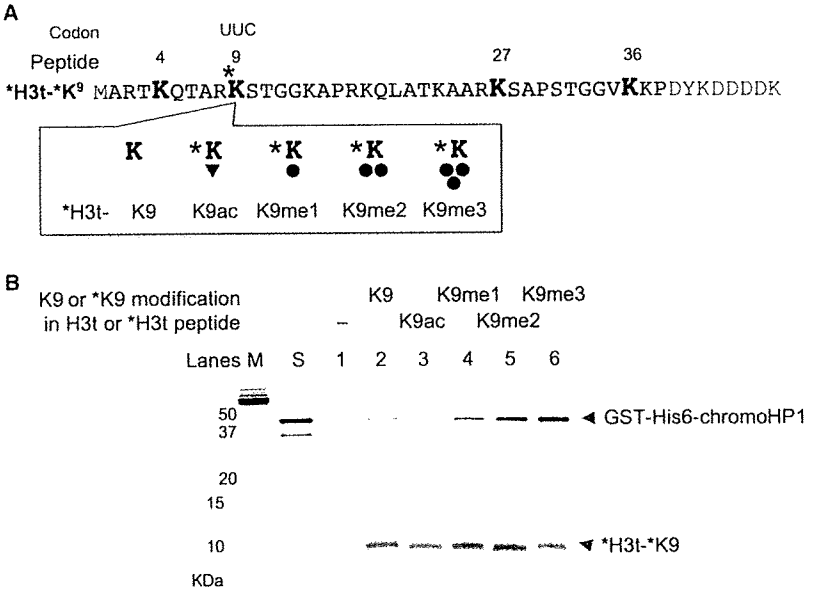


Figure 3. Modification-Specific Interaction of *H3t with the Chromodomain of HP1

(A) *H3t peptides used for the pull-down or the sensor chip for SPR analysis. Each peptide contains *K at position 9.

(B) Pull-down analysis of *H3t interaction with chromoHP1. Pulled-down prey (chromoHP1) and the baits (*H3t peptides) were probed using anti-polyhistidine and anti-Flag antibodies, respectively. M, protein size marker; S, chromoHP1 (2% of input); 1, pull-down with the bait prepared from the translation in the absence of both F and aminoacyl-tRNA$^{Asn\text{-}E2}_{GGA}$; 2–6, pull-down with *H3t-K9, -K9ac, -K9me1, -K9me2, and -K9me3, respectively.

(C) SPR analysis of *H3t interaction with chromoHP1. Respective peptides were immobilized on the sensor chip through preimmobilized anti-Flag tag antibody. Resonance units in the equilibrium with 2 μM of chromoHP1 were measured and normalized against the amount of *H3t bound to the chip. Mean values of triplicate measurements of two independent sets of experiments and the SD are shown. A scatchard plot of *H3tK9me3-chormoHP1 with varying concentrations of chromoHP1 is shown in the inset.

MALDI-TOF analysis of each Flag-purified peptide showed that the observed major peak was consistent with the expected MW for the multiple suppressions (Figure 4A, P-D, P-T, and P-Q).

To perform the ribosomal synthesis of *H3t peptides, two approaches are applicable. One is to prepare different *H3t peptides from the same template using reprogrammed genetic tables with customized *K assignments for each *H3t peptide (Figure 4B). This is possible because the dFx technology enables us to readily prepare any desired *K-tRNA$^{Asn\text{-}E2}_{NNN}$ to create a new reprogrammed genetic table. The other approach is to use a unique genetic table (e.g., as shown in Figure 1D), assigning each respective *K and then expressing the desired *H3t peptides from different mRNA sequences. In this method, swapping the positions of reassigned codons to *K in mRNA dictates the positions of *K incorporation in each *H3t peptide (Figure 4C). To show the versatility of our methodology, we employed both approaches to synthesize *H3t peptides.

In the first approach, each reprogrammed codon was assigned to K, meK, me2K, me3K, or acK by charging the respective *K onto tRNA$^{Asn\text{-}E2}_{NNN}$ (Figure 4B). In this series of experiments, the incorporation of K at preselected positions was also performed by the suppression using K-tRNA$^{Asn\text{-}E2}_{NNN}$ like other *K incorporations (as indicated by the color-coded Ks in Figure 4B). Upon expression of each peptide by the respective reprogrammed genetic table, we confirmed the product by MALDI-TOF analysis (see *H3t-II, *H3t-III, and *H3t-IV). The observed MW of the major peak was consistent with the expected values for the respective peptides in all cases, indicating that the desired *H3t peptides were synthesized.

Synthesis of H3t Peptides by Multiple Incorporations of Modified Lysines

We next directed our investigation to multiple incorporations of *Ks into H3t. Using the template shown in Figure 1C, codons at positions 4, 9, 27, and 36, coding for I, F, N, and W, respectively, were reprogrammed to assign me1K, me3K, acK, and me2K, respectively. We then suppressed the reprogrammed codons using the designated *K-tRNA$^{Asn\text{-}E2}_{NNN}$ in the *w*PURE system with proper supplementation of the cognate amino acids. The template sequence was expressed in the *w*PURE system supplemented with all amino acids to yield P-WT as a positive control, and the same template was translated in the *w*PURE system to show no detectable background expression (Figure 4A, lanes 1 and 2). When the suppressions were performed using the corresponding *K-tRNA$^{Asn\text{-}E2}_{NNN}$, a clear band appeared in each experiment (Figure 4A, lanes 3–6). Increasing the number of suppressions resulted in a concomitant decrease in the overall expression level, yet even the quadruple suppression was achieved at an expression level of 37% to that of P-WT (Figure 4A, lane 6, P-Q versus lane 1, P-WT). Most importantly,

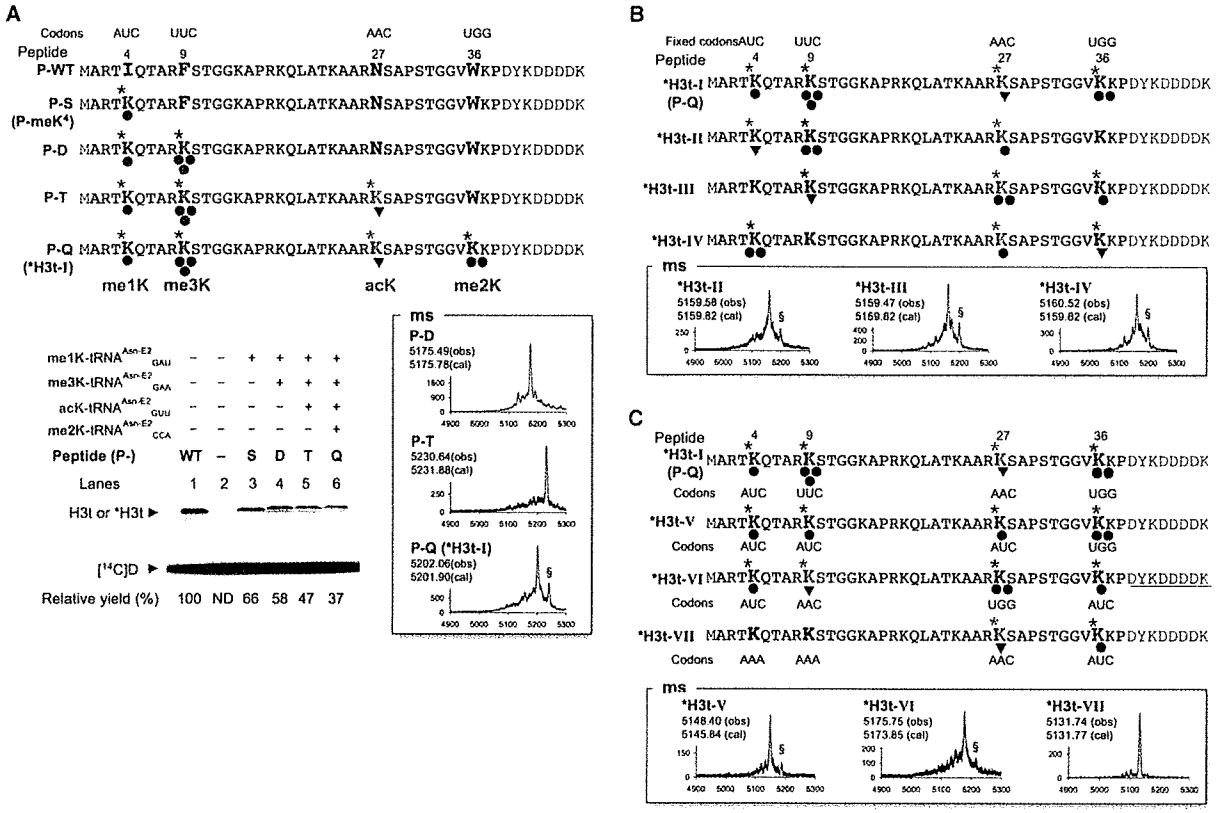


Figure 4. Multiple Incorporations of *Ks into H3t

(A) Demonstration of multiple incorporations of *Ks into H3t. Single to quadruple modifications can be coded in mRNA using the reassignment scheme shown in Figure 1C. Multiple incorporations were assessed quantitatively (left) and qualitatively (right) using 15% tricine-SDS-PAGE and MALDI-TOF, respectively.

(B) Multiple incorporations of *Ks into H3t peptides using codon reassignments using the single mRNA sequence.

(C) Multiple incorporations of *Ks into H3t peptides using codon arrangements in mRNA. § indicates a peak of the potassium adduct of me2K-containing *H3t peptides. The triangle and circle denote acetylation and methylation, respectively.

In the second approach, we prepared three mRNA templates encoding *H3t peptides (Figure 4C). Because the reprogrammed genetic codon table was fixed in this series of experiments (Figure 1D), each mRNA sequence dictated the identity and the position of *Ks incorporated, giving three different *H3t peptides with different MW. Unmodified K was assigned by the cognate AAA codon in mRNA (as indicated by black K in Figure 4C). The respective translated peptides were analyzed by MALDI-TOF, showing that the major peak of each peptide was consistent with the expected MW. Thus, this approach is also reliable for the synthesis of *H3t peptides with various combinations of K and *Ks at any designated positions.

Interplay among Lysine Modifications at Positions 9, 14, and 27 of H3t Peptides in Its binding to chromoHP1

HP1-histone H3 binding is mostly attributed to H3K9me, and, in many cases, the modification can be connected to transcriptional silencing or heterochromatin formation. However, it has also been shown that H3tK27me3 can recruit HP1 to H3t in vitro with a reduced affinity compared with that of an H3tK9me3 peptide (Fischle et al., 2003), even though the neighboring amino acids of K9 and K27 are similar (Figure 5A). However, acetylation at K14 has been generally conceived to be a transcription-facilitating epigenetic mark as opposed to K9me. Mass analyses of in vivo modifications of histone H3 revealed that these two epigenetically opposite modifications coexist in the same molecule of histone H3 (Bonenfant et al., 2007; Garcia et al., 2007; Thomas et al., 2006). Furthermore, a significant portion of histone H3 was shown to contain K9meK27me double modification or K9meK14acK27me triple modification (Garcia et al., 2007). In terms of the role of these modifications on the interaction between *H3t and HP1, it has been reported using synthetic *H3t peptides lacking the K27 modification that *H3tK9me3K14ac did not show an appreciable difference from *H3tK9me3 in qualitative pull-down experiments (Hirota et al., 2005; Mateescu et al., 2004). However, there might be interplay between the modifications of K9 and K14 along with K27 in the HP1-H3t interaction, which had not been investigated. To address the above question, we performed SPR experiments where *H3t peptides containing combinatorial modifications at the above K residues were tested for binding to chromoHP1.

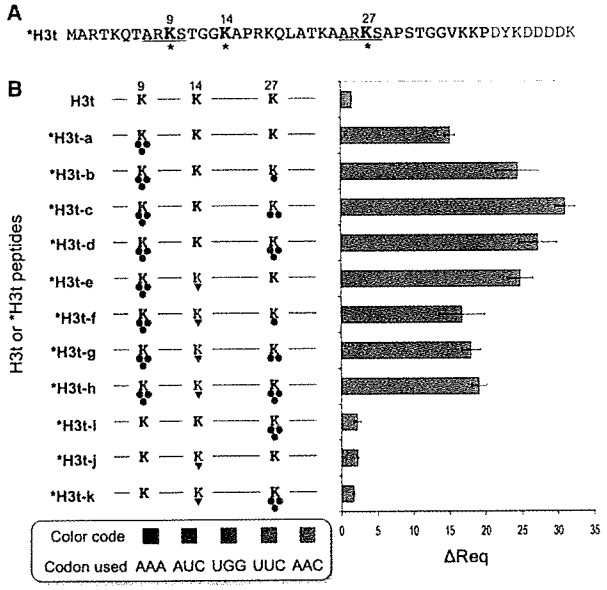


Figure 5. Interplay among Lysine Modifications at Positions 9, 14, and 27 in chromoHP1-*H3t Binding

(A) The sequence of *H3t peptides used. Positions 9, 14, and 27 contained modified Ks in various combinations. The consensus sequences in the *K9 and *K27 regions are underlined.

(B) SPR analysis of chromoHP1 interaction with *H3t peptides modified at positions 9, 14, and 27 in combinatorial manners. Modifications are summarized in the left panel. *Ks were incorporated into the peptide sequence using the codon assignments shown in color codes in the bottom left panel. The right panel shows resonance units observed in the equilibrium with 0.25 μM of chromoHP1 normalized against the amount of *H3t bound to the chip. Mean values and SD are shown in each bar, which were generated from triplicate measurements of two independent sets of experiment (*H3t-a~h) or a single set of experiment (H3t and *H3t-i~k). The triangle and circle denote acetylation and methylation, respectively.

We prepared a series of *H3t peptides with a total of 10 combinations of the methyl and acetyl modifications at positions 9, 14, and 27 (Figure 5A, *H3t-b–k; see also Figure S4 for the product verification by mass analyses) using the reprogrammed codons shown in Figure 1C (the summary of the reprogrammed codons used was also shown in the bottom panel of Figure 5B). The concentration of chromoHP1 in the mobile phase was kept at 0.25 μM to prevent saturation binding, where the amount of bound chromoHP1 (ΔReq) was expected to be proportional to the affinity of respective peptides. This let us assess the effect of modifications on chromoHP1-*H3t binding semiquantitatively. As shown earlier in Figure 3, K9me3 modification promoted the chromoHP1-*H3t binding (H3t versus *H3t-a, Figure 5B). K9me3K27me double modification resulted in elevated binding over K9me3 modification (*H3t-b~d versus *H3t-a), though K27me3 alone (*H3t-i) did not show an appreciable positive effect on chromoHP1 binding compared with H3t. However, the K14ac modification on *H3tK9me3 (*H3t-a→*H3t-e) increased the binding to a similar level observed in *H3tK9me3K27me1 (*H3t-e versus *H3t-b), whereas the same modification on *H3tK9me3K27me1 (*H3t-b→*H3t-f) suppressed chromoHP1 binding down to the level of *H3tK9me3 (*H3t-f versus *H3t-a). Once K14ac modification was placed in *H3tK9me3-K14acK27me, further methylation(s) on K27 did not affect the affinity (*H3t-f→*H3t-g or *H3t-h). Finally, neither K14ac modification alone on H3t (*H3t-j) nor K27me3 modification on *H3tK27me3 (*H3t-k) showed a significant elevation in chromoHP1 affinity from that of H3t.

DISCUSSION

In this report, we show the preparation of *H3t with multiple modifications using genetic code reprogramming and the functional use of these peptides in elucidating histone modification-effector binding relationships. Because of the biological significance of histone modifications and our insufficient knowledge of their roles in epigenetic control, it has been of great interest to develop methods to prepare histones or histone tails containing *Ks. One such method is to use enzymes that are known to modify histones (Kouzarides, 2007). This method allows for the preparation of not only modified histone tails but also the full-length histones, which represents the most significant advantage of this method. However, many of such enzymes would modify multiple sites in the target histones with incomplete modifications, leaving certain heterogeneities in the modified histones (Winter et al., 2008). This may impose difficulties in investigating specific interactions of histones or histone tails with effector proteins in certain cases. An elegant alternative approach was recently reported where a cysteine residue was introduced to H3 and then chemically modified with 2-halo-ethylamine derivatives to install a *K analog (Simon et al., 2007). The virtue of this method is its simplicity for the synthesis of the full-length *H3 analogs, whereas an inherent limitation is that only a single site modification or multiple but homogeneous *K modifications (albeit this has yet to be demonstrated) is possible; therefore it is not applicable to addressing questions regarding the importance of combinatorial modifications at the arbitrary K residues.

Chemically synthesized 20- to 30-amino-acid-long peptides containing *K at a certain specific sites has been widely and successfully used to elucidate the molecular interactions taking place on histones (Fischle et al., 2003; Jacobs et al., 2001; Lachner et al., 2001; Matthews et al., 2007; Meehan et al., 2003). Such short peptides can also be ligated to the C-terminal fragment of H3 or H4 to construct the full-length histones (Fischle et al., 2003; Shogren-Knaak et al., 2003, 2006). However, to elucidate the role of sparsely located modifications, the length of greater than 40 amino acids that span the tail would be preferable. This task is not easy to achieve using chemical synthesis, particularly if multiple *Ks with various kinds were necessary to be placed in the sequence.

Comparatively, ribosomal synthesis can be a practical and reliable method for the preparation of relatively long peptides with versatility. However, for the preparation of *H3t with multiple modifications, two prerequisites must be met: (i) the ribosome should use *Ks and (ii) the positions of these amino acids should be coded for in the mRNA precisely. Because all *Ks, including the bulkiest me3K, were incorporated into the peptide efficiently (Figure 2), the first prerequisite was easily met. The second prerequisite was solved by the use of genetic code reprogramming combined with the *w*PURE and flexizyme systems. This is, to the best of our knowledge, the first report concerning the ribosomal incorporation of *Ks with various degrees of methyl

modifications into a peptide chain. Combining these two findings, we could prepare a variety of *H3t peptides with four kinds of modifications at four different positions 4, 9, 27, and 36 in the desired combinations (Figure 4). The 47-mer Flag-tagged *H3t peptides with sparsely located such modifications were easily prepared using our approach.

Taking advantage of the fact that any kind of peptide tag can be appended to *H3t by expressing mRNA that encodes such a tag at the N or C terminus, the crude *H3t-Flag peptides from the translation can be directly immobilized on an anti-Flag antibody-coated resin or sensor chip; conveniently, the immobilized peptides can be used to confirm K9 modification-specific binding of *H3t to chromoHP1 (Figure 3). The quantitative analysis of the interaction between *H3tK9me3 and chromoHP1 using SPR showed tight binding with a dissociation constant of 0.76 μM. This value is slightly lower than a previously reported value (2.5–4.0 μM) using a short synthetic *H3tK9me3 peptide consisting of A1–A15 (Fischle et al., 2003; Jacobs and Khorasanizadeh, 2002). Although the exact reason for a decrease in the dissociation constant of our *H3tK9me3 is unclear, it is possible that the full-length *H3tK9me3 might act differently from the shorter peptide for the binding to chromoHP1. Nonetheless, the resin-bound *H3t could be used to pull down chromoHP1 (Figure 3), demonstrating the feasibility of our method for an application that allows us to pull down various and possibly novel effector proteins from cell lysates.

Finally, we have shown the use of our methodology in unveiling the possible interplay among the modifications at K9, K14, and K27. Resonance unit at equilibrium (ΔReq) of *H3t-chromoHP1 binding increased by an approximately 2-fold when the second modification of K27me was added to *H3tK9me3 (Figure 5, *H3t-a→*H3t-c). K27me3 modification alone on H3t did not positively influence binding under the conditions used here (Figure 5, *H3t-i). Despite that the neighboring residues of K9me3 and K27me3 are similar (see underlines in Figure 5A), the X-ray structure of the complex of *H3tK9me3 with chromoHP1 has revealed that T6 in the K9me3 region plays a critical role in fortifying their interaction, resulting in lower affinity of chromoHP1 to K27me3 (Fischle et al., 2003). However, a 2-fold enhancement in binding observed for *H3tK9me3K27me2 (Figure 5, *H3t-c) suggests that methylation(s) of K27 may contribute to the binding once chromoHP1 is recruited to the site of K9me3. Moreover, K14ac modification in *H3tK9me3K27me3 cancels the observed enhancement in affinity (Figure 5, *H3t-d→*H3t-h), exhibiting the binding ability similar to *H3tK9me3 (Figure 5, *H3t-a). These observations allow us to propose a model where methylation(s) on K27 retards the dissociation of chromoHP1 from the peptide by "shuttling" chromoHP1 between the K9me3 region and the K27me3 region of H3; when K14ac modification is added, the shuttling is disrupted, presumably by a conformational change of the peptide, neutralizing the enhancement by K27me3. However, this shuttling model fails to explain another observation made for the K14ac modification on *H3tK9me3 that enhances binding to chromoHP1 (Figure 5, *H3t-a→*H3t-e). Because *H3tK14ac does not bind to chromoHP1 (Figure 5, *H3t-j), this effect is clearly cooperative with the K9me3 modification. We therefore propose an alternative model that K14ac or K27me3 modification alone on *H3tK9me3 "locks" the peptide conformation with a favorable state for binding to chromoHP1. However, the simultaneous modifications in both sites unlock the favorable conformation, thereby canceling the enhancement. More biochemical and structural studies are required to reveal the molecular mechanism of the above two models in the future.

Regardless of the exact mechanism, in human cells contradictory epigenetic marks of K9me3 and K14ac do coexist in the same molecule of histone H3, and this double modification occasionally accompanies the third modification of K27me (Garcia et al., 2007). Also, K14ac was shown to be necessary together with phosphorylation (ph) at S10 to eject HP1 from histone H3K9me, during the G2/M phase of the human cell cycle (Mateescu et al., 2004); in other cases, S10ph alone was enough to break that binding (Fischle et al., 2005; Hirota et al., 2005). Thus, it can be speculated at this point that the effect of K14ac on HP1 binding to histone H3 would be K27me dependent and specific for the cell cycle and/or the position on the chromatin. Even though the crosstalk among K9me, S10ph, and K14ac modifications in histone H3 have been investigated using short peptides, our work suggests that the crosstalk of these modifications in the context with the K27 modification would be worthy to reinvestigate in a more extensive manner.

SIGNIFICANCE

Posttranslational methylation of lysines in histones poses significant impacts on several cellular functions, thus constituting one of important epigenetic marks. The effect of lysine methylation is dependent on the position and the degree (mono-, di-, and trimethylation) of modification, and can be augmented or nullified by (an)other modifications in the same and/or neighboring histone (Kouzarides, 2007). Because over 150 different combinations of lysine methylation and acetylation are known in histone H3 (Garcia et al., 2007), and such modifications are mostly found in the unstructured N terminus of histone H3 (H3t), it is required to prepare peptides with posttranslational modifications of which positions span the entire region of H3t. We here showed that the ribosome could use methylated lysine (KXme, X denotes the number of position) efficiently (regardless of its methylation state) as well as acetylated lysine (KXac). Upon genetic code reprogramming, we could prepare H3t peptides with lysine methylation and acetylation at designated sites in a combinatorial manner. Using this method, we also unveiled possible crosstalk among K9me, K14ac, and K27me of H3t upon binding to chromodomain of heterochromatin protein 1; K9me3 had a positive effect on binding, K27me augmented the positive effect, and the third modification, K14ac, nullified the augmenting effect of K27me only. Similarly, K14ac augmented K9me3-induced binding, and K27me nullified the effect of K14ac on binding. Being rapid and simple, this method would help to decipher the effect of combinatorial modifications in histones, and it can be extended to the investigations of other classes of peptides, such as p53, which contains posttranslational modifications, including lysine methylation (Chuikov et al., 2004).

EXPERIMENTAL PROCEDURES

Preparation of the Template DNA

A DNA sequence of the human H3.1 was referred to for the PCR preparation of the H3t peptide template with the addition of a C-terminal Flag tag. Briefly, K9F

(5′-CAGAC TGCCC GCTTC TCGAC CGGTG GTAAA GCA-3′) was annealed to Linker (5′-GCGAG CGGCT TTTGT AGCCA GTTGC TTCCT GGGTG CTTTA CC ACC GG-3′) and extended by Taq DNA polymerase. The extension mixture was diluted 20 times into the PCR mixture and amplified using K4I (5′-AAGAA GGAGA TATAC ATATG GCTCG TACAA TCCAG ACTGC CCGC-3′) and K27N (5′-CACCC CTCCA GTAGA GGGCG CACTG TTGCG AGCGG CTTTT GT-3′) as the 5′ and 3′ primer respectively. Similarly, the product was further extended twice using 5′T7 (5′-GTAAT ACGAC TCACT ATAGG GTTTA ACTTT AAGAA GG AGA TATAC AT-3′) and K36W (5′-CTTGT CGTCA TCGTC TTTGT AGTCA GGT TT CCACA CCCCT CCAGT AGA-3′) as the first primer pair, and 5′T7 and 3′FLAG (5′-CGAAG CTTAC TTGTC GTCAT CGTCT TTGTA-3′) as the second primer pair.

Syntheses of *K-DBEs

N^{α}-Boc-N^{ε}-methyl-L-lysine was synthesized as an N^{ε}-Boc derivative from N^{α}-Boc-L-lysine by (i) benzylation using benzaldehyde and $NaBH_4$, (ii) methylation using formaldehyde and $NaBH_3CN$, (iii) debenzylation with hydrogenation using Pd/C, and (iv) protection of the ε-methylamine with Boc group using Boc anhydride according to the reported procedure (Andruszkiewicz, 1988). Similarly, N^{α}-Boc-N^{ε},N^{ε}-dimethyl-L-lysine was synthesized from N^{α}-Boc-L-lysine using formaldehyde and $NaBH_4$ for the reductive amination. N^{α}-Boc-N^{ε},N^{ε},N^{ε}-trimethyl-L-lysine was synthesized using methyl iodide as a methylating reagent to quaternize the ε-amine of N^{α}-Boc-L-lysine according to the reported procedure (Chen and Benoiton, 1976).

Carboxylic acids of N^{α}-Boc-L-lysine derivatives were activated to 3,5-dinitrobenzyl esters, and finally, N^{α}-protecting groups were removed as described previously (Murakami et al., 2006).

Aminoacylation of $tRNA^{Asn\text{-}E2}$ with Amino Acids

$tRNA^{Asn\text{-}E2}$ and flexizyme dFx were prepared by the runoff transcription of appropriate templates as described previously (Murakami et al., 2006). Acceptor stem sequences of tRNA were changed from authentic *E. coli* $tRNA^{Asn}$ to enhance transcription (T1G) and the orthogonality toward aminoacyl-tRNA synthetases (C2G). Aminoacylation was accomplished by incubating heat-denatured/renatured tRNA with equimolar of flexizyme dFx in the presence of 0.1 M HEPES-KOH (pH 7.5), 0.6 M $MgCl_2$, and 5 mM *K-DBE for 3 hr on ice. Reactions were quenched by acidifying the reaction with NaOAc (pH 5.2), and aminoacylated tRNAs were recovered by repeated ethanol precipitations.

Translation

The PURE system (Shimizu et al., 2001) was purchased from Post Genome Institute Company, Ltd. (Tokyo, Japan) and was used according to the manufacturer's guide. The *w*PURE system contained all the aminoacyl-tRNA synthetases, but contained only 13 amino acids (C, E, F, H, I, N, and W were withdrawn, where C, cysteine; E, glutamate; H, histidine). When necessary, one or more of these amino acids were added to the translation mixture. Usually, a 3~5 μl scale reaction with the addition of 0.2 mM of amino acids of choice was used for analysis. [^{14}C]D was added to the reaction for the gel analysis of the product. Each aminoacylated $tRNA^{Asn\text{-}E2}$ was added to a final concentration of 50 μM as a mixture with the flexizyme dFx: so that the total concentration of exogenous RNAs in the translation was 100 μM.

Analyses of the Ribosomally Synthesized H3t

*H3t peptides from the 5 μl scale translation reaction were immobilized on anti-Flag-M2 agarose (Sigma-Aldrich; St. Louis, MO) by incubating 1 hr in TBS (50 mM 0.1 M HEPES-KOH, 150 mM NaCl [pH 8.0]) with rotation at room temperature. After washing the resin with TBS briefly, peptides were eluted from the resin with 0.2% TFA by incubating 30 min at room temperature. Peptides in the eluate were bound to C18 resin (ZipTip; Millipore; Billerica, MA), washed and desalted with 0.1% TFA, and eluted directly onto the MALDI target plate with saturated matrix (*R*)-cyano-4-hydroxycinnamic acid (Bruker Daltonics; Billerica, MA) in the 1:1 mixture of acetonitrile and 0.2% TFA. Average molecular masses were recorded using AutoflexII® (Bruker) in a linear positive mode. The instrument was calibrated externally with peptide and protein standards (Bruker).

Preparation of chromoHP1

The chromodomain (residues 17–76) of *Drosophila* HP1 (SWISS-PROT accession code P05205) was subcloned into *Nco* I/BamHI sites of pET42a vector (Novagen; Madison, WI) and expressed in *E. coli* strain BL21(DE3) with an N-terminal GST and hexa-His tags. The chromodomain alone was shown to be active enough to form a complex with synthetic H3t peptide (Jacobs and Khorasanizadeh, 2002). chromoHP1 in the form of N-terminal GST-His6 fusion was purified by Talon metal affinity resins (Clontech; Mountain View, CA) and dialyzed into binding buffer (20 mM imidazole, 25 mM NaCl, 2 mM DTT [pH 7.6]).

Pull-Down Analysis

A 15 μl scale translation reaction was primed with the template encoding H3t shown in Figure 4A and one of K- or *K-$tRNA^{Asn\text{-}E2}_{GAA}$ in the absence of phenylalanine in the translation system. A translation without aminoacyl-$tRNA^{Asn\text{-}E2}_{GAA}$ was prepared as a negative control for the pull-down analysis. *H3ts from the translation reaction were immobilized on 2.5 μl of the anti-Flag-M2 agarose by incubating 1 hr in TBS with rotation at room temperature. After brief washing with TBS, the resin was blocked with 3% BSA in TBS (1 hr at room temperature with rotation), washed again with TBS, and the resin immersed in 3 μl of 25 μM chromoHP1 in binding buffer for 1 hr at room temperature. The resin was washed twice with washing buffer (20 mM imidazole, 150 mM NaCl, 0.05% Tween20 [pH 7.5]) and once again with low Tris buffer (4 mM Tris-HCl, 10 mM NaCl [pH 8.0]). Elution was performed twice (10 min) using 0.1 M Gly-HCl (pH 3.0), and combined eluates were dried under reduced pressure and analyzed by western blot. Both anti-polyhistidine (Sigma) and anti-Flag tag antibodies (Sigma) were used in the same blot to visualize chromoHP1 and *H3t peptides, respectively.

In Vitro Binding Analysis by SPR

SPR assays were performed on a Biacore 2000. The instrument was maintained at 15°C, and the flow rate was 20 μl/min throughout the assay. Anti-Flag M2 antibody (Sigma) was diluted to 45 μg/ml in 10 mM sodium acetate (pH 5.0) (GE Healthcare; Buckinghamshire, England) and covalently immobilized on a sensor chip CM5 (GE Healthcare) using the Amine Coupling Kit (GE Healthcare). Translation reactions were diluted 100 times in HBS-EP (GE Healthcare) and captured on anti-Flag tag antibody-coated flow cell for 2 min. Then, varying concentrations of chromoHP1 (specified in figure legends) were injected and allowed to interact with antibody-bound *H3t for 2 min, resonance units in equilibrium were normalized by captured amounts, and molecular weights (ΔReq) were evaluated. Ten millimolars Glycine-HCl (pH 2.0; GE Healthcare) was used to regenerate the antibody-coated sensor chip.

SUPPLEMENTAL DATA

Supplemental Data include four figures and can be found with this article online at http://www.chembiol.com/cgi/content/full/15/11/1166/DC1/.

ACKNOWLEDGMENTS

acK-DBE was a kind gift from H. Murakami in the Suga Group. We thank P.C. Reid for critical proofreading. This work was supported by a grant from the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) and a research and development project in the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO) to H.S.

Received: August 22, 2008
Revised: September 25, 2008
Accepted: September 29, 2008
Published: November 21, 2008

REFERENCES

Andruszkiewicz, R. (1988). Synthesis of N-omega-*tert*-butoxycarbonyl, N-alpha-methyldiamino acids. Pol. J. Chem. *62*, 257–261.

Bonenfant, D., Towbin, H., Coulot, M., Schindler, P., Mueller, D.R., and van Oostrum, J. (2007). Analysis of dynamic changes in post-translational modifications of human histones during cell cycle by mass spectrometry. Mol. Cell. Proteomics *6*, 1917–1932.

Chen, F.C.M., and Benoiton, N.L. (1976). New method of quaternizing amines and its use in amino-acid and peptide chemistry. Can. J. Chem. *54*, 3310–3311.

Chuikov, S., Kurash, J.K., Wilson, J.R., Xiao, B., Justin, N., Ivanov, G.S., McKinney, K., Tempst, P., Prives, C., Gamblin, S.J., et al. (2004). Regulation of p53 activity through lysine methylation. Nature *432*, 353–360.

Fischle, W., Wang, Y.M., Jacobs, S.A., Kim, Y.C., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 bv Polvcomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Forster, A.C., Tan, Z.P., Nalam, M.N.L., Lin, H.N., Qu, H., Cornish, V.W., and Blacklow, S.C. (2003). Programming peptidomimetic syntheses by translating genetic codes designed de novo. Proc. Natl. Acad. Sci. U.S.A. *100*, 6353–6357.

Garcia, B.A., Pesavento, J.J., Mizzen, C.A., and Kelleher, N.L. (2007). Pervasive combinatorial modification of histone H3 in human cells. Nat. Methods *4*, 487–489.

Hirota, T., Lipp, J.J., Toh, B.H., and Peters, J.M. (2005). Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature *438*, 1176–1180.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080–2083.

Jacobs, S.A., Taverna, S.D., Zhang, Y.N., Briggs, S.D., Li, J.M., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. EMBO J. *20*, 5232–5241.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Josephson, K., Hartman, M.C.T., and Szostak, J.W. (2005). Ribosomal synthesis of unnatural peptides. J. Am. Chem. Soc. *127*, 11727–11735.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lachner, M., O'Carroll, N., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Latham, J.A., and Dent, S.Y.R. (2007). Cross-regulation of histone modifications. Nat. Struct. Mol. Biol. *14*, 1017–1024.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat. Rev. Mol. Cell Biol. *6*, 838–849.

Mateescu, B., England, P., Halgand, F., Yaniv, M., and Muchardt, C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO Rep. *5*, 490–496.

Matthews, A.G.W., Kuo, A.J., Ramon-Maiques, S., Han, S.M., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature *450*, 1106–1110.

Meehan, R.R., Kao, C.F., and Pennings, S. (2003). HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. EMBO J. *22*, 3164–3174.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural potypeptide synthesis. Nat. Methods *3*, 357–359.

Peterson, C.L., and Laniel, M.A. (2004). Histones and histone modifications. Curr. Biol. *14*, R546–R551.

Seet, B.T., Dikic, I., Zhou, M.-M., and Pawson, T. (2006). Reading protein modifications with interaction domains. Nat. Rev. Mol. Cell Biol. *7*, 473–483.

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.

Shogren-Knaak, M.A., Fry, C.J., and Peterson, C.L. (2003). A native peptide Ligation strategy for deciphering nucleosomal histone modifications. J. Biol. Chem. *278*, 15744–15748.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844–847.

Simon, M.D., Chu, F., Racki, L.R., de la Cruz, C.C., Burlingame, A.L., Panning, B., Narlikar, G.J., and Shokat, K.M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. Cell *128*, 1003–1012.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Taverna, S.D., Ueberheide, B.M., Liu, Y., Tackett, A.J., Diaz, R.L., Shabanowitz, J., Chait, B.T., Hunt, D.F., and Allis, C.D. (2007). Long-distance combinatorial linkage between methylation and acetylation on histone H3 N termini. Proc. Natl. Acad. Sci. U.S.A. *104*, 2086–2091.

Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2006). Mass spectrometric characterization of human histone H3: a bird's eye view. J. Proteome Res. *5*, 240–247.

Winter, S., Simboeck, E., Fischle, W., Zupkovitz, G., Dohnal, I., Mechtler, K., Ammerer, G., and Seiser, C. (2008). 14-3-3 proteins recognize a histone code at histone H3 and are required for transcriptional activation. EMBO J. *27*, 88–99.

DOI: 10.1002/cbic.200800439

Polymerization of α-Hydroxy Acids by Ribosomes

Atsushi Ohta,[a] Hiroshi Murakami,[b] and Hiroaki Suga*[a, b]

Over 30 years ago, Fahnestock and Rich reported intriguing data showing the capability of the ribosome to polymerize phenyllactic acid. Although the polymerization was initiated and terminated randomly on polyuridic acids, the given data convincingly suggested that the generated polymer was composed of an approximately 7:3 mixture of phenyllactic acid and phenylalanine. Despite the fact that Fahnestock's conclusion was very likely correct, there have been no reports to follow up the ribosome-catalyzed polymerization of α-hydroxy acids until very recently. At the end of 2007, we reported messenger RNA (mRNA)-directed polyester synthesis by using the new emerging method of genetic-code reprogramming in which α-hydroxy acids with various kinds of side-chains are assigned to arbitrarily chosen codons. In this work, we have achieved the ribosomal synthesis of polyesters with the sequence composition and length in a fully controlled manner according to the sequence of mRNA. This Concept article describes the background of the method development and its application to the synthesis of polyesters.

Introduction

Can ribosomes polymerize an α-hydroxy acid?

Over 30 years ago, Fahnestock and Rich published a report entitled "Ribosome-catalyzed polyester formation".[1] This report describes a landmark experiment in which it was shown that the ribosome is capable of polymerizing a non-α amino acid substrate, phenyllactic acid (F^{lac}). The classic Nirenberg method, which led to decoding the universal genetic code, was modified for their experiment. Instead of translating a synthetic polyuridic (poly-U) acid by using Phe–$tRNA^{Phe}$, [^{14}C]-labeled F^{lac}–$tRNA^{Phe}$ was added to the translation system, which was composed of a cell-free *E. coli* S-100 and minimal organic/inorganic components, such as GTP and buffer (Figure 1 A). The [^{14}C]-F^{lac}–$tRNA^{Phe}$ was chemically prepared by deamination of [^{14}C]-Phe–$tRNA^{Phe}$ by using nitrous acid. After the translation reaction, the polymerized products were precipitated with trichloroacetic acid (TCA) and the radioactivity of the precipitates was counted to quantify the yield of polymer. It was shown that the TCA-insoluble matter was formed only in the presence of all translation components, that is, S-100, poly-U, and GTP, and its recovery yield was nearly 10% of the poly-Phe synthesis. Upon treating the precipitate with alkaline, the resulting products were analyzed by paper electrophoresis. Supposedly, alkaline digestion would cleave ester bonds between F^{lac}–F^{lac}, but not amide bonds such as F^{lac}–Phe or Phe–Phe. In fact, the paper electrophoresis separated the products that originated from F^{lac}–Phe and F^{lac} with a ratio of approximately 3:7. The authors suggested that Phe likely originated in S-100, and thus such a minimal contamination of amino acids could not avoided. Nonetheless, this observation gave indirect but convincing support for the idea that the polymerization of F^{lac} occurred consecutively three or four times followed by random incorporation of Phe into the poly-F^{lac} chain.

Despite the limitations of the analytical techniques available at that time, the data convincingly suggested that the ribosome is capable of catalyzing the polymerization of F^{lac}; however, many technical as well as scientific questions were left unresolved by this demonstration. Regarding technical issues, the polyester formation was only confirmed by indirect evidence, that is, by detecting the [^{14}C]-radioisotope of the alkaline-digested sample of the acid-insoluble precipitants. Therefore, it was still unknown how long F^{lac} could be consecutively polymerized. Ideally, the polyester should be directly detected as an intact polymer. Moreover, due to unavoidable contamination of Phe in S-100, deacylated $tRNA^{Phe}$ generated by the hydrolysis of [^{14}C]-F^{lac}–$tRNA^{Phe}$ would be recharged by phenylalanyl–tRNA synthetase (PheRS). This resulted in the random incorporation of Phe into the poly-F^{lac} chain after every three or four residues. Thus, the contamination-free polymerization of F^{lac} was not, unfortunately, achieved. Moreover, it is of scientific interest whether other types of α-hydroxy acids (α-ha) can be accepted by ribosomes for polymerization. Along the same line, instead of the random initiation and termination demonstrated by Fahnestock's experiment, it is critical to utilize the full capability of the translation system, that is, to show fully controlled initiation, elongation, and termination along the mRNA sequence that leads to the generation of a defined length of polyester. Finally, it would be the most critical challenge to demonstrate the ribosomal synthesis of a variety of polyesters that contain distinct side-chains designated by mRNA templates. To this end, we set a project to challenge the ribosomal polymerization of α-hydroxy acids that would simultaneously address all of the above issues.[2]

[a] *A. Ohta, Prof. H. Suga*
Department of Chemistry and Biotechnology
Graduate School of Engineering, University of Tokyo
113-8656 Tokyo (Japan)
Fax: (+81) 3-5452-5495
E-mail: hsuga@rcast.u-tokyo.ac.jp
[b] *Dr. H. Murakami, Prof. H. Suga*
Research Center for Advanced Science and Technology
University of Tokyo, 153-8904 Tokyo (Japan)

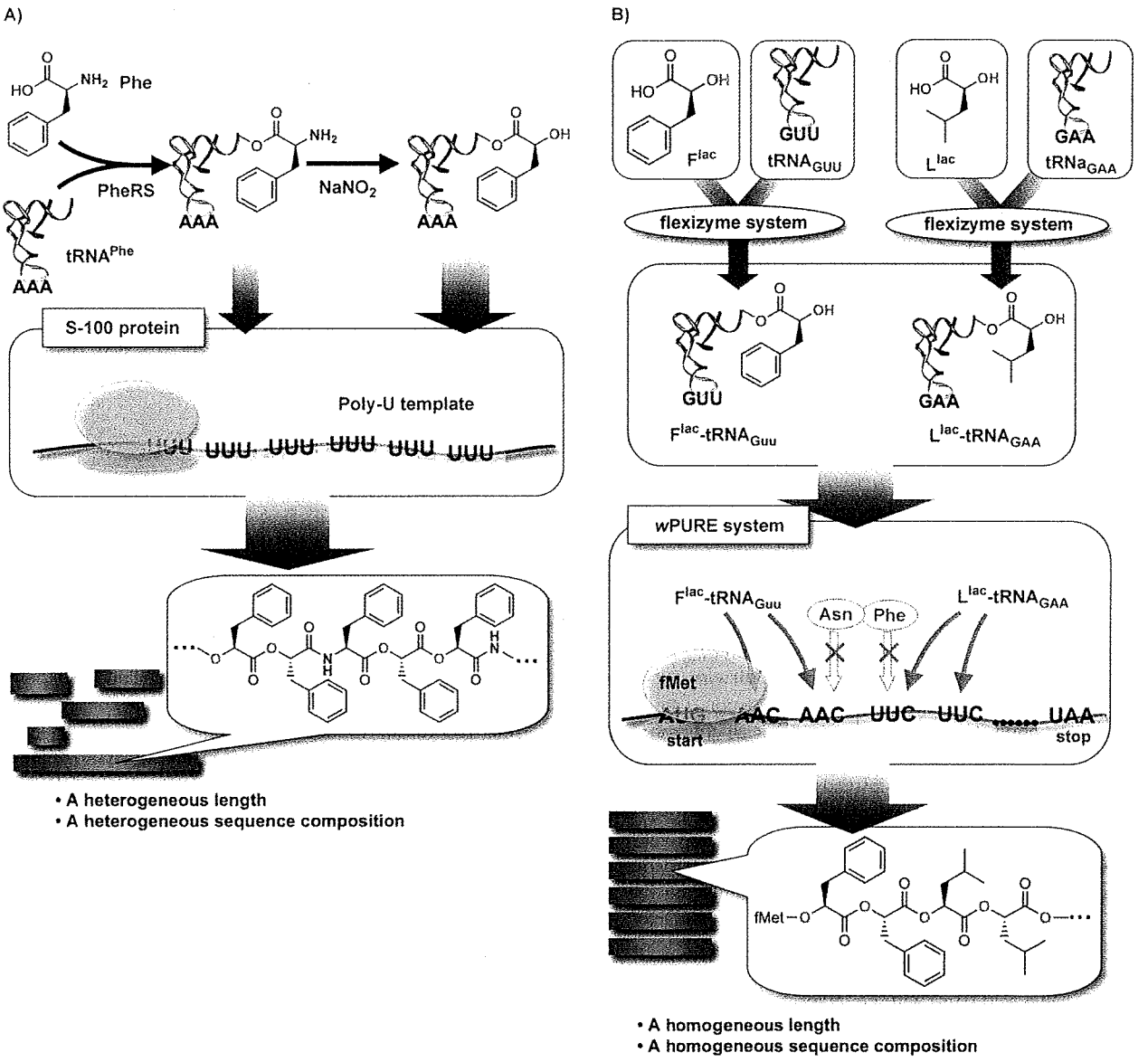


Figure 1. Two different approaches for ribosomal polyester synthesis. A) In Fahnestock's approach, in which F^{lac}–$tRNA^{Phe}$ was prepared by chemical deamination of Phe–$tRNA^{Phe}$, polymerization was randomly initiated and terminated on the poly-U template. Due to the lack of an open-reading frame in the poly-U template, the resulting product had nonhomogenous lengths of polyesters. Moreover, a trace amount of Phe–$tRNA^{Phe}$ resulted in the random incorporation of Phe into the polyester chain; B) mRNA-directed polyester synthesis by using genetic-code reprogramming. The flexizyme system facilitates the hydroxyacylation of tRNAs, and these hydroxyacyl–tRNAs were added to a *w*PURE system for the polymerization of the α-hydroxy acids; L^{lac}: isopropyllactic acid; F^{lac}: phenyllactic acid.

Genetic-Code Reprogramming for Polyester Synthesis

About ten years after Rich's report, a new methodology was developed for introducing nonproteinogenic α-amino acids into peptide chains.[3,4] A mischarged $tRNA_{CUA}$ with a nonproteinogenic amino acid was used to suppress a UAG stop codon (amber codon) and the amino acid was incorporated into a specific site of the peptide chain by using a cell-free translation system. This method is also applicable for the incorporation of α-hydroxy acid[5–9] to generate an ester bond in the peptide chain. In fact, it has been utilized to disrupt the backbone hydrogen-bonding network in a protein of interest and to specifically cleave the peptide chain at the ester site. Unfortunately, the mischarged $tRNA_{CUA}$ must inherently compete with the release factor present in the translation apparatus, that is, translation termination, and therefore the incorporation efficiency heavily depends on the kind of side chains in the α-hydroxy acid. This fact also prohibits us from performing consecutive multiple incorporations of an α-hydroxy acid(s).

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 $tRNA^{Asn}$.

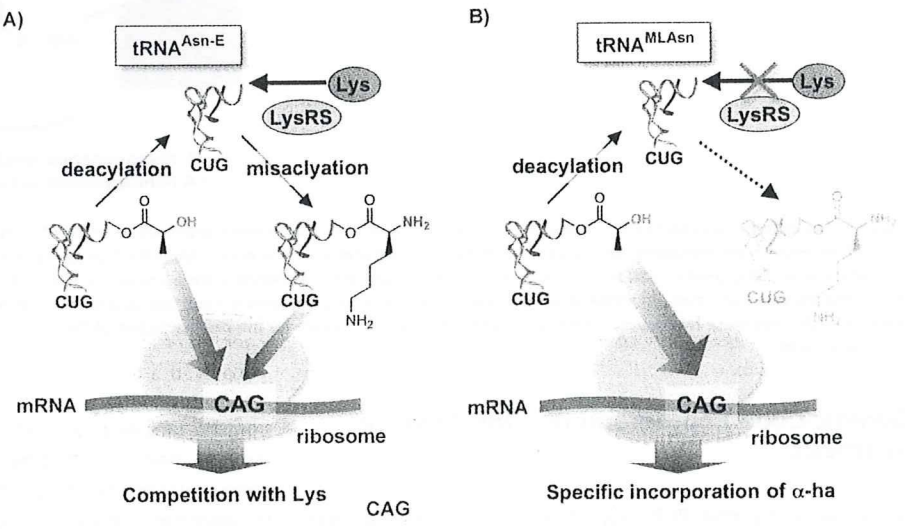


Figure 2. A strategy for controlling the competition with undesirable misincorporation into the polyester chain. A) Competition between α-hydroxy acid and Lys for $tRNA^{Asn\text{-}E}$ at the CAG codon. Because the deacylated $tRNA^{Asn\text{-}E}$ suffered from mis-lysinylation catalyzed by Lys–tRNA 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.

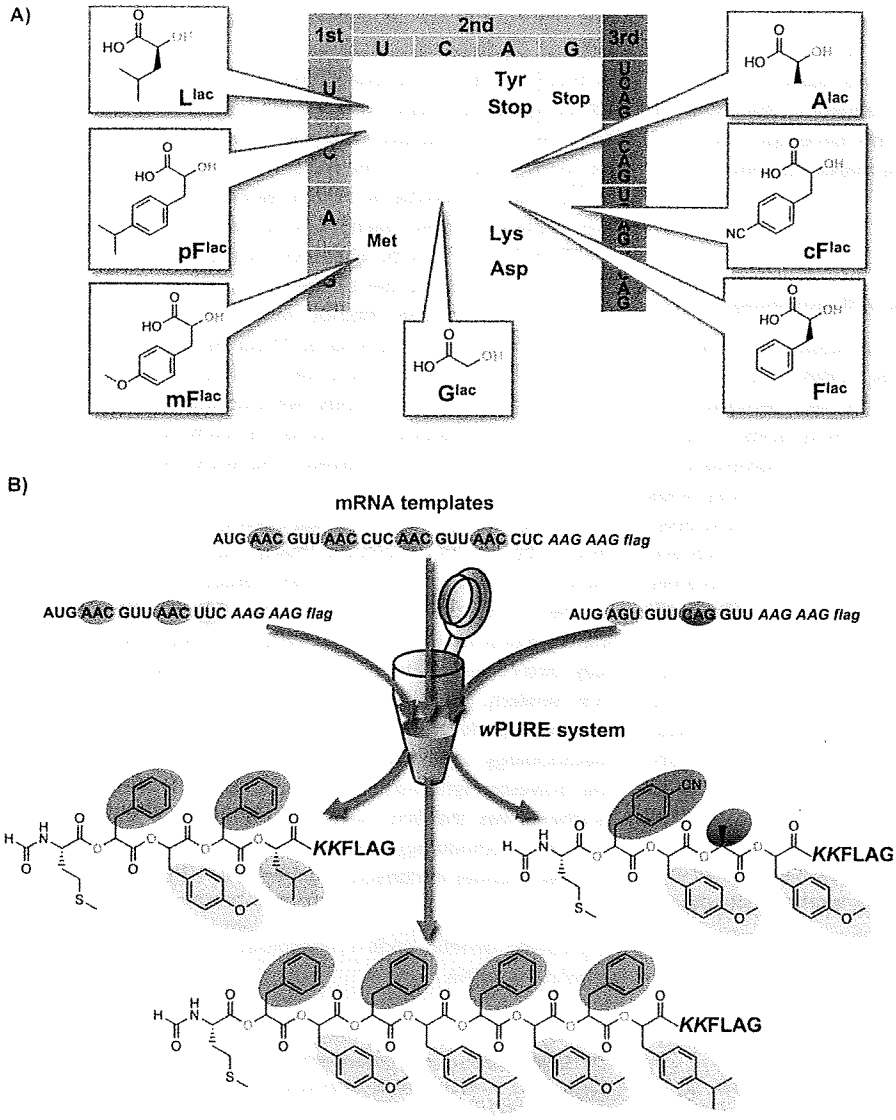


Figure 3. mRNA-directed polyester synthesis by using genetic-code reprogramming. A) The genetic-code reprogrammed by seven kinds of α-hydroxy acids; B) mRNA-directed polyester synthesis. Each template yielded respective polyester sequences under the control of the reprogrammed genetic code shown in A).

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^{lac}–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

nature

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*[1–5]. 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[8,9] 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[5]. 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.

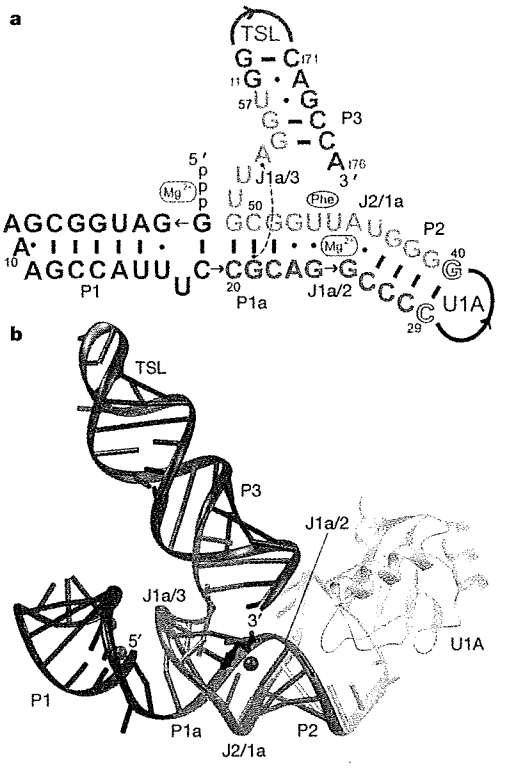


Figure 1 | Overall structure of the flexizyme–tRNA minihelix fusion. **a**, Sequence and secondary structure of the crystallization construct. Binding sites of well-ordered Mg^{2+} ions and of phenylalanine are indicated. The dashed line denotes an A-minor interaction. The minihelix (red) is numbered 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.

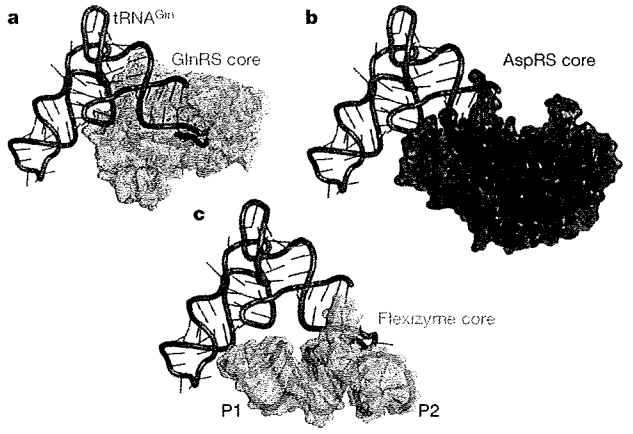
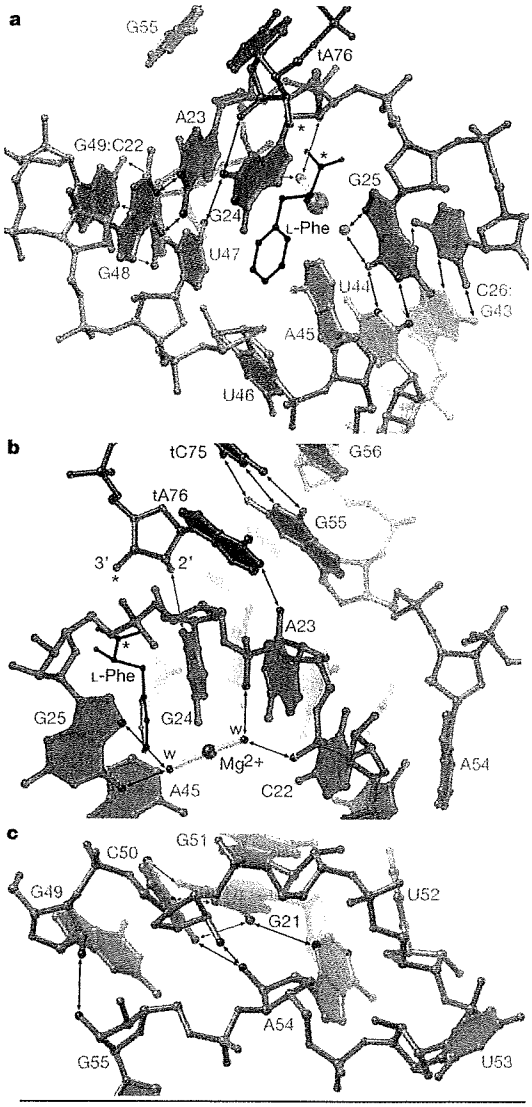


Figure 2 | Comparison of protein and RNA aminoacyl-tRNA synthetases. **a**, Core of *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS; cyan), a class I ARS, bound to tRNA^{Gln} (ref. 16). **b**, Core of *E. coli* aspartyl-tRNA synthetase (AspRS; green), a class II ARS, bound to tRNA[17]. Its cognate 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].

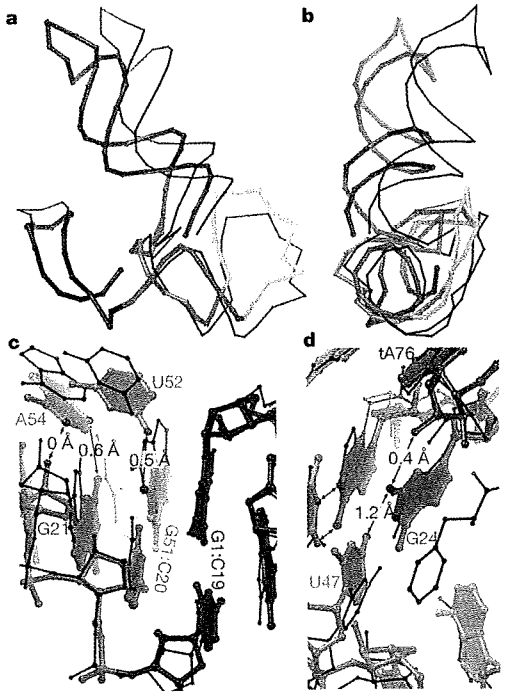


Figure 4 | Conjectural coupling of tRNA docking to active-site folding. **a**, Superposition of the two crystallographically independent flexizyme structures. Throughout, molecule B (undocked conformation) is in black. **b**, Orthogonal view. **c**, Comparison of the interface between J1a/3 and P1a in 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., Ashiai, 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).

METHODS

Protein and RNA preparation. Expression of the selenomethionyl U1A RNA-binding domain (RBD) double mutant (Y31H, Q36R) was as described[31]. Plasmid pMU40 is a pUC19 derivative encoding the flexizyme–minihelix fusion used for structure determination followed by the substrate stem-loop of the VS ribozyme[32]. The TSL (residues t3–t70 in tRNA numbering; Fig. 1a) is 5′-UGGUACGAGGUUCGAAUCCUCGUACCG-3′. The U1A-binding loop (residues 30–39) is 5′-AUUGCACUCC-3′. *In vitro* transcription reactions (310 K, 4 h) contained 75 mg l^{-1} *Bam*HI-linearized pMU40, 50 mg l^{-1} *Ssp*I-linearized pAVA (ref. 32), each NTP at 2.5 mM, 30 mM Tris-HCl pH 8.1, 40 mM $MgCl_2$, 0.1% Triton X-100, 2 mM spermidine, 10 mM dithiothreitol, 1,000 U l^{-1} *E. coli* inorganic pyrophosphatase[33] and 0.1 g l^{-1} T7 RNA polymerase. The VS ribozyme-cleaved transcript[32] was purified by denaturing PAGE, passively eluted into water, and desalted by ultrafiltration. The 2′,3′-cyclic phosphate was removed by treatment of the RNA with PNK (New England Biolabs). Reactions (310 K, 5 h) contained 68 μM RNA, PNK (1 U of enzyme per 0.14 nmol of RNA), 25 mg l^{-1} BSA, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol and 1 mM ATP (Supplementary Fig. 7). After extraction with phenol/chloroform and precipitation with ethanol, the RNA was desalted by ultrafiltration, concentrated to about 0.9 mM and stored at 277 K.

Crystallization and diffraction data collection. RNA (0.74 mM in 10 mM Hepes-KOH pH 7.5, 1 mM EDTA) was snap-cooled by heating to 363 K for 2 min and transferring to ice-water for 5 min. A 1.25-fold molar excess of selenomethionyl U1A-RBD double mutant was added to the RNA, and the solution was adjusted to 10 mM $MgCl_2$, 1 mM cobalt(III) hexammine, 0.6 mM PheEE (Sigma), 1 mM spermine, and the RNA at 0.3 mM. Sitting drops prepared by mixing 1 μl of this solution with 1 μl of a reservoir solution comprising 100 mM magnesium formate pH 7.0, 14.5–15% (w/v) PEG 3000 and 0.8–1.0 M LiCl were equilibrated by vapour diffusion at 295 K. Crystals with the symmetry of space group C2 (unit cell parameters are given in Supplementary Table 1) appeared in days and reached maximum dimensions of $0.3 \times 0.2 \times 0.14$ mm^3 over weeks. Crystals were transferred over 15 min to a solution of the same composition as the reservoir solution, but with the concentration of LiCl increased to 1.8 M, and flash-frozen by plunging into liquid nitrogen. For crystal III, the concentration of PheEE in the cryoprotectant was increased to 2.5 mM (this was limited by poor solubility of PheEE). Diffraction data (Supplementary Table 1) were collected at 100 K (with the inverse-beam method for crystal I) at beamline 5.0.2 of the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory, and reduced with the HKL package[34].

Structure determination and refinement. Six selenium sites were located and their parameters were refined, and two energy MAD phases were calculated with CNS[35] using crystal I data (Supplementary Table 1). The electron-density map resulting from density modification of these phases clearly showed the U1A-RBD and its binding site but did not reliably indicate the path of the RNA chain. The program PHASER[36] was used for molecular replacement with the use of crystal II data and two copies each (six search models) of the U1A-RBD and 15 nucleotides of its binding site, an 8-nucleotide stem-loop comprising a canonical GAAA tetraloop, and a 31-nucleotide stem-loop comprising a canonical tRNA minihelix. The six strongest peaks from an anomalous difference Fourier synthesis calculated with amplitudes from the peak data set of crystal I and model phases from the top PHASER solution coincided with the selenomethionines in the U1A-RBD models of the PHASER solution. Phase combination of the MAD phases and the PHASER model phases produced electron-density maps (Supplementary Fig. 8) that allowed unambiguous tracing of the RNA through iterative rounds of model building[37], simulated annealing, energy minimization and restrained individual isotropic *B*-factor refinement[35]. A maximum-likelihood target was employed for refinement against all structure-factor amplitudes from crystal II, and experimental phase-probability distributions. A solvent mask and an anisotropic *B*-factor correction were used throughout. The solvent mask parameters (solvent density 0.3; solvent *B*-factor 25 $Å^2$) were chosen empirically to minimize the R_{free} of the lower-resolution shells. The final crystallographic model (Supplementary Table 1) comprises two complete RNA molecules (92 residues each), two selenomethionyl U1A-RBDs (residues 1–92 and 1–91), ten Mg^{2+} ions, one K^+ ion and ten water molecules (5,346 non-hydrogen atoms). Ramachandran analysis shows that 86.8% and 12.6% of the protein residues have most favoured and additionally allowed backbone conformations, respectively; there are no residues with a disallowed conformation. All figures, except Fig. 4, are of the docked form of the RNA (molecule A) and were prepared with PyMol[38] or RIBBONS[39]. Refinement against crystal III amplitudes started with the model from crystal II refinement. Waters and ions were stripped from the model, and after iterative rounds of simulated annealing, energy minimization and tightly restrained individual isotropic *B*-factor refinement, ions were placed into the strongest features of the residual $|F_o| - |F_c|$ map. A bulk solvent mask and an anisotropic *B*-factor correction were also used throughout this refinement. The residual Fourier synthesis calculated after final energy minimization and tightly restrained individual *B*-factor refinement of this model presented features in the active site of the docked form of the RNA (chain A) that were consistent with the benzyl group of PheEE (Supplementary Fig. 3). This second model comprises both RNA molecules (92 residues each), two selenomethionyl U1A-RBDs (residues 2–92 and 1–91), ten Mg^{2+} ions, one K^+ ion and no water molecules (5,366 non-hydrogen atoms). Ramachandran analysis of the model from this second refinement shows that 78.9% and 20.5% of the protein residues have most favoured and additionally allowed backbone conformations, respectively; there are no residues with a disallowed conformation.

Mutational and *in vitro* selection analysis. Reselection has been described previously[21]. Activity assays were performed in triplicate as described[21].

31. Ferré-D'Amaré, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* 295, 541–556 (2000).
32. Ferré-D'Amaré, A. R. & Doudna, J. A. Use of *cis*- and *trans*-ribozymes to remove 5′ and 3′ heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res.* 24, 977–978 (1996).
33. Rupert, P. B. & Ferré-D'Amaré, A. R. Crystallization of the hairpin ribozyme: illustrative protocols. *Methods Mol. Biol.* 252, 303–311 (2004).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307–326 (1997).
35. Brünger, A. T. *et al.* Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* 54, 905–921 (1998).
36. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* 61, 458–464 (2005).
37. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119 (1991).
38. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, 2002).
39. Carson, M. Ribbons. *Methods Enzymol.* 277, 493–505 (1997).

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^{a}-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).

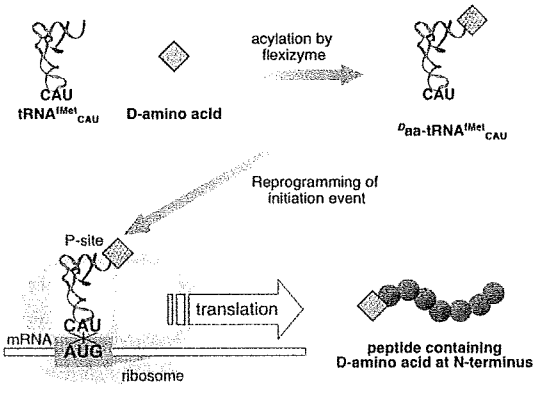


FIGURE 1. Genetic code reprogramming of the initiation event with ^{D}aa. The initiator can be reassigned to ^{D}aa instead of ^{L}Met by the integration of *w*PURE system and flexizyme system. Initiation with ^{D}aa results in the nascent peptide containing ^{D}aa at the N terminus.

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

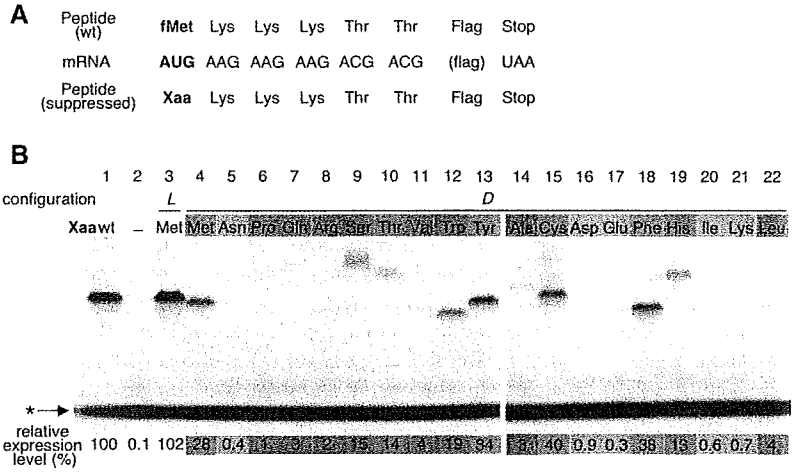


FIGURE 2. Tolerance of various ^{D}aa in initiation. (*A*) The mRNA sequence that expresses peptides initiated by various ^{D}aa. Flag in parentheses indicates the RNA sequence encoding the Flag peptide sequence (DYKDDDDK). (*B*) Tricine-SDS PAGE analysis of the translation products. (Lane *1*) expression 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