The RNA origin of transfer RNA aminoacylation and beyond

Hiroaki Suga1,*, Gosuke Hayashi1 and Naohiro Terasaka2

1Department of Chemistry, School of Science, and 2Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-0033 Tokyo, Japan

Aminoacylation of tRNA is an essential event in the translation system. Although in the modern system protein enzymes play the sole role in tRNA aminoacylation, in the primitive translation system RNA molecules could have catalysed aminoacylation onto tRNA or tRNA-like molecules. Even though such RNA enzymes so far are not identified from known organisms, in vitro selection has generated such RNA catalysts from a pool of random RNA sequences. Among them, a set of RNA sequences, referred to as flexizymes (Fxs), discovered in our laboratory are able to charge amino acids onto tRNAs. Significantly, Fxs allow us to charge a wide variety of amino acids, including those that are non-proteinogenic, onto tRNAs bearing any desired anticodons, and thus enable us to reprogramme the genetic code at our will. This article summarizes the evolutionary history of Fxs and also the most recent advances in manipulating a translation system by integration with Fxs.

Keywords: RNA world; ribozyme; flexizyme; translation; genetic code reprogramming

1. INTRODUCTION
It has been established indisputably by the X-ray structure of ribosome that its peptidyl transfer centre consists of only RNA [1]. This has given strong support to the hypothesis that primitive translation events could have been catalysed by RNA alone in the RNA world. However, tRNA aminoacylation, another essential reaction of translation, is solely catalysed by protein enzymes, aminoacyl-tRNA synthetases (ARSs), in the modern system. In the RNA world hypothesis, this discrepancy could be explained by the existence of ARS-like ribozymes [2–6]. Although such a ribozyme working in cells has never been discovered so far, over the past decade we and others have isolated such ribozymes by means of in vitro selection [2,6–9]. We herein summarize the evolutionary history of a class of such ribozymes isolated from our laboratory. This class of ribozymes, referred to as flexizymes (Fxs), is capable of charging amino acids onto the 3'-hydroxyl group of tRNA end-like protein ARSs [10]. Moreover, using the unique property of the versatility of Fxs as tRNA acylation catalysts, we and others have developed strategies of manipulating the genetic code, so-called 'genetic code reprogramming' [11–13], with the integration of a custom-made reconstituted cell-free translation system [11,14,15]. We also briefly introduce the latest research outcomes for the synthesis of non-standard peptides and perspectives in the selection of novel peptides against protein targets and engineering of the translation machinery.

2. EVOLUTION OF AMINOACYL-tRNA SYNTHETASE-LIKE RIBOZYMES (FLEXIZYMES)
Ligation of one of the hydroxyl groups at the tRNA 3'-end with the carboxyl group of amino acid involves two chemistries; (i) activation of the carboxyl group by adenosine triphosphate to yield aminoacyl-adenosine monophosphate (aminoacyl-AMP), and (ii) condensation of aminoacyl-AMP and tRNA to yield aminoacyl-tRNA. In the modern translation system, single ARS species catalyse both chemical reactions. Because of the intrinsic high energy of the aminoacyl-AMP intermediate, step (i) is a highly up-hill reaction, and thus it requires reaction with the tRNA hydroxyl group without long exposure to bulk water. Although certain RNA molecules may be able to catalyse the reaction of step (i) [16], it is however difficult to achieve both steps simultaneously. Alternatively, step (i) may rely on the formation of appropriate esters with prebiotically compatible activating groups, such as cyanomethyl ester [8,17] (the alcohol moiety can be produced by the condensation of cyanide and formaldehyde) or thioester. Thus, primitive ARS-like ribozymes might have catalysed step (ii) in a selective manner where certain amino acids are charged onto the 3'-terminal hydroxyl group of tRNA-like RNA.

We have hypothesized that primitive ARS-like ribozymes could have evolved as a part of a 5'-leader sequence. This hypothesis seems reasonable in the RNA world hypothesis since M1 RNA of ribonuclease (RNase) P, a known naturally occurring ribozyme [18,19], selectively cleaves the 5'-leader sequence from tRNA independently from tRNA species. Thus, a 5'-leader sequence could have acted as a specific self-aminoacylating catalyst to a cognate tRNA via the covalent linkage, and M1 RNA-like ribozyme would have removed it to yield the mature...
aminoacyl-tRNA (figure 1). On the basis of this hypothesis, we constructed a RNA library with 70-nucleotide (nt) random sequences attached to the 5'-end of a tRNA\textsubscript{Gln} (figure 2a). In vitro selection was performed using a phenylalanine (Phe) derivative of which the amino group was modified with biotin and the carbonyl group was activated with cyanomethyl ester, and then active species were isolated based on the ability to self-aminoacylate any site of available hydroxyl groups [9]. This selection attempt fortunately yielded a single kind of sequence, referred to as r24, capable of aminoacylating the 3'-terminal hydroxyl group (figure 2b). Importantly, r24 could be removed by the treatment of M1 RNA of RNase P, and thus it fitted well with the hypothetical view of evolution of self-aminoacylating 5'-leader ribozyme (figure 1).

We performed metal-dependent kinetics and probing, chemical probing, and nucleotide analogue interference mapping of r24, to reveal the sites necessary for Mg\textsuperscript{2+}, tRNA and Phe binding [20,21]. Complementary bases to the tRNA 3'-end region, A73–C75, were found in the 3'-terminal region of r24, and their Watson–Crick base-pair interactions were confirmed by their compensatory mutations albeit tolerance of wobble G73–U\textsuperscript{*} or U73–U\textsuperscript{*} pair (U\textsuperscript{*} denotes the r24 base at position 72). Based on the above knowledge, a shorter version of r24, called r24\textsubscript{mini}, being 57-nt long, was constructed to confirm the wild-type activity [20] (figure 2c). Accordingly, it was demonstrated that r24\textsubscript{mini} itself could charge Phe onto tRNA as a \textit{trans}-acting enzyme like protein enzymes.

To further minimize the structure of r24\textsubscript{mini}, we maintained the essential bases in the amino acid and tRNA-binding regions but the rest was re-randomized for the second activity selection (figure 2d) [10]. Sequence alignment of the active species revealed that the 3'-terminal region after the tRNA-binding site has no conservation of sequence, suggesting that this region is unimportant for activity. In fact, deletion of this region, giving a new construct only 45-nt long, did not affect the intrinsic catalytic activity of r24\textsubscript{mini}, but rather increased the fraction of correctly folded active species of the \textit{trans}-acting ribozyme, i.e. increased the final yield of aminoacyl-tRNA product. Moreover, this short ribozyme was able to charge a variety of aromatic amino acids, such as tyrosine, tryptophan and Phe analogues, onto tRNA as far as they are activated by cyanomethyl ester [22]. Because of the simplicity of the base-pair interactions between this ribozyme and tRNA, it allowed us to prepare a variety of mischarged aminoacyl-tRNAs independent of its body or anticodon sequences. Owing to the flexible utility of this ribozyme, we named it Fx (figure 2e).

Although the prototype Fx was a useful tool for the preparation of mischarged tRNAs with Phe or Phe analogues, it would become a more versatile and convenient tool if aminoacyl-tRNAs with more variety of kinds of amino acids could be prepared at the user’s will. Therefore, we turned our efforts to develop a more versatile and effective Fx tool [11]. A new Fx library was constructed, where the structural scaffold, P1 and P2, was kept the same but bases in putative small loops and junctions were re-randomized along
with insertion of a few bases (figure 2f). A drastic change was made in the substrate design. Because during the course of our studies, we learned that the methylene-aromatic ring of the Phe side chain was the major recognition element, we designed such an aromatic structure in the leaving group of the ester. In this way, the substrate recognition of Fx variants would be directed to the common leaving group of the amino acid donor substrate, and therefore any side chain of amino acids would become acceptable. In fact, selection of the above Fx library against lysine with 3,5-dinitrobenzyl ester leaving group successfully yielded a new variant, referred to as dFx (d stands for dinitrobenzyl ester), capable of effectively charging non-aromatic amino acid on tRNAs (figure 3) [11]. Meanwhile, we also repeated selection against Phe-cyanomethyl ester to re-optimize the sequence of the prototype Fx, generating a new variant, referred to as eFx (e stands for enhanced; figure 3) [11]. This eFx turns out to be a good catalyst for not only aromatic amino acids but also non-aromatic amino acids activated with 4-chlorobenzyl thioester. We have also developed another variant aFx that reacts with amino acids activated with 4-chlorobenzyl thioester. We have also developed another variant aFx that reacts with amino acids activated with an amino-derivatized benzyl thioester group, which is a useful catalyst when the substrate has poor water-solubility owing to a high hydrophobicity of its side chain or certain modifications of amino acids (figure 3) [23].

To this end, we have devised a versatile ‘Fx’ system that enables us to perform aminoacylation of any tRNAs with a wide array of proteinogenic amino acids and non-proteinogenic ones, such as d-amino acids [24], β-amino acids [11], N-methyl-amino acids [25] and N-acyl-amino acids [26]. Having this system, we initiated a new project to manipulate the genetic code, so-called genetic code reprogramming, leading us to express non-standard peptides containing multiple non-proteinogenic amino acids.
Figure 3. Flexizymes and their cognate leaving groups in substrates. Chemical structures of a benzyl ester and thioester leaving group. R represents amino acid side chains including non-proteinogenic ones. Each flexizyme recognizes the specific leaving group and charges the acyl group onto the 3'-hydroxyl group at the tRNA 3'-end. LG, leaving group; DNB, dinitrobenzyl ester; CME, cyanomethyl ester; CBT; p-chlorobenzyl thioester; ABT, amino-derivatized benzyl thioester.

3. GENETIC CODE REPROGRAMMING FACILITATED BY FLEXIBLE IN VITRO TRANSLATION SYSTEM

Genetic code reprogramming [11–13] is a concept in which some codons are reassigned from proteinogenic amino acids to non-proteinogenic ones, allowing expression of non-standard polypeptides by using an in vitro translation apparatus. Despite the simplicity of the concept, the codon reassignments require two major technical challenges; (i) preparation of aminoacyl-tRNAs with non-proteinogenic amino acids, and (ii) making vacant codons to which the above amino acids are reassigned. Although there are several different approaches available for (i), such as semi-chemical synthetic method [12,27–32] or ARS-mischarging method [13,33–37], Fxs are far superior to others in terms of flexibility of the choices of non-proteinogenic amino acids and tRNAs as well as facility and reliability [11].

In the ordinary reconstituted cell-free translation, 30 pre-purified translation factors are essential for normal protein synthesis. Removing specific amino acids and/or their cognate ARSs from the translation system pauses the expression or induces peptidyl-tRNA drop-off at the vacant codons in (ii). When the Fx system is integrated with the custom-made reconstituted translation apparatus, non-proteinogenic amino acids charged on specific tRNAs reoccupy the vacant codons. We referred to this Fx–custom-made translation integrated system as the flexible in vitro translation (FIT) system [14].

In the last few years, we have made major efforts to express a wide variety of non-standard polypeptides containing multiple non-proteinogenic amino acids in FIT systems. Non-proteinogenic amino acids (resulting peptides are in parentheses) used in these studies (table 1) are (i) N-methyl-amino acids in elongation (N-methyl-peptides) [25,38], (ii) N-alkyl glycine in elongation (peptoid–peptide hybrid) [39], (iii) N-acyl-amino acids and exotic peptides in initiation (N-terminal modified peptides) [26], (iv) α-hydroxy acids in elongation (polysters or backbone cyclic peptides) [38,40], and (v) those with unique side chains in initiation and elongation, which can react with specific partner amino acids (e.g. cyclic peptides closed by non-reducible covalent bond) [26]. Combination of any of these methods yields structurally diverse non-standard peptides such as natural product-like peptides and artificial cyclic peptides, allowing us to express unique non-standard peptide libraries from mRNA templates under reprogrammed genetic codes (figure 4).

Table 1. Representative examples for the incorporation of non-proteinogenic amino acids into a peptide chain using the FIT system.

<table>
<thead>
<tr>
<th>non-proteinogenic amino acids</th>
<th>amino acid position</th>
<th>peptides</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-methyl amino acids</td>
<td>elongation</td>
<td>N-methyl peptides</td>
<td>[25,38]</td>
</tr>
<tr>
<td>N-alkyl glycine</td>
<td>elongation</td>
<td>N-acyl peptide hybrid</td>
<td>[39]</td>
</tr>
<tr>
<td>N-acyl amino acids</td>
<td>initiation</td>
<td>N-terminal modified peptides</td>
<td>[26]</td>
</tr>
<tr>
<td>α-hydroxy acids</td>
<td>elongation</td>
<td>polysters or backbone cyclic peptides</td>
<td>[38,40]</td>
</tr>
<tr>
<td>chloro-acetyl amino acid</td>
<td>initiation</td>
<td>thioether-based cyclic peptides</td>
<td>[26]</td>
</tr>
</tbody>
</table>

Phil. Trans. R. Soc. B (2011)
dozen distinct targets with dissociation constants of low nanomolar to sub nanomolar. Moreover, some of these molecules exhibited inhibitory activities of enzymes as well as protein–protein interactions. Such results will be reported in the near future.

Another direction that we have been taking with Fx technology in our laboratory is to investigate the ribosomal translation machinery by using aminoacyl-tRNAs with a non-CCA-3'-end. We recently found that Fxs are able to catalyse aminoacylation on non-CCA-3'-end tRNAs efficiently when the Fx’s tRNA-binding bases were mutated to make base pairs with the tRNA 3'-end (G. Hayashi and H. Suga 2011, unpublished data). Furthermore, in a combination of engineered ribosomes having compensatory mutations at rRNA against non-CCA-3'-end tRNAs, we have developed an artificial ribosome–tRNAs pair that acts orthogonal to the naturally occurring ribosome–tRNAs pair, i.e. we have created an ‘orthogonal genetic code’.

In conclusion, Fxs are a unique tool to open many new doors for not only basic science but also applications involving the translation apparatus. More new ideas should follow in the future.

This work was supported by grants from the Japan Society for the Promotion of Science Specially Promoted Research (21000005), research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO), and the World Class University project of the MEST and the NRF (R31-2008-000-10103-0). G.H. was supported by a Japan Society for the Promotion of Science Research Fellowship for Young Scientists.

REFERENCES
14 Goto, Y., Katoh, T. & Suga, H. 2011 Flexizymes for
16 Kumar, R. K. & Yarus, M. 2001 RNA-catalyzed amino
19 Frank, D. & Pace, N. 1998 Ribonuclease P: unity and
22 Murakami, H., Kourouklis, D. & Suga, H. 2003 Using a
24 Goto, Y., Murakami, H. & Suga, H. 2008 Initiating