



Cite this: *Chem. Sci.*, 2021, **12**, 4063

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 10th December 2020
Accepted 29th January 2021

DOI: 10.1039/d0sc06770a
rsc.li/chemical-science

Introduction

Sequence information transfer is the basis for the evolution of living systems.¹ At the molecular level, information is transmitted from a parent to a daughter nucleic acid oligomer through a replication step, which is mediated by polymerase enzymes. A certain degree of mutation is required in this replication process for evolution to take place. Thus, the new oligomer sequences differ somewhat from the parent templates ensuring sequence variation and allowing evolutionary selection. Success in evolution requires a precise control of the mutation rate, which must be high enough to create a subtle variation in the sequences of the daughter strand population, but small enough to avoid the loss of the information encoded in the parent template.² The principles governing molecular evolution, *i.e.* replication and mutation of sequence information, have been harnessed for the development of evolutionary processes to search chemical space.^{3–6} Directed evolution methods represent a powerful strategy for tailoring the function

Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: herchelsmith.orgchem@ch.cam.ac.uk

† Electronic supplementary information (ESI) available: Materials and methods, detailed synthetic procedures, full characterization including ¹H and ¹³C NMR spectra of all compounds and the UPLC traces and MS spectra for the replication and mutation cycles are available in the supplementary material. See DOI: 10.1039/d0sc06770a

Controlled mutation in the replication of synthetic oligomers†

Diego Núñez-Villanueva and Christopher A. Hunter *

Replication of sequence information with mutation is the molecular basis for the evolution of functional biopolymers. Covalent template-directed synthesis has been used to replicate sequence information in synthetic oligomers, and the covalent base-pairs used in these systems provide an opportunity to manipulate the outcome of the information transfer process through the use of traceless linkers. Two new types of covalent base-pair have been used to introduce mutation in the replication of an oligotriazole, where information is encoded as the sequence of benzoic acid and phenol monomer units. When a benzoic acid–benzoic acid base-pairing system was used, a direct copy of a benzoic acid homo-oligomer template was obtained. When a phenol–benzoic acid base-pairing system was used, a reciprocal copy, the phenol homo-oligomer, was obtained. The two base-pairing systems are isosteric, so they can be used interchangeably, allowing direct and reciprocal copying to take place simultaneously on the same template strand. As a result, it was possible to introduce mutations in the replication process by spiking the monomer used for direct copying with the monomer used for reciprocal copying. The mutation rate is determined precisely by the relative proportions of the two monomers. The ability to introduce mutation at a controlled rate is a key step in the development of synthetic systems capable of evolution, which requires replication with variation.

of biomolecules for therapeutic or manufacturing applications.^{7–10} However, current methods rely on nucleic acid replication, so the accessible chemical space is therefore limited to nucleic acids or proteins.^{11–13}

Our aim is to extend molecular evolution to fully synthetic systems in order to access new non-natural oligomers with tailored properties defined by the sequence of the monomer building blocks. For that purpose, it is crucial to develop template-directed synthetic methods for the replication of information encoded in synthetic oligomers.^{14–21} We have recently demonstrated how covalent template-directed synthesis can be used to transfer of sequence information from a parent template to a sequence-complementary copy (Fig. 1).^{22,23}

The first step in Fig. 1 is attachment of the monomer building blocks to the template using covalent base-pairing based on the formation of an ester between a phenol (P) and a benzoic acid (A).²² Copper catalyzed azide alkyne cycloaddition (CuAAC) is used in the ZIP step to give the covalent duplex,^{24,25} and cleavage of the ester base-pairs regenerates the template along with the complementary copy. If such synthetic replication systems are eventually to be implemented in evolution experiments, the other key requirement is a method for introduction of variation through mutation. Here we describe a strategy for controlled mutation of the sequence information contained in a population of copy strands in covalent template-directed replication.



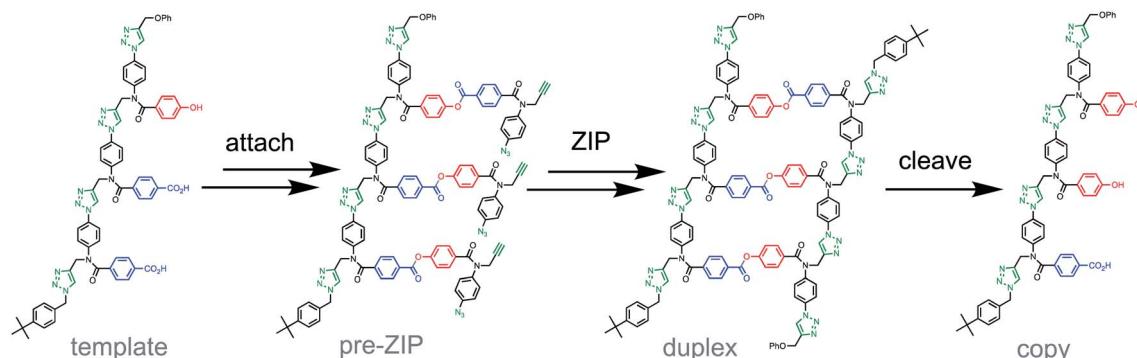


Fig. 1 Sequence information transfer using covalent template-directed synthesis. In the attach step, complementary monomers are loaded onto the template using phenol–benzoic acid ester base-pairing chemistry. In the ZIP step, intramolecular CuAAC reactions lead to oligomerization of monomers on the template followed by capping of the chain ends. In the cleave step, hydrolysis of the ester bonds connecting the new oligomer to the template regenerates the template and releases the complementary copy strand.

Approach

An interesting feature of covalent base-pairing is that the information that is transferred in the replication process can be adjusted by changing the chemical structure of the base-pair. For example, Fig. 2 shows how a traceless linker was used to make an identical copy rather than a complementary copy of the sequence information in a template strand.²³ Direct coupling between phenol and benzoic acid units leads to reciprocal replication of chemical information, analogous to nucleic acid replication (Fig. 2, top channel).²² However, if a linker is used to connect two benzoic acid units in a symmetric base-pair, then

the replication process yields an identical copy of the template strand (Fig. 2, bottom channel).²³ Here we apply this traceless linker concept to develop a new set of covalent base-pairs, which can be used to introduce mutations into a covalent template-directed replication process and to tune the mutation rate in a predictable manner.

Fig. 3 shows base-pairing systems that could be used for formation of either identical or complementary sequence copies of a template strand. There is one key difference from the base-pairing systems shown in Fig. 2, which is that the three base-pairs in Fig. 3 are all isosteric. These three base-pairs can therefore be accommodated interchangeably within the same product duplex and used simultaneously in the same replication process. By attaching monomer units to a template using mixtures of symmetric and unsymmetric base-pairs, it should be possible to obtain a mixture of reciprocal and direct copying of sequence information in the replication process, thereby introducing mutation.

Fig. 4 illustrates the principle for replication of a homooligomer. If a mixture of two different monomers is used in the attach step, all possible stochastic combinations of the symmetric and the unsymmetric base-pairs will be obtained in the pre-ZIP intermediate. After the ZIP and cleave steps, the template is regenerated along with an identical copy and all possible mutant sequences. The mutation rate can be controlled simply by changing the relative amounts of the two

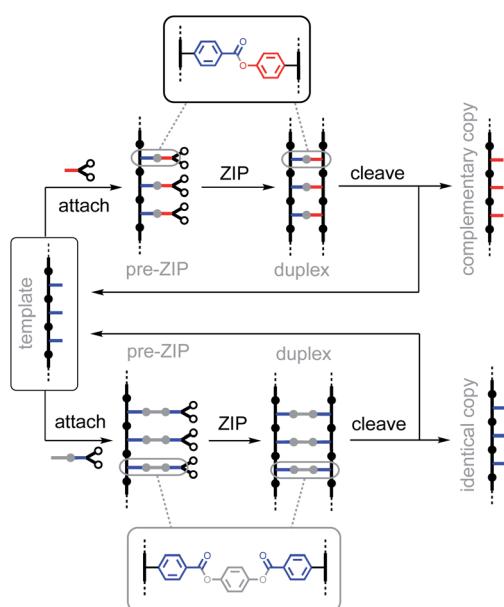


Fig. 2 The information transferred from the template to the daughter strand can be programmed by using different covalent base-pairing systems. Top channel: base pairs formed by the direct attachment of phenol and benzoic acid give rise to a copy strand which is complementary to the template. Bottom channel: base-pairs formed by connecting two benzoic acids with a linker give rise to an identical copy of the template strand.

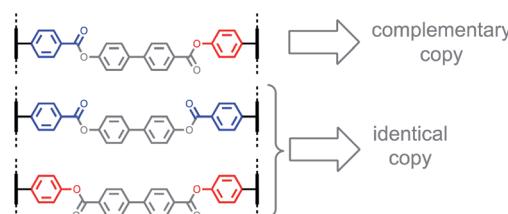


Fig. 3 Isosteric base-pairs can be used interchangeably to control how sequence information is transferred in covalent template-directed replication of phenol–benzoic acid oligomers. Unsymmetric base-pairs give rise to reciprocal replication (sequence-complementary copies), and symmetric base-pairs give rise to direct replication (identical copies).



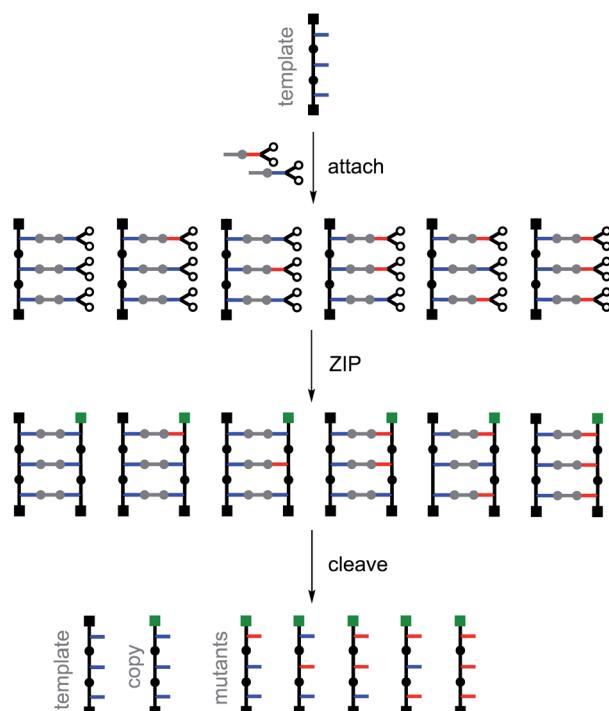


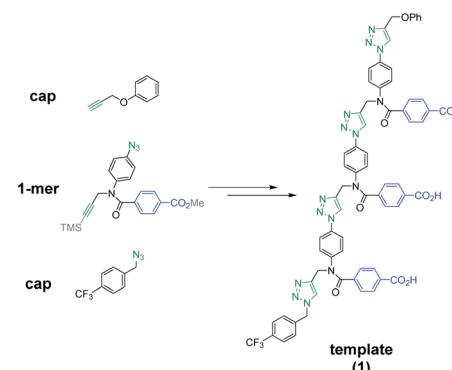
Fig. 4 Replication with mutation. A mixture of two different monomers is used in the attach step to generate a mixture of symmetric and unsymmetric base-pairs. After the ZIP and cleave steps, the template is regenerated along with a direct copy and all possible mutant sequences. The products of the replication process can be distinguished from the template by different end capping groups (green) used in the ZIP step.

monomers used in the attach step. The experiments described below demonstrate realization of the error-prone replication process illustrated in Fig. 4 by using the isosteric base-pairing system shown in Fig. 3.

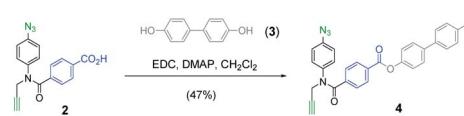
Results and discussion

Scheme 1 shows the structure of the homosequence template (1) used in this study. The 3-mer template was obtained as described previously from the corresponding TMS-protected 1-mer by CuAAC capping with phenyl propargyl ether, followed by sequential deprotection and CuAAC coupling with more monomer. The terminal alkyne was then capped with 4-trifluoromethylbenzyl azide, and the methyl esters were hydrolysed to reveal the carboxylic acids required for base-pair formation.²³

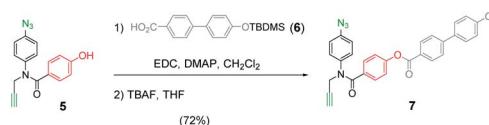
The mutation protocol shown in Fig. 4 requires two different monomers to be attached simultaneously to the template. Scheme 2 shows the synthesis of monomer 4 bearing a 4,4'-biphenol linker, which encodes for the symmetric benzoic acid–benzoic acid base-pair. Ester coupling of benzoic acid 2 with excess 4,4'-biphenol yielded 4 in moderate yield.²² Scheme 3 shows the synthesis of monomer 7 bearing a 4,4'-hydroxy-4-biphenylcarboxylic acid linker, which encodes for the unsymmetric phenol–benzoic acid base-pair. Ester coupling of phenol 5 with silyl-protected 4,4'-hydroxy-4-biphenylcarboxylic acid (6)



Scheme 1 Structure of template 1.



Scheme 2 Synthesis of 4.



Scheme 3 Synthesis of 7.

followed by TBAF-mediated removal of the phenol protecting group afforded 7 in good yield.²²

Fig. 5 shows the reaction steps carried out to mutate the information encoded in template 1 *via* covalent template-directed replication. The first step is attachment of a mixture of monomers 4 and 7 to the template *via* ester bond formation to obtain the corresponding mixture of pre-ZIP intermediates. We carried out five independent experiments with different ratios of monomer 4, which would result in a direct copy of the template sequence, and monomer 7, which leads to mutation. ¹H NMR analysis of the mixture of pre-ZIP intermediates indicated that the reactivity of the two monomers used in the ester coupling reaction is very similar. There is a difference of 0.02 ppm between the signals due to the alkyne protons of the phenol and benzoic acid monomers, so deconvolution of these signals was used to determine the loading of each monomer on the template in the attach step (see Fig. S7 and S8, and Table S1 of the ESI for details†). The proportion of mutator monomer (χ_{mutator}) found attached to the template in the pre-ZIP intermediate (0%, 38%, 54%, 75% or 100%) is similar to the proportion present in the reaction mixture (0%, 30%, 50%, 70% or 100%) in each of the five experiments.

For each of the five different mixtures of pre-ZIP intermediates, the CuAAC ZIP reaction was then carried out in the presence of an excess of 4-*t*-butylbenzyl azide, which blocks intermolecular reactions between two different oligomers and prevents intramolecular macrocyclization reactions on the template.^{24,25} After washing with petroleum ether to remove excess capping azide, phenyl propargyl ether was used to cap

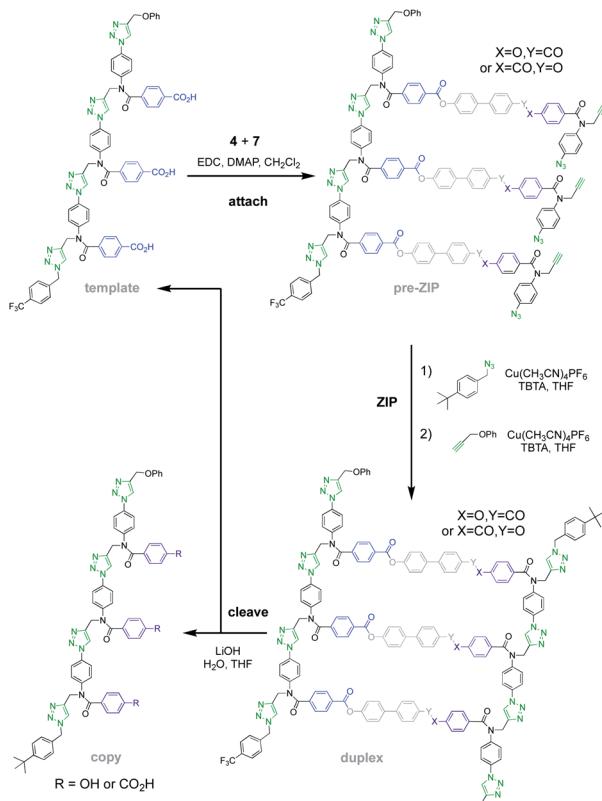


Fig. 5 Covalent template-directed replication and mutation of the chemical information encoded in template 1. Ester base-pair chemistry was used to attach a mixture of monomers 4 and 7 to the template. The resulting pre-ZIP intermediates were subjected to intramolecular CuAAC reaction in the presence of 4-t-butylbenzyl azide to block the alkyne chain end. The azide chain end was capped with phenyl propargyl ether to give the corresponding duplexes (only the antiparallel arrangement of the backbones is shown, but parallel is also possible). Hydrolysis of the ester base-pairs in the cleave step regenerated the template together with the templated products (8–15). The copied products are distinguished from the original template by the terminal *t*-butyl group.

the unreacted terminal azide groups of the daughter strands. The resulting mixture of duplexes was cleaved using basic hydrolysis to release the traceless linkers and regenerate the template 1 along with the templated products 8–15 (Fig. 6).

The product distributions from the replication cycle shown in Fig. 5 were determined using UPLC (Fig. 7). When the replication cycle was carried out using only monomer 4 in the attach step, a direct copy of the template (AAA, 8) was obtained as the only product (Fig. 7, $\chi_{\text{mutator}} = 0$). When only monomer 7 was used in the attach step, a reciprocal copy of the template (PPP, 15) was obtained as the only product (Fig. 7, $\chi_{\text{mutator}} = 1$). When mixtures of the two monomers were used, a mixture of product sequences was obtained, and the UPLC traces in Fig. 7 show clearly that the product distribution is directly related to the amount of mutator monomer 7 used in the attach step.

Deconvolution of the UPLC peaks was used to quantify the product distribution (see ESI for details[†]). For reference purposes, each of the eight possible product sequences shown in Fig. 6 was synthesized independently (see ESI for details[†]).

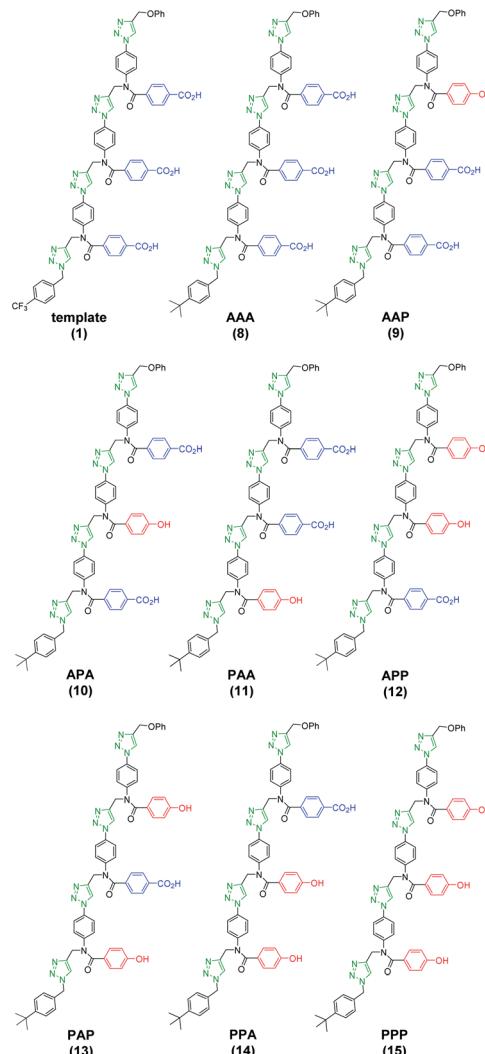


Fig. 6 Molecular structure of all possible product sequences after covalent template-directed replication of template 1 using a mixture of monomers 4 and 7. Sequences are written from the benzyl terminus using a 2-letter code: A for benzoic acid and P for phenol.

The UPLC trace labelled “Reference” in Fig. 7 shows an equimolar mixture of each of these oligomers and template 1. We were not able to fully resolve the three isomeric sequences containing two phenols and one benzoic acid (PAP, APP and PPA), so the yield of these oligomers was determined using the total area of the peak corresponding to all three isomers. Similarly, the three sequences with one phenol and two benzoic acids (PAA, APA and AAP) were not fully resolved, so these peaks were combined to determine the total yield of the three isomeric oligomers. Fig. 8 compares the experimental results with the theoretical product distribution calculated assuming statistical incorporation of the two different monomers. The agreement indicates that none of the steps in the replication process are strongly dependent on the difference between the chemical structures of the two monomers. Thus the mutation rate in the replication process can be directly determined by choosing an appropriate monomer composition used in the attach step.



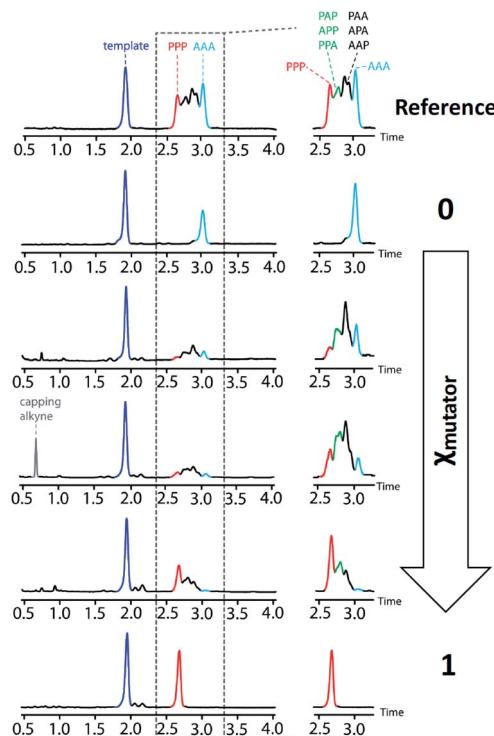


Fig. 7 UPLC chromatograms from a cycle of covalent template-directed replication and mutation of the chemical information encoded in template 1 using different amounts of mutator 7 in the attach step (χ_{mutator}). The top trace was created as a reference by mixing equimolar amounts of pure samples all eight product sequences (8–15) with template 1. Experimental details are provided in the ESI.[†]

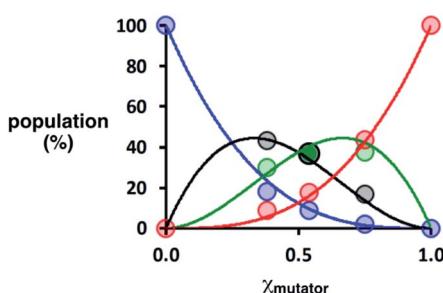


Fig. 8 Theoretical (lines) and experimental (dots) product distributions for covalent template-directed replication of template 1 in the presence of different amounts of mutator monomer 7 (χ_{mutator}). The total populations of isomeric sequences are plotted as a single point: the sum of the overlapped peaks corresponding to PAA, APA and AAP in black, and the sum of PPA, PAP and APP in green. The population of the direct copy AAA is shown in blue, and the reciprocal copy PPP in red.

Conclusions

These experiments demonstrate that the use of covalent base-pairing strategies for the replication of sequence information encoded in synthetic oligomers opens up new opportunities that are not available with the non-covalent base-pairs used in biology. More complex information transfer processes are

possible by manipulation of the chemistry used to connect the two components of a covalent base-pair. Here, isosteric covalent base-pairs with traceless linkers have been used to introduce mutation in the replication of an oligotriazole, where information is encoded as the sequence of benzoic acid and phenol monomer units. When a benzoic acid–benzoic acid base-pairing system was used, a direct copy of a benzoic acid homo-oligomer template was obtained. When a phenol–benzoic acid base-pairing system was used, a reciprocal copy, the phenol homo-oligomer, was obtained. The two base-pairing systems are isosteric, which means that they can be used interchangeably, allowing direct and reciprocal copying to take place simultaneously on the same template strand. As a result, it is possible to introduce mutations in a replication process by spiking the monomer used for direct copying with the monomer used for reciprocal copying. We show that the mutation rate is determined precisely by the relative proportions of the two monomers. The ability to introduce mutation at a controlled rate is a key step in the development of synthetic systems capable of evolution, which requires replication with variation.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the Engineering and Physical Sciences Research Council (EP/P027067/1) and the Herchel Smith Fund for funding. We thank Dr Giulia Iadevaia for the assistance with the UPLC equipment.

Notes and references

- 1 C. B. Anfinsen, *The molecular basis of evolution*, Wiley, 1959.
- 2 A. J. F. Griffiths, W. M. Gelbart, J. H. Miller and R. C. Lewontin, *Modern genetic analysis*, Freeman, 1999.
- 3 C. Tuerk and L. Gold, *Science*, 1990, **249**, 505–510.
- 4 A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818–822.
- 5 G. F. Joyce, *Angew. Chem., Int. Ed.*, 2007, **46**, 6420–6436.
- 6 M. S. Morrison, C. J. Podracky and D. R. Liu, *Nat. Chem. Biol.*, 2020, **16**, 610–619.
- 7 K. Chen and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5618–5622.
- 8 M. Famulok, G. Mayer and M. Blind, *Acc. Chem. Res.*, 2000, **33**, 591–599.
- 9 N. J. Turner, *Nat. Chem. Biol.*, 2009, **8**, 567–573.
- 10 U. T. Bornscheuer, B. Hauer, K. E. Jaeger and U. Schwaneberg, *Angew. Chem., Int. Ed.*, 2019, **58**, 36–40.
- 11 J. W. Chin, *Nature*, 2017, **550**, 53–60.
- 12 D. L. Usanov, A. I. Chan, J. P. Maianti and D. R. Liu, *Nat. Chem.*, 2018, **10**, 704–714.
- 13 J. C. W. Willis and J. W. Chin, *Nat. Chem.*, 2018, **10**, 831–837.
- 14 X. Li, Z.-Y. J. Zhan, R. Knipe and D. G. Lynn, *J. Am. Chem. Soc.*, 2002, **124**, 746–747.



15 P. K. Lo and H. F. Sleiman, *J. Am. Chem. Soc.*, 2009, **131**, 4182–4183.

16 A. Cnossen, C. Roche and H. L. Anderson, *Chem. Commun.*, 2017, **53**, 10410–10413.

17 N. Kamonsutthipaijit and H. L. Anderson, *Chem. Sci.*, 2017, **8**, 2729–2740.

18 K. R. Strom, J. W. Szostak and N. Prywes, *J. Org. Chem.*, 2019, **84**, 3754–3761.

19 N.-T. Lin, S.-Y. Lin, S.-L. Lee, C.-h. Chen, C.-H. Hsu, L.-P. Hwang, Z.-Y. Xie, C.-H. Chen, S.-L. Huang and T.-Y. Luh, *Angew. Chem., Int. Ed.*, 2007, **46**, 4481–4485.

20 Y.-Z. Ke, S.-L. Lee, C.-h. Chen and T.-Y. Luh, *Chem.-Asian J.*, 2011, **6**, 1748–1751.

21 Y.-Z. Ke, R.-J. Ji, T.-C. Wei, S.-L. Lee, S.-L. Huang, M.-J. Huang, C.-h. Chen and T.-Y. Luh, *Macromolecules*, 2013, **46**, 6712–6722.

22 D. Núñez-Villanueva, M. Ciaccia, G. Iadevaia, E. Sanna and C. A. Hunter, *Chem. Sci.*, 2019, **10**, 5258–5266.

23 D. Núñez-Villanueva and C. A. Hunter, *Org. Biomol. Chem.*, 2019, **17**, 9660–9665.

24 M. Ciaccia, D. Núñez-Villanueva and C. A. Hunter, *J. Am. Chem. Soc.*, 2019, **141**, 10862–10875.

25 D. Núñez-Villanueva, M. Ciaccia and C. A. Hunter, *RSC Adv.*, 2019, **9**, 29566–29569.

