



Cite this: DOI: 10.1039/d5nh00505a

Received 18th July 2025,
 Accepted 1st December 2025

DOI: 10.1039/d5nh00505a

rsc.li/nanoscale-horizons

Mechanism for a molecular assembler of sequence-controlled polymers using parallel DNA and a DNA polymerase

Jonathan Bath ^{*ab} and Andrew J. Turberfield ^{ab}

Construction of a molecular assembler from DNA that executes a programmed sequence of chemical reactions is a formidable challenge but worthwhile because it would allow assembly and evolution of functional polymers. We present a mechanism using parallel DNA and a DNA polymerase to address two challenges that currently block progress.

We would like to use DNA to build a molecular machine that can manufacture a polymer by executing a series of chemical reactions that add building blocks in the order encoded by an instruction tape or gene. Parallel operation of molecular assemblers, each acting on a unique instruction tape, could be used to produce large DNA-tagged libraries from which functional polymers could be selected, identified by DNA sequencing and optimised iteratively by shuffling or recombining the instruction tapes.¹

Reaction between two building blocks can be promoted by attaching them to the ends of two adaptor DNA strands and bringing them into close proximity by hybridization of the adaptors, either to each other or to a complementary instruction tape.² This is typically done in an end-of-helix configuration where one building block is attached to the 5' end of a strand and the other is attached to the 3' end of a complementary strand.³ A sequence of transfer reactions can be used to build a polymer, either by adding building blocks to the end of a growing chain⁴ or, as in the ribosome, by transferring the growing chain onto the next building block.⁵ The second approach is preferable because the chemical context is consistent for each step and, unlike addition to the end of the chain, chain elongation does not inhibit polymer growth.

There are very few examples of autonomous DNA mechanisms for stepwise polymer synthesis. He and Liu⁴ used a molecular motor to couple three building blocks in the sequence

New concepts

Construction of a molecular machine capable of stepwise assembly of sequence controlled polymers from building blocks is a significant but worthwhile challenge: it opens the way to selection and evolution of non-natural polymers from large libraries. We use a DNA instruction tape to encode the order in which building blocks are added and present a mechanism designed to ensure that an instruction can only be read after completion of the preceding programmed reaction. We use a sequence motif which allows us to juxtapose the 3' ends of two DNA strands at the end of a parallel DNA helix. This mechanism allows us to ensure that the chemical environment is consistent for successive building block coupling steps. A notable feature of our mechanism is that it involves several competing weak interactions. We show that these interactions can be fine-tuned using a combination of experiment and coarse-grain modelling.

encoded by the motor's track. Meng *et al.*⁸ used polymerisation of hairpin loops to couple building blocks and record the sequence of coupling reactions. Both studies added sequential building blocks to the distal end of the growing chain, neither had a checkpoint mechanism to ensure that the transfer chemistry was complete before the next step was initiated. McKee *et al.*⁵ showed that the growing chain can be transferred onto the incoming building block, as in the ribosome, but the mechanism required both 5'- and 3'-modified adaptor strands and was not autonomous.

We address two challenges in the design of molecular machinery that brings reactants together in a programmed sequence. First, we propose a mechanism that makes progression of the molecular machinery conditional on the successful completion of each programmed reaction. We propose to do this by covalently linking building blocks through esterification of the 3'-OH of their adaptor strands and using a DNA polymerase to recognise the 3'-OH that is revealed when the transfer reaction is complete. Second, we demonstrate that a parallel DNA duplex motif can be used to colocalise building blocks at the end of a double helix by juxtaposing the 3' ends of two adaptor strands. This avoids the requirement to recruit 5'

^a Kavli Institute for Nanoscience Discovery, Dorothy Crowfoot Hodgkin Building, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

E-mail: jonathan.bath@physics.ox.ac.uk

^b Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK



modified and 3' modified adaptors alternately which doubles the number of adaptors required and adds unnecessary complexity to the mechanism.

Design

Operation of the device is illustrated in Fig. 1. The instruction tape is a single-stranded oligonucleotide consisting of a sequence of binding sites or codons (16 nt) separated by competition sites (6 nt, fixed sequence). Building blocks are covalently coupled to the 3'-OH of adaptor strands, for example by using a flexizyme.⁶ Each building block is uniquely identified by a binding domain or anticodon at the 5' end of the adaptor. Adaptor strands deliver the correct building block to the active site by hybridization to a matching binding domain on the instruction tape (Fig. 1b).

The active site consists of two adaptor strands: the upstream adaptor carries the growing polymer chain; the downstream adaptor carries the building block that is to be added to the growing chain (Fig. 1d). Although all adaptors are present in solution, sites downstream of the active site are occupied by blocking strands that prevent adaptors from binding out of sequence. The growing chain is transferred onto the incoming building block by aminoacyl transfer which restores the 3'-OH on the spent upstream adaptor. The reaction is promoted by formation of a parallel duplex⁷ designed to hold the incoming building block and growing chain in close proximity in an end-of-helix configuration³ (Fig. 1d). The short parallel duplex domain (AG)_n is identical in all adaptors. Because the aminoacyl transfer reaction transfers the growing chain from the upstream adaptor to the incoming building block, it operates in a consistent chemical environment at every step. This is designed to avoid the decreasing yield of successive steps that

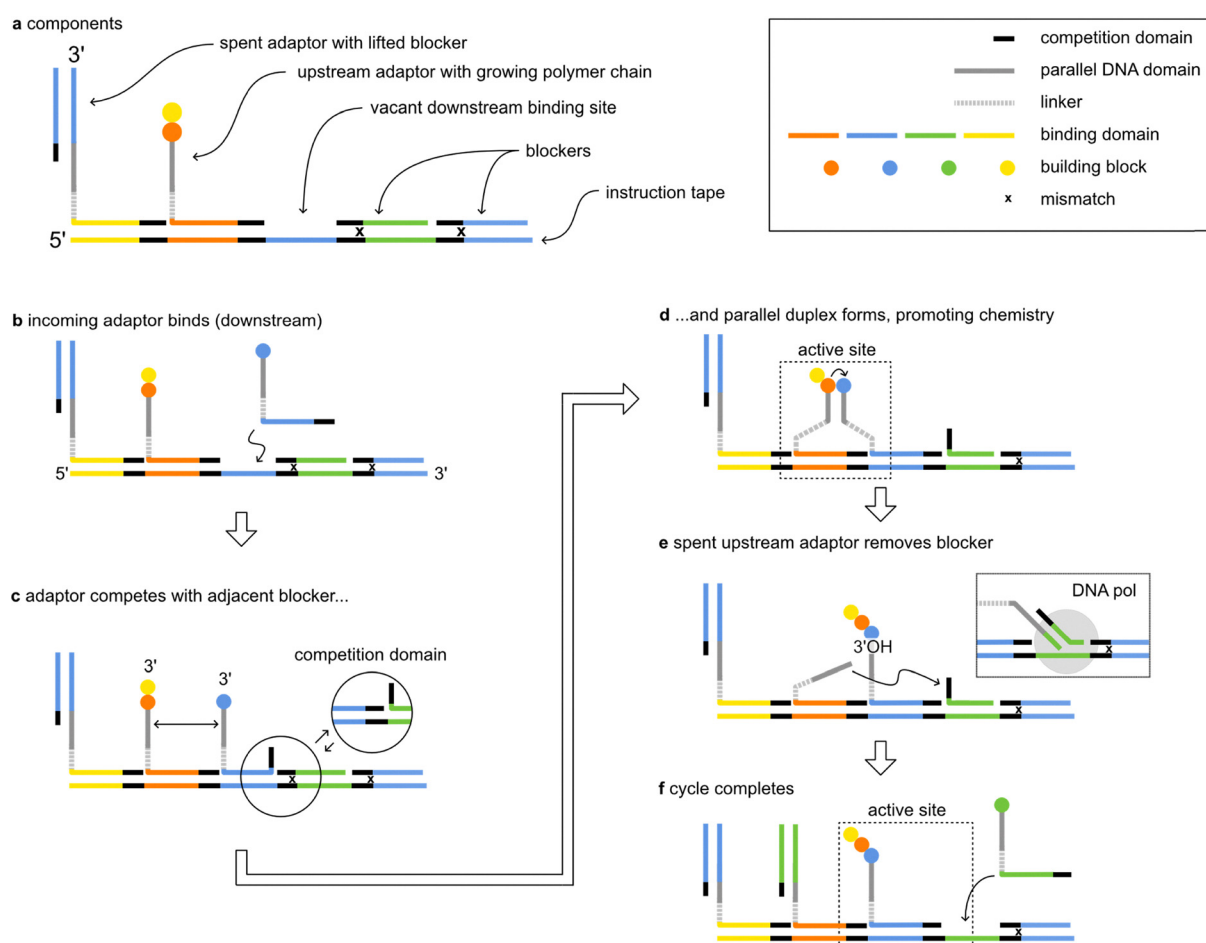


Fig. 1 Scheme for a molecular assembler. One operation cycle is shown. (a) Components are labelled. (b) An adaptor strand is recruited to the instruction tape in the downstream position delivering the incoming building block to the active site. (c) Once bound, the adaptor competes with the adjacent blocking strand for binding to a competition domain on the instruction tape; repair of a mismatch (marked x) biases the competition to expose the 3' end of the blocking strand. (d) At the same time, the 3' ends of adjacent adaptor strands can bind to form a parallel helix. The parallel helix is designed to promote a transfer reaction that localizes the incoming building block and the growing polymer chain. The transfer reaction passes the growing polymer chain onto the building block on the downstream adaptor. (e) Completion of the transfer reaction reveals the 3'-OH of the upstream adaptor, allowing it to bind to the 3' end of the blocking strand and serve as a primer for a strand displacing DNA polymerase (the inset shows a partially lifted blocking strand). (f) Polymerase extension of the upstream adaptor lifts the blocking site from the instruction tape advancing the active site one step and allowing the next adaptor to be recruited in the downstream position.



is associated with mechanisms in which building blocks are added to the end of a growing chain.^{4,8}

The structure of the instruction tape is such that the downstream adaptor strand and adjacent blocking strand compete for binding to overlapping sites (Fig. 1c). This competition is biased by a mismatch between the blocking strand and instruction tape in the competition domain; this mismatch is repaired when the adaptor strand displaces the blocking strand. The sequence of the competition domain is such that the 3' end of any adaptor can hybridise transiently to the displaced 3' end of a blocking strand. The reaction that transfers the growing product oligomer onto the incoming building block on the downstream adaptor reveals a 3' hydroxyl at the end of the spent upstream adaptor: when this binds to the displaced competition domain it serves as a primer for a strand-displacing polymerase, *e.g.* Bst DNA polymerase (Fig. 1e). Polymerase extension displaces the blocking strand from the instruction tape and reveals the binding site for the next adaptor strand in the programmed sequence. This completes the reaction cycle and advances the active site one step along the instruction tape (Fig. 1f). This mechanism ensures that binding sites are revealed one at a time, in the correct sequence, only when the previous chain extension reaction has completed.

Successful design requires a careful balance between competing interactions: the parallel duplex must be stable enough to promote chemistry by increasing the effective concentration of participating reactive groups but unstable enough that the 3' end of the spent adaptor strand is transiently released and able to bind to the blocking strand. Similarly, the interaction between the spent adaptor strand with the blocking strand must be stable enough to serve as a primer for the polymerase but not so stable that it locks the adaptor strand in place before the chain-extending chemical reaction has occurred.

We have used a combination of coarse-grained molecular modelling and experiment to characterize all aspects of the molecular assembler mechanism except the aminoacyl transfer steps and to demonstrate that it provides a viable method for coordinating progression of the molecular machinery with the successful completion of a chain-extending reaction.

Parallel DNA duplex

(GA)_n sequences are known to form parallel duplexes at neutral pH⁷ but have not previously been used as a construction motif for DNA nanotechnology. We constructed a simple test device⁹ consisting of two DNA arms attached to opposite ends of a short double-stranded linker (Fig. 2a). At the free end of each arm is a parallel DNA motif (GA)_n (*n* = 5, 6, 7, 8, 12) and, on the opposite strand, one of the reporter dyes Cy3 and Cy5. Measurement of Förster resonance energy transfer (FRET) between the dyes as a function of temperature (Fig. 2b, SI) reveals a melting transition between a high-FRET state consistent with a closed conformation formed by the parallel duplex at low temperature and a low-FRET open conformation at high temperature when the parallel duplex melts. The stability of the parallel duplex domain can be tuned by controlling the number of GA repeats.

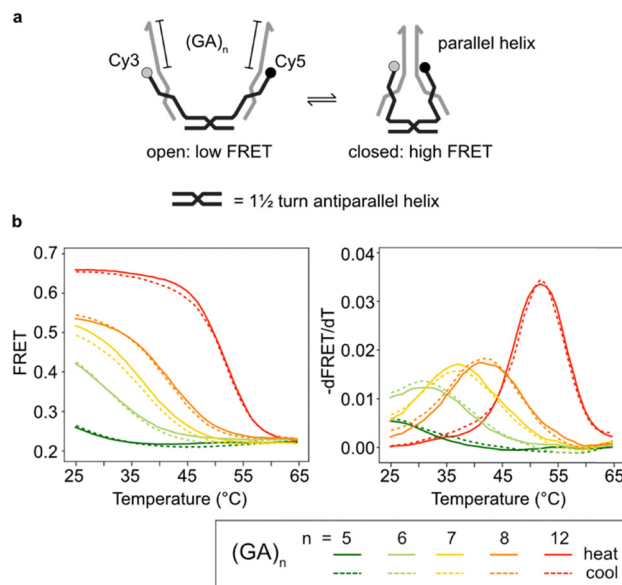


Fig. 2 Parallel duplex formation monitored by FRET. (a) Donor and acceptor fluorophores (Cy3 and Cy5) attached to the ends of antiparallel double-stranded arms are held in close proximity by formation of a parallel duplex giving a high FRET signal. (b) Measurement of FRET as a function of temperature reveals a melting transition that depends on the length of the parallel duplex (left panel). Maxima of the derivative with respect to temperature (right panel) were used to determine melting temperatures of 51.8 ± 0.1 °C, 41.6 ± 0.6 °C, 36.8 ± 0.4 °C and 30.9 ± 0.4 °C for the sequences (GA)₁₂, (GA)₈, (GA)₇ and (GA)₆. The melting temperature of (GA)₅ is less than 25 °C. See SI, for experimental details.

We measure melting temperatures of 51.8 ± 0.1 °C, 41.6 ± 0.6 °C, 36.8 ± 0.4 °C and 30.9 ± 0.4 °C for the sequences (GA)₁₂, (GA)₈, (GA)₇ and (GA)₆ respectively; the melting temperature of (GA)₅ is less than 25 °C. The sequence (GA)₆ is only marginally stable at room temperature and was therefore used to implement the molecular assembler mechanism which requires parallel duplex formation to be transient.

Competition between downstream adaptor and blocking strand

Competition between strands bound to overlapping sites can be used to expose a single-stranded toehold and thus initiate strand displacement by a complementary strand^{10,11} or, as here, enable binding of a primer for a strand-displacing polymerase. Competition between the downstream adaptor and adjacent blocking strand (Fig. 3) is achieved by overlapping their binding sites on the instruction tape to create a 6-nt competition site (Fig. 1c) which has a universal sequence. In the initial design, a single mismatch between the blocking strand and instruction tape was introduced to bias the competition toward displacement of the 3' end of the blocking strand. In subsequent designs competition was established between a mismatch in the blocking strand and a single-nucleotide gap produced by truncation of the 5' end of the adaptor strand. The single-nucleotide gap was introduced to increase the mechanical



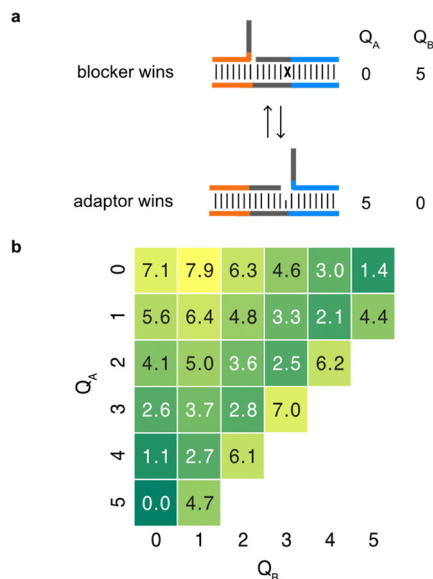


Fig. 3 Competition between adaptor and blocking strand. (a) Order parameters count the number of base pairs between the instruction tape and adaptor strand (Q_A) or blocking strand (Q_B). (b) Energy landscapes produced using oxDNA and VMMC umbrella sampling¹⁷ reveal a penalty for enclosing the G–T mismatch (moving from the first column to the second) and a small bias toward fully bound adaptor strand ($Q_A = 5$, $Q_B = 0$) over the fully bound blocker strand ($Q_A = 0$, $Q_B = 5$). The heatmap is annotated with the free energy relative to the fully bound adaptor strand in units of kcal mol⁻¹. Note that the energy cutoff for hydrogen bonding used by oxDNA allows $Q_A + Q_B > 5$. Simulation details, including weights are described in the SI.

flexibility of the system to facilitate interaction between the spent adaptor strand and downstream blocking strand. Coarse-grained simulations were conducted to quantify bias in the competition (SI). Models were constructed using oxView¹² and simulated using oxDNA^{13–15} with virtual-move Monte Carlo and umbrella sampling.¹⁶ oxDNA uses a Debye–Hückel potential to model electrostatics rather than explicitly modelling solvent and ions.^{13–15} all simulations use the oxDNA2 model with screening equivalent to 0.15 M monovalent ions. Two order parameters were defined: Q_A counts the number of base pairs between the adaptor and instruction tape; Q_B counts the number of base pairs between the blocking strand and instruction tape. A single simulation window was sampled using the umbrella module built into the oxDNA code.¹⁷ Weights required to increase the sampling of unfavourable states were improved iteratively (see SI for weights). A difference in free energy of 1.4 kcal mol⁻¹ between the competing states was observed, corresponding to a 10-fold bias toward the state in which the 3' end of the blocking strand is displaced from the instruction tape (Fig. 3).

Removal of a blocking strand by a DNA polymerase

Experiments using gel electrophoresis show that binding of an adaptor to an instruction tape is negligible when the tape is fully blocked and strongly hindered at the next binding site in sequence while the corresponding blocker has not yet been displaced (SI).

The spent upstream adaptor strand is designed to enable the removal of the adjacent blocking strand, opening the next binding domain for a matching adaptor, conditional on the transfer of the growing chain to the downstream adaptor. The 6-nt domain at the 3' end of each adaptor, which forms a parallel duplex with an adjacent adaptor, can also form an antiparallel duplex with the exposed competition domain at the 3' end of the adjacent blocking strand. When transfer of the growing chain from upstream to downstream adaptor is complete, regenerating its 3'-OH, the adaptor can act as a primer for a strand-displacing polymerase which removes the blocking strand. This mechanism is designed to couple completion of the chemical step to movement of the active site along the instruction tape.

We characterise the direct hybridization interactions between the 3' ends of adaptors and blocking strands using molecular dynamics simulations in oxDNA to generate trajectories using the default Anderson-like thermostat¹⁷ and the oxDNA analysis tool “Output bonds”.¹⁸ Simulation snapshots for configurations in which the blocker forms a duplex with the upstream adaptor, which is thereby positioned to prime displacement of the blocker by polymerase extension (correct), and with the downstream adaptor carrying the growing chain of building blocks (incorrect), are shown in Fig. 4a and b. Initial simulations showed that the correct configuration is strained,

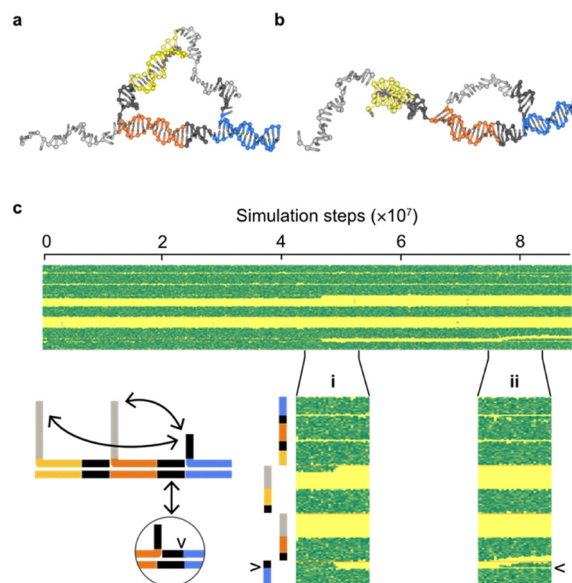


Fig. 4 Molecular dynamics simulations to model the interactions between 3' ends of adaptor strands and the blocking strand. The system comprises two adaptor strands (orange, downstream, and yellow, upstream) and one blocking strand (blue) hybridized to a three-site instruction tape. (a) Simulation snapshot showing the designed interaction between upstream adaptor and blocking strand and (b) the more strained interaction between downstream adaptor and blocking strand. (c) Base pairing (yellow for unpaired, green for paired) throughout the trajectory for each of the strands. In the initial state the 3' end of the blocking strand has been displaced from its competition domain by the downstream adaptor and is hybridized to the 3' end of the upstream adaptor (panel a). Inset (i) shows unbinding of the 3' end of the blocking strand from the 3' end of the upstream adaptor. Inset (ii) shows the 3' end of the blocking strand displacing the 5' end of the downstream adaptor strand.



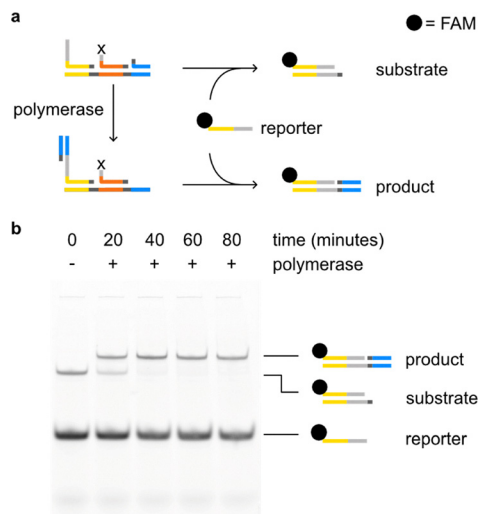


Fig. 5 Removal of blocking strand by DNA polymerase. (a) An instruction tape hybridized to a full-length upstream adaptor, a truncated downstream adaptor with the parallel duplex domain deleted and a blocking strand were incubated with Bst 3.0 DNA polymerase at 30 °C. Sample $t = 0$ was taken before addition of polymerase, the remaining samples were taken at 20-minute intervals. Samples were denatured at 80 °C for 5 minutes to inactivate the polymerase then the extent of the reaction was measured by adding an excess of a fluorescently labelled reporter strand to remove the upstream adaptor from the instruction tape by toehold-mediated strand exchange. Note that the blocking strand is modified at the 3' end with a phosphate group to prevent extension. (b) Reaction products were separated on a 15% 29:1 TAE gel imaged without staining to detect FAM fluorescence. The reporter is a 28-nt single-stranded DNA, the substrate (an adaptor bound to the reporter) is a duplex of 28 bp with a 12-nt single-stranded tail, the product (an adaptor extended by DNA polymerase bound to a blocking strand and reporter) is a nicked duplex of 49 bp with a nick on one strand 21 nt from its 5' end.

with base pairs broken at the junction between the strands bound to the instruction tape. The gaps introduced by deletion of one nucleotide at the 5' end of the adaptor, modelled in Fig. 3, were designed to release strain and favour the correct configuration (Fig. 4a). The time to half completion for unbinding of the upstream adaptor, determined by measuring the first passage time for unbinding in 128 molecular dynamics simulations, is three times longer for the correct configuration than for the incorrect configuration. Binding events were sufficiently rare in simulations that we cannot compare binding rates in the two configurations. However, gel electrophoresis demonstrates that the two configurations are approximately equally populated in equilibrium at room temperature (SI). These results suggest that the correct configuration is similar in stability to the incorrect configuration despite the larger distance between the two adaptor-binding sites. Competition between downstream adaptor and blocking strand, as designed, was observed in simulations (Fig. 4c, inset ii).

Initial experiments showed that the 6-nt primer binding site revealed by competition between the downstream adaptor and blocking strand is too short to allow removal by Bst DNA polymerase. A more stable interaction was created by extending the 3' end of the blocking strand beyond the competition

domain by 2 nt to create an additional 2 nt toehold: this was sufficient to allow the blocking strand to be displaced from the instruction tape (Fig. 5). This reaction proceeds to half completion in less than 20 minutes. It is therefore likely to be rate-limiting, significantly slower than adaptor binding and formation of the parallel duplex.

Conclusions

We have presented a design for a molecular assembler that overcomes two challenges. First we show that, by using a parallel DNA motif, two like ends of oligonucleotide adaptors (3' ends here) can be held in close proximity. This avoids the need for adaptors of alternating polarity and allows the design of a mechanism in which the local context of transfer remains consistent regardless of the product length. Second, we present a mechanism that can be used to coordinate the transfer chemistry with stepwise reading of instructions. In the absence of such a mechanism there is no guarantee that the transfer chemistry is complete before the mechanism proceeds to the next instruction. Our mechanism relies on the ability to engineer relatively weak interactions, enabling dynamic and reversible binding.^{19–22} We show that a combination of experiment and simulation can be used to fine-tune the strengths of competing interactions that are required for operation. The reaction mechanism was designed for implementation using acyl transfer reactions, with adaptor strands coupled to building blocks using flexizymes. Future work will address this challenge: preliminary results demonstrate that it is possible to use flexizymes to aminoacylate DNA strands and, after hydrolysis, to use those strands as primers for strand-displacing DNA polymerases.

Author contributions

J. B. was responsible for investigation and writing (original draft). A. J. T. was responsible for funding acquisition. J. B. and A. J. T. were responsible for conceptualization and writing (review and editing).

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information: DNA sequences, supplementary data and simulation details. See DOI: <https://doi.org/10.1039/d5nh00505a>.

Acknowledgements

This research was funded by the UKRI [EP/T000562/1].



Notes and references

- 1 R. K. O'Reilly, A. J. Turberfield and T. R. Wilks, *Acc. Chem. Res.*, 2017, **50**, 2496.
- 2 Z. J. Gartner, R. Grubina, C. T. Calderone and D. R. Liu, *Angew. Chem., Int. Ed.*, 2003, **42**, 1370.
- 3 Z. Gartner, R. Grubina, C. T. Calderone and D. R. Liu, *Angew. Chem., Int. Ed.*, 2003, **42**, 1379.
- 4 (a) Y. He and D. R. Liu, *Nat. Nanotech.*, 2010, **5**, 778; (b) M. L. McKee, P. J. Milnes, E. Stulz, A. J. Turberfield and R. K. O'Reilly, *Angew. Chem., Int. Ed.*, 2010, **49**, 7948.
- 5 M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, R. K. O'Reilly and A. J. Turberfield, *J. Am. Chem. Soc.*, 2012, **134**, 1446.
- 6 H. Murakami, H. Saito and H. Suga, *Chem. Biol.*, 2003, **10**, 655.
- 7 K. Rippe, V. Fritsch, E. Westerhof and T. M. Jovin, *EMBO J.*, 1992, **11**, 3777.
- 8 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542.
- 9 B. Yurke, A. J. Turberfield, A. P. Mills Jr, F. C. Simmel and J. L. Neumann, *Nature*, 2000, **406**, 605.
- 10 S. J. Green, J. Bath and A. J. Turberfield, *Phys. Rev. Lett.*, 2009, **101**, 238101.
- 11 J. Bath, S. J. Green and A. J. Turberfield, *Small*, 2009, **5**, 1513.
- 12 J. Bohlin, M. Matthies, E. Poppleton, J. Procyk, A. Mallya, H. Yan and P. Šulc, *Nat. Protoc.*, 2022, **17**, 1762.
- 13 B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldrige, R. Tsukanov, E. Nir, A. A. Louis and J. P. K. Doye, *J. Chem. Phys.*, 2015, **142**, 234901.
- 14 (a) P. Šulc, F. Romano, T. E. Ouldrige, L. Rovigatti, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, 2012, **137**, 135101; (b) L. Rovigatti, P. Šulc, I. Z. Reguly and F. Romano, *J. Comput. Chem.*, 2015, **36**, 1.
- 15 T. E. Ouldrige, A. A. Louis and J. P. K. Doye, *J. Chem. Phys.*, 2011, **134**, 085101.
- 16 G. M. Torrie and J. P. Valleau, *J. Comp. Physiol.*, 1977, **23**, 187.
- 17 A. Sengar, T. E. Ouldrige, O. Henrich, L. Rovigatti and P. Šulc, *Front. Mol. Biosci.*, 2021, **8**, 693710.
- 18 E. Poppleton, J. Bohlin, M. Matthies, S. Sharma, F. Zhang and P. Šulc, *Nucleic Acids Res.*, 2020, **48**, e72.
- 19 E. R. Kay, D. A. Leigh and F. Zerbetto, *Angew. Chem., Int. Ed.*, 2006, **46**, 72.
- 20 R. D. Astumian, *Chem. Sci.*, 2017, **8**, 840.
- 21 U. Sefert, *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 2011, **34**, 26.
- 22 I. M. A. Nooren and J. M. Thornton, *EMBO J.*, 2003, **22**, 3486.

