



Cite this: *Chem. Commun.*, 2021, 57, 5434

Received 10th December 2020,  
Accepted 18th February 2021

DOI: 10.1039/d0cc07922g

rsc.li/chemcomm

# Selenium-atom-modified thymidine enhances the specificity and sensitivity of DNA polymerization and detection†

Yang Li,<sup>a</sup> Yingying Zhou,<sup>a</sup> Danyan Luo,<sup>b</sup> Zhaoyi Yang,<sup>b</sup> Lillian Ruoduo Hu<sup>b</sup> and Zhen Huang<sup>\*ab</sup>

Nucleobase mismatches can jeopardize DNA polymerization specificity, causing mutations and errors in DNA replication and detection. Herein we report the first synthesis of novel 2-Se-thymidine triphosphate (<sup>Se</sup>TTP), describe the single-selenium atom-specific modification strategy (SAM) against T/G mismatches, and demonstrate SAM-assisted polymerization and detection with much higher specificity and sensitivity. SAM can effectively suppress the formation of non-specific products in DNA polymerization and detection. Thus, SAM enhances the specificity of DNA synthesis by approximately 10 000 fold, and in turn, it allows the detection of clinical COVID-19 viral RNA in low copy numbers (single-digit copies), while the conventional RT-qPCR does not.

Studying the mechanisms of nucleic acid polymerization and replication is important in molecular and cellular biology as well as nucleic acid detection and diagnostics. Detection of specific nucleic acids is essential for pathogen identification and epidemic control,<sup>1–4</sup> while nucleic acids play vital roles in living systems as genetic materials.<sup>5</sup> In addition to the enzymatic synthesis, many recent studies have demonstrated the non-enzymatic synthesis of nucleic acids.<sup>6,7</sup> Since the base pairs are the basis of nucleic acid recognition,<sup>8,9</sup> they are of great significance in studying the mechanisms of genetic information storage,<sup>10,11</sup> sequence recognition, DNA and RNA diagnosis, DNA replication,<sup>12</sup> RNA transcription<sup>13</sup> and protein translation. Furthermore, the canonical pairing and stacking of DNA nucleobases (T/A and C/G) determine the structure and stability of DNA duplex.<sup>14,15</sup> The non-canonical base pairs (*e.g.*, U/G wobble pair) in RNAs (such as ribozymes and viral RNAs) diversify RNA structures and enhance RNA functions,<sup>16–18</sup> while they (*e.g.*, T/G wobble pair) are not desired in DNA recognition

and synthesis, since they can cause mutations and reduce DNA polymerization specificity.<sup>19,20</sup> This wobble pairing needs to be addressed in order to facilitate highly-specific DNA polymerization and accurate detection.

A T/G wobble pair is formed by the alternative hydrogen-bond (Fig. 1a) through T 2-exo-oxygen, which does not participate in the T/A base pairing.<sup>21</sup> Since the atomic radius of selenium (1.16 Å) is much larger than that of oxygen (0.73 Å) in the same family group of elements, and Se has poor ability in hydrogen-bond formation,<sup>22</sup> we have hypothesized that by substituting the 2-O atom with Se, thymidine triphosphate modified with the 2-Se atom (<sup>Se</sup>TTP) can be used to discriminate against the G template during DNA polymerization, significantly increasing T specificity to the A template. The incorporated Se at the 2-position of thymine discourages the T/G wobble pairing (Fig. 1b) while keeping <sup>Se</sup>T/A the same as canonical T/A without any significant difference, indicated by the UV-melting study.<sup>23</sup> This Se atom can greatly increase the electronic and steric effects, discouraging the wobble pairing, and thereby the Se-atom substitution of thymine 2-oxygen enhances the <sup>Se</sup>T/A specificity.

Herein we report the first synthesis of <sup>Se</sup>TTP and its recognition-and-incorporation by DNA polymerase, offering

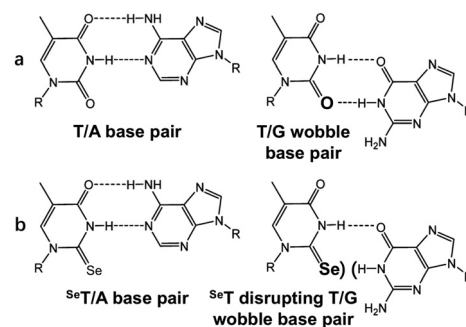


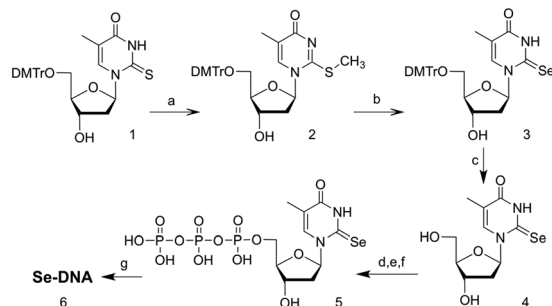
Fig. 1 The base pair structures with or without the Se replacement. (a) T/A pair and T/G wobble pair; (b) <sup>Se</sup>T/A pair without the Se disruption, while <sup>Se</sup>T/G wobble pair with the Se disruption.

<sup>a</sup> Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610000, Sichuan, P. R. China

<sup>b</sup> Szostak-CDHT Large Nucleic Acids Institute & SeNA Research Institute, Chengdu 610000, Sichuan, P. R. China. E-mail: huang@senaresearch.org

† Electronic supplementary information (ESI) available: Experimental section, NMR, Maldi-Tof, sequencing data. See DOI: 10.1039/d0cc07922g



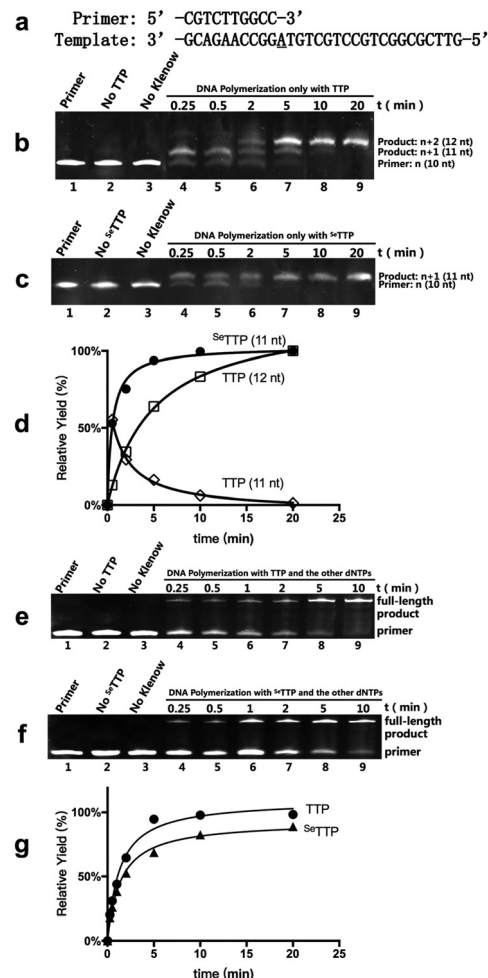


**Scheme 1** Synthesis of  $^{76}\text{TTP}$  and Se-DNA. (a)  $\text{CH}_3\text{I}$ , DIEA, toluene, DCM, room temperature; (b) Se,  $\text{NaBH}_4$ , EtOH; (c) TCA, beta-mercaptoethanol, EtOH; (d) tributyl ammonium pyrophosphate, 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one, DMF, TBA; (e)  $\text{I}_2$ ,  $\text{H}_2\text{O}$ , DMF; (f) triethylamine,  $\text{H}_2\text{O}$ ; (g) DNA polymerization.

much higher specificity and sensitivity for DNA polymerization and detection. Our synthesis started from 5'-trityl-2-thiothymidine **1** (Scheme 1), and  $\text{CH}_3\text{I}$  was used to alkylate the 2-thio-functionality in order to activate the 2-thio-moiety and offer **2**. Compound **2** was reacted with freshly prepared  $\text{NaSeH}$  in order to incorporate the Se functionality,<sup>23</sup> synthesizing **3** in 80% yield, followed by detritylation with trichloroacetic acid, offering 2-seleniothymidine **4** ( $^{76}\text{T}$ ). The generated Se-nucleoside **4** was converted to its triphosphate **5** via the one-pot and mild synthesis.<sup>24,25</sup> The Se-triphosphate (**5**,  $^{76}\text{TTP}$ ) was purified and characterized by NMR, MS and HPLC to confirm its integrity and purity, followed by DNA synthesis with  $^{76}\text{TTP}$  and DNA polymerase, such as Klenow and Taq.

To investigate the  $^{76}\text{TTP}$  recognition by DNA polymerases, we designed a DNA primer (10 nt, 5' FAM) and the "standing-start" template (30 nt) that allow incorporation of one "T" on the underlined template "A" (Fig. 2a). Indeed,  $^{76}\text{TTP}$  offered the extension with only one "T", while canonical TTP allowed the extension with up to three "T"s (Fig. 2b and c). In addition, our time-course experiments showed that in the presence of TTP or  $^{76}\text{TTP}$  and the other three dNTPs, DNA polymerase effectively synthesized the full-length products (Fig. 2e and f), while it didn't synthesize the products in the absence of TTP or  $^{76}\text{TTP}$  under the same conditions. Furthermore, it was exciting to discover that the incorporation rate with  $^{76}\text{TTP}$  was approximately 95% of TTP (Fig. 2b–d). Furthermore, it is worthy to note that in the presence of  $^{76}\text{TTP}$  or TTP and the other three dNTPs (dATP, dCTP and dGTP), the yield of the full-length Se-DNA was approximately 90% of that of the canonical DNA (Fig. 2e–g). Obviously,  $^{76}\text{TTP}$  can be recognized almost as well as the canonical TTP by DNA polymerase and its incorporation specificity is much higher than that of canonical TTP.

To investigate the discrimination of  $^{76}\text{TTP}$  on wobble pairing, we designed DNA templates with G, GG and GGG (Fig. 3a), and included "TC" before the underlined template "G, GG or GGG" in order to allow a "running-start" wobble-pairing extension. We found that in the DNA polymerization with only dATP and dGTP (a negative control), the full-length product via the primer extension was not observed. When canonical TTP was added to the polymerization with only dATP and dGTP, the T/G wobble-pairing



**Fig. 2** The  $^{76}\text{TTP}$  recognition and Se-DNA synthesis by DNA polymerase. The products of DNA extension reactions were analyzed by denaturing PAGE, offering the bands of single-stranded DNAs. (a) DNA 5'-FAM primer and template used here. (b) and (c) Primer extension by Klenow in the presence of only TTP and  $^{76}\text{TTP}$ , respectively; lanes 1, 2, and 3 (negative controls) omitted template, TTP or  $^{76}\text{TTP}$ , and Klenow, respectively. The time-course experiments in lanes 4–10 contained all, including the primer, template, Klenow, TTP or  $^{76}\text{TTP}$ . (d) The plot of the product formation in (b) and (c). (e) and (f) Primer extension by Klenow in the presence of the other three dNTPs and TTP and  $^{76}\text{TTP}$ , respectively; the other conditions were identical to (b) and (c). (g) The plot of the product formation in (e) and (f).

extension was observed (Fig. 3b). The formation of the wobble-extended full-length products was prevented only in the presence of multiple "G" on the templates (at least 3 "G"). However, when  $^{76}\text{TTP}$  was used to replace TTP (Fig. 3c), the wobble-pairing extended products were not observed. This surprising result indicated that  $^{76}\text{TTP}$  discouraged the T/G wobble pairing and polymerization, and the quantitative PAGE analysis is presented in Fig. 3d. In addition, to study the steady-state kinetics of  $^{76}\text{TTP}$  incorporation compared to TTP, the primer and template 1 were used for the investigation, in the presence of dATP and dGTP. We found that the wobble-pairing extension rate of  $^{76}\text{TTP}$  on the G-template was approximately 11 000 folds slower than that of canonical TTP (Fig. 3e–g), shedding light on the high specificity of  $^{76}\text{TTP}$ .

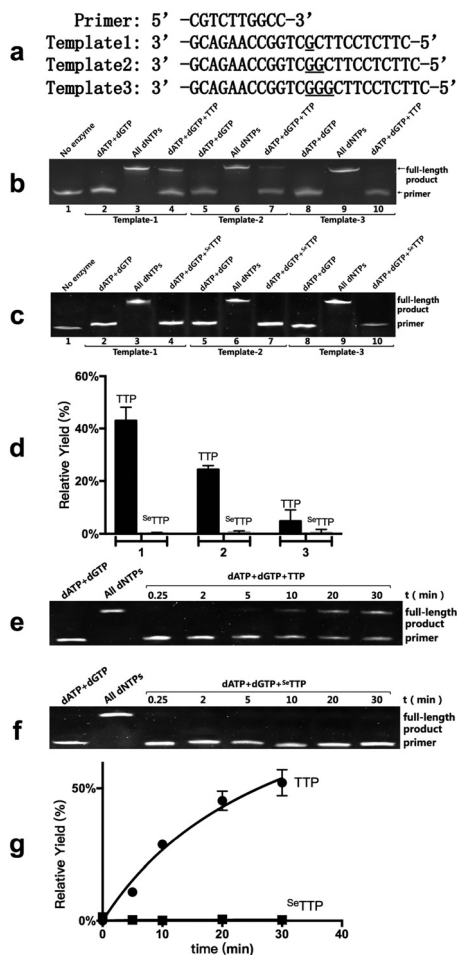


Fig. 3 Facilitation of DNA polymerization specificity by  $^{Se}TTP$ . The products of DNA extension reactions were analyzed by denaturing PAGE, offering the bands of single-stranded DNAs. (a) DNA 5'-FAM primer and templates used in (b), (c), (e) and (f). (b) Primer extension reaction by Taq DNA polymerase with the canonical dNTPs, including TTP. Lane 1: the reaction without Taq (negative control); template-1 was used in lane 2-4; lane 2: the reaction with dATP and dGTP; lane 3: with all dNTPs; lane 4: with dATP, dGTP and TTP; template-2 was used in lane 5-7; lane 5: the reaction with dATP and dGTP; lane 6 with all dNTPs; lane 7: with dATP, dGTP and TTP; template-3 was used in lane 8-10; lane 8: the reaction with dATP and dGTP; lane 9 with all dNTPs; lane 10: with dATP, dGTP and TTP. (c) Primer extension reactions with the conditions identical to (b), where  $^{Se}TTP$  was used instead of TTP. (d) Relative yields of the full-length products in lanes 4, 7 and 10 of (b) and (c). (e) and (f) The time-course experiments of DNA synthesis on T/G mispairing (template-1 for single T/G discrimination) in the presence of TTP and  $^{Se}TTP$ , respectively. (g) The plot of product formation is shown in (e) and (f).

To further demonstrate the specificity value of  $^{Se}TTP$  in DNA cloning and amplification, such as PCR, we designed two different target fragments (Fig. 4). We found that in the absence of  $^{Se}TTP$ , the PCR nonspecific amplification was a serious problem (indicated by the smeared by-products), while in the presence of  $^{Se}TTP$  (50%), the nonspecific products were almost completely suppressed, and the product yields were as high as the canonical. Further, the PCR products were analyzed by sequencing to confirm their integrity. The sequencing results (see ESI,† S15–S22) have indicated that the PCR products with

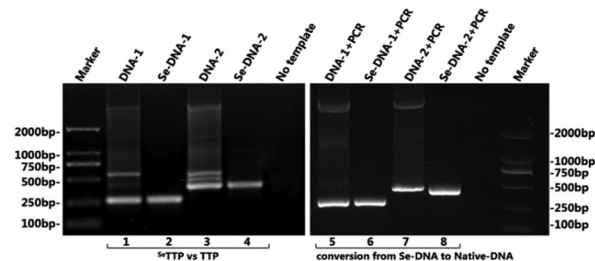


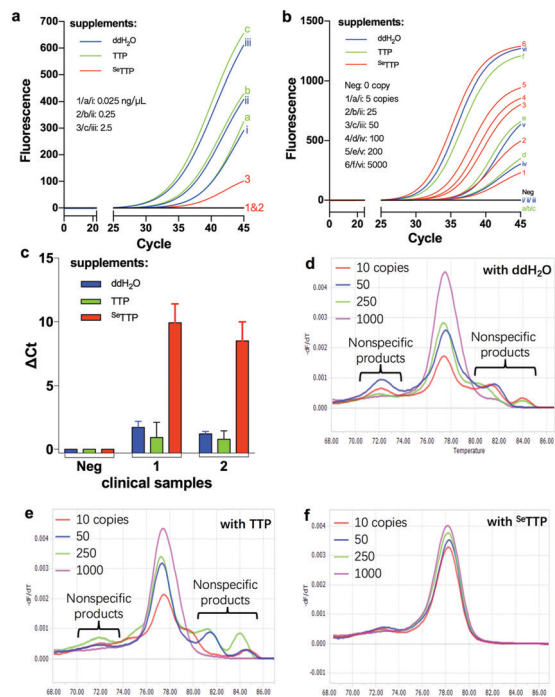
Fig. 4 Inhibition of non-specific PCR DNA amplification with  $^{Se}TTP$ . In lane 1-4, the PCR experiments were performed with two DNA templates by using TTP or  $^{Se}TTP$  with other canonical dNTPs. In lane 5-8, the PCR DNA products from lane 1-4 were diluted 50 times, followed by an additional 10 cycles of PCR with all canonical dNTPs.

TTP or  $^{Se}TTP$  had identical sequences. Furthermore,  $^{Se}T$ -DNA can be easily converted to the corresponding canonical DNA by a few cycles of PCR with canonical dNTPs. Our discovery may offer a new strategy for DNA amplification, gene cloning, reverse transcription and nucleic acid detection.

Nucleic acid detection is essential in the prevention and control of epidemics, such as the one caused by the COVID-19 virus, which has infected over 14 million people and resulted in six hundred thousand deaths worldwide. Unfortunately, it has been reported that nucleic acid detection of COVID-19 viral RNA offered a high false rate (approximately 20% false positive and negative per detection). In order to reduce the false diagnosis rate, we have decided to explore the COVID-19 detection with our unique SAM strategy (Fig. 5). First, we investigated its anti-background interference by including the viral background RNA (namely, human total RNA) in the detection. Since the typical concentration of background RNA in clinical COVID-19 detection is  $0.25 \text{ ng mL}^{-1}$ , we set the detection concentration of background RNA as 0.025, 0.25, and  $2.5 \text{ ng } \mu\text{L}^{-1}$ . Our experiment demonstrated that the Se-approach can successfully suppress the nonspecific amplification (Fig. 5a). No or low fluorescence signal was observed in the presence of supplementary  $^{Se}TTP$ , while high signals were observed similarly with supplementary water or canonical TTP, and the high signals could be mis-interpreted as false positives.

Furthermore, we have found that the Se-strategy can detect viral RNA in up to single-digit 5 copies (Fig. 5b), while the conventional kit can detect up to 100 copies, thereby indicating the higher sensitivity of the Se-strategy (20 folds better) than the conventional one. Furthermore, we examined clinical COVID-19 viral RNA samples (Fig. 5c) and found that our SAM methodology detected 1000-fold-diluted viral RNA samples, while the commercial kits (or supplementary TTP) did not. Moreover, in order to investigate the non-specific product suppression, we performed RT-qPCR and melting-temperature studies<sup>26</sup> and discovered that in the presence of  $^{Se}TTP$ , the non-specific products were not observed, while they were when supplemented with water or TTP, indicated by multiple peaks (Fig. 5d-f). Our study indicated the severe non-specific amplification by RT-qPCR, especially when the copy number of viral RNAs was low (Fig. 5d and e). However, non-specific amplification in nucleic acid detection can be significantly inhibited by supplementary  $^{Se}TTP$  (Fig. 5f).





**Fig. 5** The selenium strategy for COVID-19 detection via RT-qPCR commercial kits and melting study. (a) Inhibition of viral background RNA (human total RNA) for RT-qPCR detection with supplementary  $^{\text{Se}}\text{TTP}$ . In the presence of the background RNA (0.025, 0.25, and 2.5 ng  $\mu\text{L}^{-1}$ ), group i–iii (blue), group a–c (green) and group 1–3 curves (red) were collected with supplementary ddH<sub>2</sub>O, TTP and  $^{\text{Se}}\text{TTP}$  (0.25 mM), respectively. (b) Detection sensitivity enhancement with supplementary  $^{\text{Se}}\text{TTP}$ . In the presence of viral RNA (5, 25, 50, 100, 200 and 5000 copies), group i–vi (blue), group a–f (green) and group 1–6 curves (red) were collected with supplementary ddH<sub>2</sub>O, TTP and  $^{\text{Se}}\text{TTP}$  (0.25 mM), respectively. (c) Viral detection of COVID-19 clinical samples with supplementary ddH<sub>2</sub>O, TTP and  $^{\text{Se}}\text{TTP}$ , respectively.  $\Delta\text{Ct}$  equals Ct value minus its negative control Ct;  $\Delta\text{Ct} < 1$  is defined as negative detection result. (d–f) In the presence of human total RNA (0.25 ng  $\mu\text{L}^{-1}$ ), the RT-qPCR melting curves of COVID-19 viral RNA (10, 50, 250 and 1000 copies) with supplementary ddH<sub>2</sub>O (d), TTP (e) and  $^{\text{Se}}\text{TTP}$  (f), respectively.

In summary, we have first synthesized 2-seleno-thymidine triphosphate ( $^{\text{Se}}\text{TTP}$ ) and  $^{\text{Se}}\text{T}$ -DNAs via DNA polymerization. We have found that DNA polymerase can effectively recognize the Se-triphosphate, almost as well as the canonical one. Furthermore, we have discovered high discrimination of the 2-Se-thymine against T/G mis-pairing and polymerization, and the SAM triphosphate is over 10 000 fold more discriminative than the canonical one. Furthermore, we have demonstrated this unique SAM strategy for DNA polymerization, amplification and nucleic acid detection (such as PCR and RT-PCR) with high specificity, sensitivity and product purity. In addition, SAM provides a novel approach for investigating base-pair recognition, molecular interaction and nucleic acid polymerization, such as DNA replication, RNA transcription and reverse transcription. Moreover, this SAM strategy can

be useful in nucleic acid detection, such as COVID-19 epidemic diagnostics, control and prevention by enhancing the sensitivity and accuracy of nucleic acid detection.

This work was supported by the National Natural Science Foundation of China (21761132029 and 22077089), SeNA Research Institute, and the Fundamental Research Funds for the Central Universities.

## Conflicts of interest

There are no conflicts to declare.

## Notes and references

- 1 J.-W. Ai, H.-C. Zhang, T. Xu, J. Wu, M. Zhu, Y.-Q. Yu, H.-Y. Zhang, Z. Shen, Y. Li, X. Zhou, G.-Q. Zang, J. Xu, W.-J. Chen, Y.-J. Li, D.-S. Xie, M.-Z. Zhou, J.-Y. Sun, J.-Z. Chen and W.-H. Zhang, *medRxiv*, 2020, DOI: 10.1101/2020.02.13.20022673.
- 2 L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon and J. Lessler, *Ann. Intern. Med.*, 2020, **173**(4), 262–267.
- 3 S. Ren, Q. Wang, P. Yang, S. Cui, T. Yuan, D. Zhang, L. Long and Y. Pan, *Clin. Chem.*, 2020, **66**, 794–801.
- 4 S. Woloshin, N. Patel and A. S. Kesselheim, *N. Engl. J. Med.*, 2020, **383**(6), e38.
- 5 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
- 6 M. Sosson, D. Pfeffer and C. Richert, *Nucleic Acids Res.*, 2019, **47**, 3836–3845.
- 7 J. Niu, R. Hili and D. R. Liu, *Nat. Chem.*, 2013, **5**, 282–292.
- 8 A. M. Sismour and S. A. Benner, *Nucleic Acids Res.*, 2005, **33**, 5640–5646.
- 9 R. Laos, C. Lampropoulos and S. A. Benner, *Acta Crystallogr., Sect. C: Struct. Chem.*, 2019, **75**, 22–28.
- 10 G. Storz, *Science*, 2002, **296**, 1260–1263.
- 11 Y. Zhang, J. L. Ptacin, E. C. Fischer, H. R. Aerni, C. E. Caffaro, K. S. Jose, A. W. Feldman, C. R. Turner and F. E. Romesberg, *Nature*, 2017, **551**, 644–647.
- 12 Y. Gao, Y. Cui, T. Fox, S. Lin, H. Wang, N. de Val, Z. H. Zhou and W. Yang, *Science*, 2019, **363**(6429), eaav7003.
- 13 S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov and T. Carell, *Science*, 2019, **366**, 76–82.
- 14 G. Ferry, *Nature*, 2019, **575**, 35–36.
- 15 A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris and D. R. Liu, *Nature*, 2016, **533**, 420–424.
- 16 Z. Lan and J. A. Doudna, *Science*, 2002, **295**, 2084–2088.
- 17 C. Roost, S. R. Lynch, P. J. Batista, K. Qu, H. Y. Chang and E. T. Kool, *J. Am. Chem. Soc.*, 2015, **137**(5), 2107–2115.
- 18 C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier and T. Carell, *Angew. Chem., Int. Ed.*, 2018, **57**, 5943–5946.
- 19 O. Kennard, *J. Biomol. Struct. Dyn.*, 1985, **3**, 205.
- 20 J. R. Kiefer, C. Mao, J. C. Braman and L. S. Beese, *Nature*, 1998, **391**, 304–307.
- 21 L. E. Orgel, *J. Mol. Biol.*, 1968, **38**, 381–393.
- 22 H. Sun, S. Jiang, J. Caton-Williams, H. Liu and Z. Huang, *RNA*, 2013, **19**, 1309–1314.
- 23 A. E. Hassan, J. Sheng, W. Zhang and Z. Huang, *J. Am. Chem. Soc.*, 2010, **132**, 2120–2121.
- 24 J. Caton-Williams, M. Smith, N. Carrasco and Z. Huang, *Org. Lett.*, 2011, **13**, 4156–4159.
- 25 B. Hu, Y. Wang, S. Sun, W. Yan, C. Zhang, D. Luo, H. Deng, L. R. Hu and Z. Huang, *Angew. Chem., Int. Ed.*, 2019, **58**, 7835–7839.
- 26 T. B. Morrison, J. J. Weis and C. T. Wittwer, *Biotechniques*, 1998, **24**, 954–962.

