyme that catalyzes the
611 aminoacylation of a tRNA using an amino acid substrate to produce an
612 aminoacyl-tRNA.

613

614 **Aminoacylation:** The process of charging an amino acid onto the 3' end of a
615 tRNA molecule through an ester linkage; a reaction catalyzed by AARSs and
616 flexizymes.

617

618 **EF-Tu: Elongation factor thermo unstable.** A bacterial elongation factor that
619 recognizes aminoacyl-tRNA molecules and transports them to the active site of
620 the ribosome.

621

622 **Elongation factor:** A protein co-factor that facilitates elongation of the nascent
623 peptide chain.

624

625 **Flexible in vitro translation (FIT) system:** A genetic code reprogramming
626 technique based on the combined use of flexizyme aminoacylation with a
627 purified recombinant translation system.

628

629 **Flexizyme:** An *in vitro* evolved RNA enzyme (ribozyme) that catalyzes the
630 aminoacylation of a tRNA using an amino acid substrate to produce an
631 aminoacyl-tRNA.

632

633 **Initiation factor:** A protein co-factor that facilitates initiation of translation.

634

635 **Non-canonical amino acid:** An amino acid that differs structurally or
636 stereochemically from the 20 canonical amino acids that occur in nature.

637

638 **Purified recombinant translation system:** An *in vitro* translation technique in
639 which all required components (ribosomes, AARSs, tRNAs, protein cofactors, *etc*)
640 are purified prior and combined in the absence of other cellular factors.

641

642 **Release factor:** A protein co-factor that facilitates release of a synthesized
643 polypeptide from the ribosome.

644

645 **Stop codon suppression:** A genetic code reprogramming technique based on
646 the use of tRNAs with anti-codons complimentary to stop codons.

647

648 **Figure legends**

649

650 **Figure 1. Prokaryotic translation.** Amino acids are charged onto their cognate
651 tRNAs through the activity of aminoacyl-tRNA synthetases (AARSs) that specifically
652 recognize each amino acid--tRNA pair. The acylated tRNA (AA-tRNA) is then
653 recognized by EF-Tu, which delivers it to the A site of the ribosome, where EF-Tu is
654 released. The ribosome catalyzes the transfer of the nascent peptide onto the amino
655 acid of the acylated tRNA, increasing the length of the peptide by a single residue,
656 and the peptide--tRNA moves into the P site of the ribosome. Following the next
657 cycle of elongation the free tRNA moves to the E site of the ribosome and is released.

658

659 **Figure 2. In vitro genetic code reprogramming using AARSs.** A general scheme
660 for AARS-mediated genetic code reprogramming is shown in panel A. Components
661 of the reaction that can be altered to affect reprogramming are highlighted in red.
662 Examples of non-canonical amino acids that may be translated using such a technique
663 are shown in panel B, with the moieties differing from the canonical analogue
664 highlighted in red. The canonical codon table (upper section of panel C) and an
665 example reprogrammed codon table (lower section, panel C) are also shown. Note
666 that each non-canonical amino acid is incorporated at the same codon(s) as its
667 canonical analogue. M_a = Met analogue, R_a = Arg analogue, etc.

668

669 **Figure 3. In vitro genetic code reprogramming using flexizymes.** A general
670 scheme for flexizyme-mediated genetic code reprogramming is shown in panel A.
671 Components of the reaction that are may be altered to affect reprogramming are

672 highlighted in red. Examples of non-canonical amino acids that can be translated
673 using such a technique are shown in panel B, with the moieties differing from the
674 canonical analogue highlighted in red. The canonical codon table (upper section of
675 panel C) and an example reprogrammed codon table (lower section, panel C) are also
676 shown. MePhe = *N*-methyl-Phe, MeSer = *N*-methyl-Ser, MeGly = *N*-methyl-Gly (no
677 structure shown), MeAla = *N*-methyl-Ala, ClAc-Tyr = *N*-Chloroacetyl-D-Tyr.

678

679 **Figure 4. EF-Tu--tRNA interactions.** The overall structure of EF-Tu (white) bound
680 to aminoacylated tRNA-Phe (grey) is shown in panel A. The charged Phe residue in
681 the amino acid binding pocket of EF-Tu is highlighted in magenta. The three
682 nucleotides in the T-stem of the tRNA that contribute the majority of the sequence
683 specificity of tRNA binding to EF-Tu are highlighted in orange, with the region of
684 EF-Tu to which they bind highlighted in blue (note, however, that many other
685 nucleotides are involved in the EF-Tu/tRNA interaction but are highly conserved
686 across most tRNAs). Panel B shows a close up of the EF-Tu amino acid binding
687 pocket, including the residues involved in binding to the amino acid. Alterations
688 indicated in black have been shown to improve translation of phosphoserine [74]
689 whilst alterations in red have been shown to improve translation of large aromatic
690 non-canonical amino acids [73]. Panel C shows a close up of the sequence-specific
691 binding conferring nucleotides in the tRNA, and the residues of EF-Tu with which
692 they interact. Hydrogen bonding is indicated by dashed lines.

Figure 1

Figure 2

Figure 3

Figure 4

(A)

(B)

(C)