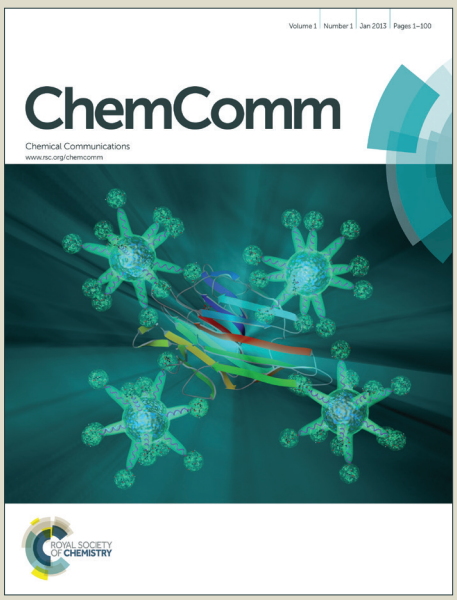
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

COMMUNICATION

Two site genetic incorporation of varying length polyethylene glycol into the backbone of one peptide

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Qingmin Zang,[a,b]† Seiichi Tada,[a]† Takanori Uzawa,[a,c] Daisuke Kiga,[a,b] Masayuki Yamamura,[a,b] and Yoshihiro Ito [a,b,c*]

Polyethylene glycol (PEG) of differing lengths was genetically incorporated into the backbone of a polypeptide using stop-anticodon and frameshift anticodon-containing tRNAs which were acylated with PEG-containing amino acids.

Synthetic polymer–protein hybrids have been developed for use as therapeutic proteins or bioreactor enzymes. Because polyethylene glycol (PEG) is nontoxic, nonimmunogenic, and highly soluble in water and organic solvents, many PEG conjugated (PEGylated) proteins have been developed and have been shown to be very beneficial.[1] However, the PEGylation of many proteins results in a measurable reduction in their biological activity, which has been attributed mainly to the non-site-specific nature of the PEGylation.[2] To overcome this, a method of more specific chemical modification has been developed, which combines conventional protein engineering with specifically modifiable amino acid incorporation and bioorthogonal chemistry using ligation enzymes.[3] A site-specific polymer attachment has also been achieved with the complete chemical synthesis of an erythropoietic protein.[4]

Alternatively, a genetic encoding approach was reported in 1989 as a method for incorporating site-specific non-canonical amino acids into peptides or proteins,[5] and various amino acids have since been incorporated in this way.[6] The method utilizes the UAG codon (i.e., the amber nonsense stop codon), which normally directs the termination of protein synthesis, to instead encode a non-canonical amino acid that is loaded onto the suppressor tRNA. Some researchers have inserted a non-canonical and reactive amino acid at a different site using the misacylated tRNA method and linked a single PEG molecule to each group as a post-translational modification using click chemistry, Staudinger-phosphite reactions, Suzuki-Miyaura reactions or Copper-free Sonogashira cross-coupling reactions.[7]

Sisido et al. developed a frameshift suppression method in which non-canonical amino acids are incorporated into proteins using four-base codon–anticodon pairs instead of a stop codon.[8] They synthesized nobel tRNAs that recognize four-base codons, for example AGGU and AGGG, and successfully incorporated nitrophenylalanine into the protein streptavidin in a site specific manner. Using this method, Shozen et al.[9] also incorporated short PEG chains, including tetraethylene glycol (PEG4) and dodecaethylene glycol (PEG12), into streptavidin via a CGGG codon. We have also incorporated a longer PEG chain, up to 1000 g/mol, using the amber stop codon suppression method.[10]

The advantage of misacylated tRNA is that it allows multiple and specific incorporation of non-canonical amino acids into peptides and proteins, and the addition of fluorescent groups via this method has been reported.[11] However, each method reported for conventional PEGylation incorporation has resulted in a significant loss in biological activity for the various proteins that have been modified. It is therefore important that the addition of PEG oligomers occurs at predetermined, specific locations in the peptide because the placement of PEG in these regions will provide the best improvement in proteolytic stability, and have the least effect on biological activity.[12] Use of the genetic method allows the precise insertion of two or more PEG chains, of differing lengths, into specific positions in individual proteins. As well as precise site incorporation it is also important that the PEG chains are relatively short, as shorter chains will have less effect on peptide and protein activity. Therefore, in this study, we attempted to site-specifically incorporate one or two PEGs by adding tRNAs carrying PEG of various lengths that recognize a stop codon and a frameshift codon via a translation system, as shown in Fig. 1.

Polyethylene glycol (PEG4, PEG8, PEG12, and PEG24, which have molecular masses of 170, 340, 510, and 1020 g/mol, respectively) were conjugated to aminophenylalanyl (AF)-tRNAs containing a CUA anticodon or a frameshift anticodon (CCCG) as previously

[a] Nano Medical Engineering Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.E-mail: y-ito@riken.jp
[b] Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan
[c] Emergent Bioengineering Materials Research Team, RIKEN Center for Emergent Matter Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
† They contribute equally to this work.
Electronic Supplementary Information (ESI) available:Materials and methods; in vitro translation; Mass data. See DOI: 10.1039/x0xx00000x

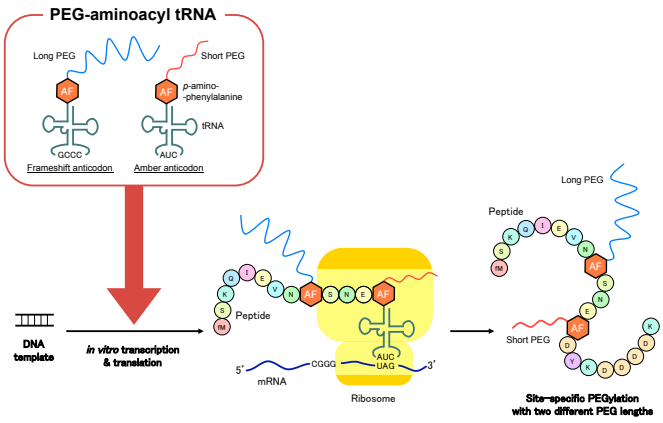


Fig. 1 Incorporation of different sized polyethylene glycols (PEGs) into a polypeptide by *in vitro* translation.

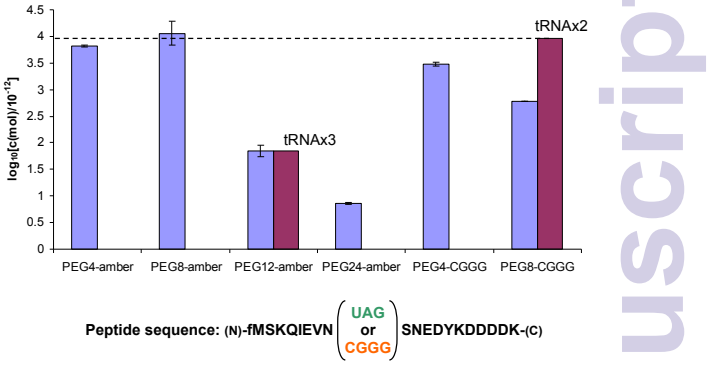


Fig. 2 Yield efficiency (logarithms of relative amount of PEGylation product to an internal standard peptide, 3xFLAG) of varying length PEGylation (PEG4, PEG8, PEG12, and PEG24) using the amber (UAG) or frameshift (CGGG) codon. fM in the peptide sequence indicates a formylmethionine. "tRNAx3" shown at PEG12-amber and "tRNAx2" shown at PEG8-CGGG indicate three- and two-fold concentrations of tRNA in the translation system, respectively.

reported.[10] The DNA templates for the cell-free translation that encode the peptide or proteins were constructed by ligating complementary oligonucleotides with the appropriate single-stranded overhangs into pROX-FL92.1amber (Protein Express, Chiba, Japan). Cell-free translation was performed using a RYTS kit (Protein Express) according to the manufacturer's protocol. The reaction mixtures were incubated at 30 °C for 2 h, unless otherwise stated. The samples were prepared for mass spectrometry as previously reported.[10] The yield efficiencies of PEG-incorporated peptides were calculated from their relative amounts compared with an added standard peptide 3xFLAG (DYKDHDGDYKDHDIDYKDDDDK, Sigma-Aldrich, St. Louis, MO, USA).

PEG was successfully incorporated by the cell-free translation system. The incorporation of the longer PEG chains resulted in lower yields of the translation product (Fig. 2). For example, in the case of the amber tRNA system, the inclusion of either PEG4 or PEG8 results in the same translation product yields, although the yields for PEG12 and 24 are 1% and 0.1% the yields obtained for the products of PEG4 and PEG8. The PEG-length dependence of the translation products of PEG8 to PEG24 may be explained by the previously reported steric hindrance between PEG and the ribosome,[10] and between PEG and elongation factor thermo unstable (EF-Tu).[13] The molecular masses of PEG4 and PEG8 are 170 and 340 g/mol, respectively, and as such, may not sterically block translation in the ribosome and/or the binding of tRNA to EF-Tu. It is noteworthy that we have confirmed that the amount of PEGylation product is not limited by the amount of tRNA used; effectively, the same yield of translation product is obtained, even when a triple concentration of tRNA (tRNA × 3) is used to incorporate PEG12.

The translation efficiency of PEG8 at the frameshift codon is over 10-fold lower compared with that at the amber codon, whereas the translation efficiency of PEG4 at the frameshift codon is only 2-fold lower compared with the amber codon (Fig. 2). The translation efficiency of PEG8 at the frameshift codon using $tRNA_{CCCG}$ is considerably lower compared with PEG4 and PEG8, which presumably results from the competitive translation of $tRNA_{CCG}$(Arg) in the translation system. Evidence for this is that a double concentration of $tRNA_{CCCG}$ (tRNA × 2) increases the yield of the translation product to a level similar to that obtained from the amber-PEG8 (Fig. 2). The comparison between amber and four-base codons has been reported by Sisido's group.[14,15] Hohsaka et al. compared the incorporation yield between amber and various four-base codons using *p*-nitrophenylalanyl-tRNA, and the yield from the use of the frameshift codon was higher than that of the amber codon.[14] However, they estimated that the incorporation efficiency of each codon was similar or sometimes lower for the four-base codon compared with the amber codon.[15] Presumably this difference is due to the dependence of the yield on the kind of non-canonical amino acids. In the case of PEG12 or PEG24 incorporation using a frameshift codon, the yield is considerably lower compared with the yield from the amber codon method.

Given the successful incorporation with either the amber or frameshift codon, we then site-specifically incorporated two PEGs into a peptide via the two codons. Two PEG4s were site-specifically incorporated into one peptide at position 9 using the amber codon and at position 13 using the frameshift codon. (entry 1, Fig. 3) Although the yield of the translation product with two PEG4s was about 2000-fold lower compared with single PEG4 incorporation, we were able to successfully observe the peptide with two

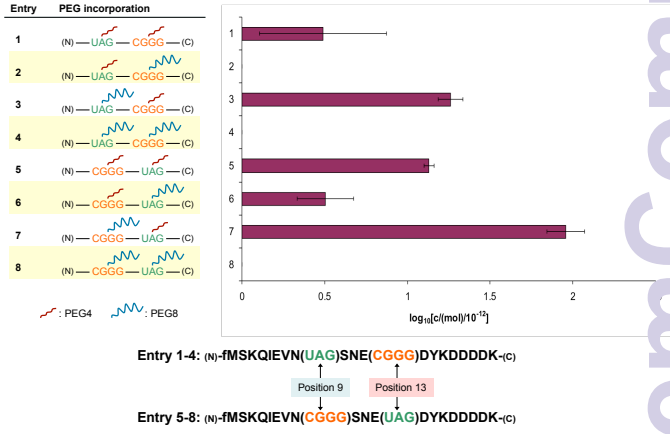


Fig. 3 Incorporation of different sized PEGs into a peptide using two different codons. PEG4 and PEG8 were incorporated using amber (UAG) or frameshift (CGGG) codons. (N) and (C) indicate the N- and C-terminals of the peptide.

incorporated PEG4 chains. In contrast, we could not incorporate two PEG8s using the same mRNA (entry 4, Fig. 3).

The simultaneous use of two different codons also allows for the synthesis of peptides containing two different sized PEGs, each incorporated into a different specific site using the amber or frameshift codon. We also found a position-dependent size preference of PEGs used in the peptide synthesis by the ribosome. As shown in entry 1 of Fig. 2, the addition of two PEG4 insertions using the same mRNA allows for the site-specific incorporation of the PEG8 amber and PEG4 frameshift codons at positions 9 and 13 (entry 3), respectively, depending on the different anticodon of the tRNAs carrying the different length PEGs. Using the same mRNA in entry 2, we could not obtain a peptide with a PEG4 at position 9 and a PEG8 at position 13. The synthesis efficiency depends on PEG size (entry 3 > entry 1 > entry 2) and is also seen in the other mRNA (entry 7 > entry 5 > entry 6), which has the frameshift codon at position 9 and the amber codon at position 13. Similarly, both mRNAs could not incorporate two PEG8s (entries 4 and 8, Fig. 3).

A preference for PEG8 over PEG4 at position 9 in these double incorporation experiments is also found in the production of single site incorporation of PEG (Fig. 2) with enough tRNA for the frameshift codon, although the preference is less different than cases of two site incorporation (Fig. 3) The difference in the translation efficiency of the two mRNAs is a function of their secondary structures. Even with the above PEG8 preference, two PEG8s within four residues could not be incorporated simultaneously. Considering the size of the ribosome tunnel from the peptidyl transferase center, the mechanism that decreased efficiency of PEG12- and PEG24-containing peptide synthesis (Fig. 2) appears to inhibit this two PEG8 incorporation.

In characterizing the peptides with incorporated PEG, we also found a mass corresponding to truncated peptides (lack of fMSKQIEVN or SKQIEVN, Fig. 4). These products arise because translation is apparently terminated just before the incorporation of PEG. The attempted incorporation of longer PEGs results in a higher yield of the truncated peptides. In the case of the amber tRNA system, the yield of the truncated peptide was only about 2% for the PEG4, while for the PEG8 the yield of truncated peptide was sharply increased to about 17%. For the longer PEG12 and PEG24,

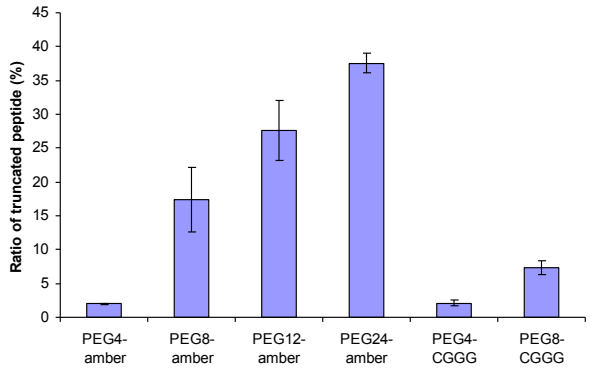


Fig. 4 Yield of truncated peptides using different lengths of PEG with a single codon. The truncated peptides are XSNEDYKDDDDK, where X is PEG (4, 8, 12, or 24). The ratio of truncated peptides is the percentage in all peptides containing PEG.

the yields of the truncated peptides were increased to about 2[illegible] and 36.5%, respectively. For the frameshift codon system, the yield[illegible] of the truncated peptides were about 2% and 7%, respectively. As such, it appears that the longer PEGs might block the peptidyl transferase reaction in the ribosome. A similar phenomenon is also reported by Abe et al.[16], when a large fluorescent, unnatural amino acid was incorporated into proteins.

In conclusion, we have demonstrated the site-specific incorporation of one or two different sized PEGs into one peptide. For the incorporation of one PEG, the yield of the PEGylation product decreased as the molecular weight of the PEG chain increased. In addition, the yields of the truncated PEGylation peptide increased as the lengths of PEG increased. For the incorporation of two PEGs of different lengths, the PEGylation efficiency depends on the length of the PEG, the codon, and the incorporation sites. In this study a cell-free translation system was used to generate the PEGylated product. A cell-free translation system is generally useful for production of highly pure proteins, but currently the cost is relatively high because of the low yields of the cell-free translation system. However, considering recent improvement in the translation system, the cost will come down over time. We successfully incorporated relatively shorter PEG chains of PEG4 and PEG8. The shorter PEGs affect the active site of the proteins to a lesser extent compared with the longer PEGs. Additionally, short PEG chains can be used for the analysis of PEGylation influence on the peptide/protein structure depending on the difference in incorporation sites. Lawrence et al. have reported a relationship between the PEGylation site and the peptide structure using a PEG4 molecule.[12] Overall, this new method enables control over both of the size and site of PEG incorporation, and thus, should provide more active and stable PEGylated proteins as bioconjugate drugs.

Notes and references

1. R. M. Broyer, G. N. Grover and H. D. Maynard, *Chem. Commun.*, 2011, **47**, 2212-2226; H. A. Klok, *Macromolecules,* 2009, **42**, 7990-8000; A. J. De Graaf, M. Kooijiman, W. E. Hennink and J. Mastrobattista, *Bioconjug. Chem.,* 2009, **20**, 1281-1295; M. C. Joralemon, S. McRae and T. Emrick, *Chem. Commun.,* 2010, **46**, 1377-1393; L. A. Canalle, D. W. P. M. Lowik and J. C. M. Van Hest, *Chem. Soc. Rev.*, 2010, **39**, 329-353; T. L. Cheng, K. F. Chuang, B. M. Chen and S. R. Roffler, *Bioconj Chem*, 2012, **23**, 881-899; I. W. Hamley, *Biomacromolecules*, 2014, **15**, 1543-1559.
2. N. Nischan and C.P. R. Hackenberger, *J. Org. Chem.,* 2014, **79**, 10727-10733.
3. Y. Yamamoto, Y. Tsutsumi, Y. Yoshioka, T. Nishibata, K. Kobayashi, T. Okamoto, Y. Mukai, T. Shimizu, S. Nakagawa, S. Nagata and T. Mayumi, *Nat. Biotechnol.,* 2003, **21**, 546-552; J. H. Wang, S. C. Tam, H. Huang, D. Y. Ouyang, Y. Y. Wang and Y. Zheng, *Biochim. Biophys. Res. Comm.,* 2004, **317**, 965-971; D. F. Doherty, M. S. Rosendahl, D. J. Smith, J. M. Hughes, E. A. Chlipala and G. N. Cox, *Bioconjug. Chem.,* 2005, **16**, 1291-1298; H. Wang, L. Z. He, M. Lensch, H. J. Gabius, C. J. Fee and A. P. J. Middelberg, *Biomacromolecules*, 2008, **9**, 3223-3230; J. Hu, V. Duppatla, S. Harth, W. Schmitz and W. Sebald, *Bioconjug. Chem.,* 2010, **21**, 1762-1772; F. F. Schumacher, M. Nobles, C. P. Ryan, M. E. B. Smith, A. Tinker, S. Caddick and J. R. Baker,

Bioconjug. Chem., 2011, **22**, 132-136; N. S. Joshi, L. R. Whitaker and M. B. Francis, *J. Am. Chem. Soc.,* 2004, **126**, 15942-15943; S. D. Tilley and M. B. Francis, *J. Am. Chem. Soc.,* 2006, **128**, 1080-1081; C. S. Cazalis, C. A. Haller, L. Sease-Cargo and E. L. Chaikof, *Bioconjug. Chem.,* 2004, **15**, 1005-1009; Y. H. Cong, E. Pawlisz, P. Bryant, S. Balan, E. Laurine, R. Tommasi, R. Singh, S. Dubey, K. Peciak, M. Bird, A. Sivasankar, J. Swierkosz, M. Muroni, S. Heidelberger, M. Farys, F. Khayrzad, J. Edwards, G. Badescu, I. Hodgson, C. Heise, S. Somavarapu, J. Liddell, K. Powell, M. Zloh, J. W. Choi, A. Godwin and S. Brocchini, *Bioconjug. Chem.,* 2012, **23**, 248-263; N. Toda, S. Asano and C. F. Barbas, *Angew. Chem.,* 2013, **52**, 12592-12596 ; M. Wendler, L. Grinberg, X. Y. Wang, P. E. Dawson and M. Baca, *Bioconj Chem.,* 2014, **25**, 93-101; P. M. Levine, T. W. Craven, R. Bonneau and K. Kirshenbaum, *Chem. Commun.,* 2014, **50**, 6909-6912; Z. Zhou, J. Zhang, L. J. Sun, G. H. Ma and Z. G. Su, *Bioconj. Chem.,* 2014, **25**, 138-146; Y. M. Li, Y. T. Li, M. Pan, X. Q. Kong, Y. C. Huang, Z. Y. Hong and L. Liu, *Angew, Chem.,* 2014, **53**, 2198-2202.

4. G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Y. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wiken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent and J. A. Bradburne, *Science,* 2003, **299**, 884-887; S. Y. Chen, S. Cressman, F. Mao, H. Shao, D. W. Low, H. S. Beilan, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, L. Savatski, J. W. Adamson, C. E. Bozzini, A. Kung, S. B. Kent, J. A. Bradburne and G. G. Kochendoerfer, *Chem. Biol.,* 2005, **12**, 371-383.
5. C. J. Noren, S. J. AnthonyCahill, M. C. Griffith and P. G. Schultz, *Science,* 1989, **244**, 182-188; J. D. Bain, E. S. Diala, C. G. Glabe, T. A. Dix and A. R. Chamberlin*, J. Am. Chem. Soc.,* 1989, **111**, 8013-8014.
6. H. Koide, S. Yokoyama, G. Kawai, J. M. Ha, T. Oka, S. Kawai, T. Miyake, T. Fuwa and T. Miyazawa, *Proc. Natl. Acad. Sci., U.S.A.*, 1988, **85**, 6237-6241; K. Josephson, M. C. Hartman and J. W. Szostak, *J. Am. Chem. Soc.,* 2005, **127**, 11727-11735; L. Wang, J. Xie and P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struc.,* 2006, **35**, 225-249; C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.,* 2010, **79**, 413-444; L. Wang, A. Brock, B. Hererich and P. G. Schultz, *Science*, 2001, **292**, 498-500; F. Wang, S. Robbins, J. T. Guo, W. J. Shen and P. G. Schults, *PLoS ONE*, 2010, **5**, e9354.
7. A. Deiters, T. A. Cropp, D. Summerer, M. Mukherji and P. G, Schultz, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 5743-5745; N. Li, R. K. V. Lim, S. Edwardraja and Q. Lin, *J. Am. Chem. Soc.,* 2011, **133**, 15316-15319; P. Majkut, V. Böhrsch, R. Serwa, M. Gerrits, and C. P. Hackenberger, *Methods Mol. Biol.,* 2012, **794**: 241-249; A. Dumas, C. D. Spicer, Z. Gao, T. Takehana, Y. A. Lin, T. Yasukohchi and B. G. Davis, *Angew. Chem.,* 2013, **52**, 3916-3921.
8. T. Hohsaka, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.,* 1996, **118**, 9778-9779.
9. N. Shozen, I. Iijima and T. Hohsaka, *Bioorg. Med. Chem. Lett.,* 2009, **19**, 4909-4911.
10. S. Tada, T. Andou, T. Suzuki, N. Dohmae, E. Kobatake and Y. Ito, *PLoS ONE*, 2012, **8**, e49235.
11. D. Kajihara, R. Abe, I. Iijima, C. Komiyama, M. Sisido and T. Hohsaka, *Nat. Methods,* 2006, **3**, 923-929; S. X. Chen, N. E. Fahmi, L. Wang, C. Bhattacharya, S. J. Benkovic and S. M. Hecht, *J. Am. Chem. Soc*., 2013, **135**, 12924-12927.
12. P. B. Lawrence, Y.Gavrilov, S. S. Matthews, M. I. Langlois, D. Shental-Bechor, H. M. Greenblatt, B. K. Pandey, M. S. Smith, R. Paxman, C. D. Torgerson, J. P. Merrell, C. C. Ritz, M. B. Prigozhin, Y. Levy and J. L. Price. *J. Am. Chem. Soc*., 2014, **136**, 17547-17560.
13. J. Mittelstaet, A. L. Konevega and M. V. Rodnina, *J. Am. Chem. Soc.*, 2013, **135**, 17031-17038.
14. T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami and M. Sisi[illegible], *Biochemistry*, 2001, **40**, 11060-11064.
15. M. Sisido, M. Taki, T. Ohtsuki and T. Hohsaka, *Protein, Nucle[illegible] acid, Enzyme*, 2006, **51**, 399-407.
16. R. Abe, K. Shiraga, S. Ebisu, H. Takagi and T. Hohsaka, *J. Biosci. Bioeng*., 2010, **110**, 32-38.