ARTICLES
PUBLISHED ONLINE: 11 APRIL 2016 | DOI: 10.1038/NCHEM.2495
nature chemistry

An autonomous molecular assembler for programmable chemical synthesis

Wenjing Meng[1†], Richard A. Muscat[1†], Mireya L. McKee[1], Phillip J. Milnes[2], Afaf H. El-Sagheer[3,4], Jonathan Bath[1], Benjamin G. Davis[3], Tom Brown[3], Rachel K. O'Reilly[2] and Andrew J. Turberfield[1*]

Molecular machines that assemble polymers in a programmed sequence are fundamental to life. They are also an achievable goal of nanotechnology. Here, we report synthetic molecular machinery made from DNA that controls and records the formation of covalent bonds. We show that an autonomous cascade of DNA hybridization reactions can create oligomers, from building blocks linked by olefin or peptide bonds, with a sequence defined by a reconfigurable molecular program. The system can also be programmed to achieve combinatorial assembly. The sequence of assembly reactions and thus the structure of each oligomer synthesized is recorded in a DNA molecule, which enables this information to be recovered by PCR amplification followed by DNA sequencing.

Synthetic molecular machinery made by DNA self-assembly can link natural and unnatural building blocks in a defined sequence by covalent bonds[1–10]. Sequential covalent synthesis, with a sequence hard-wired into the molecular machinery, has also been demonstrated using a synthetic rotaxane[11]. There are close parallels between these synthetic molecular assemblers and the ribosome. The ribosome makes use of non-covalent interactions between nucleotides both to stabilize its three-dimensional structure and to enable each codon to recruit the corresponding tRNA adapter to the active site[12,13]. The same interactions are used in synthetic nanostructures to control the assembly of rationally designed DNA and RNA components[14–16]. In the ribosome, peptide bond formation occurs when amino acids are held together by their tRNA adapters in a catalytic pocket: Liu and co-workers have developed DNA-templated synthesis (DTS), a simpler form of catalysis in which DNA adapters enhance reaction rates by tethering reactants together[17]. The ribosome can select building blocks from a pool of tRNA-linked amino acids and concatenate them in a sequence defined by the mRNA molecular program. The principle of DTS can be extended to allow multiple reactions to be carried out in a defined sequence, in the presence of all reactants, by using the addition of DNA signalling strands[6,7] or temperature-dependent changes in secondary structure[10] to bring successive reactants together.

Unlike the ribosome, the synthetic systems described above require an externally supplied signal to initiate each new reaction. Systems based on DNA can, however, be designed to assemble and evolve autonomously according to DNA-encoded programs. Synthetic molecular motors, fuelled either by the formation of base pairs[18–21] or by hydrolysis of the nucleic acid backbone[22–27], have stepped along pre-assembled tracks, and a hydrolysis-driven motor has been used to concatenate reactants bound to a track in a predetermined sequence[28].

Here, we report a programmable and autonomous DNA-based assembler. This molecular system can be programmed to synthesize oligomers with defined sequences and then reprogrammed for combinatorial synthesis. Alternate steps in a DNA hybridization chain reaction[29–32] are coupled to acyl transfer[1,2,4,28,33] or Wittig[1,7–10] reactions that link specified building blocks to a growing covalently bonded oligomer by peptide or carbon–carbon double bonds, respectively. Both chemistries involve transfer reactions[1,11] in which the formation of a covalent bond to the active end of the growing oligomer is coordinated with cleavage of the building block from its DNA adapter, analogous to the transfer of amino-acid residues from aminoacyl tRNAs to the growing polypeptide chain in the ribosome. Both chemistries can accommodate building blocks with a wide range of side chains, creating the possibility of automated molecular synthesis of a wide range of sequence-controlled products, from biomimetic peptides to completely unnatural oligomers.

The assembler program consists of a set of DNA instruction strands that defines the reaction sequence. This can be reprogrammed simply by changing the instruction set. A DNA-encoded record of the sequence of each oligomer produced is created by ligation of instruction strands after synthesis. This record can be amplified and read by DNA sequencing, creating the possibility of *in vitro* selection from libraries of products[2,3,34], each created by an individual molecular assembler.

Results

The oligonucleotide components of the programmable molecular assembler are shown in Fig. 1a (for sequences see Supplementary Table 1). **Cargo**, which carries the growing oligomer at its 3′ end, is initially hybridized to strand **I** to form the initiator duplex **I**:**Cargo**. The other components are hairpins of two types: chemistry hairpins (for example, **A** and **B**), which carry the reactive building blocks of the polymer, and instruction hairpins (for example, **I>A** and **A>B**), which control their assembly. The name of an instruction hairpin reflects its function: hairpin **A>B** recognizes **A** and links it to **B**. The stems of all hairpins comprise the same sequence of base pairs, which is also present in the initiator duplex. Hairpins are distinguished by two single-stranded toehold domains[35] (coloured in Fig. 1), which are used to initiate hybridization reactions and thus control the sequence of assembly. One toehold is external

[1]Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK. [2]Department of Chemistry, University of Warwick, Coventry, West Midlands CV4 7AL, UK. [3]Department of Chemistry, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA, UK. [4]Chemistry Branch, Department of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez University, Suez 43721, Egypt. †These authors contributed equally to this work. *e-mail: a.turberfield@physics.ox.ac.uk

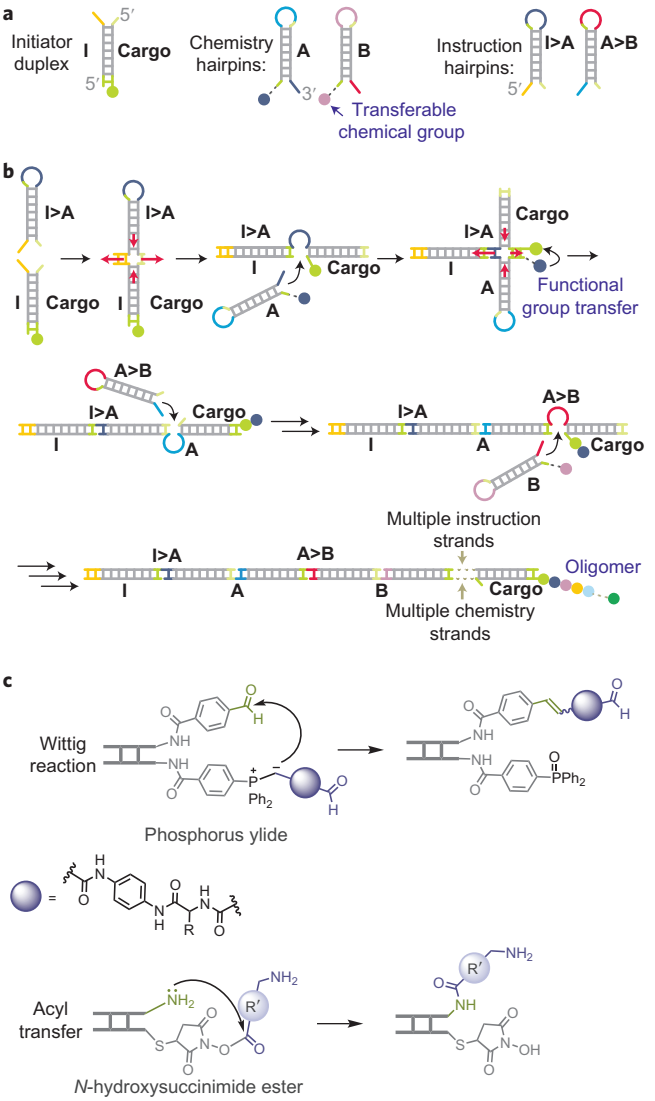

Figure 1 | Operation of the assembler. a, Molecular components. **Cargo** starts in the initiating duplex, which carries an identifying toehold on strand **I** (yellow). 'Chemistry' hairpins (for example, **A** and **B**) carry building blocks for DTS. 'Instruction' hairpins (for example, **I>A** and **A>B**) program the assembly sequence. Each hairpin carries external and internal address toeholds (coloured) and one of two generic external toeholds (green). **b**, Assembler mechanism. Hybridization of the initiating duplex (**I:Cargo**) to the matching toeholds of the first instruction **I>A** creates a Holliday junction whose migration (indicated by red arrows) leads to the opening of the loop domain of the hairpin[31]. Loop opening activates the sequestered toehold domain (blue) that specifies that the chemistry hairpin **A** is to be added next. **A** is recruited to the chain by hybridization of its external toehold to this newly exposed domain. Junction migration completes insertion and reveals a sequestered toehold that is recognized by and triggers insertion of instruction **A>B**. Chain extension proceeds by alternate addition of instruction and chemistry hairpins. As each chemistry hairpin is added to the chain, its building block is held in close proximity to the growing oligomer attached to **Cargo**, enhancing the rate of transfer of the incoming building block from its DNA adapter to the oligomer. **c**, Chemistries used for oligomer propagation. Curly arrows show the initial nucleophilic attack of the transfer reactions. R and R′ are variable moieties that distinguish building blocks.

(a single-stranded overhang, readily available for hybridization) and one is sequestered within the loop[18,29,30,32,36]. Hairpins also display one of two generic external toeholds (one is carried by all chemistry hairpins and one by all instruction hairpins), which are complementary to toehold domains at either end of the **Cargo**.

Each building block is covalently coupled to the 5′ end of a specific chemistry hairpin that serves as an adapter analogous to tRNA. A variable moiety analogous to a natural amino-acid side chain (R and R′, Fig. 1c) distinguishes building blocks. We have used two different chemistries for oligomer synthesis. For the Wittig reaction, leading to the formation of a C=C double bond, each bifunctional chain-extending building block incorporates a phosphonium ylide and an aldehyde[1,7–10]. The chain-initiating building block that remains coupled to the **Cargo** has aldehyde functionality only. Building blocks with ylide functionality only can be used as chain terminators because they lack the aldehyde group required for chain propagation. For polypeptide synthesis, the building blocks are functionalized with *N*-hydroxysuccinimide (NHS) ester and amine moieties. At the beginning of the reaction the first building block is conjugated to the **Cargo**, resulting in the formation of a peptide bond and leaving a free amine for chain extension.

The DNA assembly mechanism (Fig. 1b) is a one-dimensional hairpin hybridization chain reaction (HCR)[29–32]. The product of DNA assembly is a linear duplex formed by opened hairpins hybridized in a staggered pattern; the **Cargo** is transferred to each incoming hairpin such that it remains bound at the growing end of the duplex[31]. Each strand of the duplex is discontinuous: one strand consists of instruction hairpins and the other of chemistry hairpins. The sequence of assembly is programmed through interactions between pairs of complementary toeholds (Fig. 1b); these toeholds serve as addresses and routing instructions[21,32]. A pair of toeholds (together, a 'split toehold'[21,31]) is displayed at the end of the initiator and, at later stages of assembly, near the end of the growing DNA duplex. Half of the split toehold is always formed by one of two toehold domains at the ends of the **Cargo**. The other half, an instruction toehold, is responsible for selecting the next hairpin to be added to the chain. The first instruction toehold is part of the initiator. At later stages, the active instruction toehold is part of the last added hairpin: this was initially sequestered in the loop domain of the hairpin and activated when the hairpin was opened[29–32]. A hairpin with two complementary external toeholds can hybridize to this split toehold to form a four-armed branched junction. This junction can then be resolved by branch migration to form a more stable complex in which the **Cargo** is transferred to the newly added hairpin at the end of the growing duplex. Junction formation is driven by the free energy of toehold hybridization, and junction resolution is associated with a further decrease in free energy that includes the increase in entropy associated with opening the loop of the hairpin. As each hairpin is added to the chain its loop is opened, making the previously sequestered toehold accessible for hybridization to initiate the next stage of chain growth. The **Cargo** is carried forward at the end of the duplex, being transferred to each newly added hairpin.

Instruction and chemistry hairpins are added alternately at the active end of the growing DNA duplex. Chemistry hairpins contain no information about the reaction sequence. Both of the unique toehold domains of a chemistry hairpin are used as addresses to identify it and thus the building block that it carries. The two unique toeholds of an instruction hairpin are routing instructions: the external (input) toehold recognizes the preceding chemistry hairpin, and the initially sequestered (output) toehold signals the next chemistry hairpin to be added. The set of instruction hairpins thus constitutes a program that determines the sequence of hairpin addition (Supplementary Fig. 2). For example, the instructions **I>A**, **A>B**, **B>C** specify concatenation of three chemistry hairpins in the sequence **A**, **B**, **C**. This program is deterministic, because there is only one possible assembly sequence. Non-deterministic instruction sets can also be executed: addition of instruction **C>A**

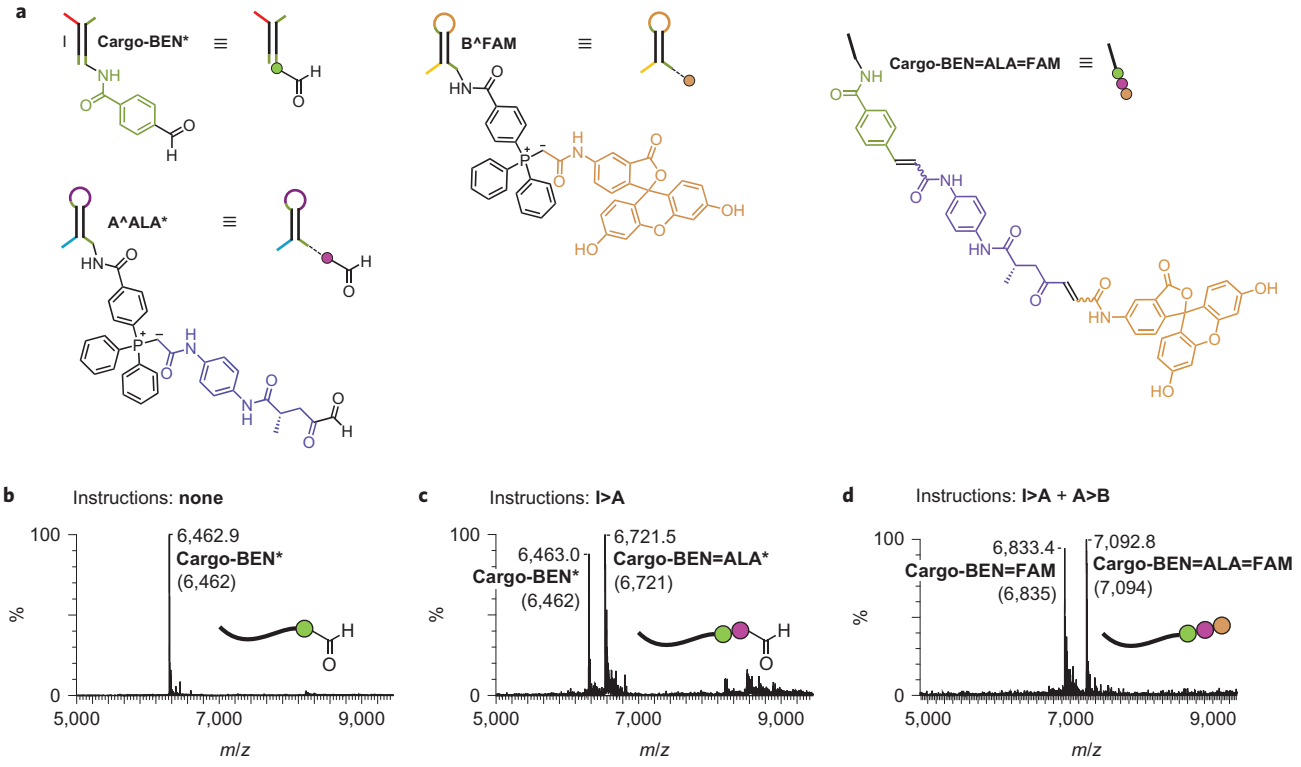

Figure 2 | Ordered synthesis of a diolefin product. a, Building blocks and hairpin adapters used to demonstrate ordered DTS using Wittig chemistry. In naming building blocks, ^ denotes a cleavable linker with phosphonium ylide functionality, * denotes aldehyde functionality, and = denotes a carbon–carbon double bond between building blocks. **–BEN*** is an initiator, **^ALA*** a bifunctional chain propagator and **^FAM** a fluorescent terminator. The structure of programmed DTS product **Cargo-BEN=ALA=FAM** is also shown. **b**–**d**, ESI(–) mass spectra of the **Cargo** strand allowing identification of the covalently attached products. Expected masses of products (Supplementary Fig. 5) are shown in brackets. In **b**, a mix of chemistry strands **I:Cargo-Ben***, **A^ALA*** and **B^FAM** in the absence of instructions yields no significant reactions, and only the **Cargo-BEN*** strand is observed. In **c**, addition of the instruction **I>A** directs synthesis of the monoolefin **Cargo-BEN=ALA***, but there is no significant non-templated reaction with FAM, which is also present. In **d**, addition of both instruction strands **I>A** and **A>B** directs synthesis of the diolefin **Cargo-BEN=ALA=FAM**.

would allow continuous polymerization $(\mathbf{A},\mathbf{B},\mathbf{C})_n$, and further addition of instruction **C>B** would form a random polymer.

As each chemistry hairpin is added to the growing chain it forms a blunt duplex with the **Cargo**. At the free end of this duplex the incoming building block is held adjacent to the reactive end of the oligomer carried by the **Cargo**, an arrangement that favours the covalent reaction transferring the building block to the growing oligomer (Fig. 1c). For building blocks coupled by the Wittig reaction, nucleophilic attack of the ylide group of the incoming building block on the unreacted aldehyde at the distal end of the growing oligomer initiates cleavage of the building block from its DNA adapter and formation of an olefin bond to the oligomer. The spent adapter is left with an unreactive phosphine oxide at its 5′ end[7–10,37]. For peptide bond formation, nucleophilic attack on the NHS ester by the free amine group at the end of the growing oligomer initiates cleavage of the building block from its DNA adapter and formation of an amide bond to the oligomer, leaving an *N*-hydroxysuccinimide moiety at the 5′ end of the spent adapter.

The covalent-bond-forming reactions and the non-covalent DNA hybrization reactions are not coordinated in this mechanism. The yield of oligomer synthesis therefore depends, in part, on the ratio between covalent coupling and DNA HCR rates. The average time for hairpin addition must be tuned to be longer than the average reaction time by adjusting the hairpin concentration: it is ~20 min per step at a hairpin concentration of 5 μM (Supplementary Fig. 3). The completion time for a DNA-templated Wittig reaction has been measured to be less than 5 min under similar conditions[8], and the half-time for an acyl transfer reaction is 5 min (Supplementary Fig. 4). A further limitation of this mechanism is that, as the oligomer grows, the separation between its reactive end and an incoming building block increases and hence it is expected that the coupling yield decreases[7,28,38].

Ordered templated synthesis can be demonstrated by adding different molecular programs (different combinations of instruction hairpins) to generate different products from a common pool of chemistry hairpins. Figure 2 shows an example of programmed synthesis using Wittig chemistry (see Supplementary Method 1 for details of the synthesis of reactants). Mixing two chemistry hairpins (the chain-extending building block **A^ALA*** and the chain terminator **B^FAM**) with the initiating duplex **I:Cargo-BEN*** in the absence of instruction strands results in no detectable coupling reactions (Fig. 2b; see figure caption for an explanation of the symbols). Addition of the instruction strand **I>A** moves the **Cargo** from strand **I** to **A**, templating formation of the product **Cargo-BEN=ALA*** (Fig. 2c). Only small quantities of non-programmed **FAM**-linked products are detected (Supplementary Fig. 5). The addition of both instructions **I>A** and **A>B** results in sequential, programmed DTS of the product **Cargo-BEN=ALA=FAM** (Fig. 2d). Products **Cargo-BEN=FAM** and **Cargo-BEN*** are also detected (Supplementary Fig. 5), indicating that Wittig reactions (although very probably not a DNA assembly step) have been missed, either because the **Cargo** moved on before the coupling reaction occurred or because chemistry strands had lost activity as a result of background phosphine oxidation[39]. Chain extension and the covalent attachment of **FAM** to the **Cargo**-attached olefin product can be demonstrated using gel electrophoresis (Supplementary Fig. 6).

Table 1 | Detected and expected masses of products of programmed synthesis.

Reactants (Cargo-BEN* +)	Instructions	Product	Expected	Detected
A^ALA* + B^FAM	–	Cargo-BEN	6,462.3	6,462.9
A^ALA* + B^FAM	I>A	Cargo-BEN=ALA*	6,721.5	6,721.5
A^ALA* + B^FAM	I>A + A>B	Cargo-BEN=ALA=FAM	7,092.9	7,092.8
		Cargo-BEN=FAM	6,833.6	6,833.4
A^ALA* + C^PHE	I>A + A>C	Cargo-BEN=ALA=PHE	7,133.0	7,135.7
		Cargo-BEN=PHE	6,873.7	6,873.6
C^UNI* + A^ALA	I>C + C>A	Cargo-BEN=UNI=ALA	7,080.9	7,080.8
		Cargo-BEN=ALA	6,797.6	6,799.5
C^UNI* + D^BIO	I>C + C>D	Cargo-BEN=UNI=BIO	7,142.1	7,142.4
		Cargo-BEN=BIO	6,858.8	6,860.6
A^ALA* + D^BIO	I>A + A>D	Cargo-BEN=ALA=BIO	7,118.1	7,118.3
		Cargo-BEN=BIO	6,858.8	6,857.9
C^UNI* + A^ALA* + D^BIO	I>C + C>A + A>D	Cargo-BEN=UNI=ALA=BIO	7,401.3	7,401.6
		Cargo-BEN=UNI=BIO	7,142.1	7,138.6
		Cargo-BEN=BIO	6,858.8	6,858.3

See Supplementary Figs 5 and 8 for structures, and see the caption to Fig. 2 for an explanation of the notation.

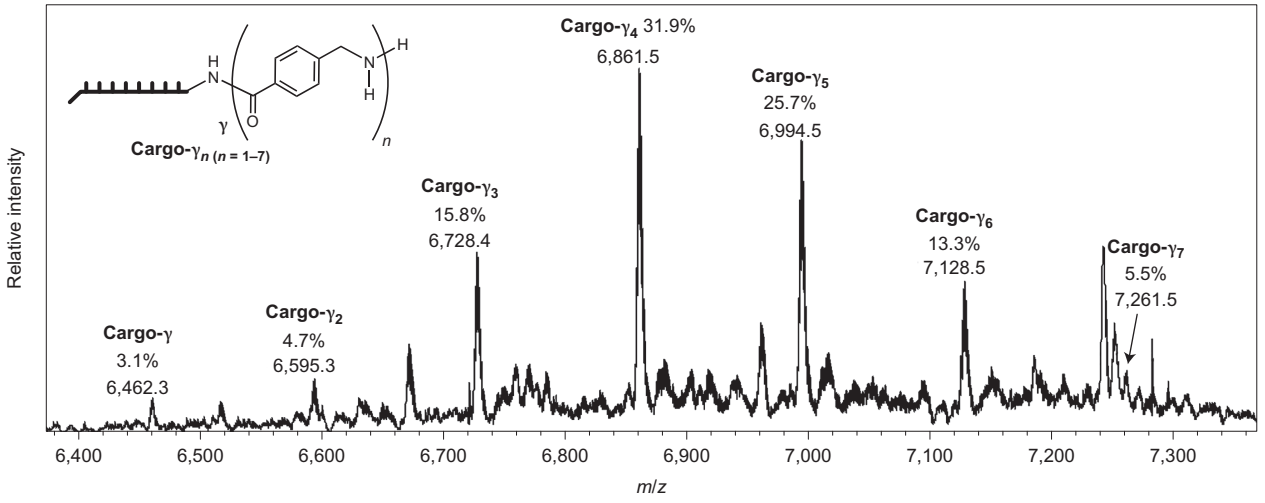

Figure 3 | Programmed synthesis of a polypeptide with a single, repeated building block. The mass spectrum shows **Cargo**-linked products. The reaction was carried out by mixing **I** (1 equiv. = 10 μM), **Cargo-γ** (1 equiv.), **I>C** (1 equiv.), **C-γ** (15 equiv.) and **C>C** (14 equiv.) in 1 M NaCl and 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7). The reaction mixture was incubated on ice (to minimize hydrolysis of NHS esters) for 18 h. The product distribution, inferred from peak intensities, is shown as percentages.

A different programmed sequence of assembly is achieved when instruction **I>B** is added to the same reactants (that is, **I:Cargo-BEN**, **A^ALA*** and **B^FAM**) under the same reaction conditions, resulting in the synthesis of **Cargo-BEN=FAM** (Supplementary Fig. 7).

Other two- and three-step programmed syntheses were carried out (Table 1) using adapters with alanine (**ALA**), phenylalanine (**PHE**), alkyne (**UNI**) and biotin (**BIO**) side-chain functionalities (Supplementary Fig. 8). **^ALA*** and **^UNI*** were used as chain-extending building blocks, and **^ALA**, **^FAM**, **^PHE** and **^BIO**, with protected aldehyde functionality, were used as chain terminators. The 'universal' alkyne adapter **UNI** confers the additional option of orthogonal functionalization, through copper-catalysed azide-alkyne cycloaddition ('click' chemistry), post-DTS[9]. The reactants (chemistry and instruction hairpins and the initiator duplex) were mixed simultaneously. Programmed DTS of three diolefins (**Cargo-BEN=ALA=PHE**, **Cargo-BEN=UNI=ALA** and **Cargo-BEN=UNI=BIO**) and a triolefin (**Cargo-BEN=UNI=ALA=BIO**) was observed (Table 1 and Supplementary Fig. 8). Truncated products, corresponding to missed reactions, were also observed. However, there was no significant yield of non-programmed reactions (cf. Fig. 2b).

Similar results were obtained when the same molecular machinery was used for peptide synthesis (see Supplementary Methods for the synthesis of reaction components). Supplementary Fig. 9 shows the products of programmed synthesis of a tetrapeptide using **γ** (4-(aminomethyl)benzoic acid), **β** (trans-4-(aminomethyl)cyclohexane-1-carboxylic acid) and chain terminator **α** (isovaleric acid) as amino-acid building blocks. Truncated by-products were observed, corresponding to missed reactions which are consistent with degradation of the building block by hydrolysis of the NHS ester or premature progression of the DNA machinery before the acyl transfer reaction could take place. However, in this system, too, there was no evidence of significant non-programmed reactions.

The synthesis of longer oligomers is also possible. Figure 3 shows the results of an acyl transfer reaction with a recursive program, **I>C** and **C>C**, and chemistry hairpin **C-γ** (hairpin **C** conjugated to the NHS-ester of amino acid **γ**, Supplementary Method 2), designed to produce oligomers $\boldsymbol{\gamma_n}$ with a single repeated unit. The product distribution shown in Fig. 3 is inferred from the mass spectrum of **Cargo**-based species (assuming that the ionization efficiencies in negative ion mode electrospray ionization (ESI(−)) of all **Cargo**-based species are dominated by the negatively charged DNA and are therefore approximately equal[28]). Formation of up to six peptide bonds was observed, with the tetrapeptide (the product of three acyl transfer reactions) being the most abundant product.

Figure 4 shows the results of programmed combinatorial synthesis using this mechanism. Two programs were used: chemistry

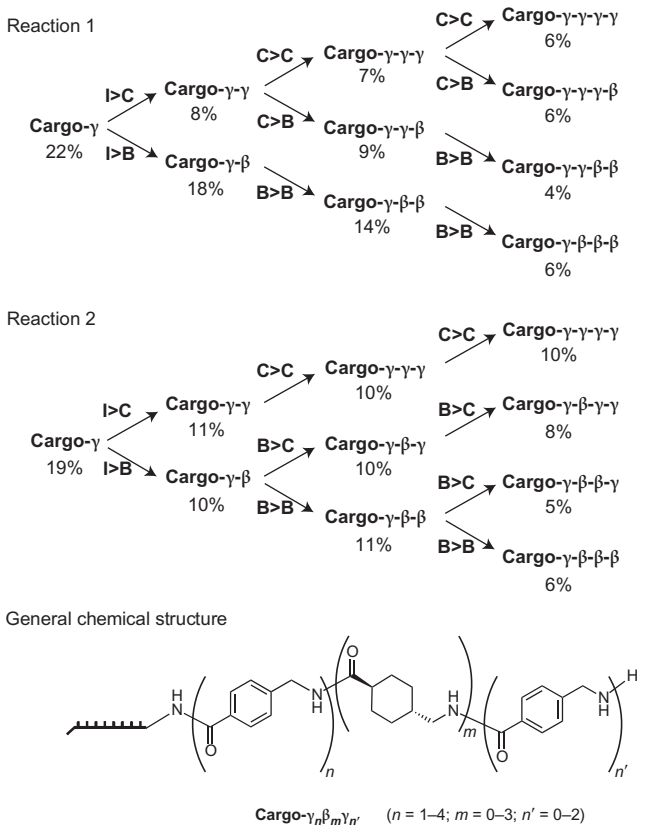

Figure 4 | Programmed combinatorial synthesis of a polypeptide. Reaction of chemistry hairpins **C-γ** and **B-β** with program **I>C**, **I>B**, **C>C**, **B>B**, **C>B** (reaction 1) or **I>C**, **I>B**, **C>C**, **B>B**, **B>C** (reaction 2) was initiated with amino acid **γ** conjugated to **Cargo**. Product distributions inferred from peak intensities in the mass spectrum of **Cargo**-based species (Supplementary Fig. 10) are shown as percentages.

hairpins **C-γ** and **B-β** were combined with instruction hairpins **I>C**, **I>B**, **C>C**, **B>B** and either **C>B** (reaction 1) or **B>C** (reaction 2). Both reactions give synthetic pathways that branch. In reaction 1, for example, addition of **γ** can be followed by addition of **γ** or **β**, whereas addition of **β** is always followed by addition of **β**. The distribution of products (percentages shown in Fig. 4) indicates differences in reactivity between the building blocks and a decrease in reaction efficiency as the oligomer extends.

The DNA duplex formed by the hybrization of chemistry and instruction strands remains attached to the product oligomer and provides a record of the sequence of reactions carried out by the assembler. To make this record readable, DNA components were altered to add primer-binding sequences at the beginning and end of the duplex. The instruction strands that form half of the duplex were ligated to produce a permanent record of the reaction sequence (Fig. 5) using either T4 DNA ligase or 'click' ligation. The triazole linkage produced by this click ligation can be read through by a polymerase[40], and enzyme-free ligation expands the range of reaction conditions accessible to the molecular assembler. The record was amplified by PCR, cloned and sequenced (see Methods). Using both ligation methods, the programmed reaction sequence was successfully read back from the instruction toehold sequences embedded in the reaction record (for sequencing data see Supplementary Fig. 11).

Discussion

Sequence-controlled synthesis is an important and elusive goal of polymer chemistry[41]. We have demonstrated synthetic molecular machinery that is capable of assembling an oligomer in a sequence

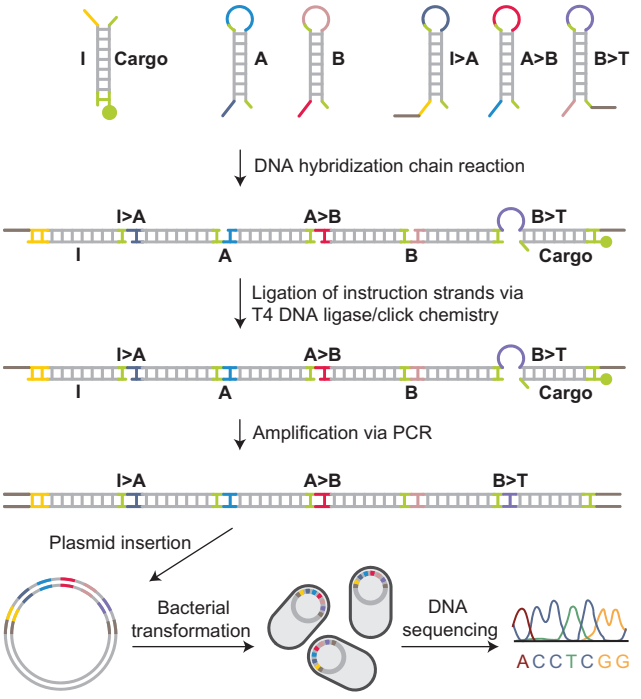

Figure 5 | Recording the reaction sequence. During programmed covalent synthesis a discontinuous DNA duplex is formed by hybridization of instruction and chemistry hairpins. A permanent record of the reaction sequence is obtained by ligation, PCR amplification, cloning and sequencing of the concatenated instruction hairpins. An additional instruction strand, **B>T**, was used to signal the end of the reaction record: this strand and a modified first instruction strand **I>A**, carry PCR primer domains (brown).

specified by a reconfigurable program encoded in DNA. The assembler is modular and can be generalized to other DTS reactions that coordinate bond formation with cleavage from an adapter. We have demonstrated the formation of polyolefin and polypeptide chains. The assembler can operate autonomously in the presence of a pool of reactants, and to change the sequence of the oligomer it is only necessary to change the molecular program. A DNA record of the reaction sequence can be amplified and read, and an assembly program incorporating non-deterministic steps could thus be used to generate a library of different products to be used for *in vitro* selection experiments[2,3,34]. Enzymatic restriction of the amplified record to recover copies of the original instruction strands would make it possible to automate multi-round selection. The number of distinguishable toehold sequences could be increased, at the expense of a reduced hairpin incorporation rate, by the use of a spacer between the toehold and hairpin neck[42], allowing the use of large pools of monomers. The fact that a selected oligomer can be identified by amplifying and sequencing its attached DNA record, analogous to 'ribosome display'[43], means that single-molecule detection sensitivity is, in principle, possible.

Autonomous, programmable covalent synthesis greatly expands the potential functionalities of molecular robotic systems. Molecular sensing and computation have already been integrated to create autonomous systems capable of displaying the result of a computation based on environmental inputs[44–46]. The further integration of autonomous covalent synthesis creates the possibility of molecular systems capable of both sensing and altering their environments, including the possibility of medical molecular robotic devices capable of local diagnosis and treatment.

Methods

HCR-programmed synthesis. The DNA sequences used for HCR are listed in Supplementary Table 1. Preparation of **A^ALA**, **B^FAM**, **D^PHE**, **D^UNI**, **E^BIO** and **Cargo-BEN*** for Wittig reactions and **C-γ**, **B-β**, **A-α** and **Cargo-γ** for acyl

transfer reactions and subsequent purification by HPLC is described in the Supplementary Methods. Mass spectra for modified DNA strands are shown in Supplementary Fig. 1.

Wittig reactions were performed in TAPS buffer (100 mM TAPS, 1 M NaCl, pH 8.5). Reactions at 0.5 μM (Table 1 and Supplementary Figs 6 and 7) were prepared using 20 μM hairpin stocks in deionized H_2O. The aldehyde groups on **A^ALA** and **D^UNI** were deprotected by adding 0.5 volumes of sodium periodate (50 mM $NaIO_4$, 0.5 M NaOAc, pH 3.5). For experiments at 5 μM (Fig. 2 and Supplementary Figs 4 and 5), stocks in deionized H_2O were concentrated *in vacuo* and, in the case of **A^ALA**, deprotected as above, then resuspended at 50 μM in TAPS buffer. The initiator duplex was prepared by annealing equal quantities of strands **I** and **Cargo-BEN*** in TAPS buffer: samples were cooled from 95 °C to room temperature over 10 min. Before use, hairpins were incubated at 95 °C for 2 min then quenched at −20 °C for 1 min. When present, instruction strands were added in 10% excess over other components. On completion of a reaction performed at 0.5 μM, the reaction mixture was desalted using a NAP-10 column (GE Healthcare), equilibrated with deionized H_2O then concentrated *in vacuo*. A second desalting step was performed using a Micro Bio-Spin 6 column (Bio-Rad) before a final concentration *in vacuo*. Samples were analysed by liquid chromatography-mass spectrometry (LC-MS).

Acyl transfer reactions were performed in 1 M NaCl with 50 mM NaH_2PO_4/Na_2HPO_4, pH 7.0, with **Cargo** at 10 μM and equal amounts of other DNA reagents (unless otherwise specified). The reaction mixture was shaken at 4 °C overnight then desalted using a NAP-10 column. **Cargo** was isolated by denaturing HPLC (Supplementary Method 3, HPLC B) before analysis by ultra performance liquid chromatography-mass spectrometry (UPLC-MS).

Mass spectrometry. LC-MS was performed by eluting samples from a Waters XBridge OST C18 (Supplementary Method 3, HPLC C) directly into a Waters LCT Premier reflectron time-of-flight (TOF) mass spectrometer. Data were acquired in the negative-ion mode and analysed and deconvolved using the manufacturer's software (MassLynx V4.1, Waters).

UPLC-MS was performed by eluting samples from a Waters Acquity UPLC OST C18 (Supplementary Method 3, UPLC D) directly into a Bruker μTOF mass spectrometer. Data were acquired in the negative-ion mode and analysed and deconvoluted using MaxEnt.

Recording assembly: enzymatic ligation. Hairpins and initiator duplex were mixed at a concentration of 5 μM (each component) for 2 h. The products of the DNA assembly reaction were then diluted to 250 nM in T4 DNA ligase reaction buffer with 20 units μl^{-1} T4 DNA ligase (all enzymes and their reaction buffers from NEB). The mixture was incubated at room temperature for 1 h, followed by heat inactivation for 20 min at 65 °C.

PCR reactions containing 0.03 units μl^{-1} VentR(exo−), 120 pM ligated sample, 200 μM dNTPs and 0.5 μM primers were amplified with 25 cycles of 96 °C for 30 s, 56 °C for 30 s and 72 °C for 30 s. A product concentration of 30 ng μl^{-1} was obtained after removal of nucleotides and primers using a Qiagen PCR purification kit.

PCR product was cloned into pcDNA5/FRT (ThermoFisher) modified to express the fluorescent proteins citrine and mCherry using Gibson assembly reactions containing 30 pg each of plasmid and insert. HCR primer overhangs were designed to insert into an intronic region of the citrine gene[47]. Transformation of *Escherichia coli* (NEB electrocompetent) with the reaction mix gave ampicillin-resistant colonies. Miniprep plasmid DNA from two colonies was isolated using a Qiagen miniprep kit and sequenced.

Recording assembly: click ligation. Instruction hairpins were modified with azide and alkyne (Supplementary Table 1). Hairpins and initiator duplex at a concentration of 10 μM were incubated for 18 h in the presence of $CuSO_4$ (240 nmol), sodium ascorbate (2.4 μmol) and tris(3-hydroxypropyltriazolylmethyl) amine (1.68 μmol), as previously described[40]. The reaction mixture was desalted by passing through a Micro Bio-Spin 6 column (Bio-rad). The sample was purified by denaturing gel electrophoresis. PCR products (prepared as above) were cut with HindIII and EcoRI and ligated into pUC19 with T4 DNA ligase. Transformation of *E. coli* (DH5α, NEB C2987) with the reaction mix gave ampicillin-resistant colonies. Miniprep plasmid DNA from two colonies was isolated using a Qiagen miniprep kit and sequenced.

Received 19 June 2014; accepted 24 February 2016; published online 11 April 2016

References

1. Gartner, Z. J., Kanan, M. W. & Liu, D. R. Multistep small-molecule synthesis programmed by DNA templates. *J. Am. Chem. Soc.* **124,** 10304–10306 (2002).
2. Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305,** 1601–1605 (2004).
3. Hansen, M. H. *et al.* A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **131,** 1322–1327 (2009).
4. He, Y. & Liu, D. R. A sequential strand-displacement strategy enables efficient six-step DNA-templated synthesis. *J. Am. Chem. Soc.* **133,** 9972–9975 (2011).
5. Li, X., Gartner, Z. J., Tse, B. N. & Liu, D. R. Translation of DNA into synthetic *N*-acyloxazolidines. *J. Am. Chem. Soc.* **126,** 5090–5092 (2004).
6. Liao, S. & Seeman, N. C. Translation of DNA signals into polymer assembly instructions. *Science* **306,** 2072–2074 (2004).
7. McKee, M. L. *et al.* Programmable one-pot multistep organic synthesis using DNA junctions. *J. Am. Chem. Soc.* **134,** 1446–1449 (2012).
8. McKee, M. L. *et al.* Multistep DNA-templated reactions for the synthesis of functional sequence controlled oligomers. *Angew. Chem. Int. Ed.* **49,** 7948–7951 (2010).
9. Milnes, P. J. *et al.* Sequence-specific synthesis of macromolecules using DNA-templated chemistry. *Chem. Commun.* **48,** 5614–5616 (2012).
10. Snyder, T. M. & Liu, D. R. Ordered multistep synthesis in a single solution directed by DNA templates. *Angew. Chem. Int. Ed.* **44,** 7379–7382 (2005).
11. Lewandowski, B. *et al.* Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science* **339,** 189–193 (2013).
12. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289,** 905–920 (2000).
13. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289,** 920–930 (2000).
14. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
15. Seeman, N. C. DNA in a material world. *Nature* **421,** 427–431 (2003).
16. Chworos, A. *et al.* Building programmable jigsaw puzzles with RNA. *Science* **306,** 2068–2072 (2004).
17. Li, X. & Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem. Int. Ed.* **43,** 4848–4870 (2004).
18. Turberfield, A. J. *et al.* DNA fuel for free-running nanomachines. *Phys. Rev. Lett.* **90,** 118102 (2003).
19. Green, S. J., Bath, J. & Turberfield, A. J. Coordinated chemomechanical cycles: a mechanism for autonomous molecular motion. *Phys. Rev. Lett.* **101,** 238101 (2008).
20. Omabegho, T., Sha, R. & Seeman, N. C. A bipedal DNA Brownian motor with coordinated legs. *Science* **324,** 67–71 (2009).
21. Muscat, R. A., Bath, J. & Turberfield, A. J. A programmable molecular robot. *Nano Lett.* **11,** 982–987 (2011).
22. Bath, J., Green, S. J. & Turberfield, A. J. A free-running DNA motor powered by a nicking enzyme. *Angew. Chem. Int. Ed.* **44,** 4358–4361 (2005).
23. Tian, Y., He, Y., Chen, Y., Yin, P. & Mao, C. A DNAzyme that walks processively and autonomously along a one-dimensional track. *Angew. Chem. Int. Ed.* **44,** 4355–4358 (2005).
24. Bath, J., Green, S. J., Allen, K. E. & Turberfield, A. J. Mechanism for a directional, processive, and reversible DNA motor. *Small* **5,** 1513–1516 (2009).
25. Lund, K. *et al.* Molecular robots guided by prescriptive landscapes. *Nature* **465,** 206–210 (2010).
26. Wickham, S. F. J. *et al.* Direct observation of stepwise movement of a synthetic molecular transporter. *Nature Nanotech.* **6,** 166–169 (2011).
27. Wickham, S. F. J. *et al.* A DNA-based molecular motor that can navigate a network of tracks. *Nature Nanotech.* **7,** 169–173 (2012).
28. He, Y. & Liu, D. R. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nature Nanotech.* **5,** 778–782 (2010).
29. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **101,** 15275–15278 (2004).
30. Lubrich, D., Green, S. J. & Turberfield, A. J. Kinetically controlled self-assembly of DNA oligomers. *J. Am. Chem. Soc.* **131,** 2422–2423 (2009).
31. Venkataraman, S., Dirks, R. M., Rothemund, P. W. K., Winfree, E. & Pierce, N. A. An autonomous polymerization motor powered by DNA hybridization. *Nature Nanotech.* **2,** 490–494 (2007).
32. Yin, P., Choi, H. M. T., Calvert, C. R. & Pierce, N. A. Programming biomolecular self-assembly pathways. *Nature* **451,** 318–322 (2008).
33. McKee, M. L. *et al.* Peptidomimetic bond formation by DNA-templated acyl transfer. *Org. Biomol. Chem.* **9,** 1661–1666 (2011).
34. Halpin, D. R. & Harbury, P. B. DNA display II genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol.* **2,** 1022–1030 (2004).
35. Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406,** 605–608 (2000).
36. Green, S. J., Lubrich, D. & Turberfield, A. J. DNA hairpins: fuel for autonomous DNA devices. *Biophys. J.* **91,** 2966–2975 (2006).
37. Dambacher, J. *et al.* Water is an efficient medium for Wittig reactions employing stabilized ylides and aldehydes. *Tetrahedron Lett.* **46,** 4473–4477 (2005).
38. Gartner, Z. J., Grubina, R., Calderone, C. T. & Liu, D. R. Two enabling architectures for DNA-templated organic synthesis. *Angew. Chem. Int. Ed.* **42,** 1370–1375 (2003).

39. Schnell, A., Dawber, J. G. & Tebby, J. C. The mechanism of hydrolysis of phosphonium ylides. *J. Chem. Soc. Perkin Trans.* **2,** 633–636 (1976).
40. El-Sagheer, A. H., Sanzone, A. P., Gao, R., Tavassoli, A. & Brown, T. Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108,** 11338–11343 (2011).
41. Badi, N. & Lutz, J.-F. Sequence control in polymer synthesis. *Chem. Soc. Rev.* **38,** 3383–3390 (2009).
42. Genot, A. J., Zhang, D. Y., Bath, J. & Turberfield, A. J. Remote toehold: a mechanism for flexible control of DNA hybridization kinetics. *J. Am. Chem. Soc.* **133,** 2177–2182 (2011).
43. Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA* **94,** 4937–4942 (1997).
44. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335,** 831–834 (2012).
45. Muscat, R. A., Bath, J. & Turberfield, A. J. Small molecule signals that direct the route of a molecular cargo. *Small* **8,** 3593–3597 (2012).
46. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R. & Shapiro, E. An autonomous molecular computer for logical control of gene expression. *Nature* **429,** 423–429 (2004).
47. Gibson, D. G. *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).

Acknowledgements

The authors thank E. Stulz and L. Gong for discussions. This research was supported by Biotechnology and Biological Sciences Research Council sLOLA grants BB/J00054X/1 and BB/J001694/2, Engineering and Physical Sciences Research Council grants EP/F056605/1, EP/F008597/1, EP/I016651/1, EP/F055803/1, EP/F009062/1, EP/G037930/1 and EP/P504287/1, and a Royal Society–Wolfson Research Merit Award (to A.J.T.).

Author contributions

W.M., R.A.M., M.L.M. and P.J.M. performed the experiments. W.M., R.A.M., M.L.M., P.J.M., J.B., R.K.O'R. and A.J.T. contributed to experimental design, interpretation of data and preparation of the manuscript. A.E.S., T.B. and B.G.D. provided critical materials and advice on covalent reactions and product purification.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.J.T.

Competing financial interests

The authors declare no competing financial interests.