Clearly, it is necessary to use another methodology that allows us to control the undesirable competition in translation so as to polymerize many kinds of α-hydroxy acids. More recently an alternative technology, referred to as genetic-code reprogramming, has been devised to resolve the above problem.

In genetic-code reprogramming some of the proteinogenic amino acids and/or other components are withdrawn from the translation system so as to break the tight relationship between the amino acids and cognate codons in the genetic code. In 2003, Forster et al. introduced this concept by demonstrating that three kinds of nonproteinogenic amino acids were reassigned to three different codons and incorporated into a peptide in succession by sense suppression.[10] Despite the fact that this early work used a translation system that was unable to turnover, two significant benefits over the classical method that used amber suppression[3,4] are evident. First, because the proteinogenic amino acids, the codons of which are aimed at reprogramming, are withdrawn from the translation system, there are no direct competitors against the desired suppression. This is in sharp contrast to the amber suppression in which the release factor competes with a suppressor $tRNA_{CUA}$ that is charged with a nonproteinogenic amino acid to terminate the translation. Therefore, it is expected that the efficiency of sense suppression would be higher than that of amber suppression. Second, because the assignment of nonproteinogenic amino acids can be achieved by choosing any desired codons, and is not restricted to stop codons, the number of nonproteinogenic amino acids for the codon reassignments is—in principle—unlimited. If these two concepts were achieved, we expect that nonstandard peptides or even other types of biopolymers could be synthesized with the translation machinery. Stimulated by Foster's experiment, several groups have started work on further development of genetic-code reprogramming.[11–25]

For nearly ten years, we have engaged in a project to develop artificial ribozymes, called "flexizymes", which are capable of charging amino acids onto tRNA.[16,26–31] The latest version of the flexizyme system enables us to charge virtually any amino acid, including nonproteinogenic ones, onto tRNA that bear various anticodons.[16] Importantly, we have found that it is able to charge a variety of α-hydroxy acids onto tRNAs, to yield hydroxyacyl-tRNAs (ha-tRNAs).[2,16] Thus, the use of this system should facilitate the study of the ribosomal synthesis of polyesters to address the questions that were raised earlier when it was combined with the genetic-code reprogramming methodology.

Despite the fact that some α-hydroxy acids were successfully incorporated into a peptide (protein) chain at a specific site by the nonsense suppression,[5–9,32,33] to the best of our knowledge there is no report for the successive incorporations of α-hydroxy acids by any means except for Fahnestock's experiment. To successfully achieve the polymerization of α-hydroxy acids by ribosome catalysis, two technical improvements turned out to be critical. First, we needed to use a special reconstituted *E. coli* cell-free translation system,[25,34–37] referred to as *w*PURE, in which both amino acids and cognate aminoacyl-tRNA synthetases (ARSs) for the reprogramming codons were *withdrawn* from the ordinary PURE translation system. Because α-hydroxy acids are intrinsically poorer substrates for ribosome than α-amino acids,[38–40] the *w*PURE system that lacked only α-amino acids was insufficient to control the background level of competitive incorporation of α-amino acids into the polyester chain. Second, we developed two types of engineered, orthogonal tRNAs that are not aminoacylated by *E. coli* ARSs, $tRNA^{Asn\text{-}E}$ and $tRNA^{MLAsn}$ (Figure 2A and B), to carry the α-hydroxy acids—in the experiments, we used three tRNAs, but two of these belonged to the family of $tRNA^{Asn\text{-}E}$ and behaved virtually the same. We performed the reassignment of seven codons to seven α-hydroxy acids (Figure 3A), but one of the codons, CAG (Gln), suffered from minor misincorporation of Lys as the LysRS catalyzed the mischarge of Lys onto $tRNA^{Asn\text{-}E}_{CUG}$, the body sequence of which was our standard scaffold for orthogonal tRNAs with various anticodons (Figure 2A). We thus screened potential body sequences of orthogonal tRNA and found $tRNA^{MLAsn}$, derived from mycobacteriophage L5 $tRNA^{Asn}$.

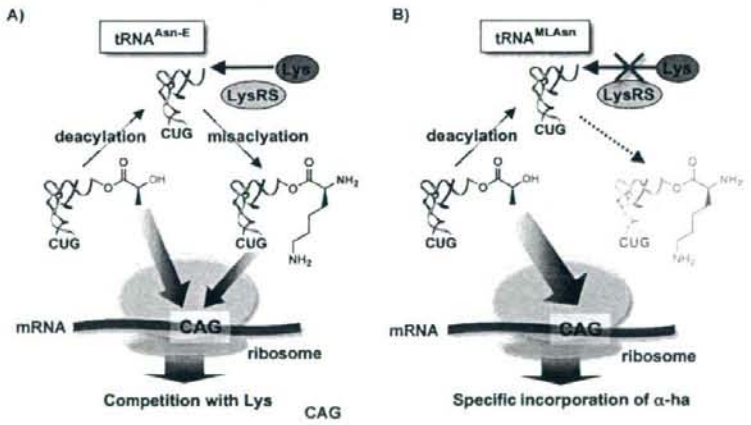


Figure 2. A strategy for controlling the competition with undesirable misincorporation into the polyester chain. A) Competition between α-hydroxy acid and Lys for $tRNA^{Asn\text{-}E}$ at the CAG codon. Because the deacylated $tRNA^{Asn\text{-}E}$ suffered from mis-lysinylation catalyzed by Lys-tRNA synthetase (LysRS), the resulting Lys-$tRNA^{Asn\text{-}E}$ competed with ha-tRNA for CAG decoding; this resulted in misincorporation of Lys into the polyester chain. B) Specific incorporation of α-hydroxy acid on $tRNA^{MLAsn}$ at the CAG codon. The deacylated $tRNA^{MLAsn}$ was inert to LysRS, which prevented its mis-lysinylation. Thus, the CAG codon was exclusively decoded by ha-$tRNA^{MLAsn}$; this led to the programmed incorporation of the α-hydroxy acid.

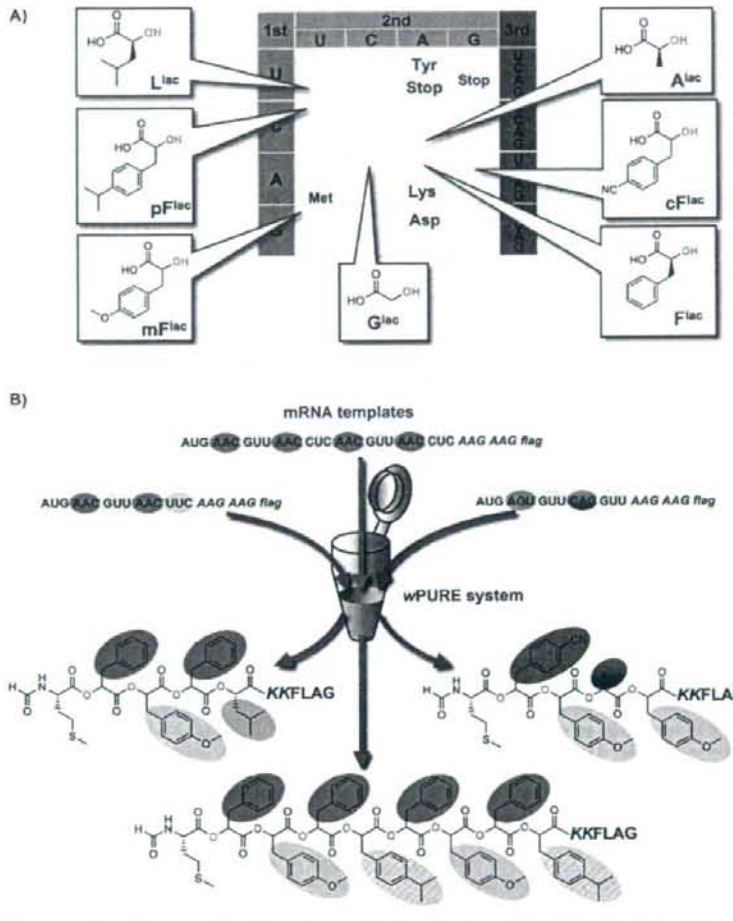


Figure 3. mRNA-directed polyester synthesis by using genetic-code reprogramming. A) The genetic-code reprogrammed by seven kinds of α-hydroxy acids; B) mRNA-directed polyester synthesis. Each template yielded respective polyester sequences under the control of the reprogrammed genetic code shown in A).

This $tRNA^{MLAsn}$ was inert to LysRS and other *E. coli* ARSs, and therefore the CAG codon could also be used for reprogramming (Figure 2B).

Messenger RNA-Directed Polyester Synthesis

Using the above-described *w*PURE and flexizyme systems, we attempted to express polyesters in a mRNA-dependent manner (Figure 1B). We designed the mRNA template sequence to encode a polyester–peptide hybrid. In this polymer, the polymerization was initiated from formylated methionine assigned by AUG, elongated with α-hydroxy acids assigned by seven different codons, and further elongated by three amino acids (K, D, and Y) yielding a modified sequence of flag peptide, KKDYKDDDDK (Figure 3B). This flag peptide facilitated isolation as well as detection/quantification of the products by means of tricine-SDS-PAGE when [^{14}C]-D was included in the translation mixture to yield the [^{14}C]-flag peptide at the C terminus of polyester. It should be noted that *w*PURE is a coupled transcription–translation system, and therefore the actual template is the corresponding DNA that bears a T7 promoter sequence. Thus, the DNA template was added to the *w*PURE system along with ha–tRNAs that were prepared by the flexizyme system, and the resulting product was assayed by MALDI-TOF mass spectrometry and tricine-SDS-PAGE.

We demonstrated the expression of a trimer-polyester and four tetramer-polyesters with a variety of compositions of α-hydroxy acids. In the case of these polyesters, their expression level was comparable to wild-type peptide expression, and gave a quantity that ranged from 5–15 pmol per 5 μL. Most importantly, in all cases the MALDI-TOF analysis revealed a single major peak that was consistent with the molecular mass (MS) of the anticipated composition of polyesters. Therefore, this was the first direct evidence that showed ribosomal synthesis of a polyester, the sequence and length of which were fully controlled by the mRNA template. We also extended the polymerization of α-hydroxy acids to pentamer, hexamer, octamer, and dodecamer. In all cases, we observed a band on tricine-SDS-PAGE in the individual expression; this suggests that the product contained the corresponding [^{14}C]-flag peptide, that is, it is likely that the polyester-flag hybrids were expressed. The MALDI-TOF analysis of the polyester-flag hybrids, however, gave no peak; we examined a variety of analytical conditions by altering the supporting matrixes and/or laser powers, but the expected MS that corresponded to the full-length polyester–flag hybrid was not observed. We assumed that the product was lost during sample manipulation, for example, by aggregation or precipitation of polyesters. We thus treated the crude product under basic condition to hydrolyze it, and analyzed the resulting sample by MALDI-

TOF and found the peaks for $(ha)_2$–flag and ha–flag in all cases. Most importantly, the observed MS for each of the $(ha)_2$–flag was consistent with the MS value that was expected from that expressed from the respective mRNA sequence. Therefore, we concluded that it was very likely that the full-length polyesters were expressed in accordance with the mRNA sequences.

Outlook

Challenges successfully achieved and still remaining

We have thus far resolved the following questions: 1) Can the polyester sequence be programmed by mRNA? Yes, it can be done by genetic-code reprogramming. We have demonstrated the reassignment of seven different α-hydroxy acids to seven codons and successfully polymerized them to polyesters according to the mRNA templates. 2) How many α-hydroxy acids can be successively polymerized? We have demonstrated polyester synthesis up to tetramer length as confirmed by MS evidence of the full-length products, and up to the dodecamer length as confirmed by MS evidence of the fragmented products.

Although the above two achievements resolved most issues that remained unanswered in the Fahnestock experiment, some issues are still unresolved. The expression level of polyesters longer than dodecamers was very low, which made it difficult to ensure their expression. Moreover, we were only able to detect hydrolyzed products, rather than the full-length polyesters, when their lengths were longer than five. Thus, from a technical point of view, we need to achieve better polymerization efficiency of α-hydroxy acids than the current system, and to develop a method to detect the polyesters as intact full-length products. How can we solve these problems? Unfortunately, we currently do not have a definitive approach. For the former improvement, we might be able to engineer elongation factor Tu (EF-Tu)[41] or orthogonal tRNAs[42,43] for higher affinity of ha–tRNA because it is known that EF-Tu binds F^{Nc}–tRNA nearly 300-fold poorer than Phe–tRNA.[39] Also, it would be critical to engineer the ribosome itself[44–47] to increase the acyl-transfer rate for the α-hydroxy acids in the active site, because the transfer rate of α-hydroxy acids has been estimated to be at least tenfold slower than that of α-amino acids.[38,40] By combining these approaches, it might be possible to solve the former problem. The latter problem might be solved by an appropriate manipulation and isolation of the products to avoid their loss, and also by MALDI-TOF analysis under the conditions that are specially designed for the analysis of labile polyesters.

Despite the above technical problems, our technology as a whole has great potential to serve as a platform to new research directions. First, mRNA-programmed polyester synthesis allows us to polymerize multiple α-hydroxy acids that have various side-chains with a desired sequence. Moreover, combining ribosomal polymer synthesis with some in vitro display techniques[48,49] facilitates the selection of functional sequences of polyesters or polyester–peptide hybrids from the corresponding libraries with high complexities. Considering the fact that the ester bond has unique properties—in terms of plasticity and rigidity—that are distinct from the peptide bond, it is of great interest to generate functional polyesters and polyester–polypeptide hybrid biopolymers and investigate their details. Second, α-hydroxy acids are also useful for investigating elongation chemistry in the ribosome. In fact, the use of an α-hydroxy acid (phenyllactic acid) in elongation slows the rate of the peptidyl-transfer reaction in ribosomes;[40] this allows researchers to ask specific questions of the chemical event that are not readily accessible by the use of standard amino acids. Although Rich's classic method was used to prepare the ha–tRNA in the reported work so far,[40] our flexizyme system readily expands the usable repertoire of α-hydroxy acids (and also nonproteinogenic amino acids) without limiting the tRNA species. This would provide researchers a nearly unlimited capability to investigate the elongation chemistry that occurs in the active site of the ribosome.

On the other hand, the methodology developed through this study has already given many fruits. For instance, the same methodology has been fully utilized in our recent work for the synthesis of *N*-methyl-peptides.[22] In this work, because the analytical problem did not exist, we were able to successfully detect up to dodecamer *N*-methyl-peptides by MALDI-TOF. Similarly, we recently reported the ribosomal synthesis of a cyclic peptide that is highly resistant to peptidases, and the methodology used was an adaptation of that developed for the polyester synthesis.[21,23,24] Thus, mRNA-directed polyester synthesis was the first example of a series of studies,[21–24,50] and the methodology described here provides a foundation for new avenues in ribosomal synthesis of biopolymers.

Keywords: genetic code · ribosomes · template synthesis · translation · tRNA

[1] S. Fahnestock, A. Rich, *Science* **1971**, *173*, 340–343.
[2] A. Ohta, H. Murakami, E. Higashimura, H. Suga, *Chem. Biol.* **2007**, *14*, 1315–1322.
[3] J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, E. S. Diala, *J. Am. Chem. Soc.* **1989**, *111*, 8013–8014.
[4] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188.
[5] J. D. Bain, E. S. Diala, C. G. Glabe, D. A. Wacker, M. H. Lyttle, T. A. Dix, A. R. Chamberlin, *Biochemistry* **1991**, *30*, 5411–5421.
[6] J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **1992**, *255*, 197–200.
[7] J. T. Koh, V. W. Cornish, P. G. Schultz, *Biochemistry* **1997**, *36*, 11314–11322.
[8] P. M. England, H. A. Lester, D. A. Dougherty, *Biochemistry* **1999**, *38*, 14409–14415.
[9] P. M. England, Y. Zhang, D. A. Dougherty, H. A. Lester, *Cell* **1999**, *96*, 89–98.
[10] A. C. Forster, Z. Tan, M. N. Nalam, H. Lin, H. Qu, V. W. Cornish, S. C. Blacklow, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6353–6357.
[11] A. Frankel, S. W. Millward, R. W. Roberts, *Chem. Biol.* **2003**, *10*, 1043–1050.
[12] C. Merryman, R. Green, *Chem. Biol.* **2004**, *11*, 575–582.
[13] Z. Tan, A. C. Forster, S. C. Blacklow, V. W. Cornish, *J. Am. Chem. Soc.* **2004**, *126*, 12752–12753.
[14] K. Josephson, M. C. Hartman, J. W. Szostak, *J. Am. Chem. Soc.* **2005**, *127*, 11727–11735.
[15] M. C. Hartman, K. Josephson, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4356–4361.

[16] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357–359.
[17] F. P. Seebeck, J. W. Szostak, *J. Am. Chem. Soc.* **2006**, *128*, 7150–7151.
[18] M. C. Hartman, K. Josephson, C. W. Lin, J. W. Szostak, *PLoS ONE* **2007**, *2*, e972.
[19] S. Sando, K. Abe, N. Sato, T. Shibata, K. Mizusawa, Y. Aoyama, *J. Am. Chem. Soc.* **2007**, *129*, 6180–6186.
[20] B. Zhang, Z. Tan, L. G. Dickson, M. N. Nalam, V. W. Cornish, A. C. Forster, *J. Am. Chem. Soc.* **2007**, *129*, 11316–11317.
[21] Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 120–129.
[22] T. Kawakami, H. Murakami, H. Suga, *Chem. Biol.* **2008**, *15*, 32–42.
[23] Y. Sako, Y. Goto, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 241–249.
[24] Y. Sako, J. Morimoto, H. Murakami, H. Suga, *J. Am. Chem. Soc.* **2008**, *130*, 7232–7234.
[25] A. O. Subtelny, M. C. Hartman, J. W. Szostak, *J. Am. Chem. Soc.* **2008**, *130*, 6131–6136.
[26] H. Saito, D. Kourouklis, H. Suga, *Embo J.* **2001**, *20*, 1797–1806.
[27] H. Murakami, D. Kourouklis, H. Suga, *Chem. Biol.* **2003**, *10*, 1077–1084.
[28] H. Murakami, H. Saito, H. Suga, *Chem. Biol.* **2003**, *10*, 655–662.
[29] K. Ramaswamy, H. Saito, H. Murakami, K. Shiba, H. Suga, *J. Am. Chem. Soc.* **2004**, *126*, 11454–11455.
[30] D. Kourouklis, H. Murakami, H. Suga, *Methods* **2005**, *36*, 239–244.
[31] M. Ohuchi, H. Murakami, H. Suga, *Curr. Opin. Chem. Biol.* **2007**, *11*, 537–542.
[32] H. H. Chung, D. R. Benson, P. G. Schultz, *Science* **1993**, *259*, 806–809.
[33] S. W. Millward, T. T. Takahashi, R. W. Roberts, *J. Am. Chem. Soc.* **2005**, *127*, 14142–14143.
[34] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, *Nat. Biotechnol.* **2001**, *19*, 751–755.
[35] Y. Shimizu, T. Kanamori, T. Ueda, *Methods* **2005**, *36*, 299–304.
[36] Z. Tan, S. C. Blacklow, V. W. Cornish, A. C. Forster, *Methods* **2005**, *36*, 279–290.
[37] A. C. Forster, H. Weissbach, S. C. Blacklow, *Anal. Biochem.* **2001**, *297*, 60–70.
[38] S. Fahnestock, H. Neumann, V. Shashoua, A. Rich, *Biochemistry* **1970**, *9*, 2477–2483.
[39] K. H. Derwenskus, M. Sprinzl, *FEBS Lett.* **1983**, *151*, 143–147.
[40] P. Bieling, M. Beringer, S. Adio, M. V. Rodnina, *Nat. Struct. Mol. Biol.* **2006**, *13*, 423–428.
[41] Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, M. Sisido, *J. Am. Chem. Soc.* **2007**, *129*, 14458–14462.
[42] F. J. LaRiviere, A. D. Wolfson, O. C. Uhlenbeck, *Science* **2001**, *294*, 165–168.
[43] H. Asahara, O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3499–3504.
[44] L. M. Dedkova, N. E. Fahmi, S. Y. Golovine, S. M. Hecht, *J. Am. Chem. Soc.* **2003**, *125*, 6616–6617.
[45] L. M. Dedkova, N. E. Fahmi, S. Y. Golovine, S. M. Hecht, *Biochemistry* **2006**, *45*, 15541–15551.
[46] K. Wang, H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Biotechnol.* **2007**, *25*, 770–777.
[47] L. Cochella, R. Green, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3786–3791.
[48] S. W. Millward, S. Fiacco, R. J. Austin, R. W. Roberts, *ACS Chem. Biol.* **2007**, *2*, 625–634.
[49] W. W. Ja, A. P. West, Jr., S. L. Delker, P. J. Bjorkman, S. Benzer, R. W. Roberts, *Nat. Chem. Biol.* **2007**, *3*, 415–419.
[50] A. Ohta, Y. Yamagishi, H. Suga, *Curr. Opin. Chem. Biol.* **2008**, *12*, 159–167.

Received: June 30, 2008
Published online on November 4, 2008

Vol 454|17 July 2008|doi:10.1038/nature07033

LETTERS

Structural basis of specific tRNA aminoacylation by a small *in vitro* selected ribozyme

Hong Xiao[1], Hiroshi Murakami[2], Hiroaki Suga[2,3] & Adrian R. Ferré-D'Amaré[1]

In modern organisms, protein enzymes are solely responsible for the aminoacylation of transfer RNA. However, the evolution of protein synthesis in the RNA world required RNAs capable of catalysing this reaction. Ribozymes that aminoacylate RNA by using activated amino acids have been discovered through selection *in vitro*[1–3]. Flexizyme is a 45-nucleotide ribozyme capable of charging tRNA *in trans* with various activated L-phenylalanine derivatives. In addition to a more than 10^5 rate enhancement and more than 10^4-fold discrimination against some non-cognate amino acids, this ribozyme achieves good regioselectivity: of all the hydroxyl groups of a tRNA, it exclusively aminoacylates the terminal 3′-OH[5–7]. Here we report the 2.8-Å resolution structure of flexizyme fused to a substrate RNA. Together with randomization of ribozyme core residues and reselection, this structure shows that very few nucleotides are needed for the aminoacylation of specific tRNAs. Although it primarily recognizes tRNA through base-pairing with the CCA terminus of the tRNA molecule, flexizyme makes numerous local interactions to position the acceptor end of tRNA precisely. A comparison of two crystallographically independent flexizyme conformations, only one of which appears capable of binding activated phenylalanine, suggests that this ribozyme may achieve enhanced specificity by coupling active-site folding to tRNA docking. Such a mechanism would be reminiscent of the mutually induced fit of tRNA and protein employed by some aminoacyl-tRNA synthetases[4,8] to increase specificity.

Flexizyme was selected as a *cis*-acting 5′-leader sequence of tRNA that catalyses the aminoacylation of the 3′-OH of tRNA employing activated L-phenylalanine derivatives such as L-phenylalanyl-cyanomethyl ester and L-phenylalanyl-adenylate[3]. Flexizyme can be separated from its attached tRNA by RNase P. The released 45-nucleotide ribozyme is active *in trans*, with an efficiency comparable to that shown *in cis*[7]. Previous analyses indicate that, like some protein aminoacyl-tRNA synthetases (ARSs)[10], flexizyme recognizes solely the tRNA acceptor stem and is fully active with minihelices (comprising the tRNA acceptor stem and T stem-loop)[11]. To understand how this compact ribozyme achieves high regioselectivity, we solved the crystal structure of a flexizyme–tRNA minihelix fusion. This was facilitated by replacing a functionally dispensable loop[7] of flexizyme with the binding site for the U1A protein. Crystallization of the reaction product (L-phenylalanyl-RNA) was precluded by its hydrolytic instability. Instead, L-phenylalanyl-ethyl ester (PheEE), a mimic of L-phenylalanyl-cyanomethyl ester, was soaked into crystals (Methods).

The flexizyme core consists of a coaxial stack of four helices, one irregular and three A-form (paired regions P1, P1a and P2), and a 3′ extension (Fig. 1). The irregular helix comprises two segments (joining regions J1a/2 and J2/1a) that have previously been shown biochemically to participate in amino-acid recognition but were thought to be single-stranded[12] (Supplementary Fig. 1). Residues 52–54 near the 3′ end of the ribozyme form a hairpin-shaped turn (J1a/3), helping project the last three residues (55–57) of flexizyme away from the ribozyme helical stack. These nucleotides form helix P3 by base-pairing with the complementary residues tG73–tC75 of the minihelix ('t' denotes tRNA residues), juxtaposing tA76 (the site of aminoacylation) with the broad minor groove of the irregular helix. Flexizyme contacts the minihelix only on its acceptor end, and our structure is compatible with full-length tRNA binding to the ribozyme in the same orientation (Supplementary Fig. 2). Natural ribozymes, including the ribosome[13] and RNase P (ref. 14), also recognize tRNA by base-pairing with the CCA terminus.

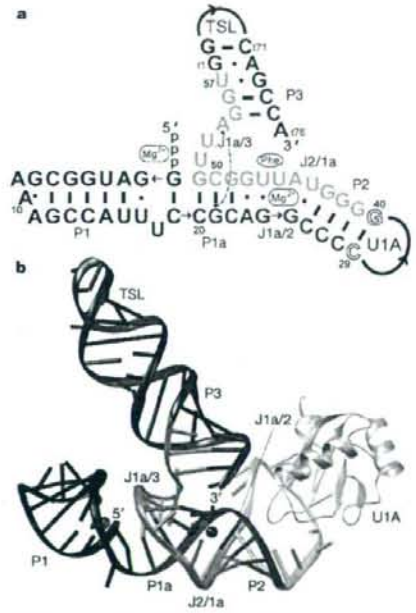


Figure 1 | Overall structure of the flexizyme–tRNA minihelix fusion. **a**, Sequence and secondary structure of the crystallization construct. Binding sites of well-ordered Mg^{2+} ions and of phenylalanine are indicated. The dashed line denotes an A-minor interaction. The minihelix (red) is numbered by using the tRNA convention ('t' precedes residues). T stem-loop (TSL) and U1A-binding loop are described in Methods. **b**, Diagram of the three-dimensional structure. Phenylalanine is depicted in green, and Mg^{2+} ions in magenta.

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109-1024, USA. [2]Research Center for Advanced Science and Technology, The University of Tokyo, 153-8904 Tokyo, Japan. [3]Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 113-8656, Tokyo, Japan.

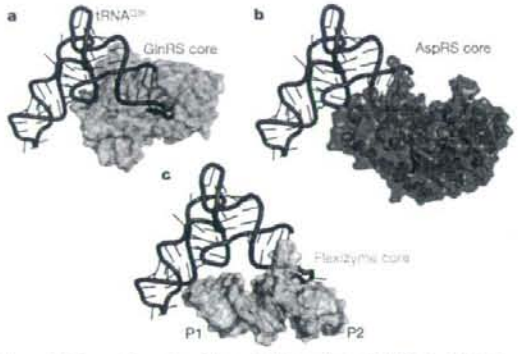
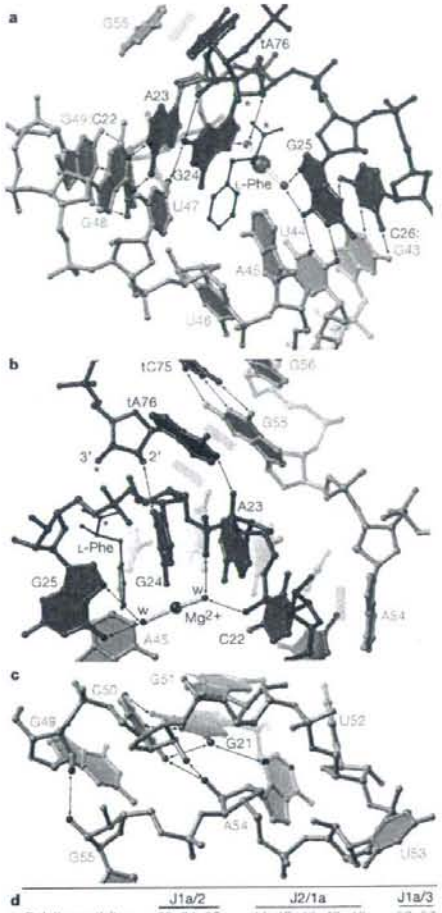


Figure 2 | Comparison of protein and RNA aminoacyl-tRNA synthetases. **a**, Core of *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS; cyan), a class I ARS, bound to $tRNA^{Gln}$ (ref. 16). **b**, Core of *E. coli* aspartyl-tRNA synthetase (AspRS; green), a class II ARS, bound to $tRNA^{17}$. Its cognate tRNA was superimposed on $tRNA^{Gln}$, and the latter is shown. **c**, Flexizyme core (omitting the U1A-binding loop) docked onto $tRNA^{Gln}$ by superimposing the minihelix on the tRNA acceptor stem.

Protein ARSs belong to two structurally distinct families[15]. Class I enzymes approach the acceptor stem from the minor groove and initially aminoacylate the terminal 2′-OH of tRNA. Class II enzymes approach from the major groove and initially aminoacylate the 3′-OH (Fig. 2a, b). After charging, the acyl group equilibrates between the 2′-OH and the 3′-OH. Superposition of flexizyme on representative class I and II ARS–tRNA complexes shows that flexizyme approaches tRNA in a manner analogous to that of class II ARSs (Fig. 2c). It was shown previously[6] that flexizyme aminoacylates tRNA analogues lacking the terminal 2′-OH group 2.4-fold more slowly than it does wild-type tRNA. In contrast, the ribozyme showed a 200-fold decrease in activity when presented with a tRNA analogue lacking a 3′-OH. Thus, flexizyme preferentially aminoacylates the tRNA 3′-OH. Mimicry of class II ARSs by flexizyme seems to result from both a common direction of approach to the tRNA substrate, and the conformation of the enzyme-bound CCA terminus. Unlike class I ARSs, which impose a kink on this single-stranded segment of tRNA[16], class II ARSs bind to a CCA in a near-helical conformation[17]. Flexizyme base-pairs to the CCA, imposing a helical conformation on it. In double-helical RNA, the 2′-OH is buried in the minor groove, making it less accessible to an enzyme than the 3′-OH.

The irregular helix harbouring the active site of flexizyme is formed by three non-canonical base pairs (A23•G48, G24•U47 and G25•U44) and two unpaired nucleotides (A45 and U46; Fig. 3a). The pair formed between the Watson–Crick faces of A23 and G48 widens the minor groove, allowing only the O4 carbonyl oxygen of U47 to interact with G24. Juxtaposition of these two 'stretched' non-canonical pairs and the two unpaired nucleotides results in the amino-acid-binding pocket. A hydrated Mg^{2+} ion in the major groove of the active-site helix (Fig. 3b) bridges the Hoogsteen face of G25 and the backbone phosphates of A23 and G24, stabilizing the active site.

Flexizyme precisely positions the acceptor end of its substrate tRNA by four methods. First, the J1a/3 hairpin (Fig. 3b, c) helps to place G55, which pairs with tC75 to form the bottom base pair of P3, perpendicular to the main helical stack of the ribozyme. A54 of J1a/3 packs against the C50•G21 base pair of P1a through a canonical class I A-minor interaction[18]. U52 stacks on A54 and also packs against P1a, and U53 is extruded from the turn. Mutation of A54 to U abrogates 95% of ribozyme activity[12], suggesting that the A-minor interaction is required for activity. Second, in addition to base-pairing with tC75, G55 makes a partial cross-strand stack with the base of tA76 (Fig. 3a, b). Replacement of G55•tC75 with the corresponding C55•tG75



d	J1a/2			J2/1a					J1a/3	
Relative activity	23	24	25	44	45	46	47	48	52	53
1	A	G	G	U	A	U	U	G	U	U
1.00 ± 0.07	A	G	G	U	A	G	U	G	U	A
0.98 ± 0.05	A	G	G	A	G	A	U	G	U	U
0.98 ± 0.04	G	G	C	G	A	U	U	A	U	U
0.97 ± 0.06	G	G	C	G	A	U	U	A	U	A
0.93 ± 0.03	A	G	G	A	G	G	U	G	U	A
0.55 ± 0.03	A	G	U	C	G	U	U	G	U	U
0.34 ± 0.03	A	G	U	C	G	U	U	G	U	G

Figure 3 | Structure and sequence requirements of the active site. **a**, View into the minor groove. The ethyl group of the bound L-phenylalanine-ethyl ester, lacking electron density, is omitted. Asterisks denote the two reactive atoms; grey rectangles indicate some stacking interactions. Tb^{3+} cleavage assays[30] indicate the presence, close to J2/1a, of an outer-sphere coordinated divalent cation required for activity. This cation may correspond to the Mg^{2+} ion shown. Two inner-sphere coordination water molecules are resolved crystallographically. **b**, View from the major groove. The Mg^{2+} ion lies about 8 Å away from the reactive carbonyl; its electrostatic contribution to transition-state stabilization is probably modest. **c**, The J1a/3 hairpin turn. **d**, Summary of a of reselection experiment in which the ten listed residues were randomized. The activity of reselected sequences (mean ± s.d. for three independent measurements) is normalized to that of the parental sequence. Nucleotides that varied between the reselected and parental sequences are in red, and those that did not in black.

mutations decreases flexizyme activity by 90% (ref. 12). Because G forms the most stable stacking interactions[19], our structure suggests that its partial cross-strand stacking on tA76 is needed for flexizyme function. Third, the base of tA76 makes van der Waals contact with the ribose of G24, and a hydrogen bond between its N1 and the 2′-OH of A23 (Fig. 3b). Mutation of tA76 to G or U results in 85% and 90% loss of activity, respectively[12], suggesting that the specific interactions made by the adenine are important for activity. Fourth, the 2′-OH of tA76 hydrogen bonds with N2 of G24 (Fig. 3a). None of these interactions require a covalent bond between the two RNAs. Because the activities of *cis*- and *trans*-flexizymes are comparable[7], the structure of the *trans*-flexizyme–tRNA complex is probably similar to what we observe.

Although weak, which is consistent with the high K_m of flexizyme for activated phenylalanine (more than 5 mM)[7], residual $|F_o| - |F_c|$ electron density for the benzyl group of the inhibitor PheEE is present in the pocket formed by juxtaposition of the U47•G24 pair with the unstacked U46 in the active site of one of the two crystallographically independent conformers of flexizyme in our crystals (Fig. 3a and Supplementary Fig. 3). The phenyl group stacks on G24, maximizing the favourable interaction[20] of the O6 of the purine with the partial positive charge at the centre of the phenyl ring (Fig. 3b). The importance of this oxygen–aromatic interaction is supported by the marked preference of flexizyme for aromatic substrates (Supplementary Figs 4 and 5). An adjacent depression formed between the splayed bases of G24 and G25 accommodates the α-carbon and the carbonyl group of the amino acid (Fig. 3a). No electron density is observed for the ethyl group of PheEE, but space is present for it to project away from the RNA, which is consistent with the tolerance of flexizyme for bulky leaving groups, including AMP[5]. The amine of the phenylalanine points directly out of the groove, which is consistent with the lack of recognition of this functional group by flexizyme[5].

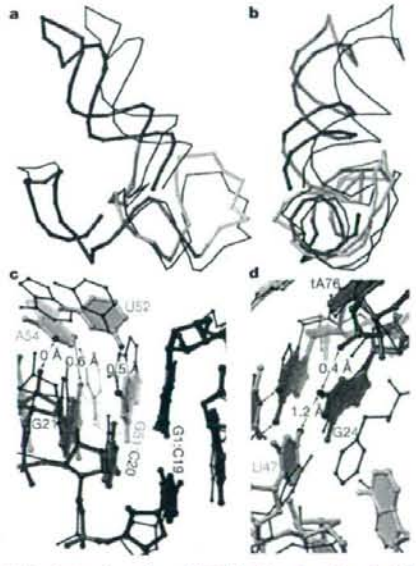


Figure 4 | Conjectural coupling of tRNA docking to active-site folding. **a**, Superposition of the two crystallographically independent flexizyme structures. Throughout, molecule B (undocked conformation) is in black. **b**, Orthogonal view. **c**, Comparison of the interface between J1a/3 and P1a in the two molecules. P1 adopts similar conformations in both molecules; J1a/3 residues (U52 and A54) are farther from P1a in molecule B. Red numbers denote the increase in length of the indicated hydrogen bonds from molecule A to molecule B. Differences are mostly comparable to the precision of the structures (Methods); however, distances are systematically larger for molecule B. **d**, Buckling of U47 into the phenylalanine-binding site in molecule B obstructs the binding of the amino acid.

To verify the functional importance of flexizyme core nucleotides, we reselected seven new active variants (Fig. 3d) from a pool of flexizyme–microhelix RNAs containing random sequences in the active site and the hairpin turn[21]. The reselection indicates that the identities of only two nucleotides in the irregular active site helix, G24 and U47, are essential. The unstacked U46 can be replaced with a purine, and position 45 can be either purine. The expansion of the minor groove produced by the A23•G48 pair is important, because it or the reverse G23•A48 pair is conserved. U52 from the apex of the hairpin turn is also absolutely conserved. This underscores the importance of J1a/3 for activity.

In the two flexizyme conformers in our structure, J1a/3 functions as a hinge between the ribozyme helical stack and the minihelix (Fig. 4a, b). In molecule B, the J1a/3 hairpin is not fully docked with P1a (Fig. 4c), the 2′-OH of tA76 does not engage the base of G24, and the functionally critical G24•U47 pair of the active site is not formed (Fig. 4d and Supplementary Fig. 6). Also in molecule B, U47 buckles towards U46, obstructing the binding of PheEE. On the basis of our crystallographic and reselection studies, we speculate that adoption by the active site of flexizyme of a conformation capable of binding and positioning activated phenylalanine is intimately coupled to the docking of J1a/3 and P1a. This coupling of tRNA binding and active-site folding would be reminiscent of the indirect readout of acceptor stem conformation by some protein ARSs[22].

To function in translation, tRNA needs to be aminoacylated exclusively at its 3′-terminal residue. This regioselective aminoacylation may have evolved in the RNA world to recruit the replicase ribozyme to the tRNA-like ends of the genome, thereby promoting full-length replication[23,24]. Alternatively, it may have evolved to expand the chemical diversity of primitive ribozyme active sites, in an analogous manner to the prosthetic groups of proteins[25]. Either model requires ARS ribozymes capable of regioselective aminoacylation. Together with previous analyses, our structures and reselection show that remarkably few active-site and J1a/3 residues are needed for flexizyme activity. The simplicity of flexizyme supports the hypothesis that ribozymes were responsible for all biochemical catalysis before the evolution of translation. It is possible that this small ribozyme achieves regioselectivity through an induced-fit mechanism. Protein enzymes, including ARSs[8,9], frequently attain enhanced specificity by coupling active-site folding to substrate binding, but previous structural studies of catalytic RNAs have mostly revealed rigid, pre-organized small-molecule binding sites[26–28]. Our demonstration that an ARS ribozyme evolved *in vitro* achieves specificity by interacting only with the acceptor stem of tRNA is consistent with the proposal[29] that class II ARS that are active with tRNA minihelices are closely related evolutionarily to the primordial ARSs.

METHODS SUMMARY

The flexizyme–minihelix fusion RNA was cleaved with the VS ribozyme to generate a homogeneous 3′ end. The 2′,3′-cyclic phosphate was opened and removed with T4 polynucleotide kinase (PNK) (Supplementary Fig. 7). The RNA was co-crystallized with selenomethionyl U1A protein and the structure was determined by a combination of multiwavelength anomalous diffraction (MAD) and molecular replacement (Supplementary Fig. 8). A second structure was solved with amplitudes from a crystal soaked in a cryoprotectant including 2.5 mM PheEE (the solubility limit of the compound). The models have been refined to R_{free}/R_{work} of 27.9%/22.3% and 30.4%/25.4% against all data extending to 2.8 Å and 3.0 Å, respectively (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 January; accepted 28 April 2008.
Published online 11 June 2008.

1. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* 267, 643–647 (1995).

2. Illangasekare, M. & Yarus, M. A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA* **5**, 1482–1489 (1999).
3. Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl Acad. Sci. USA* **96**, 5470–5475 (1999).
4. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Struct. Biol.* **7**, 28–33 (2000).
5. Saito, H., Kourouklis, D. & Suga, H. An *in vitro* evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20**, 1797–1806 (2001).
6. Saito, H. & Suga, H. A ribozyme exclusively aminoacylates the 3′-hydroxyl group of the tRNA terminal adenosine. *J. Am. Chem. Soc.* **123**, 7178–7179 (2001).
7. Murakami, H., Saito, H. & Suga, H. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **10**, 655–662 (2003).
8. Cusack, S., Yaremchuk, A. & Tukalo, M. The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with $tRNA^{Ser}$ and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J.* **15**, 2834–2842 (1996).
9. Moulinier, L. *et al.* The structure of an AspRS-$tRNA^{Asp}$ complex reveals a tRNA-dependent control mechanism. *EMBO J.* **20**, 5290–5301 (2001).
10. Schimmel, P., Giegé, R., Moras, D. & Yokoyama, S. An operational code for amino acids and possible relationship to genetic code. *Proc. Natl Acad. Sci. USA* **90**, 8763–8768 (1993).
11. Ramaswamy, K., Saito, H., Murakami, H., Shiba, K. & Suga, H. Designer ribozymes: programming the tRNA specificity into flexizyme. *J. Am. Chem. Soc.* **126**, 11454–11455 (2004).
12. Saito, H., Watanabe, K. & Suga, H. Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA* **7**, 1867–1878 (2001).
13. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
14. Kirsebom, L. A. & Svard, S. G. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13**, 4870–4876 (1994).
15. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206 (1990).
16. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ at 2.8 Å resolution: implications for tRNA discrimination. *Science* **246**, 1135–1142 (1989).
17. Ruff, M. *et al.* Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science* **252**, 1682–1689 (1991).
18. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA* **98**, 4899–4903 (2001).
19. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
20. Burley, S. K. & Petsko, G. A. Weakly polar interactions in proteins. *Adv. Protein Chem.* **39**, 125–192 (1988).
21. Murakami, H., Ohta, A., Ashial, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **3**, 357–359 (2006).
22. Perona, J. J. & Hou, Y. M. Indirect readout of tRNA for aminoacylation. *Biochemistry* **46**, 10419–10432 (2007).
23. Weiner, A. M. & Maizels, N. tRNA-like structures tag the 3′ ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl Acad. Sci. USA* **84**, 7383–7387 (1987).
24. Orgel, L. E. The origin of polynucleotide-directed protein synthesis. *J. Mol. Evol.* **29**, 465–474 (1989).
25. Wong, J. T. Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. *Orig. Life Evol. Biosph.* **21**, 165–176 (1991).
26. Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282**, 259–264 (1998).
27. Serganov, A. *et al.* Structural basis for Diels–Alder ribozyme catalyzed carbon–carbon bond formation. *Nature Struct. Mol. Biol.* **12**, 218–224 (2005).
28. Klein, D. J. & Ferré-D'Amaré, A. R. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752–1756 (2006).
29. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017–5035 (1998).
30. Saito, H. & Suga, H. Outersphere and innersphere coordinated metal ions in an aminoacyl-tRNA synthetase ribozyme. *Nucleic Acids Res.* **30**, 5151–5159 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff at ALS beamline 5.0.2 and J. Bolduc for assistance with synchrotron and in-house X-ray data collection, respectively, and T. Edwards, T. Hamma, D. Klein, J. Pitt, J. Posakony, A. Roll-Mecak and B. Shen for discussions. A.R.F. is a Distinguished Young Scholar in Medical Research of the W. M. Keck Foundation. This work was supported by grants from research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (to H.S.) and the W. M. Keck Foundation (to A.R.F.).

Author Information Atomic coordinates and structure factor amplitudes for aminoacyl-tRNA synthetase ribozyme–minihelix fusions refined against crystal II and crystal III data have been deposited with the Protein Data Bank with accession codes 3CUL and 3CUN, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.R.F. (aferre@fhcrc.org).

RNA
A PUBLICATION OF THE RNA SOCIETY

Initiating translation with D-amino acids

Yuki Goto, Hiroshi Murakami and Hiroaki Suga

RNA 2008 14: 1390-1398
Access the most recent version at doi:10.1261/rna.1020708

Initiating translation with *D*-amino acids

YUKI GOTO,[1,2] HIROSHI MURAKAMI,[1] and HIROAKI SUGA[1,2,3]

[1]Research Center of Advanced Science and Technology, The University of Tokyo, Tokyo, 153-8904, Japan
[2]Department of Advanced Interdisciplinary Studies, Graduate School of Engineering, The University of Tokyo, Tokyo, 153-8904, Japan
[3]Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, Tokyo, 113-8656, Japan

ABSTRACT

Here we report experimental evidence that the translation initiation apparatus accepts *D*-amino acids (^{D}aa), as opposed to only *L*-methionine, as initiators. Nineteen ^{D}aa, as the stereoisomers to their natural *L*-amino acids, were charged onto initiator $tRNA^{fMet}_{CAU}$ using flexizyme technology and tested for initiation in a reconstituted *Escherichia coli* translation system lacking methionine, i.e., the initiator was reprogrammed from methionine to ^{D}aa. Remarkably, all ^{D}aa could initiate translation while the efficiency of initiation depends upon the type of side chain. The peptide product initiated with ^{D}aa was generally in a nonformylated form, indicating that methionyl-tRNA formyltransferase poorly formylated the corresponding ^{D}aa-$tRNA^{fMet}_{CAU}$. Although the inefficient formylation of ^{D}aa-$tRNA^{fMet}_{CAU}$ resulted in modest expression of the corresponding peptide, preacetylation of ^{D}aa-$tRNA^{fMet}_{CAU}$ dramatically increased expression level, implying that the formylation efficiency is one of the critical determinants of initiation efficiency with ^{D}aa. Our findings provide not only the experimental evidence that translation initiation tolerates ^{D}aa, but also a new means for the mRNA-directed synthesis of peptides capped with ^{D}aa or acyl-^{D}aa at the N terminus.

Keywords: *D*-amino acid; translation; initiation; genetic code reprogramming; flexizyme

INTRODUCTION

The translation machinery polymerizes α-amino acids according to the sequence information encoded in the open reading frame of the mRNA, designating the length and sequence of the synthesized polypeptide composed of 20 proteinogenic α-amino acids with *L*-stereo configuration (^{L}aa). The main player that governs the strict use of ^{L}aa is aminoacyl-tRNA synthetases (aaRSs) that are able to discriminate cognate ^{L}aa against not only noncognate proteinogenic ^{L}aa but also nonproteinogenic ones including *D*-amino acids (^{D}aa); thus aaRSs play a central role in refusing noncognate amino acids from the incorporating elements (Söll 1990; Sankaranarayanan and Moras 2001). However, even if the tRNA aminoacylation step is circumvented, ^{D}aa cannot be efficiently incorporated into the nascent peptide chain during elongation. For instance, a variety of ^{D}aa precharged onto an "amber" suppressor $tRNA_{CUA}$ have been examined for elongation; they are either modestly or in many cases not at all incorporated into the nascent peptide chain (Roesser et al. 1989; Bain et al. 1991; Ellman et al. 1992; Starck et al. 2003; Tan et al. 2004; Murakami et al. 2006). This has been attributed to the fact that either elongation factor (EF-Tu) or ribosome (or possibly both) does not allow ^{D}aa-$tRNA_{CUA}$ to read the amber stop codon, resulting in the undesired termination of peptide synthesis executed by release factor. A more recent attempt to use ribosome mutants has given a modest increase in efficiency for the incorporation of ^{D}Met and ^{D}Phe (Dedkova et al. 2003, 2006), but it is yet unclear how generally this approach is applicable to a variety of ^{D}aa.

Like elongation, the initiation event is also strictly governed by MetRS and initiation factors (IFs) (Kozak 1983; Gold 1988; Gualerzi and Pon 1990). In the prokaryotic translation system, peptide synthesis is exclusively initiated with N^{α}-formyl methionine (f-^{L}Met) (Kozak 1983). To circumvent this limitation, we have recently shown that upon using precharged aminoacyl-$tRNA^{fMet}_{CAU}$ in a reconstituted *Escherichia coli* cell-free translation system (more details are discussed below), initiator Met can be reassigned to other proteinogenic ^{L}aa and peptide synthesis successfully

Abbreviations: DMSO, dimethyl sulfoxide; HEPES, 2-[4-(2-hydroxyethyl)-1-piperadinyl]ethansulfonic acid; EDTA, ethylenediamine tetraacetic acid; Tris, Tris(hydroxymethyl)aminomethane; TFA, trifluoroacetic acid; MeCN, acetonitrile.

Reprint requests to: Hiroaki Suga, Research Center of Advanced Science and Technology, The University of Tokyo, 4-6-1, Komaba, Meguro, Tokyo, 153-8904, Japan; e-mail: hsuga@rcast.u-tokyo.ac.jp; fax: 81-3-5452-5495.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1020708.

initiated (Goto et al. 2008). This suggests that the initiation governance can be overridden by such a genetic code reprogramming strategy. On the other hand, we have no knowledge of whether or not ^{D}aa can adapt to the initiation event.

Here we report for the first time, to the best of our knowledge, that translation can be initiated with ^{D}aa. We have demonstrated that the translation apparatus tolerates any of 19 ^{D}aa at initiation with efficiencies depending upon the type of their side chains. The most intriguing discovery is that peptides initiated with ^{D}aa are not formylated in most cases. This result is in sharp contrast to the fact that peptides initiated with ^{L}aa are generally formylated. Moreover, the use of pre-*N*-acetylated ^{D}aa (Ac-^{D}aa) significantly enhances the expression of peptides initiated with ^{D}aa, suggesting that formylation efficiency is one of the critical determinants of expression yield initiated with ^{D}aa. Our study represents the first evidence that translation initiation tolerates a variety of amino acids independent of stereochemistry and also offers a new means to generate various peptides containing ^{D}aa at the N terminus by translation.

RESULTS

^{D}Met acts as an initiator upon the reprogrammed initiation

To facilitate the reprogramming of the translation initiation, we integrated two systems, PURE and flexizyme. The former system is a reconstituted *E. coli* cell-free translation system (PURE stands for protein synthesis using recombinant elements) (Shimizu et al. 2001). Using this system, we are able to withdraw methionine (Met) from the translation components (referred to as *w*PURE system), making the initiation codon (AUG) vacant (Fig. 1). The flexizyme system consists of artificially evolved ribozymes, enabling us to charge a wide variety of amino acids onto any desired tRNAs, including the initiator $tRNA^{fMet}_{CAU}$ (Murakami et al. 2003a,b, 2006). Indeed, we have previously reported that various ^{L}aa-$tRNA^{fMet}_{CAU}$ molecules prepared by the flexizyme system can function as an initiator in the Met-withdrawn *w*PURE system (Goto et al. 2008). Importantly, the flexizyme system tolerates not only ^{L}aa with nonproteinogenic side chains but also ^{D}aa with a variety of side chains (Murakami et al. 2006). Thus, the integration of these two systems facilitates the reassignment of the vacant initiation codon to ^{D}aa (Fig. 1).

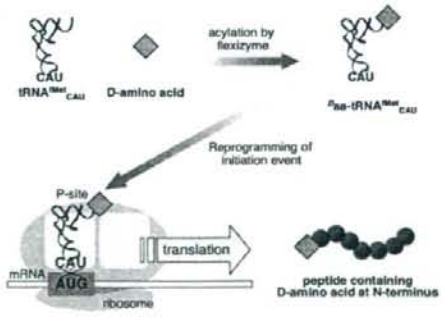


FIGURE 1. Genetic code reprogramming of the initiation event with ^{D}aa. The initiator can be reassigned to ^{D}aa instead of ^{L}Met by the integration of *w*PURE system and flexizyme system. Initiation with ^{D}aa results in the nascent peptide containing ^{D}aa at the N terminus.

To examine whether the ^{D}aa-$tRNA^{fMet}_{CAU}$ molecules prepared by the flexizyme system can initiate the translation reaction in the *w*PURE system, we designed a mRNA sequence for the expression of a 14-mer peptide (Fig. 2A). The flag peptide sequence (DYKDDDDK; D = Asp, Y = Tyr, K = Lys) (Brizzard et al. 1994) included in this peptide acted as a [^{14}C]-Asp-labeling tag for detecting the expression level in tricine-SDS PAGE upon addition of [^{14}C]-Asp and also as a purification tag for isolating the full-length peptide for MALDI-TOF analysis.

Prior to examining ^{D}aa-initiation, we performed control experiments to ensure that the reprogrammed initiation would work as planned. Tricine-SDS PAGE analysis (Schagger and von Jagow 1987) of the peptide expressed in the ordinary PURE system yielded a single evident band, whereas the same assay using the *w*PURE system yielded only negligible background bands originated from "in-frame" misinitiations (see below). Most importantly, no band corresponding to the full-length peptide appeared in this analysis, indicating that ^{L}Met was in fact depleted in the *w*PURE system (Fig. 2B, lanes 1,2). On the other hand, when ^{L}Met-$tRNA^{fMet}_{CAU}$ was added to the *w*PURE system, an evident band appeared with the same mobility and intensity as that observed in lane 1 (Fig. 2B, lane 3). MALDI-TOF analysis of the flag-purified cold (nonradiolabeled) peptide expressed under the same conditions as lane 3 showed the N-terminal formylated peptide with the expected molecular weight (Fig. 3, ^{L}Met), implying that N^{α}-formylation of ^{L}Met-$tRNA^{fMet}_{CAU}$ occurred by methionyl-tRNA formyltransferase (MTF) present in the translation system, and the resulting f-^{L}Met-$tRNA^{fMet}_{CAU}$ exclusively initiated the translation. These control results were consistent with our previously reported results (Goto et al. 2008).

Next, the same DNA template was translated with the *w*PURE system in the presence of ^{D}Met-$tRNA^{fMet}_{CAU}$ (Fig. 2B, lane 4). Tricine-SDS PAGE analysis of the product gave a single evident band, but the band moved slightly faster than that produced by ^{L}Met-initiation (Fig. 2B, cf. lanes 4 and 3). Its molecular weight analysis by MALDI-TOF mass spectrometry indicated two peaks consistent with those of the nonformylated peptide (H-peptide) and formylated peptide (f-peptide) in an ~1:1 ratio (Fig. 3, ^{D}Met; * and † indicate H- and f-peptides, respectively). Because ^{L}Met-initiation yielded the expected peptide as a fully formylated

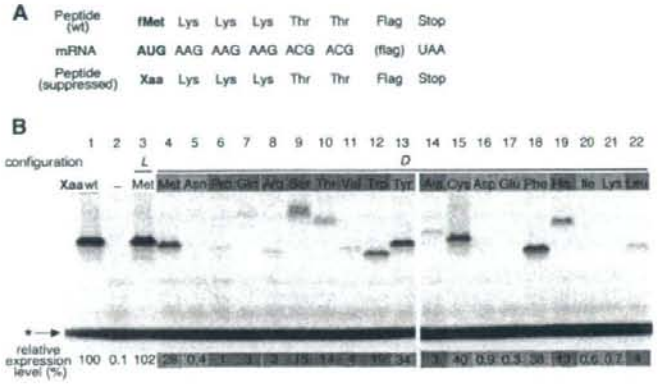


FIGURE 2. Tolerance of various ^{D}aa in initiation. (*A*) The mRNA sequence that expresses peptides initiated by various ^{D}aa. Flag in parentheses indicates the RNA sequence encoding the Flag peptide sequence (DYKDDDDK). (*B*) Tricine-SDS PAGE analysis of the translation products. (Lane *1*) expression of wild type; (lane *2*) in the absence of Met; (lane *3*) reprogrammed initiation with ^{L}Met; (lanes *4–22*) reprogrammed initiation with various ^{D}aa. Each expression level relative to the wild type is determined by a mean score of duplicates or more. The ^{D}aa giving >25%, 10%–25%, 1%–10%, and <1% of the wild-type expression level are highlighted in orange, pink, cyan, and gray, respectively. The band indicated by an asterisk corresponds to free [^{14}C]-Asp that remained unincorporated into the Flag peptide.

form, this result left us with two questions: (1) Was the partial formylation observed in the peptide caused by an incomplete formylation of ^{D}Met-tRNA$^{fMet}_{CAU}$ catalyzed by MTF, resulting in the mixture of H-^{D}Met-peptide and f-^{D}Met-peptide? As an alternative scenario, (2) was ^{D}Met-tRNA$^{fMet}_{CAU}$ racemized during the translation, yielding the mixture of H-^{D}Met-peptide and f-^{L}Met-peptide (and possibly f-^{D}Met-peptide)?

Translation can be initiated with various kinds of ^{D}aa

Before addressing the above questions by additional sets of experiments, we decided to survey the expressions of the same DNA template initiated with the rest of 18 types of ^{D}aa-tRNA$^{fMet}_{CAU}$ prepared by the flexizyme system (Fig. 2B, lanes 5–22). We found an apparent band of the product initiated with three ^{D}aa (Fig. 2B, ^{D}Tyr, ^{D}Cys, and ^{D}Phe [highlighted in orange]) with >25% initiation efficiency compared with that observed in wild-type expression (f-^{L}Met-peptide) and a faint (10%–25%) yet evident band of the product initiated with four ^{D}aa (Fig. 2B, ^{D}Ser, ^{D}Thr, ^{D}Trp, and ^{D}His [highlighted in pink]). Moreover, the expression initiated with six ^{D}aa (Fig. 2B, ^{D}Pro, ^{D}Gln, ^{D}Arg, ^{D}Val, ^{D}Ala, and ^{D}Leu [highlighted in cyan]) yielded a very faint (1–10%) yet clearly observable band. We isolated the full-length peptides from all translation samples by flag purification, and their molecular weight was analyzed by MALDI-TOF mass spectrometry. To our surprise, all samples gave the molecular weight corresponding to the respective peptide initiated with the expected amino acid (Fig. 3), including those that did not yield a detectable band on tricine-SDS PAGE analysis (Fig. 2B, ^{D}Asn, ^{D}Asp, ^{D}Glu, ^{D}Ile, and ^{D}Lys [highlighted in gray]). Even more surprisingly, the N terminus of the majority of peptides was the nonformylated form, i.e., H-peptides. Two exceptions were, however, observed; initiation with ^{D}Ser yielded a mixture of H-Ser- and f-Ser-peptides (Ser denotes a single *D*-stereoisomer or a mixture of *D*- and *L*-stereoisomers), while ^{D}Cys yielded three peaks, two of which were consistent with H-Cys- and f-Cys-peptides.

The MALDI-TOF analysis of ^{D}aa-initiated peptides that did not yield a clear band (Fig. 2B, highlighted in gray) generally gave a poor signal/noise ratio with one or occasionally two peaks originating from the background expression of mRNA by "in-frame" misinitiations (see Fig. 3, ^{D}Ile; Supplemental Fig. S1). However, the most important outcome was that 16 out of 19 ^{D}aa-tRNA$^{fMet}_{CAU}$ initiated peptide synthesis without formyl modification on their N terminus, giving the corresponding H-peptides only. This was in sharp contrast to our previously reported result that ^{L}aa-initiated peptides were generally formylated at their N terminus (Goto et al. 2008). This unmistakable difference in occurrence of the N-terminal formylation of peptide initiated with ^{L}aa or ^{D}aa suggested that ^{D}aa-tRNA$^{fMet}_{CAU}$ in most cases was a poor substrate for MTF due to the *D*-stereochemistry of its α-carbon, so that it initiated the translation without N^{α}-formylation. This in turn suggested that ^{D}aa-tRNA$^{fMet}_{CAU}$ that gave the H-peptide in MALDI-TOF analysis (Fig. 3, peaks labeled with *) was unlikely to have racemized during the translation; consequently, the observed respective H-peptide is H-^{D}aa-peptide.

No racemization of ^{D}aa-tRNA$^{fMet}_{CAU}$ occurs during translation

In order to solidify the above idea, we set competition experiments of initiation using ^{D}aa-tRNA$^{fMet}_{CAU}$ against ^{L}aa-tRNA$^{fMet}_{CAU}$ to see how much contamination of ^{L}aa-tRNA$^{fMet}_{CAU}$ over ^{D}aa-tRNA$^{fMet}_{CAU}$ would result in visualizing f-^{L}aa-peptide. This experiment aimed at mimicking the situation in which the initiator ^{D}aa partially racemized to ^{L}aa during the translation reaction. We predicted that the competition efficiency would depend upon the type of ^{D}aa. Therefore, we chose three ^{D}aa (^{D}Tyr, ^{D}Trp, and ^{D}Leu) giving a high, moderate, and low efficiency, respectively, observed in the

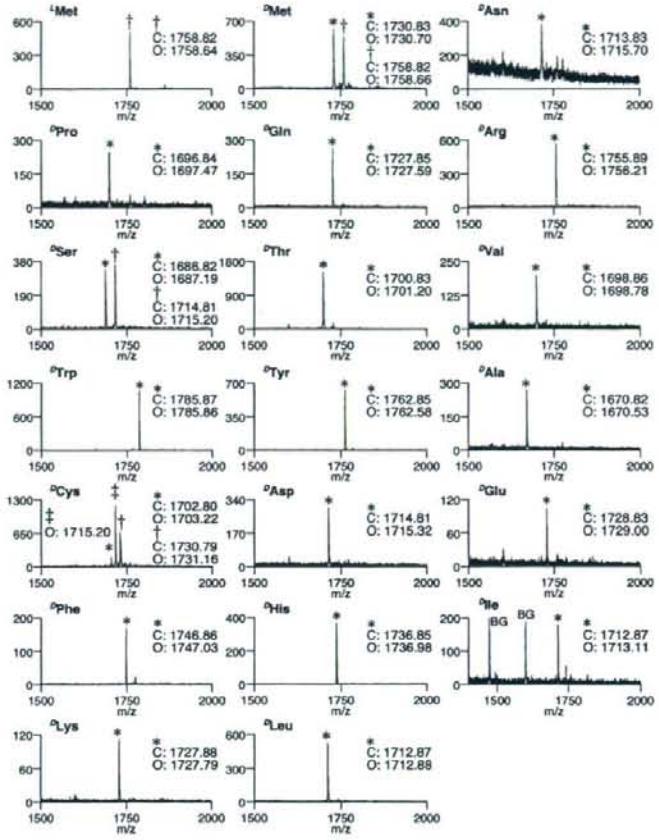


FIGURE 3. MALDI-TOF mass spectra of the translated peptides. * and † indicate a peak corresponding to the H- and f-peptide, respectively. "BG" indicates peaks corresponding to the background expression originated in-frame misinitiation. ‡ indicates the peak corresponding to f-^{D}Ser-peptide proposed in this work. The calculated mass (C) and observed mass (O) are shown in each spectrum.

tricine-SDS PAGE analysis (Fig. 2B). Three different ratios, 99:1, 95:5, and 80:20, of the mixtures of the ^{D}aa-tRNA$^{fMet}_{CAU}$ and ^{L}aa-tRNA$^{fMet}_{CAU}$ were prepared for the expression of peptides and each translation product was analyzed by MALDI-TOF mass spectrometry (Fig. 4). A 1% *L*-contamination to the sample of ^{D}Tyr and ^{D}Trp gave a single major peak corresponding to H-peptide, but a very tiny peak of f-peptide was accompanied in the case of ^{D}Trp-initiation. On the other hand, even with a 1% contamination of ^{L}Leu, ^{D}Leu suffered from a nearly 1:1 mixture of intense peaks of H- and f-peptides. A 5% *L*-contamination gave a significant increase in the relative intensity of the peaks corresponding to f-peptides in the cases of both ^{D}Trp and ^{D}Leu, while ^{D}Tyr-initiation yielded only a minor peak of f-peptide. In the case of 20% L-contamination, f-^{L}Trp- and f-^{L}Leu-initiations almost competed out the corresponding ^{D}aa-initiation, and even for ^{D}Tyr the peak ratio of H- and f-peptides became nearly 1:1. These results clearly indicated that even if a small amount, as little as 1%, of ^{L}aa-tRNA$^{fMet}_{CAU}$ is contaminated in ^{D}aa-tRNA$^{fMet}_{CAU}$, the peptide product would suffer from the formation of f-peptide. Since we observed a single peak of H-peptide in the 16 cases in Figure 3, these products should be assigned to H-^{D}aa-peptides. We thus concluded that the racemization of ^{D}aa-tRNA$^{fMet}_{CAU}$ did not occur during the translation in these cases; even if it occurred, its degree should be far less than 1%, which would be negligible.

^{D}Met-, ^{D}Ser-, and ^{D}Cys-tRNA$^{fMet}_{CAU}$ act as modest substrates of MTF

The above studies so far supported that racemization of ^{D}aa-tRNA$^{fMet}_{CAU}$ does not generally occur during the translation. However, in the case of initiation with ^{D}Met-, ^{D}Cys-, or ^{D}Ser-tRNA$^{fMet}_{CAU}$ we did observe two peaks corresponding to H-peptide and f-peptide (for ^{D}Cys there was an additional peak). Therefore, it was still necessary to verify the possibility of their racemization by another method. Accordingly, we decided to use the well-known fact that ^{D}aa are generally poor substrates in the elongation event using the amber suppression method (Noren et al. 1989; Bain et al. 1991; Ellman et al. 1992; Starck et al. 2003; Tan et al. 2004; Murakami et al. 2006). Since the flexizyme system is able to afford the ^{D}aa-tRNA molecules for initiation and elongation (charged onto tRNA$^{fMet}_{CAU}$ and tRNA$^{AsnE-2}_{CUA}$ [Ohta et al. 2007], respectively) with exactly the same quality, we should be able to verify the occurrence of racemization by running the peptide synthesis with the ^{D}aa- or ^{L}aa-tRNA$^{fMet}_{CAU}$ initiation and ^{D}aa- or ^{L}aa-tRNA$^{AsnE-2}_{CUA}$ elongation side by side. Thus, in parallel to the translation of the mRNA previously designed for initiating with the ^{D}aa-tRNA$^{fMet}_{CAU}$, we designed an mRNA template containing the amber codon and expressed the peptide in the presence of ^{D}aa-tRNA$^{AsnE-2}_{CUA}$ (Fig. 5A).

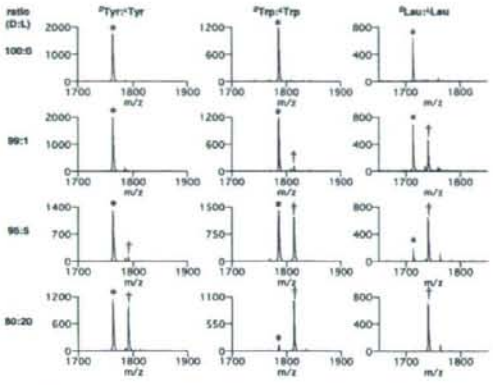


FIGURE 4. MALDI-TOF mass spectra of the peptides initiated with the mixture of ^{D}aa-tRNA$^{fMet}_{CAU}$ and ^{L}aa-tRNA$^{fMet}_{CAU}$. * and † indicate a peak corresponding to the H- and f-peptide, respectively.

Both ^{L}Met-tRNA$^{fMet}_{CAU}$ and ^{L}Met-tRNA$^{AsnE\text{-}2}_{CUA}$ functioned as the translation initiator and elongator, respectively, for the cognate mRNA templates, giving the expected full-length peptides (Fig. 5B, lanes 1,3). On the other hand, ^{D}Met-tRNA$^{fMet}_{CAU}$ could initiate the translation whereas ^{D}Met-tRNA$^{AsnE\text{-}2}_{CUA}$ could not suppress the amber codon (Fig. 5B, lanes 2,4). This clearly indicated that ^{D}Met-tRNA$^{AsnE\text{-}2}_{CUA}$ was not racemized during the translation, and therefore it was reasonable to assume that ^{D}Met-tRNA$^{fMet}_{CAU}$ was not either. Likewise, ^{D}Cys was not incorporated into the nascent peptide chain by amber suppression (Fig. 5B, lanes 5–8), suggesting that ^{D}Cys-tRNA$^{AsnE\text{-}2}_{CUA}$ was not racemized. This result also indicated that ^{D}Cys-tRNA$^{fMet}_{CAU}$ was not racemized and primed the translation with ^{D}Cys. To our surprise, the amber suppression by ^{D}Ser-tRNA$^{AsnE\text{-}2}_{CUA}$ gave an evident band as intense as that by ^{L}Ser-tRNA$^{AsnE\text{-}2}_{CUA}$; however, their mobility were slightly different (Fig. 5B, cf. lanes 11 and 12). MALDI-TOF analysis of the respective peptide products expressed in the presence of ^{D}Ser-tRNA$^{AsnE\text{-}2}_{CUA}$ and ^{L}Ser-tRNA$^{AsnE\text{-}2}_{CUA}$ showed the same expected molecular weight, implying that these peptides have the same compositions of sequence (Supplemental Fig. S2). These results suggested that the observed difference in mobility between lanes 11 and 12 (Fig. 5B) could be attributed to the difference in chirality of the Ser residue, i.e., ^{D}Ser-tRNA$^{AsnE\text{-}2}_{CUA}$ elongated the peptide chain without racemization. This result assured that ^{D}Ser-tRNA$^{fMet}_{CAU}$ initiated the translation without racemization.

To this end, we concluded that neither ^{D}Met-, ^{D}Cys-, nor ^{D}Ser-tRNA$^{fMet}_{CAU}$ were racemized, but they were partially formylated by MTF and competitively initiated the translation. Even though the f-^{D}aa-tRNA$^{fMet}_{CAU}$ was formed presumably in only a small amount, its higher initiation efficiency compared to ^{D}aa-tRNA$^{fMet}_{CAU}$ could affect the outcome of translation. Consequently, we obtained a mixture of the respective H-^{D}aa-peptide and f-^{D}aa-peptide.

MALDI-TOF analysis of the ^{D}Cys-primed peptides showed two peaks, of which the observed molecular weights were consistent with H-^{D}Cys- and f-^{D}Cys-peptides like ^{D}Met and ^{D}Ser. However, there is an additional peak with the molecular weight of 1715.20 Da (Fig. 3, ^{D}Cys, see peak indicated by ‡). Although it is difficult to define what exactly this peptide is, the molecular weight is consistent with f-^{D}Ser-peptide (the calculated molecular weight is 1714.81 Da). We therefore suggest the following mechanism as a possible scenario to convert ^{D}Cys-primed peptide to f-^{D}Ser-peptide (Supplemental Fig. S3). The α-amino group of ^{D}Cys-tRNA$^{fMet}_{CAU}$ could be formylated by MTF similar to ^{D}Met and ^{D}Ser, while the sulfhydryl side-chain group could also be formylated because of its inherent low pKa and high nucleophilicity. The formylated sulfhydryl group then might become a good leaving group so that the oxygen of N^{α}-formyl group likely could attack the β-carbon of the side chain to yield an oxazoline-containing peptide. Hydrolysis of the oxazoline ring consequently might yield f-^{D}Ser-peptide. Note that this unusual event occurred in only the case of initiation with ^{D}Cys, not ^{L}Cys. Therefore, the *D*-configuration of ^{D}Cys likely played a critical role in processing this unusual and interesting chemistry.

Preacylation of the α-amino group on ^{D}aa enhances initiation efficiency

In our previous work on reprogramming the initiation using various ^{L}aa, we have found that preacylation on ^{L}aa-tRNA$^{fMet}_{CAU}$, e.g., N^{α}-acetylation, significantly enhances the initiation efficiency (Goto et al. 2008). Therefore, we wondered whether the same trend could be observed for ^{D}aa-tRNA$^{fMet}_{CAU}$ initiation. Three ^{D}aa with moderate initiation efficiencies (^{D}Met, ^{D}Trp, and ^{D}Phe; 28%, 19%, and 38%, respectively) and two ^{D}aa with low efficiencies (^{D}Asn and ^{D}Ala; 0.4% and 3%, respectively) were chosen for the synthesis of the corresponding Ac-^{D}aa substrates and charged onto tRNA$^{fMet}_{CAU}$ using the flexizyme system. The resulting Ac-^{D}aa-tRNA$^{fMet}_{CAU}$ was used to initiate the translation to produce each peptide in parallel to the initiation with the corresponding ^{D}aa-tRNA$^{fMet}_{CAU}$ (Fig. 6).

In all cases, the initiation with Ac-^{D}aa-tRNA$^{fMet}_{CAU}$ dramatically increased the expression level compared to that with ^{D}aa-tRNA$^{fMet}_{CAU}$; particularly, the preacetylation of ^{D}Asn and ^{D}Ala increased the expression level from nearly invisible band intensity to a clearly visible intensity (Fig. 6, lanes 8–11, 0.4%–31% and 3%–56%, respectively). MALDI-TOF analyses of the respective peptides were also consistent with the expected molecular weights of the Ac-^{D}aa-peptides (Supplemental Fig. S4), indicating that Ac-^{D}aa-tRNA$^{fMet}_{CAU}$ exclusively initiated the translation.

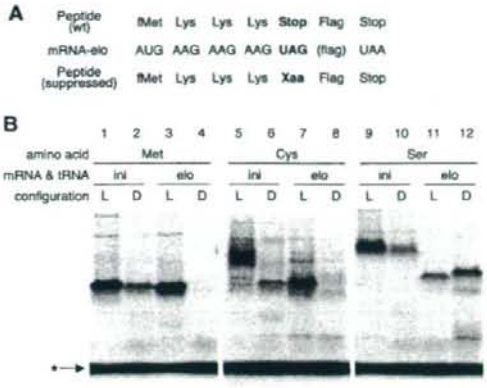


FIGURE 5. Determination of the configuration of the N-terminal residue. (*A*) The mRNA sequence used for the expression with amber suppression. UAG codon indicated in bold was suppressed with various amino acids. Flag in parentheses indicates the RNA sequence encoding the Flag peptide sequence (DYKDDDDK). (*B*) Tricine-SDS PAGE analysis of the translation products. (Lanes *1,5,9*) reprogrammed initiation with ^{L}aa; (lanes *2,6,10*) reprogrammed initiation with ^{D}aa; (lanes *3,7,11*) amber suppression with ^{L}aa; (lanes *4,8,12*) amber suppression with ^{D}aa. The types of side chain are Met in lanes 1–4, Cys in lanes 5–8, and Ser in lanes 9–12. The band indicated by an asterisk corresponds to free [^{14}C]-Asp that remained unincorporated into the Flag peptide.

DISCUSSION

By means of the genetic code reprogramming for initiation, we have shown that the translation apparatus is able to use ^{D}aa in the initiation event even though its efficiency varies depending upon the type of side chain. The most intriguing finding is that in most cases the peptide product had a free N terminus, i.e., without a formyl group, indicating that MTF does not generally formylate the ^{D}aa-tRNA$^{fMet}_{CAU}$ (Fig. 7). Moreover, the use of Ac-^{D}aa-tRNA$^{fMet}_{CAU}$, an analog of formylated initiator, is able to significantly increase the initiation efficiency, giving the Ac-^{D}aa-peptide with a higher yield than that initiated with ^{D}aa-tRNA$^{fMet}_{CAU}$. This suggests that the preacylation of ^{D}aa-tRNA$^{fMet}_{CAU}$ is able to overcome the inherent modest initiation efficiency of ^{D}aa controlled by MTF.

We have shown that the translation apparatus tolerates ^{D}aa for initiation to afford H-^{D}aa-peptide or Ac-^{D}aa-peptide when ^{D}aa-tRNA$^{fMet}_{CAU}$ or Ac-^{D}aa-tRNA$^{fMet}_{CAU}$ is given to the Met-withdrawn *w*PURE system. This is in sharp contrast to the little success achieved in the incorporation of ^{D}aa into the nascent peptide chain via amber-suppression elongation (Fig. 7). Why so? To provide a definitive answer(s) to this question, more detailed mechanistic investigations are certainly required; but at the present stage we are able to suggest the following three potential reasons.

The first reason can be attributed to the nature of the mechanism for the fidelity control, which relies upon the function of initiation factor 2 (IF-2). Despite the fact that MetRS charges ^{L}Met onto both initiator tRNA$^{fMet}_{CAU}$ and elongator tRNA$^{Met}_{CAU}$, IF-2 recognizes only f-^{L}Met-tRNA$^{fMet}_{CAU}$ over ^{L}Met-tRNA$^{Met}_{CAU}$ (Schmitt et al. 1996; Boelens and Gualerzi 2002). The differences between these two ^{L}Met-tRNAs lies in the tRNA's body sequence (particularly that the 5′-terminal nucleotide is unpaired in the initiator whereas it is paired in the elongator) (Mayer et al. 2001) and the formyl group on ^{L}Met. Our previous and current studies of the reprogrammed initiation have shown that the formylated aminoacyl-tRNA$^{fMet}_{CAU}$ very likely recruits IF-2 more efficiently than nonformylated, yet the formyl group is not the essential selection element. Apparently, a more critical selection element is the structural features of the initiator tRNA$^{fMet}_{CAU}$ distinct from the elongator tRNA$^{Met}_{CAU}$ (Mayer et al. 2003). Our results show that IF-2 likely tolerates not only various ^{L}aa but also even ^{D}aa (particularly when N^{α}-acylated) as far as they are charged onto tRNA$^{fMet}_{CAU}$, suggesting that its recognition of side chains and the chirality of amino acids are less strict compared with that of the tRNA's body sequence. On the other hand, in the elongation event, it has been firmly established that EF-Tu discriminates cognate pairs of ^{L}aa and elongator tRNA against noncognate pairs upon recruiting them to the ribosome elongation complex with the strict control of a nearly uniform affinity toward the cognate pairs (Stanzel et al. 1994; Ibba and Söll 1999; LaRiviere et al. 2001). This mechanism is also applicable to pairs of nonproteinogenic amino acids and amber or other possible elongator tRNAs; particularly, the opposite chirality in ^{D}aa is forcefully rejected by this EF-Tu selection filter, resulting in the

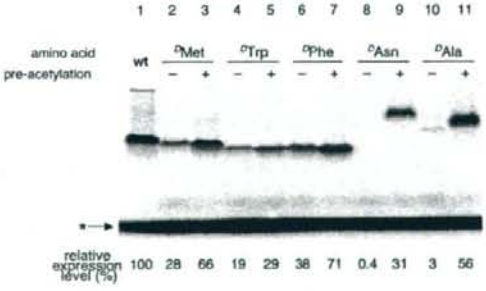


FIGURE 6. Tricine PAGE analysis of the translation products with N^{α}-acylated initiators. (Lane *1*) expression of wild type; (lanes 2–11) reprogrammed initiation with Na-free amino acids or Na-acetyl amino acids. Each expression level relative to wild type is determined by a mean score of duplicates or more. The band indicated by an asterisk corresponds to free [^{14}C]-Asp that remained unincorporated into the Flag peptide.

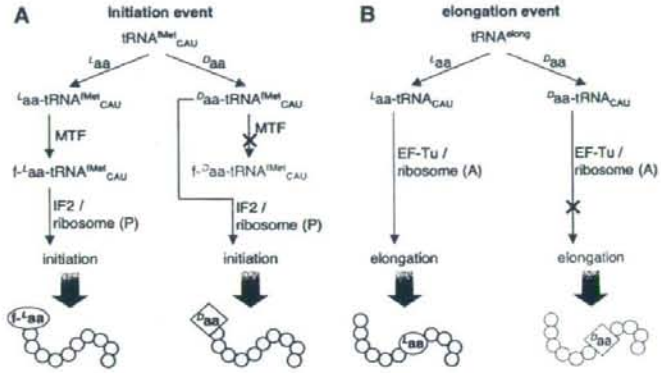


FIGURE 7. Initiation versus elongation with ^{L}aa and ^{D}aa. (*A*) ^{L}aa charged onto tRNA$^{fMet}_{CAU}$ is formylated by MTF to give f-^{L}aa-tRNA$^{fMet}_{CAU}$; then IF2 brings it to the P site of the ribosome initiation complex and the translation is started to yield f-^{L}aa-peptide. On the other hand, ^{D}aa charged onto tRNA$^{fMet}_{CAU}$ is not generally formylated, i.e., it bypasses the formylation step; yet IF2 is able to bind and bring ^{D}aa-tRNA$^{fMet}_{CAU}$ to the P site of the ribosome initiation complex; thus the translation reaction can be initiated with ^{D}aa, generally giving H-^{D}aa-peptide. (*B*) ^{L}aa charged onto amber tRNA$_{CUA}$ binds to EF-Tu and goes to the ribosome P site, and ^{L}aa is incorporated into the peptide nascent chain. ^{D}aa charged onto amber tRNA$_{CUA}$ presumably binds to EF-Tu poorly and also is incompatible with the ribosome A site and in most cases fails to suppress the amber codon.

observed poor incorporation efficiency of ^{D}aa-elongation in general.

As the second reason, we suggest that the recognition of initiator ^{D}aa-tRNA$^{fMet}_{CAU}$ by P site of ribosome is less strict than that of elongator ^{D}aa-tRNAs by the A site. Presumably, since the P site needs to accommodate peptidyl-tRNA, it has a more spacious pocket than the A site (Noller et al. 2005). In fact, our recent studies using various peptidyl-tRNAs for initiation imply that the P site has a surprising tolerance toward the nonproteinogenic peptidyl groups (Y. Goto, H. Murakami, and H. Suga, in prep.). Moreover, in the P site, the aminoacyl carbonyl carbon of initiator f-^{L}Met-tRNA$^{fMet}_{CAU}$ or peptidyl-tRNAs acts as an electrophile to the α-amino group of elongator ^{L}aa-tRNAs, so that it is only necessary to set its corresponding carbonyl group at the appropriate position; this positioning would be unlikely to be influenced by the chirality of the α-carbon of electrophile amino acids. In the A site, on the other hand, in order to efficiently process the nucleophilic attack of the α-amino group of elongator ^{L}aa-tRNA to the carbonyl group of the initiator or peptidyl-tRNAs, precise positioning of the α-amino group is critical. This positioning should be largely influenced by the chirality of the α-carbon of nucleophile amino acids, in contrast to that of electrophile in the P site. We thus propose that the spacious and functional differences in ribosome's P and A sites play a critical role in accepting and discriminating ^{D}aa in the initiation and elongation events, respectively.

The last reason involves a technical issue; since the background initiation was nearly completely suppressed by depleting Met in the *w*PURE system, undesired competition of f-Met-tRNA$^{fMet}_{CAU}$ against ^{D}aa-tRNA$^{fMet}_{CAU}$ does not occur, leading to the exclusive initiation of ^{D}aa-tRNA$^{fMet}_{CAU}$. On the other hand, the background of the amber suppression cannot be completely repressed due to the inclusion of RFs in the present methodology. This may suggest that if the background level of competing elongations were controlled by genetic code reprogramming, ^{D}aa can be incorporated into nascent peptide chain.

Although, as stated earlier, knowing the exact mechanism in the ^{D}aa-initiation requires more detailed investigations, particularly at the molecular level, the present work has given us an intriguing question for a general mechanism of the fidelity controls at initiation. It has been established that some aaRSs, such as TyrRS, TrpRS, and AspRS, mischarge the corresponding ^{D}aa onto the cognate elongator tRNAs (Calendar and Berg 1966; Soutourina et al. 2000). This mischarging event gives a negative impact on cell growth via two possible mechanisms. A small portion of such a mischarged ^{D}aa may be incorporated into a nascent peptide chain even though ^{D}aa is generally a poor substrate for elongation (Roesser et al. 1989; Bain et al. 1991; Ellman et al. 1992; Dedkova et al. 2003; Murakami et al. 2006). Alternatively, undesirable generation of the mischarged elongator tRNAs may decrease the concentration of available elongator tRNAs for the innate function. At least in bacteria and yeast, it has been found that ^{D}Tyr-tRNA deacylase plays a role in discharging such mischarged ^{D}aa-tRNAs, and thus the harmful accumulation of ^{D}aa-tRNAs are avoided (Calendar and Berg 1967; Soutourina et al. 2004). On the other hand, we found in the present work that ^{D}Met-tRNA$^{fMet}_{CAU}$ could be formylated by MTF and the resulting f-^{D}Met-tRNA$^{fMet}_{CAU}$ could rather efficiently initiate the translation. Therefore, even if MetRS mischarged only a small fraction of tRNA$^{fMet}_{CAU}$ with ^{D}aa, the resulting f-^{D}Met-tRNA$^{fMet}_{CAU}$ would compete with f-^{L}Met-tRNA$^{fMet}_{CAU}$ at the initiation of protein synthesis. Hence, it is intriguing to study whether MetRS mischarges ^{D}Met onto tRNA$^{fMet}_{CAU}$. If so, it would be important to test whether ^{D}Tyr-tRNA deacylase is able to discharge ^{D}Met-tRNA$^{fMet}_{CAU}$, since it has been shown that this enzyme is able to discharge not only ^{D}Tyr-tRNA^{Tyr} but also ^{D}Trp-tRNA^{Trp} and ^{D}Asp-tRNA^{Asp} (Soutourina et al. 2000). If MetRS is able to mischarge ^{D}Met and if ^{D}Tyr-tRNA deacylase is unable to discharge

^{D}Met-tRNA$^{fMet}_{CAU}$, it will be interesting to investigate the mechanism of how the generation of f-^{D}Met-peptide is avoided or how f-^{D}Met-peptide is processed, e.g., whether peptide deformylase and methionine aminopeptidase are able to remove the f-^{D}Met group from the f-^{D}Met-peptide.

Nevertheless, the technical merit of our work is apparent. The reprogrammed initiation with ^{D}aa and acyl-^{D}aa in translation enables us to synthesize a variety of H-^{D}aa and acyl-^{D}aa-peptides, respectively. The ^{D}aa-capping would grant resistance against proteolytic degradation to the peptide as demonstrated in previous works using chemical synthesis (Hong et al. 1999; Tugyi et al. 2005). We have reported here a new method for the ribosomal synthesis of ^{D}aa-capped or acyl-^{D}aa-capped peptides. The mRNA-programmed synthesis of such peptide libraries should provide a new avenue to discover novel physiologically stable peptidic drug candidates against various therapeutic targets, and such investigations are underway.

MATERIALS AND METHODS

Materials

Chemicals were purchased from Watanabe Chemical Industries, Nacalai Tesque, Kanto Chemical, Sigma-Aldrich Japan, or Wako Pure Chemical Industries unless noted otherwise and used without further purification. All oligonucleotides were purchased from Operon Biotechnologies. Flexizyme and tRNA molecules were synthesized using the same procedure as previously described (Murakami et al. 2006; Ohta et al. 2007; Goto et al. 2008).

General protocol of translation

^{D}aa-tRNA$^{fMet}_{CAUU}$ was prepared by the following procedure. We heated 40 μM tRNA$^{fMet}_{CAU}$ in 0.2 M HEPES-K buffer (pH 7.5), 0.2 M KCl (7.5 μL) at 95°C for 3 min and cooled to 25°C for 5 min. $MgCl_2$ (3 M, 3 μL) and flexizyme (dFx or eFx; see Murakami et al. 2006) (200 μM, 1.5 μL) were added and the mixture was incubated at 25°C for 5 min. The reaction was initiated by addition of 3 μL of 25 mM ^{D}aa substrate in DMSO and incubated on ice for the optimized times, generally 2–6 h (Murakami et al. 2006). After acylation, the reaction was stopped by addition of 45 μL of 0.6 M sodium acetate at pH 5, and the RNA was recovered by ethanol precipitation. The pellet was rinsed twice with 70% ethanol with 0.1 M sodium acetate at pH 5.0, and once with 70% ethanol. The Xaa-tRNA$^{fMet}_{CAU}$ was dissolved in 0.5 μL of 1 mM sodium acetate just before adding to translation mixture.

The *w*PURE system containing all necessary components for translation except for all 20 standard amino acids was used in this study. Translation was carried out using *w*PURE system with 0.04 μM mDNA1 containing 200 μM each Thr, Tyr, Lys, 50 μM [^{14}C]-Asp, and 120 μM of various ^{D}aa-tRNA$^{fMet}_{CAU}$ molecules. The wild-type expression was carried out with *w*PURE system with 0.04 μM mDNA1 containing 200 μM each Met, Thr, Tyr, Lys and 50 μM [^{14}C]-Asp. The translation mixture (2.5 μL) was incubated at 37°C for 1 h and analyzed by Tricine-SDS PAGE and autoradiography (FLA-5100, Fuji).

Analysis of peptides by MALDI-TOF

For MALDI-TOF analysis, the translation reaction (5 μL) was performed in the presence of Asp, instead of [^{14}C]-Asp. The translation product from mDNA1 was immobilized with FLAG-M2 agarose (Sigma). After washing the resin with 30 μL of W buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl), the immobilized peptides were eluted with 10 μL of 0.2% TFA. The purified peptide was desalted with ZipTip$_{μ\text{-}C18}$ (Millipore), and eluted with 1 μL of a 50% MeCN, 0.1% TFA solution saturated with the matrix (*R*)-cyano-4-hydroxycinnamic acid. MALDI-TOF mass spectrometry was performed using an autoflex II TOF/TOF (BRUKER DALTONICS) under the linear/positive mode and externally calibrated with peptide calibration standard II (BRUKER DALTONICS).

Competition of ^{D}aa-tRNA$^{fMet}_{CAU}$ and ^{L}aa-tRNA$^{fMet}_{CAU}$

^{D}aa-tRNA$^{fMet}_{CAU}$ and ^{L}aa-tRNA$^{fMet}_{CAU}$ (aa = Tyr, Trp, and Leu) were prepared by the flexizyme system with the same procedure as described in the general protocol of translation. After completing the flexizyme reaction, the reaction mixture for ^{D}aa-tRNA$^{fMet}_{CAU}$ and that for the corresponding ^{L}aa-tRNA$^{fMet}_{CAU}$ were mixed with three different ratio (*D*:*L* = 99:1, 95:5, and 80:20). Then the reaction was stopped by addition of 45 μL of 0.6 M sodium acetateat pH 5, and the RNA mixture was recovered by ethanol precipitation. The pellet was rinsed twice with 70% ethanol with 0.1 M sodium acetate (pH 5.0), and once with 70% ethanol. The Xaa-tRNA$^{fMet}_{CAU}$ was dissolved in 0.5 μL of 1 mM sodium acetate just before adding to translation mixture. Translation reaction was carried out with the same procedure as described above, except for adding the mixture of ^{D}aa-tRNA$^{fMet}_{CAU}$ and ^{L}aa-tRNA$^{fMet}_{CAU}$ to be 120 μM as total concentration of tRNAs in the translation mixture. MALDI-TOF analysis of the products was carried out in the same manner as above.

SUPPLEMENTAL DATA

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank Dr. Patrick C. Reid for proofreading. This work was supported by grants from the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) to H.S., Grants-in-Aid for JSPS Fellows (18-10526) to Y.G., a research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO) to H.S., and the Industrial Technology Research Grant Program in NEDO (05A02513a) to H.M.

Received January 28, 2008; accepted March 30, 2008.

REFERENCES

Bain, J.D., Diala, E.S., Glabe, C.G., Wacker, D.A., Lyttle, M.H., Dix, T.A., and Chamberlin, A.R. 1991. Site-specific incorporation of nonnatural residues during in vitro protein biosynthesis with semisynthetic aminoacyl-tRNAs. *Biochemistry* **30:** 5411–5421.

Boelens, R. and Gualerzi, C.O. 2002. Structure and function of bacterial initiation factors. *Curr. Protein Pept. Sci.* **3:** 107–119.

Brizzard, B.L., Chubet, R.G., and Vizard, D.L. 1994. Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* **16:** 730–735.

Calendar, R. and Berg, P. 1966. The catalytic properties of tyrosyl ribonucleic acid synthetases from *Escherichia coli* and *Bacillus subtilis*. *Biochemistry* **5:** 1690–1695.

Calendar, R. and Berg, P. 1967. D-Tyrosyl RNA: Formation, hydrolysis and utilization for protein synthesis. *J. Mol. Biol.* **26:** 39–54.

Dedkova, L.M., Fahmi, N.E., Golovine, S.Y., and Hecht, S.M. 2003. Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.* **125:** 6616–6617.

Dedkova, L.M., Fahmi, N.E., Golovine, S.Y., and Hecht, S.M. 2006. Construction of modified ribosomes for incorporation of d-amino acids into proteins. *Biochemistry* **45:** 15541–15551.

Ellman, J.A., Mendel, D., and Schultz, P.G. 1992. Site-specific incorporation of novel backbone structures into proteins. *Science* **255:** 197–200.

Gold, L. 1988. Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu. Rev. Biochem.* **57:** 199–233.

Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. 2008. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **3:** 120–129.

Gualerzi, C.O. and Pon, C.L. 1990. Initiation of mRNA translation in prokaryotes. *Biochemistry* **29:** 5881–5889.

Hong, S.Y., Oh, J.E., and Lee, K.H. 1999. Effect of D-amino acid substitution on the stability, the secondary structure, and the activity of membrane-active peptide. *Biochem. Pharmacol.* **58:** 1775–1780.

Ibba, M. and Söll, D. 1999. Quality control mechanisms during translation. *Science* **286:** 1893–1897.

Kozak, M. 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47:** 1–45.

LaRiviere, F.J., Wolfson, A.D., and Uhlenbeck, O.C. 2001. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* **294:** 165–168.

Mayer, C., Stortchevoi, A., Kohrer, C., Varshney, U., and RajBhandary, U.L. 2001. Initiator tRNA and its role in initiation of protein synthesis. *Cold Spring Harb. Symp. Quant. Biol.* **66:** 195–206.

Mayer, C., Kohrer, C., Kenny, E., Prusko, C., and RajBhandary, U.L. 2003. Anticodon sequence mutants of *Escherichia coli* initiator tRNA: Effects of overproduction of aminoacyl-tRNA synthetases, methionyl-tRNA formyltransferase, and initiation factor 2 on activity in initiation. *Biochemistry* **42:** 4787–4799.

Murakami, H., Kourouklis, D., and Suga, H. 2003a. Using a solid-phase ribozyme aminoacylation system to reprogram the genetic code. *Chem. Biol.* **10:** 1077–1084.

Murakami, H., Saito, H., and Suga, H. 2003b. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **10:** 655–662.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. 2006. A highly flexible tRNA acylation method for nonnatural polypeptide synthesis. *Nat. Methods* **3:** 357–359.

Noller, H.F., Hoang, L., and Fredrick, K. 2005. The 30S ribosomal P site: A function of 16S rRNA. *FEBS Lett.* **579:** 855–858.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., and Schultz, P.G. 1989. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244:** 182–188.

Ohta, A., Murakami, H., Higashimura, E., and Suga, H. 2007. Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14:** 1315–1322.

Roesser, J.R., Xu, C., Payne, R.C., Surratt, C.K., and Hecht, S.M. 1989. Preparation of misacylated aminoacyl-tRNA^{Phe}s useful as probes of the ribosomal acceptor site. *Biochemistry* **28:** 5185–5195.

Sankaranarayanan, R. and Moras, D. 2001. The fidelity of the translation of the genetic code. *Acta Biochim. Pol.* **48:** 323–335.

Schagger, H. and von Jagow, G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166:** 368–379.

Schmitt, E., Guillon, J.M., Meinnel, T., Mechulam, Y., Dardel, F., and Blanquet, S. 1996. Molecular recognition governing the initiation of translation in *Escherichia coli*. A review. *Biochimie* **78:** 543–554.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. 2001. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19:** 751–755.

Söll, D. 1990. The accuracy of aminoacylation—Ensuring the fidelity of the genetic code. *Experientia* **46:** 1089–1096.

Soutourina, J., Plateau, P., and Blanquet, S. 2000. Metabolism of D-aminoacyl-tRNAs in *Escherichia coli* and *Saccharomyces cerevisiae* cells. *J. Biol. Chem.* **275:** 32535–32542.

Soutourina, O., Soutourina, J., Blanquet, S., and Plateau, P. 2004. Formation of D-tyrosyl-tRNA^{Tyr} accounts for the toxicity of D-tyrosine toward *Escherichia coli*. *J. Biol. Chem.* **279:** 42560–42565.

Stanzel, M., Schon, A., and Sprinzl, M. 1994. Discrimination against misacylated tRNA by chloroplast elongation factor Tu. *Eur. J. Biochem.* **219:** 435–439.

Starck, S.R., Qi, X., Olsen, B.N., and Roberts, R.W. 2003. The puromycin route to assess stereo- and regiochemical constraints on peptide bond formation in eukaryotic ribosomes. *J. Am. Chem. Soc.* **125:** 8090–8091.

Tan, Z., Forster, A.C., Blacklow, S.C., and Cornish, V.W. 2004. Amino acid backbone specificity of the *Escherichia coli* translation machinery. *J. Am. Chem. Soc.* **126:** 12752–12753.

Tugyi, R., Uray, K., Ivan, D., Fellinger, E., Perkins, A., and Hudecz, F. 2005. Partial D-amino acid substitution: Improved enzymatic stability and preserved Ab recognition of a MUC2 epitope peptide. *Proc. Natl. Acad. Sci.* **102:** 413–418.

ELSEVIER

Journal of Hepatology 49 (2008) 207–216

Journal of Hepatology

www.elsevier.com/locate/jhep

Identification of novel candidate tumour marker genes for intrahepatic cholangiocarcinoma☆

Ryuhei Nishino[1], Masao Honda[1], Taro Yamashita[1], Hajime Takatori[1], Hiroshi Minato[2], Yoh Zen[2], Motoko Sasaki[3], Hiroyuki Takamura[4], Katsuhisa Horimoto[5], Tetsuo Ohta[4], Yasuni Nakanuma[3], Shuichi Kaneko[1,*]

[1]*Department of Gastroenterology, Kanazawa University Graduate School of Medical Science, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[2]*Pathology Section, Kanazawa University Hospital, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[3]*Department of Human Pathology, Kanazawa University Graduate School of Medical Science, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[4]*Department of Gastroenterologic Surgery, Kanazawa University Graduate School of Medical Science, Kanazawa University, 13-1 Takara-Machi, Kanazawa 920-8641, Japan*
[5]*Biological Network Team, Computational Biology Research Centre, National Institute of Advanced Industrial Science and Technology, Japan*

See Editorial, pages 160–162

***Background/Aims*: Specific markers are required for early detection and diagnosis of intrahepatic cholangiocarcinoma (ICC); however, the tumour markers currently in use are not specific for ICC.**

***Methods*: We compared an ICC cDNA library with that of hepatocellular carcinoma (HCC) by serial analysis of gene expression (SAGE). The expression patterns in each were confirmed by quantitative real-time reverse transcriptase-polymerase chain reaction (RT-PCR), immunoblotting and immunohistochemical analysis of 74 samples including 16 ICC samples.**

***Results*: A comparison of the two libraries revealed distinct gene expression patterns for each type of liver cancer. In addition to the known tumour markers, we detected nine novel genes associated with ICC. By comparing the mean transcript abundance in the ICC library with those in other libraries, including gastric, colon, prostate and breast cancer, together with our RT-PCR results, we identified three genes as specific markers of ICC: biglycan, insulin-like growth factor-binding protein 5 and claudin-4. Immunoblotting and immunohistochemical analyses showed that claudin-4 was highly expressed in ICC. Moreover, discrimination analysis revealed that a combination of these genes could be used to distinguish ICC from HCC or metastatic adenocarcinoma.**

***Conclusions*: We identified novel marker genes of ICC that are potentially useful for the diagnosis of liver cancer.**

***Keywords*: Differential diagnosis; Hepatocellular carcinoma; Intrahepatic cholangiocarcinoma; SAGE**

Received 5 October 2007; received in revised form 10 March 2008; accepted 24 March 2008; available online 5 May 2008
Associate Editor: J.M. Llovet

☆ The authors declare that they do not have anything to disclose regarding funding from industries or conflict of interest with respect to this manuscript.

* Corresponding author. Tel.: +81 76 265 2231; fax: +81 76 234 4250.
E-mail address: skaneko@m-kanazawa.jp (S. Kaneko).

Abbreviations: ICC, intrahepatic cholangiocarcinoma; HCC, hepatocellular carcinoma; AFP, alpha-foetoprotein; KRT, keratin; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; SAGE, serial analysis of gene expression; HBV, hepatitis B virus; HCV, hepatitis C virus; NL, normal liver; CLD, chronic liver disease; RT-PCR, reverse transcriptase-polymerase chain reaction; polyA-RNA, polyadenylated RNA; mRNA, messenger RNA; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; CITED4, Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 4; GSTP1, glutathione-*S*-transferase pi; BGN, biglycan; IGFBP5, insulin-like growth factor-binding protein 5; CLDN4, claudin-4; PFKP, platelet type phosphofructokinase; TM4SF1, transmembrane 4 L six family member 1; CAPN1, calpain 1 (μ/l) large subunit; CLDN10, claudin-10; S100A6, S100 calcium-binding protein A6; Wnt, wingless-type MMTV integration site; TGF-beta, transforming growth factor-beta.

0168-8278/$34.00
doi:10.1016/j.jhep.2008.03.025

1. Introduction

Intrahepatic cholangiocarcinoma (ICC), which arises from bile duct cells in the liver, currently accounts for approximately 15% of primary liver cancer [1]. Though it is less familiar than hepatocellular carcinoma (HCC), the incidence of ICC is increasing, especially in the United States, the United Kingdom and Australia [2–5].

The differential diagnosis of liver cancer depends on a combination of serological, radiological and histological examinations; however, the process can be difficult, especially in advanced stages. In addition, liver cancer may include metastasis from other organs, making the diagnosis even more difficult. Nevertheless, a correct diagnosis is essential to select the proper therapy and to determine a patient's prognosis.

Tumour markers are routinely used in the differential diagnosis of liver cancer. For example, alpha-foetoprotein (AFP) is a sensitive and specific marker of HCC [6]; keratin (KRT) 7 and KRT19 are amongst the markers of ICC. KRT7 and KRT19, which are expressed in the bile duct epithelium and ICC [7–10], can be used to distinguish HCC from ICC immunohistochemically. These are not specific to ICC, however, and are expressed in other cancers such as non-small-cell lung carcinoma [11]. Other proteins such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) have been used as serum markers for ICC [12,13], but these are also not specific to ICC because they are overexpressed in other malignant tumours. Thus, identifying ICC-specific markers will be valuable for differentially diagnosing liver cancer to characterise ICC at the molecular level and to develop improved therapies for patients with ICC. Recently, Lodi et al. cited claudin-4 (CLDN4) expression as a possible marker in differentiating biliary tract cancers from HCC [14]. However, the gene's ability to distinguish metastatic liver cancer has not been evaluated.

In recent years, serial analysis of gene expression (SAGE) [15] and DNA microarrays have been used to comprehensively analyse gene expression, including comparisons of the expression patterns in normal liver [16] and HCC [17–19]. We used SAGE to construct cDNA libraries of ICC and HCC and performed comprehensive analyses of the expression patterns in these tumours to identify novel markers of ICC.

2. Materials and methods

2.1. Tissue samples

For SAGE analysis, we prepared one ICC sample and three HCC samples (Table 1). The ICC sample was obtained by the surgical resection of a solitary cancerous lesion in the liver, which was diagnosed histopathologically as a moderately differentiated cholangiocellular carcinoma (Fig. 1A). The HCC samples were obtained from three surgically resected cancerous lesions, all of which were diagnosed histopathologically as well-differentiated HCCs (Fig. 1B). The clinical characteristics of the patients are provided in Table 1.

In total, 74 samples (i.e., 16 ICC, seven normal liver (NL), 20 chronic liver disease (CLD), 26 HCC (Tables 1 and 2) and five extrahepatic adenocarcinoma) were used for quantitative real-time reverse transcriptase-polymerase chain reaction detection (RT-PCR) and discrimination analyses. The NL samples were obtained by the surgical resection of colon tumours from patients with metastatic liver cancer, whereas the CLD samples were non-cancerous liver samples obtained from patients with HCC. The ICC samples were taken from nine patients by surgical resection and six cadavers at autopsy, in addition to the sample used for SAGE. The characteristics of the patients are provided in Table 2. As samples of extrahepatic adenocarcinoma, we used BD Premium Total RNA™ (BD Biosciences Clontech, Palo Alto, CA) taken from adenocarcinomas of human breast (Cat. No.636635), colon (636634), stomach (636629), uterus (636628) and lung (636633).

The study protocol conformed to the ethical guidelines of the Declaration of Helsinki (1975). All patients provided written informed consent for the analysis of the biopsy specimens and the hospital ethics committee approved the study.

2.2. SAGE

Each tissue sample was homogenised in liquid nitrogen and total RNA was extracted using a ToTALLY RNA™ kit (Ambion, Austin, TX). The polyadenylated RNA (polyA-RNA) was subsequently purified using a MicroPoly(A)Pure™ kit (Ambion). Aliquots of polyA-RNA (3 μg) from each ICC and HCC sample were used for SAGE, as described previously [15–17]. The NL tissue and HCC cDNA libraries have also been described previously [16,17]. To compare liver cancer with cancers from other organs, we obtained 16 SAGE libraries from the National Center for Biotechnology Information (NCBI) SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE/). These libraries originated from the tumours of the stomach (G189 and G234), colon (Tu98 and Tu102), prostate (PrCA-1 and Chen_Tumour_Pr) and breast (95-259, 95-347, DCIS, DCIS-2, DCIS-3, DCIS-4, DCIS-5, IDC-3, IDC-4 and IDC-5).

For the analysis of molecular functions, we used all databases registered in MetaCore™ from GeneGo Inc. (http://portal.genego.com/cgi/index.cgi). Statistical significance in this database was calculated using the basic equation (Supplementary Fig. 1).

2.3. Quantitative real-time RT-PCR detection

Template cDNA was synthesised from 1 μg of total RNA using SuperScript™ II RT (Invitrogen, San Diego, CA). The quantitative real-time RT-PCR detection was performed using a TaqMan® Gene Expression Assay kit (Applied Biosystems, Foster City, CA). The amount of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) (Hs99999905) mRNA in each sample was used to standardise the quantity of each of the following mRNA: Cbp/p300-interacting transactivator with Glu/Asp-rich carboxy-terminal domain 4 (*CITED4*) (Hs00388363), glutathione-*S*-transferase pi (*GSTP1*) (Hs00168310), biglycan (*BGN*) (Hs00156076), insulin-like growth factor-binding protein 5 (*IGFBP5*) (Hs00181213), claudin-4 (*CLDN4*) (Hs00533616), phosphofructokinase platelet type (*PFKP*) (Hs00242993), transmembrane 4 L six family member 1 (*TM4SF1*) (Hs00371997), calpain 1 (mu/l) large subunit (*CAPN1*) (Hs00559804) and claudin-10 (*CLDN10*) (Hs00199599).

2.4. Immunoblot analysis

Each tissue sample was homogenised in liquid nitrogen and a protein extract was prepared using radioimmunoprecipitation assay (RIPA) buffer. The extracts (7 μg/lane) were subsequently electrophoresed on SDS–10% polyacrylamide gels and transferred onto polyvinylidene fluoride. An extract prepared from KMBC cells, a human extrahepatic bile duct carcinoma cell line [21], was used as a positive control. The blots were then incubated for 1 h with goat polyclonal antibodies against claudin-4 (1:200 dilution; Santa Cruz Biotechnol-