

Molecular replication using covalent base-pairs with traceless linkers

Diego Núñez-Villanueva^a and Christopher A. Hunter^{*a}

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

DOI: 10.1039/x0xx00000x

Formation of esters between phenols and benzoic acids has been used as a base-pairing strategy for sequence information transfer in template-directed synthesis of linear oligomers, but the copy strand produced by this process has the complementary sequence to the template strand. A unique feature of kinetically inert covalent base-pairing is that the nature of the chemical information that is transferred can be modulated by changing the chemical connectivity between the two bases. It is possible to form a base-pair between two benzoic acids by using a hydroquinone linker, which is eliminated when the product duplex is hydrolysed. Using this approach, covalent template-directed synthesis was carried out using a benzoic acid 3-mer template to produce an identical copy. This direct replication process was used in iterative rounds of replication leading to an increase of the population of the copied oligomer.

Introduction

Sequence information transfer is the basis for the evolution of living systems on Earth. The replication, transcription and translation of the molecular information encoded by nucleic acids is achieved through non-covalent template-directed synthesis.^{1,2} The high fidelity of these processes is achieved by complex enzymatic machinery that controls binding of the correct monomer to the template and covalent connection to the growing chain.³⁻⁵ Molecular replication is the basis of the directed evolution strategy, which has been used to search biopolymer sequence space for functional molecules.^{6,7} As nucleic acids are the only molecules capable of sequence information transfer, directed evolution methods are currently limited to nucleic acids, or polymers that can be interfaced with

them such as proteins.⁸⁻¹¹ Although synthetic replicating systems have been reported, these are literally two-bit information systems, because replication is based on the coupling of two components.¹²⁻²⁰ Molecular replication based on oligomerisation reactions requires a different approach.

In order to extend evolutionary search methods to non-natural oligomers, we have been investigating a chemical approach to the template-directed synthesis of oligomers.^{21,22} Figure 1 shows how covalent base-pairs can be used as the basis for sequence information transfer between parent and daughter strands.²¹ The oligomer backbone was assembled using copper catalysed azide-alkyne cycloaddition (CuAAC) reactions, and esters formed between the phenol (red) and benzoic acid (blue) side-chains provided the covalent base-pairing chemistry.²³⁻³² In the first step of the copying process,

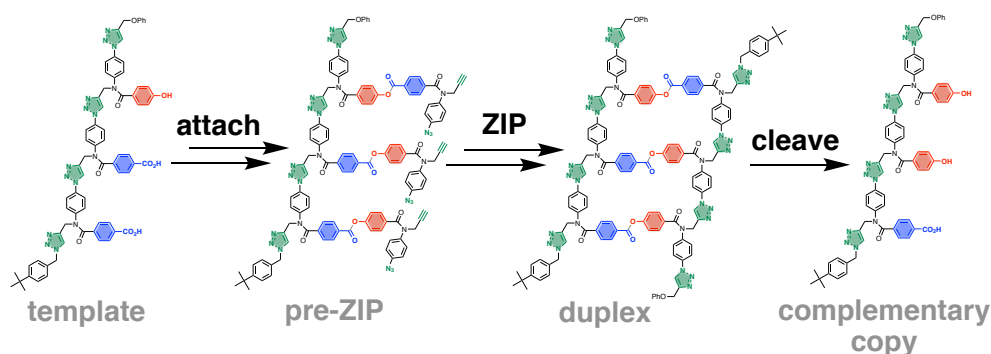


Figure 1. Sequence information transfer using covalent template-directed synthesis. In the **attach** step, monomers are coupled with complementary groups on the template using phenol-benzoic acid ester base-pairing chemistry. In the **ZIP** step, intramolecular CuAAC reactions lead to oligomerisation of monomers on the template and then the chain ends are capped. In the **cleave** step, the ester bonds connecting the new oligomer to the template are broken to regenerate the template and release the copy strand, which has a complementary sequence to the template.

^a 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 UPLC traces for full replication cycles. See DOI: 10.1039/x0xx00000x

monomer building blocks were attached to complementary bases on the template strand by a sequence of phenol protection, attachment of phenol monomers by ester coupling, deprotection, and attachment of benzoic acid monomers by ester coupling. In the ZIP step, the monomer units were oligomerised on the template using a high dilution CuAAC and in the presence of a chain capping agent to prevent competing intermolecular processes. Hydrolytic cleavage of the ester base-pairs regenerated the template and released the complementary copy strand.

The use of covalent base-pairing provides a number of interesting features that are not available to non-covalent systems.³³⁻⁴⁴ All of the intermediates in the process are kinetically stable, which allows isolation and characterisation, providing a straightforward method for identifying the origin of any side-products.²¹ The formation of base-pairs can be carried out at high concentrations to ensure quantitative yields in the attach step, while the ZIP reaction can be carried out at low dilution to minimise competing intermolecular processes. Here we show that the introduction of a linker between the two bases in a covalent base-pair can be used to modulate the nature of the information that is transferred in the copying process.

The phenol-benzoic acid base-pairing strategy used in the process shown in Figure 1 gives a copy strand with a sequence which is complementary to the template.⁴⁵ Figure 2 illustrates the use of a traceless linker to connect two identical bases, which leads to a copy strand with a sequence which is the same as the original template. Incorporation of a hydroquinone linker (grey) allows the formation of a base-pair between two benzoic

acids (blue), which leads to direct replication of benzoic acid oligomers. Multiple cycles of direct replication should exponentially amplify the template (*cf* PCR amplification of DNA).⁴⁶ Here we describe the direct replication of an oligomer using a disposable hydroquinone linker to connect benzoic acid bases on the parent and daughter strand.

Results and discussion

Synthesis

Scheme 1 shows the synthesis of a 3-mer template **7** bearing three benzoic acid bases. The synthesis of building blocks **1** and **2** have been described previously.^{21,22} CuAAC coupling of **1** with **2** followed by TBAF-mediated deprotection of the TMS group gave the 2-mer **3** in excellent yield. Application of the same CuAAC coupling and TMS deprotection procedure to **3** afforded the 3-mer **4** in excellent yield. CuAAC capping of **4** with *p*-trifluoromethylbenzyl azide **5** provided the fully capped 3-mer **6**, and hydrolysis of the methyl esters using LiOH gave template **7** in good yield.

Scheme 2 shows the synthesis of **10**, a benzoic acid monomer equipped with the hydroquinone linker required for direct replication via benzoic acid-benzoic acid base-pairing. The synthesis of benzoic acid monomer **8** was reported previously.²¹ Ester coupling of **8** with TBDMS-protected hydroquinone **9**⁴⁷ followed by deprotection of the silyl group provided **10** in good yield.

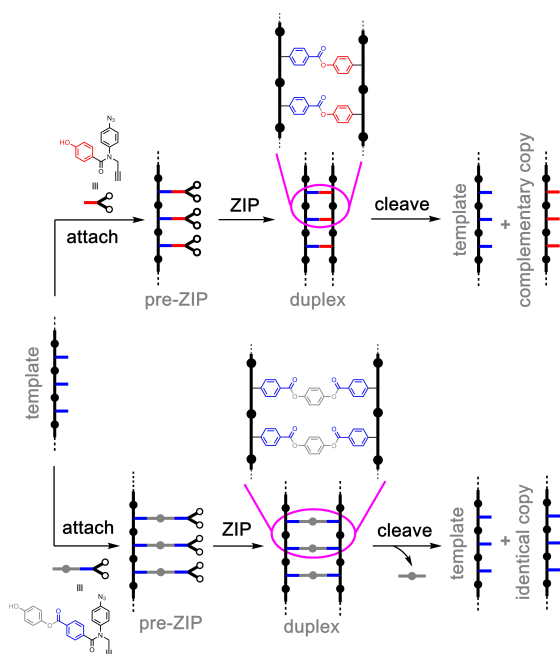
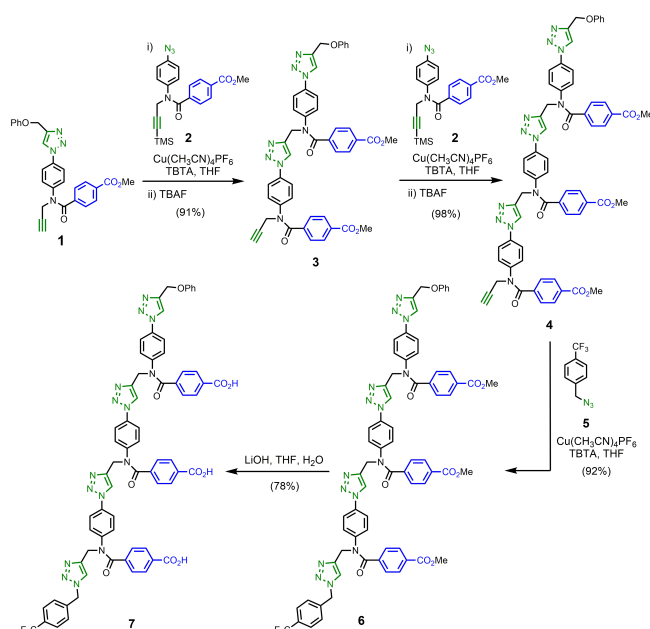
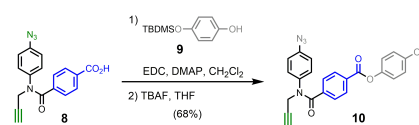


Figure 2. Two different base-pairing strategies to control information transfer in covalent template-directed synthesis of linear oligomers. Top: phenol-benzoic acid base-pairs give rise to a copy strand which is complementary to the template. Bottom: benzoic acid-benzoic acid base-pairs connected via a traceless hydroquinone linker give rise to a copy strand which is identical to the template.



Scheme 1. Synthesis of 3-mer template **7**.



Scheme 2. Synthesis of linker-enhanced monomer **10**.

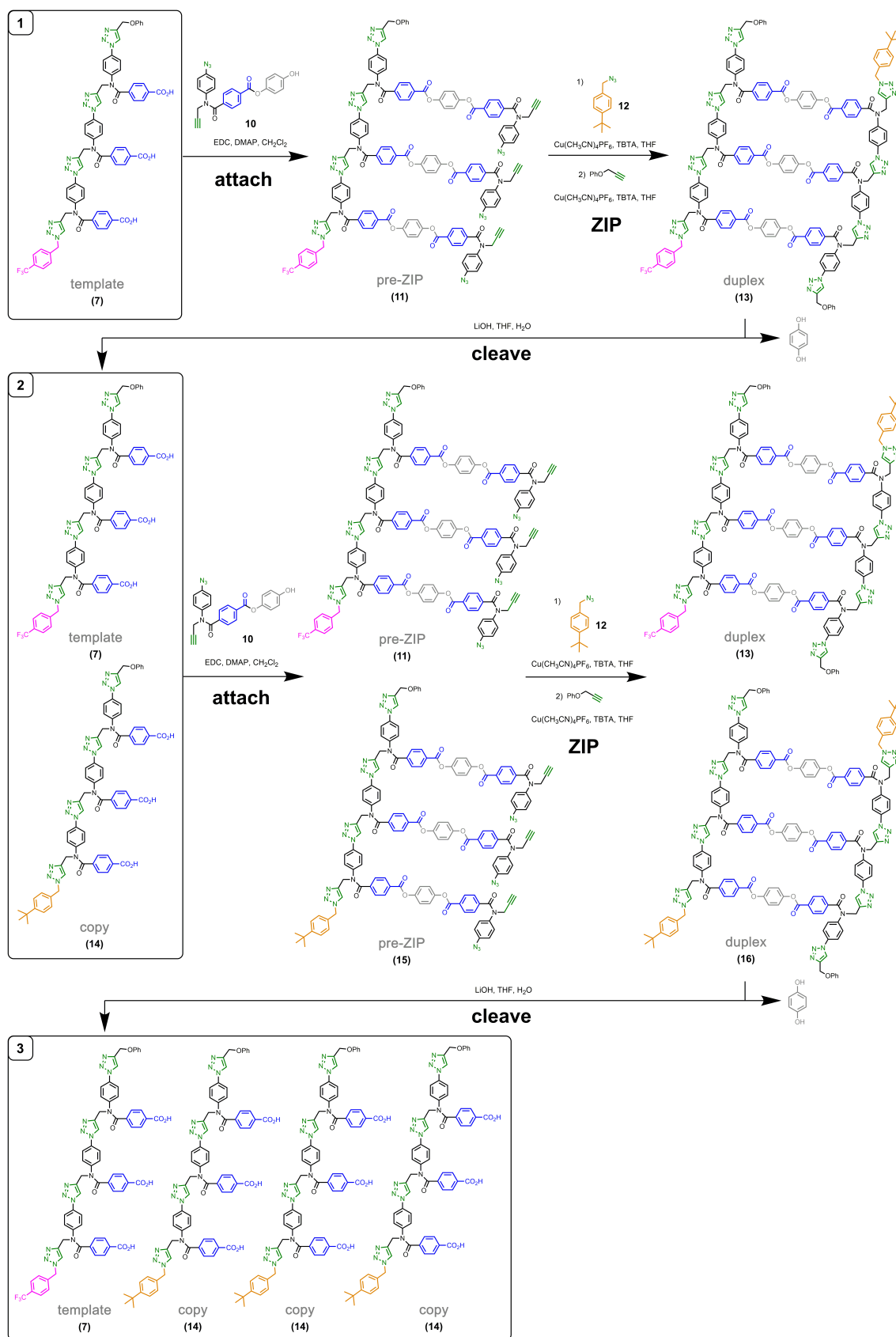


Figure 3. Two cycles of direct replication on a 3-mer template. Ester base-pair chemistry was used to load the monomers onto the template. The resulting pre-ZIP intermediate was subjected to a CuAAC reaction to give the duplex. Hydrolysis of the ester base-pairs in the cleave step regenerated the template together with the copy strand and the hydroquinone linker. The original template strand is distinguished from the copy strands by using a trifluoromethyl group in place of the *t*-butyl group on the chain capping azide.

Template-Directed Replication Cycles.

Figure 3 shows two cycles of direct replication that were carried out using template **7**. First, the benzoic acid-benzoic acid base-pairs were formed by EDC coupling of **7** with **10**, which loaded three monomers onto the template. The UPLC trace in Figure 4(b) shows quantitative conversion of the template strand to the pre-ZIP intermediate **11**. Then CuAAC was used in the ZIP step to oligomerise the monomers on the template. This reaction was carried under dilute conditions (25 μ M) to minimize competing intermolecular reactions, and in the presence of a large excess of 4-*tert*-butylbenzyl azide (1 mM), which acts as a chain capping agent. We have previously shown that this concentration of capping agent successfully intercepts the unreacted terminal alkyne group of the copy strand to prevent undesired macrocyclisation or intermolecular oligomerisation reactions, but without competing with the desired ZIP reaction.^{22,45} The 4-*tert*-butylbenzyl azide used in the ZIP step (orange in Figure 3) is different from the terminal *p*-trifluoromethylbenzyl group present in the template (pink in Figure 3), providing a spectroscopic handle that distinguishes the parent and daughter strands and allows quantification of the efficiency of the copying process. Figure 4(c) shows the UPLC trace of the crude reaction mixture of the ZIP reaction obtained after aqueous work-up and precipitation with petroleum ether to remove the excess of capping azide. Apart from residual reagents and capping azide, there is a single major species, which has a mass that corresponds to the mono-capped duplex.

The crude reaction mixture from the ZIP step was treated with phenyl propargyl ether to cap the terminal azide of the copy strand via another CuAAC reaction and give the fully capped duplex **13**. The crude reaction mixture from this reaction was hydrolysed using lithium hydroxide to cleave the esters that connect the two oligomers. Figure 4(d) shows the UPLC trace for the crude mixture obtained after the hydrolysis step. The three major species are the template **7**, the copy strand **14** and the hydroquinone linker. There are also residues of the copper ligand (asterisks), and the product of a CuAAC reaction between phenyl propargyl ether and capping azide carried through the previous steps. It is worth noting that the template and copy strands are structurally very similar, differing only in the capping group, but this subtle difference is sufficient to resolve the two products in the UPLC trace. After purification by flash chromatography to remove the linker and reagents, the product of the replication cycle was obtained as a mixture of the template **7** and copy strand **14** (see box 2 in Figure 3).

The mixture of **7** and **14** obtained from this first replication cycle was subjected to a second cycle of the same series of reactions as illustrated in Figure 3. In this case, the template is a mixture of the original template **7** and the copy **14**, so the ZIP step gives a mixture of two duplexes, **13** and **16**, which differ in the terminal trifluoromethyl or *t*-butyl group. If all of the reactions proceed in quantitative yield, the second replication cycle should yield the original template **7** and three times as much of the copy strand **14** (box 3 in Figure 3). Figure 5 shows the UPLC traces and ¹H NMR spectra of the product mixtures

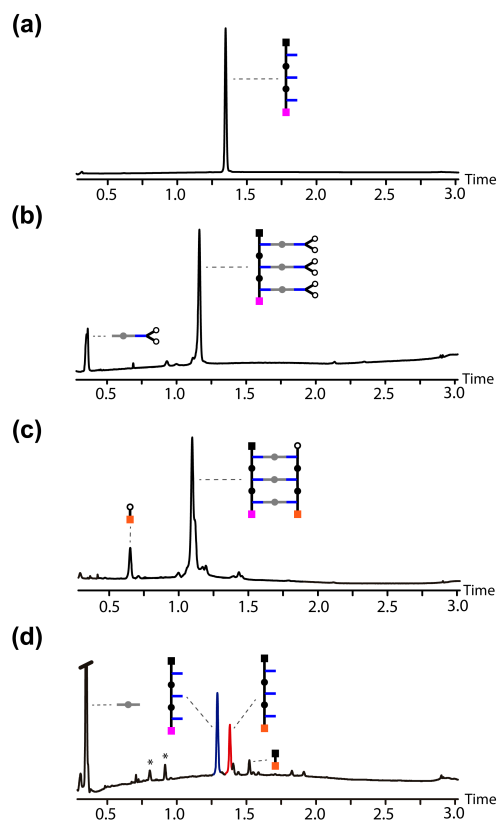


Figure 4. UPLC traces for key steps of the replication cycle: (a) the starting template **7**; (b) the crude reaction mixture of the attach step contains pre-ZIP intermediate **11** and the excess of 1-mer **10**; (c) the crude product obtained after the ZIP reaction contains duplex **13** and residual capping 4-*tert*-butylbenzyl azide; (d) crude product obtained after CuAAC capping of the ZIP products with phenyl propargyl ether, followed by hydrolysis with LiOH. The traceless linker, hydroquinone, is shown in grey while TBTA ligand and its N-oxide derivative are marked with asterisks. Minor quantities of the CuAAC product formed by reaction of residual azide with the capping alkyne are indicated.

obtained after each replication cycle. For reference purposes, a pure sample of the copy strand **14** was obtained by direct synthesis using a similar route to that shown in Scheme 1 for preparation of **7** (see ESI for details), and the data for pure samples of the starting template **7** and the copy strand **14** are shown in Figure 5. The difference between the *p*-trifluoromethyl and *t*-butyl terminal groups leads to a 0.2 ppm difference in the ¹H NMR chemical shifts of the signals due to the adjacent methylene groups, so the yield of the replication cycles can be quantified by integration. The ratios between template and copy obtained from the ¹H NMR spectra are identical to the ratios measured from the UPLC peak areas, showing that the capping groups do not significantly affect the UV/Vis extinction coefficients.

Visual inspection of the traces in Figure 5 show that repeated rounds of replication amplify the amount of the copy strand relative to the amount of original template. After the first replication cycle, the ratio of template to copy is 100:60. This result suggests that there are side reactions in the ZIP step that lead to incomplete assembly of the copy strand. The most likely cause is intermolecular reactions with the capping azide that can compete with the intramolecular monomer-monomer

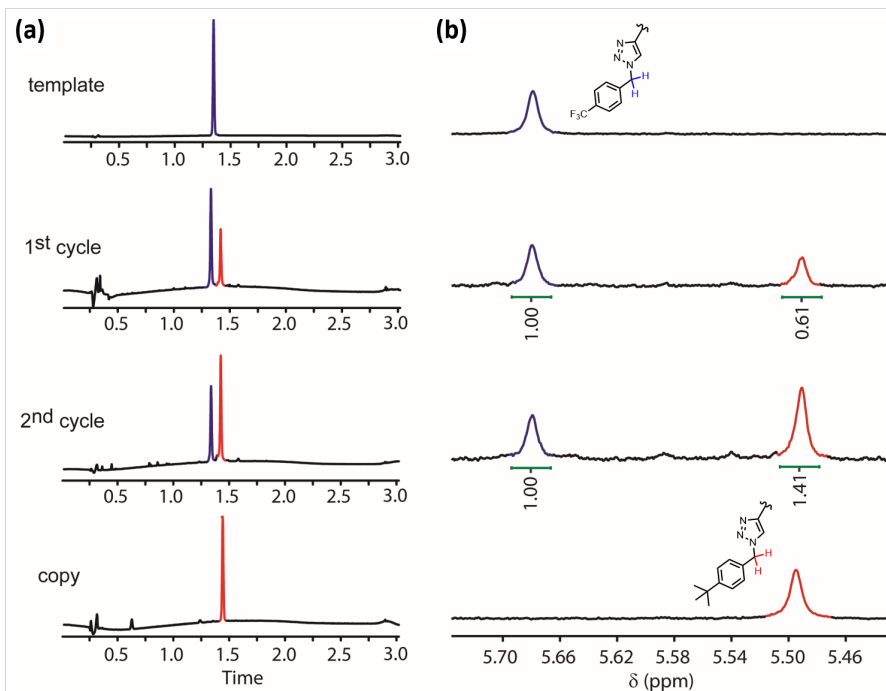


Figure 5. (a) UPLC traces of pure samples of the template **7** (blue) and copy **14** (red), and the products obtained after the first and second replication cycles. *Conditions:* C18 column at 40°C using a gradient of water/formic acid (0.1%) and CH₃CN/formic acid (0.1%). (b) Partial 500 MHz ¹H NMR spectra (DMSO-*d*₆, 298 K) showing the methylene signals due the capping groups for pure samples of the template **7** (blue) and copy **14** (red), and the products obtained after the first and second replication cycles.

coupling reactions. This would lead to truncated copy strands that would be lost on hydrolysis of the duplex, whereas the template strand from these truncated copies would be recovered intact. Figure 4(c) shows that there are small additional peaks in the UPLC trace of the crude reaction mixture from the ZIP reaction, but these peaks could not be reliably assigned. However in the UPLC trace in Figure 4(d), which was obtained after hydrolysis of the duplex, the small additional peaks close to the copy strand were assigned as the capped monomer and capped 2-mer that would arise from multiply capped products from the ZIP reaction. The second replication cycle from the mixture of **7** and **14** obtained in the first cycle amplifies the population of the copy, and the ratio of template to copy switches to 100:140. This result is consistent with a 60% efficiency for the ZIP reaction, so the fraction of copy strand obtained after two cycles of replication (60%) is slightly lower than the theoretical maximum illustrated in box 3 of Figure 3 (75%).

Conclusions

In summary, we have developed an alternative chemical method based on covalent template-directed synthesis for the direct replication of synthetic oligomers. By choosing an appropriate base-pairing system, different types of replication process can be programmed into these oligomers. Base-pairs formed by direct coupling between phenol and benzoic acid units lead to reciprocal replication of sequence information, because the copy strand is the sequence complement of the

template.²¹ Here we demonstrate that direct replication is also possible by using a traceless hydroquinone linker to connect two benzoic acid units in a base-pair. A benzoic acid 3-mer template was successfully used to template the synthesis of an identical copy, and we show that iterative replication steps can be performed leading to an increase of the population of the copy. The use of linkers to change the nature of the chemical information that is transferred by covalent template-directed synthesis opens interesting possibilities not only for cyclical replication processes but also for transcription of sequence information between chemically distinct oligomer families.

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.

Notes and references

- 1 R. R. Sinden, *DNA Structure and Function*. Academic Press, 1994.
- 2 A. J. F. Griffiths, W. M. Gelbart, J. H. Miller and R. C. Lewontin, *Modern Genetic Analysis*. Freeman, 1999.
- 3 C. M. Joyce and S. J. Benkovic, *Biochemistry*, 2004, **43**, 14317.
- 4 A. J. Berdis, *Chem. Rev.*, 2009, **109**, 2862.

- 5 E. Nudler, *Ann. Rev. Biochem.*, 2009, **78**, 335.
- 6 A. D. Ellington and J. W. Szostak, *Nature*, 1990, **346**, 818.
- 7 L. E. Orgel, *Acc. Chem. Res.*, 1995, **28**, 109.
- 8 K. Chen and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5618.
- 9 J. W. Chin, *Nature*, 2017, 550, 53.
- 10 D. L. Usanov, A. I. Chan, J. P. Maianti and D. R. Liu, *Nat. Chem.*, 2018, **10**, 704.
- 11 J. C. W. Willis and J. W. Chin, *Nat. Chem.*, 2018, **10**, 831.
- 12 G. von Kiedrowski, *Angew. Chem.*, 1986, **25**, 932.
- 13 T. Tjivikua, P. Ballester and J. Rebek Jr., *J. Am. Chem. Soc.*, 1990, **112**, 1249.
- 14 D. N. Reinhoudt, D. M. Rudkevich and F. de Jong, *J. Am. Chem. Soc.*, 1996, **118**, 6880.
- 15 D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin and M. R. Ghadiri, *Nature*, 1996, **382**, 525.
- 16 D. Philp and A. Robertson, *Chem. Commun.*, 1998, 879.
- 17 L. J. Prins, D. N. Reinhoudt and P. Timmerman, *Angew. Chem. Int. Ed.*, 2001, **40**, 2382.
- 18 S. Xu and N. Giuseppone, *J. Am. Chem. Soc.*, 2008, **130**, 1826.
- 19 A. Robertson, A. J. Sinclair and D. Philp, *Chem. Soc. Rev.*, 2000, **29**, 141.
- 20 K. R. Strom, J. W. Szostak and N. Prywes, *J. Org. Chem.*, 2019, **84**, 3754.
- 21 D. Núñez-Villanueva, M. Ciaccia, G. Iadevaia, E. Sanna and C. A. Hunter, *Chem. Sci.*, 2019, **10**, 5258.
- 22 M. Ciaccia, D. Núñez-Villanueva and C. A. Hunter, *J. Am. Chem. Soc.*, 2019, **141**, 10862.
- 23 S. Höger, A.-D. Meckenstock and H. Pellen, *J. Org. Chem.*, 1997, **62**, 4556.
- 24 S. Höger and A.-D. Meckenstock, *Chem. Eur. J.*, 1999, **5**, 1686.
- 25 K. Hiratani, M. Kaneyama, Y. Nagawa, E. Koyama and M. Kanesato, *J. Am. Chem. Soc.*, 2004, **126**, 13568.
- 26 C. Schweez, P. Shushkov, S. Grimme and S. Höger, *Angew. Chem. Int. Ed.*, 2016, **55**, 3328.
- 27 L. Steemers, M. J. Wanner, M. Lutz, H. Hiemstra and J. H. Van Maarseveen, *Nat. Commun.*, 2017, **8**, 15392.
- 28 S. C. Zimmerman, M. S. Wendland, N. A. Rakow, I. Zharov and K. S. Suslick, *Nature*, 2002, **418**, 399.
- 29 S. C. Zimmerman, I. Zharov, M. S. Wendland, N. A. Rakow and K. S. Suslick, *J. Am. Chem. Soc.*, 2003, **125**, 13504.
- 30 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.
- 31 Y.-Z. Ke, S.-L. Lee, C.-h. Chen and T.-Y. Luh, *Chem.-Asian J.*, 2011, **6**, 1748.
- 32 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.
- 33 C. J. Pedersen, *J. Am. Chem. Soc.*, 1967, **89**, 7017.
- 34 C. O. Dietrich-Buchecker and J.-P. Sauvage, *Chem. Rev.*, 1987, **87**, 795.
- 35 D. B. Amabilino and J. F. Stoddart, *Chem. Rev.*, 1995, **95**, 2725.
- 36 S. Anderson, H. L. Anderson and J. K. M. Sanders, *Acc. Chem. Res.*, 1993, **26**, 469.
- 37 R. L. E. Furlan, S. Otto and J. K. M. Sanders, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4801.
- 38 X. Li, Z.-Y. J. Zhan, R. Knipe and D. G. Lynn, *J. Am. Chem. Soc.*, 2002, **124**, 746.
- 39 P. K. Lo and H. F. Sleiman, *J. Am. Chem. Soc.*, 2009, **131**, 4182.
- 40 L. Fang, M. A. Olson, D. Benitez, E. Tkatchouk, W. A. Goddard III and J. F. Stoddart, *Chem. Soc. Rev.*, 2010, **39**, 17.
- 41 R. S. Forgan, J.-P. Sauvage and J. F. Stoddart, *Chem. Rev.*, 2011, **111**, 5434.
- 42 F. B. L. Cougnon and J. K. M. Sanders, *Acc. Chem. Res.*, 2012, **45**, 2211.
- 43 J.-F. Ayme, J. E. Beves, C. J. Campbell and D. A. Leigh, *Chem. Soc. Rev.*, 2013, **42**, 1700.
- 44 P. S. Bols and H. L. Anderson, *Acc. Chem. Res.*, 2018, **51**, 2083.
- 45 D. Núñez-Villanueva, M. Ciaccia and C. A. Hunter, *RSC Adv.*, 2019, **9**, 29566.
- 46 R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, N. Arnheim, *Science*, 1985, **230**, 1350.
- 47 D. Núñez-Villanueva, M. A. Jinks, J. Gómez Magenti and C. A. Hunter, *Chem. Commun.*, 2018, **54**, 10874.