JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

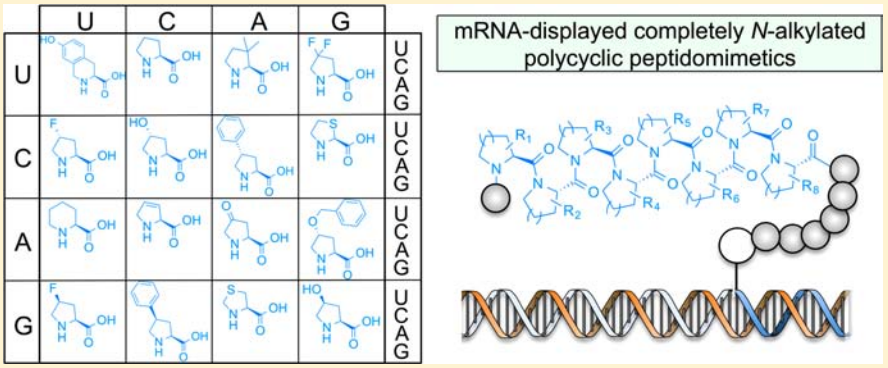
pubs.acs.org/JACS

Extensive Reprogramming of the Genetic Code for Genetically Encoded Synthesis of Highly N-Alkylated Polycyclic Peptidomimetics

Takashi Kawakami,* Takahiro Ishizawa, and Hiroshi Murakami*

Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

Supporting Information



ABSTRACT: Cyclic structures can increase the proteolytic stability and conformational rigidity of peptides, and N-alkylation of the peptide backbone can make peptides more cell-permeable and resistant to proteolysis. Therefore, cyclic N-alkyl amino acids are expected to be useful building blocks to increase simultaneously these pharmacological properties of peptides. In this study, we screened various cyclic N-alkyl amino acids for their ribosomal incorporation into peptides and identified cyclic N-alkyl amino acids that can be efficiently and successively incorporated. We also demonstrated genetic code reprogramming for reassigning 16 NNU codons to 16 different cyclic N-alkyl amino acids with high fidelity to synthesize highly N-alkylated polycyclic peptidomimetics and an mRNA-displayed library of completely N-alkylated polycyclic peptidomimetics by using our recently developed TRAP (transcription/translation coupled with association of puromycin linker) display. In vitro selection from a highly diverse library of such completely N-alkylated polycyclic peptidomimetics could become a powerful means to discover small-molecule ligands such as drug candidates that can be targeted to biomolecules inside living cells.

INTRODUCTION

Peptide selection using an in vitro display technique, such as ribosome display[1,2] or mRNA display,[3,4] is a powerful approach for discovering novel functional peptides from highly diverse peptide libraries. The poor pharmacokinetics of proteinogenic peptides have been improved by ribosomally synthesizing peptides bearing nonproteinogenic structures that increase the proteolytic stability, membrane permeability, and conformational rigidity of the peptide.[5] For example, cyclic structures can increase the proteolytic stability and conformational rigidity of peptides; thus, various posttranslational redox-insensitive macrocyclization strategies have been applied to in vitro peptide selection.[5−7] In fact, it was shown that some in vitro selected cyclic peptides had much higher affinity for their target proteins than their linear counterparts.[8,9]

N-Alkylation of the peptide backbone is also particularly interesting because it can make peptides more membrane-permeable[10−13] and stable to proteolysis.[11,14] To give these pharmacological properties to ribosomally synthesized peptides, the ribosomal synthesis of N-methyl peptides and peptoids has been developed.[15−22] Furthermore, based on the ribosomal N-methyl peptide synthesis, partially N-methylated macrocyclic peptide libraries were created by use of a ribosomal translation system for in vitro display selection.[23,24] Significantly, it was shown that N-methylation and cyclic structure of an in vitro-selected macrocyclic N-methyl peptide were important for its proteolytic resistance.[23] The superior pharmacological properties of the cyclic structure and the N-alkyl peptide structure suggest that cyclic N-alkyl amino acids (CNAs) are extremely useful building blocks that can increase not only the cell permeability and proteolytic stability of peptides, similar to noncyclic N-alkyl amino acids, but also their conformational rigidity.

In addition to the beneficial pharmacological properties of CNAs, the compatibility of diverse CNAs with a ribosomal translation system is also important for the selection of in vivo-functioning peptides containing multiple CNAs. Classical in vitro nonsense suppression using a crude lysate translation system showed that the ribosomal translation machinery

Received: May 22, 2013
Published: July 30, 2013

   dx.doi.org/10.1021/ja405044k | J. Am. Chem. Soc. 2013, 135, 12297−12304

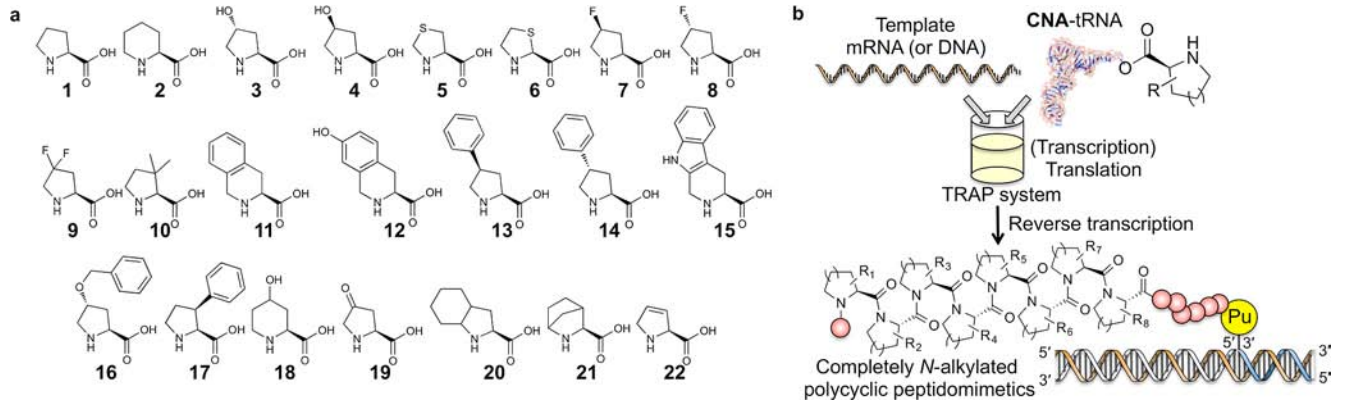


Figure 1. DNA-programmed synthesis of highly N-alkylated polycyclic peptidomimetics by genetic code reprogramming. (a) Chemical structures of CNAs used in this study: **1**, L-proline; **2**, L-pipecolic acid; **3**, *trans*-4-hydroxy-L-proline; **4**, *cis*-4-hydroxy-L-proline; **5**, L-thiazolidine-4-carboxylic acid; **6**, L-thiazolidine-2-carboxylic acid; **7**, *cis*-4-fluoro-L-proline; **8**, *trans*-4-fluoro-L-proline; **9**, 4,4-difluoro-L-proline; **10**, 3,3-dimethyl-L-proline; **11**, L-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid; **12**, L-7-hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid; **13**, *cis*-4-phenyl-L-proline; **14**, *trans*-4-phenyl-L-proline; **15**, L-2-tryptoline-3-carboxylic acid; **16**, *O*-benzyl-*trans*-4-hydroxy-L-proline; **17**, *cis*-3-phenyl-L-proline; **18**, 4-hydroxy-L-pipecolic acid; **19**, 4-oxo-L-proline; **20**, (2*S*,3a*S*,7a*S*)-2,3,3a,4,5,6,7,7a-octahydroindole-2-carboxylic acid; **21**, (1*R*,3*S*,4*S*)-2-azabicyclo[2.2.1]heptane-3-carboxylic acid; **22**, 3,4-dehydro-L-proline. (b) mRNA-templated synthesis of a library of completely N-alkylated polycyclic peptidomimetics displayed on mRNAs via a puromycin (Pu)−DNA linker by using various CNA-tRNAs in the TRAP system. Completely N-alkylated polycyclic peptidomimetics translated from mRNA were spontaneously displayed on their encoding mRNAs in the TRAP system.

accepts several CNAs, such as **2**, **3**, and **4** (Figure 1a), as its substrates.[25] The ribosomal incorporation of CNAs (**7**, **8**, and **22**) charged onto tRNA by wild-type ProRS in vivo has been also demonstrated.[26−28] However, these approaches did not give "intrinsic" quantitative translation compatibility, that is, incorporation efficiency of the CNAs, because of competing translation events such as termination by endogenous release factor 1 (RF1) or natural proline incorporation.

A new strategy, called genetic code reprogramming, has been developed to completely reassign a codon to a nonproteinogenic amino acid, which avoids the competition caused by endogenous translation components.[29−31] Genetic code reprogramming is achieved by using a reconstituted cell-free translation system[32,33] in which some competing components, such as amino acids, aminoacyl-tRNA synthetases (aaRSs), tRNAs, and RFs, are omitted. Another advantage of genetic code reprogramming is that multiple (more than three) distinct nonproteinogenic amino acids can be simultaneously incorporated into ribosomally synthesized peptides.[5,34−36]

The ribosomal incorporation of nonproteinogenic CNAs by use of a reconstituted cell-free translation system has been investigated by two groups. In 2007, Forster and co-workers[17] reported that 3-*trans*-hydroxyproline (**3**) chemoenzymatically charged onto tRNAs was cotranslationally incorporated into peptides with an efficiency comparable to that for alanine and phenylalanine. Significantly, they also suggested that CNAs (proline and 3-*trans*-hydroxyproline) are better suited than noncyclic N-alkyl amino acids (N-methyl amino acids and N-butyl amino acids) for incorporation into peptides by the translation apparatus. However, they only demonstrated the single incorporation of 3-*trans*-hydroxyproline; multiple nonproteinogenic CNA-tRNAs were not investigated for ribosomal translation,[37] perhaps because of the synthetic difficulty of chemoenzymatic tRNA aminoacylation.[38]

Szostak and co-workers have shown that four CNAs, including thiazolidine-2-carboxylic acid (**6**), thiazolidine-4-carboxylic acid (**5**), and 3,4-dehydroproline (**22**), are ProRS substrates[39] and are excellent substrates for the translation apparatus.[34] In addition, thiazolidine-4-carboxylic acid (**5**) was successfully used as a building block in mRNA display selection, which demonstrates the availability of the nonproteinogenic CNA for in vitro peptide selection.[9] However, it is not possible to simultaneously incorporate multiple distinct CNAs at different codons by this ProRS-catalyzed tRNA aminoacylation method because the CNAs can be incorporated only at Pro codons. Moreover, because several CNAs were not aaRS substrates,[39] a translation assay with many CNAs was limited by the number of CNAs that are compatible with the aaRS enzymes.

Hence, although these pioneering studies by the two groups demonstrated the single incorporation of a limited number of CNAs, comprehensive compatibility screening and the multiple incorporation of many different CNAs have not yet been reported because of the limitations in the used tRNA aminoacylation methods.

In this study, we comprehensively screened various CNAs for their ribosome-catalyzed incorporation into peptides by using flexible tRNA-aminoacylation ribozymes (flexizymes)[31,40−42] and identified CNAs that can be efficiently and successively incorporated into the backbone of peptides. Then, we investigated the fidelity of the reprogrammed genetic code in which 16 NNU codons were reassigned to 16 different CNAs by expressing diverse highly N-alkylated polycyclic peptidomimetics in a template DNA-dependent manner. Finally, we determined the display efficiency of a random library of completely N-alkylated polycyclic peptidomimetics on their encoding mRNAs by using our recently developed in vitro TRAP (transcription/translation coupled with association of puromycin-linker) display[43] (Figure 1b).

EXPERIMENTAL SECTION

Materials. Preparation of the translation system, DNA templates, and aminoacyl-tRNAs is described in the Supporting Information.

Ribosomal Synthesis of Peptides Containing Single Cyclic N-Alkyl Amino Acids. A translation reaction mixture containing 0.04 μM DNA template; 0.5 mM each Met, Tyr, and Lys; 50 μM [^{14}C]Asp; 0.03 μM MetRS; 0.02 μM TyrRS; 0.11 μM LysRS; 0.13 μM AspRS; and 200, 100, 50, 25, or 12.5 μM cyclic N-alkyl aminoacyl-tRNA$^{\text{Asn-E2}}_{\text{GGA}}$ was incubated for 60 min at 37 °C. The products were analyzed by

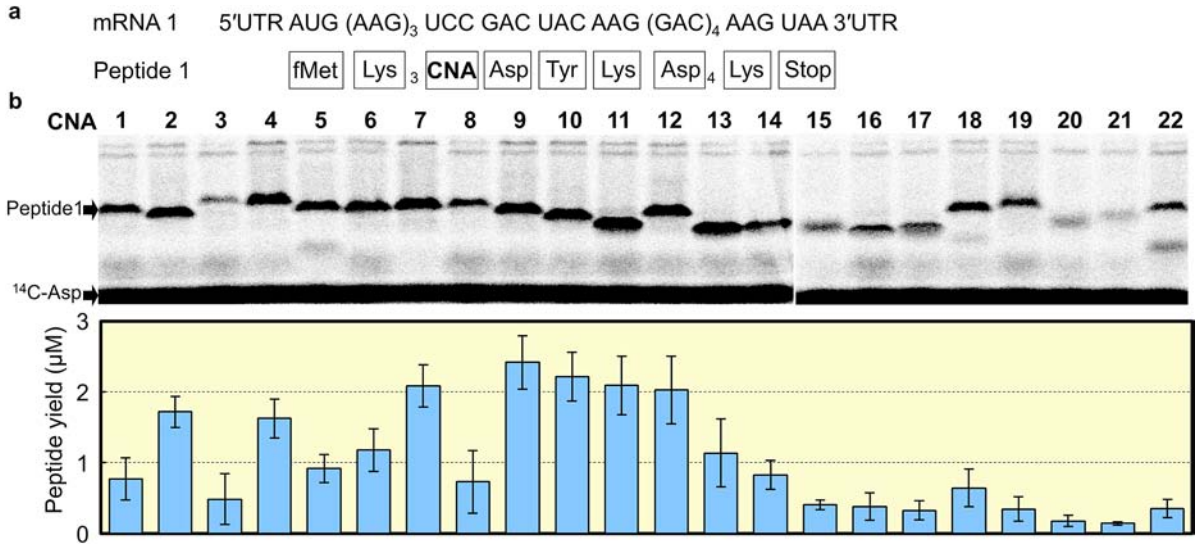


Figure 2. Single incorporation of CNAs. (a) Sequences of mRNA 1 and peptide 1 encoded by mRNA 1. The CNAs were reassigned to the vacant UCC codon. (b) Tricine SDS−PAGE analysis of the expressed peptides labeled with [^{14}C]Asp detected by autoradiography. The peptides were expressed in the presence of 100 μM of the designated cyclic N-alkyl aminoacyl-tRNA$^{Asn\text{-}E2}_{GGA}$ prepared by flexizymes. The yield of each peptide, based on its observed radioisotope count, is shown in the graph. Error bars represent the standard deviation calculated from the experiments, which were performed in triplicate.

tricine sodium dodecyl sulfate−polyacrylamide gele electrophoresis (SDS−PAGE) and autoradiography (Pharox FX, Bio-Rad). For matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) mass analysis, the reaction was carried out with Asp instead of [^{14}C]Asp and 100 μM cyclic N-alkyl aminoacyl-tRNA$^{Asn\text{-}E2}_{GGA}$, and the translation product was desalted with C-TIP (Nikkyo Technos), eluted with 80% acetonitrile and 0.5% acetic acid saturated with CHCA, and analyzed on an Autoflex II (Bruker Daltonics) operated in the linear positive mode.

Ribosomal Synthesis of Di-N-alkyl Bicyclic Peptides, Tri-N-alkyl Tricyclic Peptides, and Tetra-N-alkyl Tetracyclic Peptides. A translation reaction mixture containing 0.04 μM DNA template; 0.5 mM each Met, Tyr, and Arg; 0.03 μM MetRS; 0.02 μM TyrRS; 0.03 μM ArgRS; 50 μM of the respective cyclic N-alkyl aminoacyl-tRNA$^{Asn\text{-}E2}$, and 5 μM each tRNA$^{fMet}_{CAU}$, tRNA$^{Tyr}_{CUA}$, and tRNA$^{Arg}_{CCG}$ instead of native base-modified tRNA mixtures was incubated for 60 min at 37 °C. MALDI-TOF mass analysis was performed as described above.

Streptavidin Pull-down of Biotinylated Random Peptide/mRNA/cDNA Complexes for Display Efficiency Analysis of Proteinogenic Peptides and N-Alkyl Polycyclic Peptides. A random mRNA library was translated to a biotinylated N-alkyl polycyclic peptide library in the TRAP system lacking native base-modified tRNA mixtures and containing 0.5 mM each Ser and Trp, 0.04 μM SerRS, 0.03 μM TrpRS, 2.5 μM mRNA library, 2.5 μM puromycin−DNA linker, 20 μM biotinyl-Phe-tRNA$^{fMet}_{CAU}$, 10 μM elongation factor P (EF-P), 10 μM each 16 cyclic N-alkyl aminoacyl-tRNA$^{Asn\text{-}E2}$, and 5 μM each tRNA$^{Ser}_{CAU}$ and tRNA$^{Trp}_{CCA}$ for 25 min at 37 °C. For the proteinogenic peptide library, the TRAP system lacking Met, Gln, Lys, and Glu as amino acids and 20 kinds of aaRSs, 2.5 μM mRNA library, 2.5 μM puromycin−DNA linker, 20 μM biotinyl-Phe-tRNA$^{fMet}_{CAU}$, native base-modified tRNA mixtures, and 5 μM tRNA$^{Ser}_{CAU}$ as tRNAs was used instead. After the dissociation of ribosomes with ethylenediaminetetraacetic acid (EDTA), reverse transcription of the mRNA library was performed with an appropriate primer (Table S6 in Supporting Information) and RNase H-inactivated reverse transcriptase. After quenching the reverse transcription reaction with EDTA and neutralizing the solution with *N*-(2-hydroxyethyl)piperazine-*N*′-ethanesulfonic acid (HEPES), cDNA/mRNA complexes displaying biotinylated peptides were recovered with streptavidin-coated magnetic beads and quantified by quantitative polymerase chain reaction (PCR). Alternatively, these recovered cDNAs as well as the reverse-transcribed cDNAs from the initial random mRNA library were sequenced by high-throughput DNA sequencing on an Ion Torrent 318 Chip (Life Technologies) according to the manufacturer's protocol.

RESULTS

Screening of Cyclic N-Alkyl Amino Acids Used for Genetic Code Reprogramming. In the two previous studies of nonproteinogenic amino acid mutagenesis, CNAs bearing a positively or negatively charged side chain were poor substrates for ribosomal translation.[44,45] In our previous studies, we also found that noncyclic N-alkyl amino acids (N-methyl amino acids and N-substituted glycines) bearing a charged side chain cannot be ribosomally incorporated into peptides.[18,19] Therefore, we decided to exclude CNAs with a charged side chain from translation compatibility screening. In addition, because chemical synthesis is usually required to obtain large quantities of peptides after in vitro selection, we considered the availability of the corresponding fluorenylmethyloxycarbonyl (Fmoc) building blocks when we chose the CNAs for screening. On the basis of these criteria, 21 commercially available nonproteinogenic CNAs were selected (Figure 1a), including 13 CNAs that had not been explored previously in any translation system.

All 21 CNAs were chemically derivatized to 3,5-dinitrobenzyl esters or cyanomethyl esters in order to convert them to the corresponding flexizyme substrates. The flexizyme-catalyzed tRNA-aminoacylation conditions for the CNAs were optimized by using these activated ester derivatives and a microhelix RNA, which is a tRNA analogue, prepared by in vitro transcription. The product yields of the CNA−microhelix RNAs were quantified by separation on acid-denaturing PAGE, showing that all of the CNAs could be charged onto microhelix RNA with yields of over 25% under optimized conditions (Figure S1 in Supporting Information). The yields were sufficient for the translation assay, according to previous studies of flexizyme-based translation substrate screening.[18,19]

Next, an engineered orthogonal *Escherichia coli* tRNA^{Asn}-based tRNA (tRNA$^{Asn\text{-}E2}$)[35] bearing a GGA anticodon was aminoacylated with the CNAs by the flexizymes under

dx.doi.org/10.1021/ja405044k | J. Am. Chem. Soc. 2013, 135, 12297−12304

optimized conditions. The CNA-tRNA$^{Asn\text{-}E2}_{GGA}$ products were assayed by their single incorporation at the UCC codon directed by DNA encoding a model peptide (Figure 2a). Peptide expression was performed in the reconstituted transcription/translation coupling system that contained only four proteinogenic amino acids (Met, Lys, Asp, and Tyr) and their cognate aaRSs. For comparison, the same DNA was transcribed and translated with Pro-tRNA$^{Asn\text{-}E2}_{GGA}$ prepared from proline-3,5-dinitrobenzyl ester and the flexizyme. The peptide yields were calculated based on the incorporation of [^{14}C]Asp, which was located downstream of the CNA residue in the model peptide. The 60 min translation reaction was terminated by the addition of SDS, and the peptide products were separated from [^{14}C]Asp by tricine SDS−PAGE and detected by autoradiography.

Figure 2b shows that nearly half of the CNAs were incorporated into the peptide with a higher efficiency than the proteinogenic proline. Although the remaining CNAs were less efficiently incorporated, the concentration of all the expressed peptides that contained CNAs was still more than 0.1 μM in the presence of 100 μM of the corresponding CNA-tRNA. Titrating the concentration of the CNA-tRNAs in the translation reaction showed that a concentration of around 100 μM was adequate for incorporation of the CNAs (Figure S2 in Supporting Information). We also found that less translation-compatible CNAs generally possessed more bulky or branched side chains (Figures 1a and 2b; CNAs **15**, **16**, **17**, **20**, and **21**). A similar tendency was observed for the noncyclic N-alkyl amino acids (N-methyl amino acids and N-alkyl glycines) in our previous studies.[18,19] Thus, the efficient incorporation of these N-alkyl amino acids bearing bulky or branched side chains may be hampered by the same mechanism caused by ribosome and/or elongation factor Tu. More importantly, we found that 13 previously unexamined CNAs were compatible with a ribosomal translation system.

Subsequent MALDI-TOF mass spectrometry (MS) analysis of the desalted translation products synthesized in the presence of CNA-tRNA confirmed that all of the CNAs were ribosomally incorporated into the peptide (Figure S3 in Supporting Information). However, in the case of CNAs **15** and **21**, C-terminal truncated peptides were also detected as side products. These truncated peptides were probably produced by the following mechanism. The fMet-(Tyr)$_3$-tRNA dropped off the ribosomal peptidyl-tRNA binding site (P-site) before its peptidyl transfer to the CNA-tRNA in the aminoacyl-tRNA binding site, presumably because of its poor peptidyl-transfer efficiency. However, translocation of the CNA-tRNA to the P-site could have occurred, and thus the C-terminal truncated peptide could be produced by continuous peptide elongation.

To test the incorporation of the CNAs more rigorously, we prepared a DNA encoding an fMet-(Tyr)$_3$-(CNA)$_2$-Arg-(Tyr)$_3$ peptide to check for the successive double incorporation of CNAs at two UCC codons (Figure S4 in Supporting Information). The translation system for expressing the peptide was constructed from three proteinogenic amino acids (Met, Tyr, and Arg) and their cognate aaRSs. MALDI-TOF-MS analysis of the desalted fMet-(Tyr)$_3$-(CNA)$_2$-Arg-(Tyr)$_3$ translation products revealed that 15 CNAs (**2**−**10**, **12**−**14**, **16**, **19**, and **22**) were successively incorporated into the peptide (Figure S4 in Supporting Information). The successive double incorporation of CNAs **11** and **18** was not detected, and only single incorporation of the CNAs at the fifth position was observed instead. This was perhaps caused by the following mechanism, which is similar to ribosome hopping.[46] fMet-(Tyr)$_3$-CNA-tRNA$_{GGA}$ dissociated from the first UCC codon, and then the ribosome/fMet-(Tyr)$_3$-CNA-tRNA$_{GGA}$ complex traversed the downstream mRNA. After the pairing between the adjacent second UCC codon and GGA anticodon of fMet-(Tyr)$_3$-CNA-tRNA$_{GGA}$ was re-established, regular translation resumed to produce the peptide containing a single CNA. This was unexpected because CNA **18** and, particularly, CNA **11** were efficiently incorporated singly into the peptide; however, similar results were observed during the incorporation of D-amino acids in our recent study.[47] Incorporation of the four other CNAs (**15**, **17**, **20**, and **21**) was undetectable, perhaps because of their relatively low incorporation efficiency even for single incorporation. Therefore, the 15 successively incorporated CNAs were used together with proline for all subsequent experiments.

Reprogramming the Genetic Code for Synthesis of Highly N-Alkylated Polycyclic Peptidomimetics. We investigated the fidelity of the reprogrammed genetic code in which 16 NNU codons were reassigned to the 16 distinct CNAs. A major concern for correctly decoding the 16 NNU codons to the corresponding CNAs is misreading of the codons through G−U base pairing instead of the correct A−U and G−C pairs at the first or second codon base (Figure S5 in Supporting Information).[48−52] Thus, we designed mRNAs containing two different codons chosen from four codons (UUU, UCU, CUU, and CCU; Figure 3a) and prepared the corresponding template DNAs. All four test codons are near-cognate with one another and three of them may be misread by the corresponding near-cognate aminoacyl-tRNAs through G−U base pairs (Figure S5, purple). We also prepared four types of tRNA with the same tRNA$^{Asn\text{-}E2}$ body and different anticodons to read the four codons, and each tRNA was charged with a CNA (**12**, **1**, **8**, or **3**) according to the reprogrammed genetic code shown in Figure 3b. The less efficiently incorporated CNA **3** (Figure 2) was reassigned to the CCU codon, which is not misread by any near-cognate tRNA through G−U base pairs (Figure S5, purple). The more efficiently incorporated CNA **12** was reassigned to the UUU codon, which may be misread by the other near-cognate tRNAs. The translation system was reconstituted from in vitro-transcribed tRNA$^{fMet}_{CAU}$, tRNA$^{Tyr}_{CUA}$, and tRNA$^{Arg}_{CCG}$, instead of a native base-modified tRNA mixture, so that all 16 NNU codons are vacant, and tRNA$^{Tyr}_{CUA}$ is tyrosylated by TyrRS and reads a UAG codon (Figure 3b). MALDI-TOF-MS analysis of desalted samples from transcription/translation using the six DNA templates showed that all six programmed di-N-alkyl bicyclic peptides were synthesized as the main products (Figure 3c, Figure S6 in Supporting Information).

Similarly, the successive triple incorporation of three distinct CNAs was examined by translating mRNAs containing three different codons chosen from the same four codons (UUU, UCU, CUU, and CCU) as those used in di-N-alkyl bicyclic peptide synthesis (Figure 4a). MALDI-TOF-MS analysis of the desalted translation products showed that the translation reaction with each template DNA yielded the expected tri-N-alkyl tricyclic peptide as the main product for all four templates (Figure 4b, Figure S7 in Supporting Information). The DNA-programmed synthesis of the desired di-N-alkyl bicyclic peptides and tri-N-alkyl tricyclic peptides showed correct decoding of the four near-cognate codons (UUU, UCU, CUU, and CCU) to the designated CNAs without misreading.

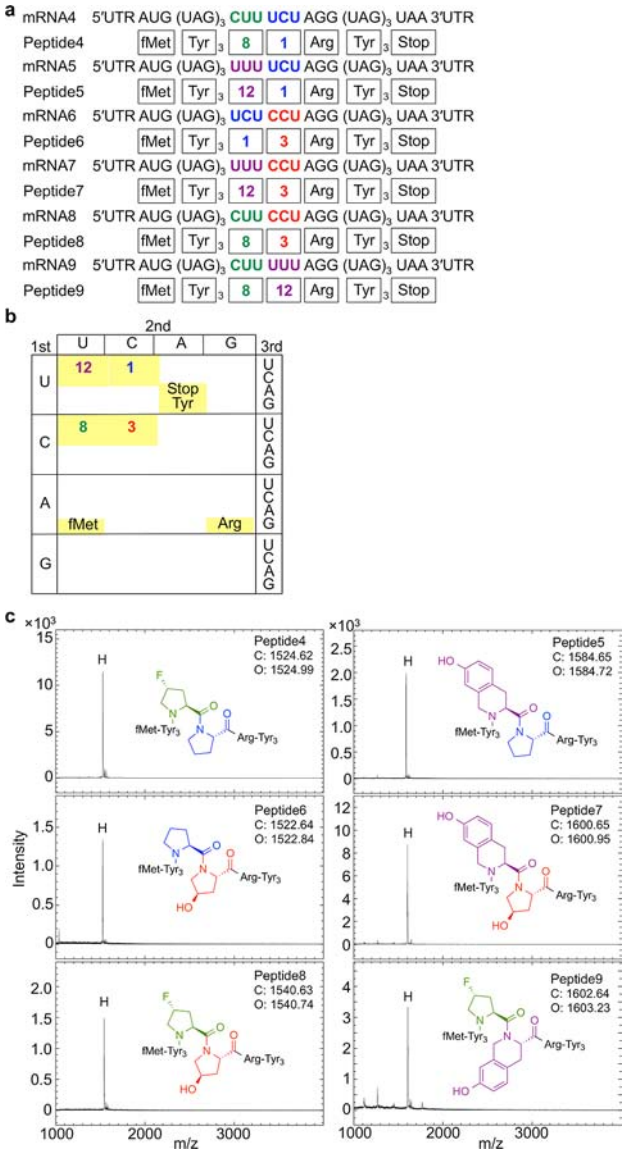


Figure 3. Genetic code reprogramming for DNA-programmed synthesis of di-N-alkyl bicyclic peptides. (a) Sequences of mRNA templates 4–9 and their encoded di-N-alkyl bicyclic peptides 4–9. CNAs and their reassigned codons are colored as in the reprogrammed genetic code shown in panel b. (b) Reprogrammed genetic code for DNA-programmed synthesis of di-N-alkyl bicyclic peptides. Four NNU codons were reassigned to four different CNAs. (c) MALDI-TOF mass spectra and structures of di-N-alkyl bicyclic peptides. The calculated (C) and observed (O) molecular masses for singly charged species $[M + H]^+$ are shown in each spectrum. H denotes peaks corresponding to the respective di-N-alkyl bicyclic peptides.

The correct decoding of the near-cognate four codons in the DNA-programmed synthesis of highly N-alkylated polycyclic peptidomimetics prompted us to attempt the reassignment of all 16 NNU codons to 16 different CNAs on the reprogrammed genetic code. Template mRNA 14, which contained the four different NNU codons (UUU, UCU, CUU, and CCU) used in the previous experiments, was designed (Figure 5a,b, purple). The corresponding DNA template was transcribed and translated in the presence of four types of CNA (**12**, **1**, **8**, and **3**) charged onto the respective tRNAs according to the reprogrammed genetic code (Figure 5b, purple). MALDI-TOF-MS analysis showed that template

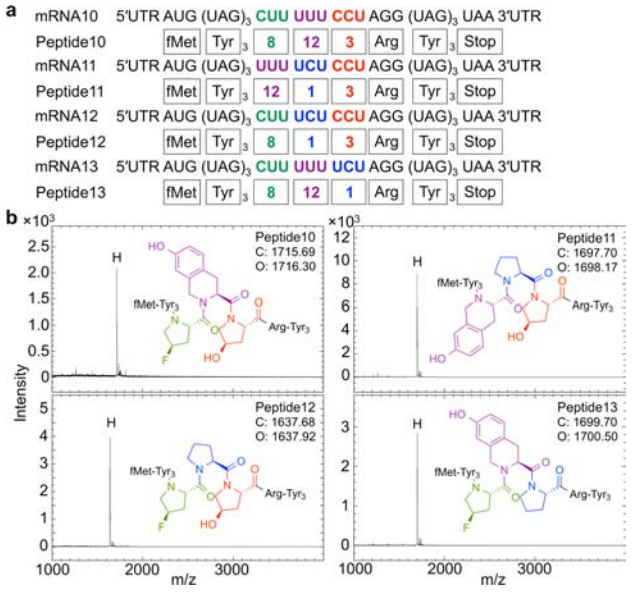


Figure 4. Genetic code reprogramming for DNA-programmed synthesis of tri-N-alkyl tricyclic peptides. (a) Sequences of mRNA templates 10–13 and their encoded tri-N-alkyl tricyclic peptides 10–13. CNAs and their reassigned codons are colored as in the reprogrammed genetic code shown in Figure 3b. (b) MALDI-TOF mass spectra and structures of tri-N-alkyl tricyclic peptides. The calculated (C) and observed (O) molecular masses for the singly charged species $[M + H]^+$ are shown in each spectrum. H denotes peaks corresponding to the tri-N-alkyl tricyclic peptides.

mRNA 14 was successfully translated into the desired tetra-N-alkyl tetracyclic peptide 14 with the correct mass (Figure 5c, Figure S8 in Supporting Information). Three extra template mRNAs 15, 16, and 17, each of which contained a different set of four near-cognate NNU codons chosen from the other 12 NNU codons (Figure 5a, Figure S5 in Supporting Information, shown in blue, green, and red) were designed to express three different tetra-N-alkyl tetracyclic peptides. In addition, 12 further tRNAs with the same $tRNA^{Asn\text{-}E2}$ body sequence and different anticodons were prepared and aminoacylated with the CNAs according to the reprogrammed genetic code (Figure 5b, shown in blue, green, and red). The less efficiently incorporated CNAs (**22**, **14**, and **19**) were reassigned to codons that are not misread by any near-cognate tRNAs (ACU, CAU, and AAU, respectively), whereas the more efficiently incorporated CNAs (**7**, **9**, and **4**) were reassigned to codons that may be misread by the other near-cognate tRNAs (GUU, UGU, and GGU, respectively). MALDI-TOF-MS analysis of the desalted translation products revealed that all three tetra-N-alkyl tetracyclic peptides were expressed as programmed in the DNA templates (Figure 5c).

The DNA-programmed synthesis of the desired four tetra-N-alkyl tetracyclic peptides as the main products show that the 16 NNU codons can be correctly reassigned to the 16 different CNAs without detecting near-cognate misreading (Figure 5b, Figure S5 in Supporting Information).

TRAP Display of Completely N-Alkylated Polycyclic Peptidomimetic Random Library. After the fidelity of the reprogrammed genetic code in which 16 NNU codons were reassigned to the 16 different CNAs was confirmed, we evaluated the display efficiency of a completely N-alkylated polycyclic peptidomimetics library on a random NNU mRNA library by using our recently developed TRAP display.[43] The

dx.doi.org/10.1021/ja405044k | J. Am. Chem. Soc. 2013, 135, 12297–12304

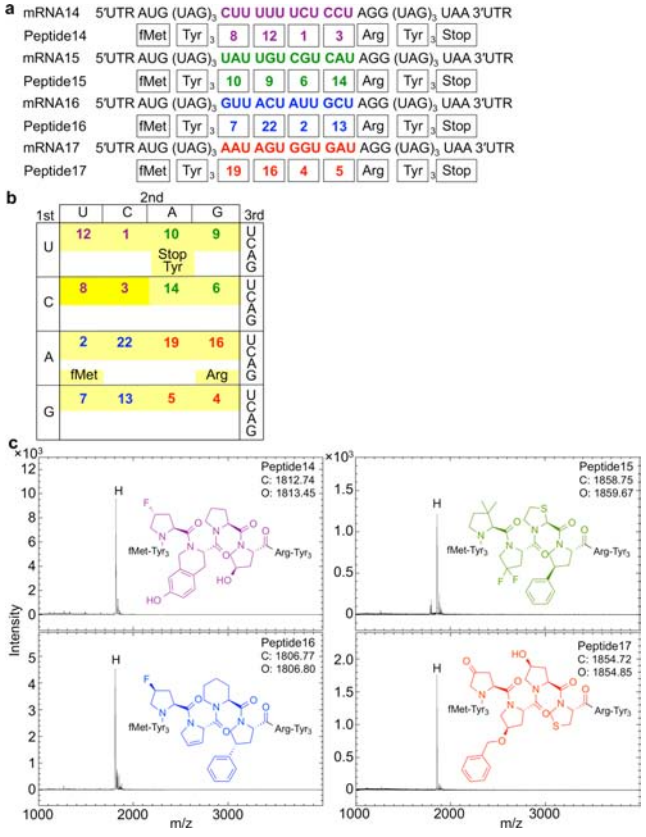


Figure 5. Genetic code reprogramming for DNA-programmed synthesis of tetra-N-alkyl tetracyclic peptides. (a) Sequences of mRNA templates 19–23 and their encoded tetra-N-alkyl tetracyclic peptides 19–23. CNAs and their reassigned codons are colored as in the reprogrammed genetic code shown in panel b. (b) Reprogrammed genetic code for DNA-programmed synthesis of tetra-N-alkyl tetracyclic peptides. All 16 NNU codons were reassigned to 16 different CNAs. (c) MALDI-TOF mass spectra and structures of tetra-N-alkyl tetracyclic peptides. The calculated (C) and observed (O) molecular masses for the singly charged species $[M + H]^+$ are shown in each spectrum. H denotes peaks corresponding to the respective tetra-N-alkyl tetracyclic peptides.

TRAP display uses a transcription/translation coupling system containing puromycin attached to the 3′-end of an oligo-DNA linker that is complementary to the 3′-end of the mRNA. In the TRAP system, DNA is continuously transcribed and translated to a peptide, and the expressed peptide is spontaneously displayed ("trapped") on its encoding mRNA via the puromycin–DNA linker. In the TRAP system, there is no RF1 to stall the ribosome at a vacant UAG codon immediately before the puromycin–DNA linker annealing region on mRNAs, and thus efficient peptidyl transfer to the puromycin on the annealing DNA linker can be induced.

In our previous TRAP display selections using random NNK libraries, the spacer sequence between the random NNK codons and the puromycin–DNA linker annealing region on mRNAs contained a GGU codon.[24,43] Therefore, before the display efficiency of the completely N-alkylated polycyclic peptidomimetic library was evaluated, several spacer sequences containing no XXU codons were investigated by using streptavidin pull-down of an NNU mRNA library displaying a biotinylated proteinogenic peptide library (Figure S9a,b in Supporting Information). The undesired N-terminal truncated peptides could be displayed by puromycin attacking the premature peptidyl-tRNAs before the translation of the desired open reading frame was complete. Thus, a biotin-modified amino acid was incorporated in the C-terminal constant region (Figure S9b in Supporting Information) in order to recover selectively mRNAs displaying the full-length peptide and leave behind mRNAs displaying the N-terminal truncated peptide, which allows the intrinsic display efficiency to be determined. The quantification of cDNA reverse-transcribed from the pulled-down mRNA/biotinylated peptide complex showed that the $(AUG)_4$ spacer exhibited the best peptide display efficiency (Figure S9c in Supporting Information).

Similar pull-down analysis showed that peptide display efficiency decreased as the distance increased between the first vacant UAG codon, where puromycin attacked the peptidyl-tRNA, and the puromycin–DNA linker annealing region on the mRNAs (Figure S10 in Supporting Information). This indicates that it is unlikely that the truncated N-terminal peptides, which were produced by premature attack of puromycin during translation upstream of the spacer region on the mRNAs, were displayed. On the other hand, expression of C-terminal truncated peptides was observed when the inefficiently incorporated CNAs (**15** and **21**) were used in the previous experiment (Figure S3 in Supporting Information). Thus, the C-terminal truncated peptides could be displayed and cause the peptide display efficiency to be misevaluated. Therefore, we decided to incorporate a biotin-modified amino acid at the N-terminus rather than in the C-terminal constant region for the selective recovery of mRNA-displayed full-length peptides consisting of the N-terminal biotin-amino acid, all internal random amino acids, and C-terminal spacer amino acids (Figure 6b).

The display efficiency of completely N-alkylated polycyclic peptidomimetics was evaluated by constructing an mRNA library containing the $(AUG)_4$ spacer for the expression of a peptide library composed of eight random CNA residues (Figure 6b). An *N*-biotinyl-Phe-tRNA^{ini} was used to label the peptides with biotin at their N-terminus (Figure 6a). The TRAP system was prepared without Met to reassign the start AUG codon to *N*-biotinyl-Phe (^{biot}F) and with in vitro transcribed $tRNA^{Ser}_{CAU}$, which is serylated by SerRS, to reassign the spacer AUG codons to Ser (Figure 6b,c). The mRNA template was added instead of the DNA template to the TRAP system because the use of an mRNA template had previously showed better peptide display efficiency than that of a DNA template.[43] Streptavidin pull-down from the control translation that contained the proteinogenic amino acids charged onto tRNAs by aaRSs in situ (Figure 6c, left) showed that 10.7% of the mRNA displayed their encoded biotinylated proteinogenic peptides (Figure 6d). On the other hand, a similar streptavidin pull-down experiment for translation that contained 16 precharged CNA-tRNAs (Figure 6c, right) showed that the biotinylated completely N-alkylated polycyclic peptidomimetics were displayed on their mRNA with a display efficiency of 0.4%. This difference may be due to the multiple turnover of aminoacyl-tRNA generation with aaRSs and faster peptide bond formation of the non-N-alkyl amino acids than the N-alkyl amino acids.[53,54] Very recent studies reported that translation elongation factor P (EF-P) enhances the translation of poly(proline)-containing proteins.[55,56] Therefore, we examined whether EF-P improves the display efficiency of completely N-alkylated polycyclic peptidomimetics. A streptavidin pull-down experiment for translation that contained 16 precharged CNA-tRNAs in the presence of EF-P showed that

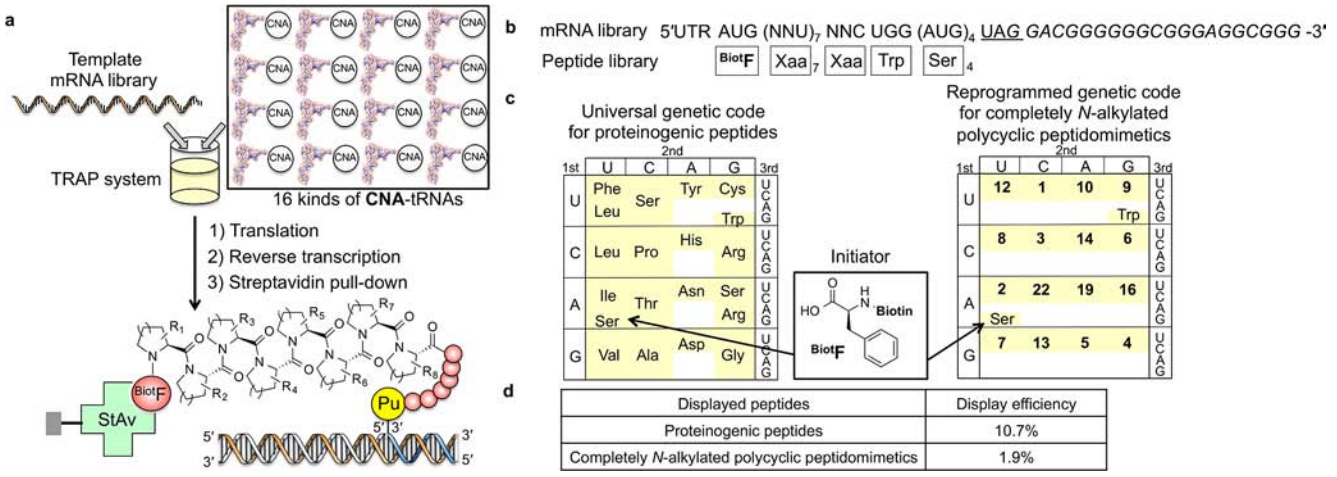


Figure 6. Display efficiency of completely N-alkylated polycyclic peptidomimetics for the TRAP display, determined by a streptavidin (StAv) pull-down assay. (a) Scheme for StAv pull-down of peptide/mRNA/cDNA complexes. In the TRAP system, the expressed peptides are spontaneously displayed on their encoding mRNA via a Pu−DNA linker in the translation system. mRNA displaying its encoding peptide, which contained *N*-biotinylphenylalanine (^{biot}F), was separated from mRNA displaying no peptide by use of StAv-immobilized beads. (b) Sequences of the mRNA library and peptide library used to quantify display efficiency. The start AUG codon was reassigned to ^{biot}F. The vacant UAG codon for stalling the ribosome to allow efficient peptidyl transfer to puromycin is underlined. The puromycin−DNA linker annealing region is italicized. Xaa indicates random CNAs or random proteinogenic amino acids. (c) Genetic code for expression of a library of proteinogenic peptides and completely N-alkylated polycyclic peptidomimetics. (d) Display efficiency of proteinogenic peptides and completely N-alkylated polycyclic peptidomimetics. Display efficiency was calculated by dividing the amount of recovered cDNA after StAv pull-down by that before StAv pull-down. cDNAs were quantified by quantitative PCR. Experiments were performed in triplicate.

the biotinylated completely N-alkylated polycyclic peptidomimetics were displayed on their mRNA with a display efficiency of 1.9% (Figure 6d). The display efficiency of the completely N-alkylated polycyclic peptidomimetics indicates that a 0.5 mL translation reaction containing 2.5 μM mRNA could yield approximately 24 pmol, that is, 10^{13} unique completely N-alkylated polycyclic peptidomimetics displayed on their encoding mRNAs.

To know whether this decreasing display efficiency for completely N-alkylated polycyclic peptidomimetics was caused by inefficient decoding of specific codons assigned to less translation-compatible CNAs, we performed high-throughput sequencing of cDNA reverse-transcribed from the mRNA library displaying the proteinogenic peptides or the completely N-alkylated polycyclic peptidomimetics together with the initial mRNA library. A comparison of the number of each codon in more than 10^6 sequenced cDNAs revealed that there was no significant difference in the relative appearance frequency of each codon between mRNAs displaying the proteinogenic peptides and completely N-alkylated polycyclic peptidomimetics (Figure S11 in Supporting Information). This result suggested that the decreased display efficiency of N-alkylated polycyclic peptidomimetics was caused by decreased decoding efficiency of all CNAs and not by specific CNAs.

■ DISCUSSION

Genetic code reprogramming utilizing nonproteinogenic aminoacyl-tRNAs and a reconstituted cell-free translation system has been used previously for simultaneous incorporation of multiple nonproteinogenic amino acids into ribosomally synthesized peptides.[9,18−21,23,24,29−31,34−36] However, because the number of nonproteinogenic amino acids reassigned to codons in these studies was limited, many remaining codons in the reprogrammed genetic code were assigned to either cognate proteinogenic amino acids or as blanks. For example, in the first demonstration of genetic code reprogramming, only three nonproteinogenic amino acids were reassigned to three different sense codons.[29] In a study of highly modified peptide selection, even 12 nonproteinogenic amino acids were reassigned, but the remaining codons were still reassigned to eight cognate proteinogenic amino acids, which produced a peptide library containing nonproteinogenic and proteinogenic amino acids.[9] In addition, the combination of tRNA preaminoacylation and in situ aminoacylation by aaRS could make it difficult to control the homogeneity of the expressed peptides and thus limit the number of simultaneously used nonproteinogenic amino acids since the time-dependent increase or decrease of aminoacyl-tRNA concentration in these two aminoacylation reactions is completely different.[18,20,22,23,31]

In contrast, we have reassigned all 16 NNU codons to 16 different CNAs to produce a library of completely N-alkylated polycyclic peptidomimetics by using 16 different precharged CNA-tRNAs. The high decoding fidelity of the 16 NNU codons was confirmed by translating four types of mRNA containing a different set of four near-cognate NNU codons into tetra-N-alkyl tetracyclic peptides as the sole products (Figure 5). Moreover, we quantitatively evaluated the display efficiency of the library and found that 10^{13} different completely N-alkylated polycyclic peptidomimetics displayed on their mRNAs could be prepared from a 0.5 mL translation reaction (Figure 6). Such completely N-alkylated polycyclic peptidomimetics selected from a highly diverse library via in vitro display selection could be used as drug candidates that can be targeted to intracellular disease-related biomolecules and as small-molecule affinity reagents for monitoring and controlling intracellular endogenous biomolecules for research without the need for an invasive injection.[57]

dx.doi.org/10.1021/ja405044k | J. Am. Chem. Soc. 2013, 135, 12297−12304

■ ASSOCIATED CONTENT

Supporting Information

Additional text with detailed experimental procedures, and six tables and 11 figures as described in the main text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

kawakami@bio.c.u-tokyo.ac.jp; murah@bio.c.u-tokyo.ac.jp

Notes

The authors declare the following competing financial interest(s): Author Hiroshi Murakami declares competing financial interests, as one of the founders and shareholders in PeptiDream Inc, a peptide drug discovery company using similar technology.

■ ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Young Scientists (B25879416 to T.K.) and Funding Program for Next Generation World-Leading Researchers (LR011 to H.M.) from the Japan Society for the Promotion of Science, an HFSP long-term fellowship (LT000939/2010 to T.K.), and a research grant from the Kato Memorial Bioscience Foundation (to T.K.).

■ REFERENCES

(1) Mattheakis, L. C.; Bhatt, R. R.; Dower, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9022.
(2) Hanes, J.; Pluckthun, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4937.
(3) Nemoto, N.; MiyamotoSato, E.; Husimi, Y.; Yanagawa, H. *FEBS Lett.* **1997**, *414*, 405.
(4) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297.
(5) Kawakami, T.; Murakami, H. *J. Nucleic Acids* **2012**, *2012*, No. 713510.
(6) Hayashi, Y.; Morimoto, J.; Suga, H. *ACS Chem. Biol.* **2012**, *7*, 607.
(7) Morimoto, J.; Hayashi, Y.; Suga, H. *Angew. Chem., Int. Ed.* **2012**, *51*, 3423.
(8) Hofmann, F. T.; Szostak, J. W.; Seebeck, F. P. *J. Am. Chem. Soc.* **2012**, *134*, 8038.
(9) Schlippe, Y. V.; Hartman, M. C.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10469.
(10) Kwon, Y. U.; Kodadek, T. *J. Am. Chem. Soc.* **2007**, *129*, 1508.
(11) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* **2008**, *41*, 1331.
(12) Tan, N. C.; Yu, P.; Kwon, Y. U.; Kodadek, T. *Bioorg. Med. Chem.* **2008**, *16*, 5853.
(13) White, T. R.; Renzelman, C. M.; Rand, A. C.; Rezai, T.; McEwen, C. M.; Gelev, V. M.; Turner, R. A.; Linington, R. G.; Leung, S. S.; Kalgutkar, A. S.; Bauman, J. N.; Zhang, Y.; Liras, S.; Price, D. A.; Mathiowetz, A. M.; Jacobson, M. P.; Lokey, R. S. *Nat. Chem. Biol.* **2011**, *7*, 810.
(14) Baeriswyl, V.; Heinis, C. *Protein Eng. Des. Sel.* **2013**, *26*, 81.
(15) Merryman, C.; Green, R. *Chem. Biol.* **2004**, *11*, 575.
(16) Tan, Z.; Forster, A. C.; Blacklow, S. C.; Cornish, V. W. *J. Am. Chem. Soc.* **2004**, *126*, 12752.
(17) Zhang, B.; Tan, Z.; Dickson, L. G.; Nalam, M. N.; Cornish, V. W.; Forster, A. C. *J. Am. Chem. Soc.* **2007**, *129*, 11316.
(18) Kawakami, T.; Murakami, H.; Suga, H. *Chem. Biol.* **2008**, *15*, 32.
(19) Kawakami, T.; Murakami, H.; Suga, H. *J. Am. Chem. Soc.* **2008**, *130*, 16861.
(20) Subtelny, A. O.; Hartman, M. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2008**, *130*, 6131.
(21) Kawakami, T.; Ohta, A.; Ohuchi, M.; Ashigai, H.; Murakami, H.; Suga, H. *Nat. Chem. Biol.* **2009**, *5*, 888.
(22) Subtelny, A. O.; Hartman, M. C.; Szostak, J. W. *Angew. Chem., Int. Ed.* **2012**, *50*, 3164.
(23) Yamagishi, Y.; Shoji, I.; Miyagawa, S.; Kawakami, T.; Katoh, T.; Goto, Y.; Suga, H. *Chem. Biol.* **2011**, *18*, 1562.
(24) Kawakami, T.; Ishizawa, T.; Fujino, T.; Reid, P. C.; Suga, H.; Murakami, H. *ACS Chem. Biol.* **2013**, *8*, 1205.
(25) Cornish, V. W.; Mendel, D.; Schultz, P. G. *Angew. Chem., Int. Ed.* **1995**, *34*, 621.
(26) Deming, T. J. F., M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1996**, *29*, 1442.
(27) Deming, T. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Macromol. Sci. A: Pure Appl. Chem.* **1997**, *34*, 2143.
(28) Steiner, T.; Hess, P.; Bae, J. H.; Wiltschi, B.; Moroder, L.; Budisa, N. *PLoS One* **2008**, *3*, No. e1680.
(29) Forster, A. C.; Tan, Z.; Nalam, M. N.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6353.
(30) Josephson, K.; Hartman, M. C. T.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 11727.
(31) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. *Nat. Methods* **2006**, *3*, 357.
(32) Forster, A. C.; Weissbach, H.; Blacklow, S. C. *Anal. Biochem.* **2001**, *297*, 60.
(33) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751.
(34) Hartman, M. C.; Josephson, K.; Lin, C. W.; Szostak, J. W. *PLoS One* **2007**, *2*, No. e972.
(35) Ohta, A.; Murakami, H.; Higashimura, E.; Suga, H. *Chem Biol* **2007**, *14*, 1315.
(36) Forster, A. C. *Nucleic Acids Res.* **2009**, *37*, 3747.
(37) Forster, A. C. *Biotechnol. J.* **2012**, *7*, 835.
(38) Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1984**, *23*, 1468.
(39) Hartman, M. C.; Josephson, K.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4356.
(40) Ohuchi, M.; Murakami, H.; Suga, H. *Curr. Opin. Chem. Biol.* **2007**, *11*, 537.
(41) Xiao, H.; Murakami, H.; Suga, H.; Ferre-D'Amare, A. R. *Nature* **2008**, *454*, 358.
(42) Niwa, N.; Yamagishi, Y.; Murakami, H.; Suga, H. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3892.
(43) Ishizawa, T.; Kawakami, T.; Reid, P. C.; Murakami, H. *J. Am. Chem. Soc.* **2013**, *135*, 5433.
(44) Karginov, V. A.; Mamaev, S. V.; An, H.; Van Cleve, M.; Hecht, S. M.; Komatsoulis, G. A.; Abelson, J. N. *J. Am. Chem. Soc.* **1997**, *119*, 8166.
(45) Choudhury, A. K.; Golovine, S. Y.; Dedkova, L. M.; Hecht, S. M. *Biochemistry* **2007**, *46*, 4066.
(46) Kane, J. F.; Violand, B. N.; Curran, D. F.; Staten, N. R.; Duffin, K. L.; Bogosian, G. *Nucleic Acids Res.* **1992**, *20*, 6707.
(47) Fujino, T.; Goto, Y.; Suga, H.; Murakami, H. *J. Am. Chem. Soc.* **2013**, *135*, 1830.
(48) Cochella, L.; Green, R. *Science* **2005**, *308*, 1178.
(49) Johansson, M.; Lovmar, M.; Ehrenberg, M. *Curr. Opin. Microbiol.* **2008**, *11*, 141.
(50) Ledoux, S.; Olejniczak, M.; Uhlenbeck, O. C. *Nat. Struct. Mol. Biol.* **2009**, *16*, 359.
(51) Murakami, H.; Ohta, A.; Suga, H. *Nat. Struct. Mol. Biol.* **2009**, *16*, 353.
(52) Zaher, H. S.; Green, R. *Cell* **2009**, *136*, 746.
(53) Wohlgemuth, I.; Brenner, S.; Beringer, M.; Rodnina, M. V. *J. Biol. Chem.* **2008**, *283*, 32229.
(54) Pavlov, M. Y.; Watts, R. E.; Tan, Z.; Cornish, V. W.; Ehrenberg, M.; Forster, A. C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 50.
(55) Doerfel, L. K.; Wohlgemuth, I.; Kothe, C.; Peske, F.; Urlaub, H.; Rodnina, M. V. *Science* **2013**, *339*, 85.
(56) Ude, S.; Lassak, J.; Starosta, A. L.; Kraxenberger, T.; Wilson, D. N.; Jung, K. *Science* **2013**, *339*, 82.
(57) Kawakami, T.; Cheng, H.; Hashiro, S.; Nomura, Y.; Tsukiji, S.; Furuta, T.; Nagamune, T. *ChemBioChem* **2008**, *9*, 1583.