# Transfer of Sequence Information and Replication of Diimine Duplexes

Kyle R. Strom,*[ID] Jack W. Szostak,[ID] and Noam Prywes*,†[ID]

Howard Hughes Medical Institute, Department of Molecular Biology, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, United States

Ⓢ *Supporting Information*

ABSTRACT: The ability of the biopolymers RNA and DNA to store and transfer information is essential to life. Herein, we demonstrate template-directed replication in a set of dimer duplexes that use reversible covalent bonds to form base-pairing interactions. Binary sequence information was encoded as a sequence of aniline and benzaldehyde subunits linked together by a diethynyl benzene backbone. These dimers formed sequence-specific, imine-linked duplexes, which could be separated and used as templates for the synthesis of daughter duplexes with identical sequences.
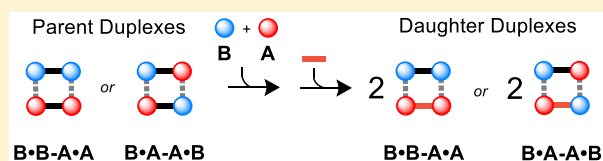
INTRODUCTION

Genetic polymers store and copy information. All known life on Earth is based on functional information stored in the genetic polymers DNA and RNA. In cells, the replication of these polymers is aided by a multitude of enzymes. However, this machinery is not thought to be strictly necessary for replication; RNA and some RNA-like polymers can undergo nonenzymatic templated copying.[1−4] Evidently, the molecular structure alone can allow for both information storage and copying. Modifications to the backbone and nucleobases of the natural biopolymers have been explored extensively, and a range of alternative genetic polymers have been produced.[5,6] These molecules are, to varying degrees, bioinspired. They are structural analogues of natural biopolymers and employ base-pairing interactions like the purine-pyrimidine base pairing found in DNA and RNA.[7−14]

The chemical space of genetic polymers beyond nucleotides has only just begun to be explored in depth. The structural constraints necessary for information storage and replication in a molecule or polymer are undefined. However, three general features are necessary: a sequence-defined template, sequence-specific base pairing of monomers to the template, and a way to link the template bound monomers.

Sequence-definition allows for arbitrary information storage in any polymer with at least two different subunits.[15] Many synthetic, sequence-defined polymers which bear little resemblance to biopolymers have been constructed,[16] suggesting that the diversity of accessible informational polymers is likely to be vast. Notably, these polymers are typically constructed by the stepwise addition of monomer subunits, with no obvious mechanism for replication. In biopolymers, replication is enabled by the formation of duplexes. Potentially, any two monomers that recognize each other and interact reversibly can function as a base-pair in a genetic polymer. While many synthetic duplexes have been described with non-nucleic acid base-pairing interactions,[17−21] only a few have

been demonstrated to facilitate template-directed polymerization. Luh and co-workers have shown that double-stranded polynorbornenes held together by hydrolyzable ferrocene diesters function as templates for the synthesis of daughter polynorbornenes via ring-opening metathesis polymerization (ROMP).[22] Kamonsutthipaijit and Anderson have reported the template-directed synthesis of linear alkyne-terminated porphyrin oligomers using the coordination of pyridine to Zn-porphyrins as a base pair and Glasser−Hay coupling to link monomers.[23] While these examples show exceptional control of sequence length and polydispersity, they do not show the transfer of sequence information as both the starting template and synthesized polymer are homopolymers without sequence definition.

There are only a few examples of sequence-defined duplexes which do not rely on nucleic acid base-pairing. Lehn and co-workers have shown that sequence-specific mixed bipyridine and terpyridine trimers form complementary duplexes by coordination to specific metal ions.[17] Gong and co-workers have shown that sequence-defined polyamide duplexes associate more strongly with complementary H-bonding donor−acceptor pairs.[18] Hunter and co-workers have shown trimer duplexes held together by hydrogen-bonding interactions between phenol and *N*-oxo pyridine subunits form preferentially between complementary sequences.[20] While these examples demonstrate that sequence recognition and the formation of sequence-specific duplexes is not limited to nucleic acids, none have been shown to undergo templated copying. To the best of our knowledge, molecules unrelated to biomolecules, that both store sequence information and undergo template-directed transfer of that sequence information, have not been demonstrated.

Templated reactions utilizing dimer templates have been exploited very successfully in autocatalytic self-replicators like those pioneered by Rebek[6,7] and expanded by others.[24,25] In these systems, monomers and templates are in dynamic equilibrium such that the products of a templated reaction can function as a new template in subsequent reactions in the same pot. While these minimal self-replicators impressively recapitulate the behavior of much more complicated replicases, they do not contain or replicate sequence information like their biological self-replicating counterparts.

Work by Moore and co-workers has shown polyaniline and polyaldehyde homopolymers reversibly form polyimine duplexes.[26–28] While these homopolymers do not encode sequence information and were not shown to facilitate templated copying, they demonstrate the feasibility of imine bonds as a base pair. We expected that heteropolymeric materials with sequence-defined aniline and benzaldehyde subunits would from duplexes between complementary sequences, and that these duplexes could function as templates for the copying of sequence information from parent duplexes to daughter duplexes.

The simplest model system with which to demonstrate both information storage and replication in these polymers is a set of dimers (Figure 1). Binary information storage in these dimers can b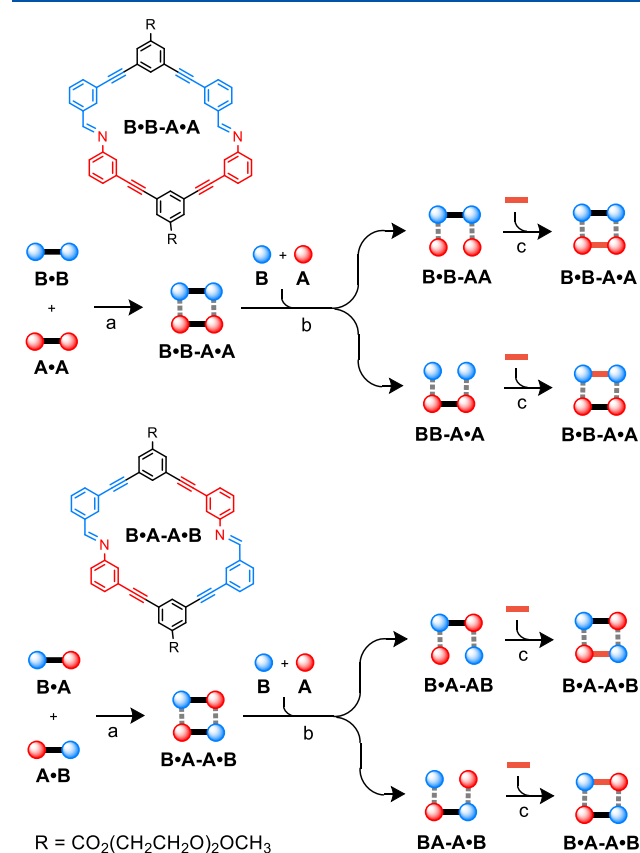e readily visualized as a two-letter alphabet with benzaldehyde subunits as "B" (blue) and aniline subunits as "A" (red). There are three possible dimeric sequences, dianiline A·A, dibenzaldehyde B·B, and and aniline–aldehyde A·B. Because the backbone of A·B is symmetric, there is no difference between A·B and B·A. Using this notation, A·A—B·B is the duplex formed from the condensation of the complementary sequences A·A and B·B, and B·A—A·B is the duplex formed from the condensation of the self-complementary sequence A·B (Figure 1, step a).

Two steps are required for the replication of these duplexes. First, addition of excess monomers A and B separates the duplexes to form two strands with monomers templated in a sequence specific fashion (Figure 1, step b). Second, linking of these monomers produces two daughter complexes with the same A·A—B·B, or B·A—A·B, sequences as the corresponding parent sequences (Figure 1, step c).

Herein, we report the synthesis of information-containing synthetic duplexes A·A—B·B and B·A—A·B and demonstrate their template-directed replication. These duplexes feature reversible covalent imine bonds as base pairs, a fully conjugated ethynylbenzene backbone, with a short PEG group for solubility, and undergo replication in organic solvent via imine formation followed by Pd(0)/Cu(I) catalyzed Sonogashira cross coupling.[29]

RESULTS AND DISCUSSION

Homodimers B·B and A·A were synthesized in a single step via a double Sonogashira reaction of 3-ethynylbenzaldehyde B or 3-ethynylaniline A with 0.2 equiv of diglymyl 3,5 diiodobenzaote 1 (Figure 2). Adding a solution of dialdehyde B·B in benzene to a solution of its dianiline complement A·A in benzene and allowing the mixture to sit undisturbed at 0 °C for 3 days afforded pure crystals of the double-condensation product B·B—A·A in quantitative yield. The structure of B·B—A·A was confirmed by X-ray crystallography (Supplementary
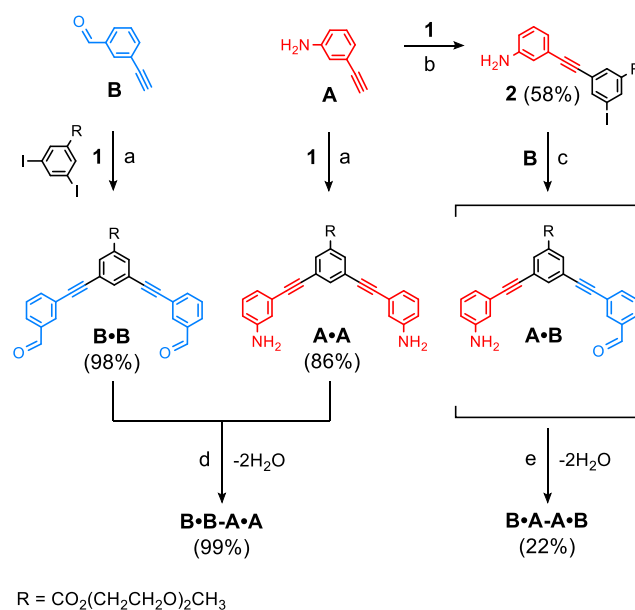

Figure 1. Formation and replication of dimeric dimine duplexes. (a) Complementary sequences B·B and A·A, or A·B, form sequence-specific duplexes B·B—AA or B·A—A·B. (b) Addition of monomers B and A separates the duplex and forms two dimers templated with monomers strands. (c) Coupling of the templated monomers gives two daughter duplexes with the same B·B—A·A or B·A—A·B, sequence as the parent.

Figure 2. Synthesis of duplexes A·A—B·B and B·A—A·B: (a) 0.2 equiv of 1, cat. Pd(PPh₃)₄ and CuI; (b) 5 equiv of 1, cat. Pd(PPh₃)₄, and CuI; (c) 1.1 equiv of B, cat. Pd(PPh₃)₄, and CuI; (d) 0.1 M A·A and B·B in C₆H₆ with 0.5% TFA; (e) pure mixture of residue was precipitated from hot 4:1, C₆H₆/CHCl₃.
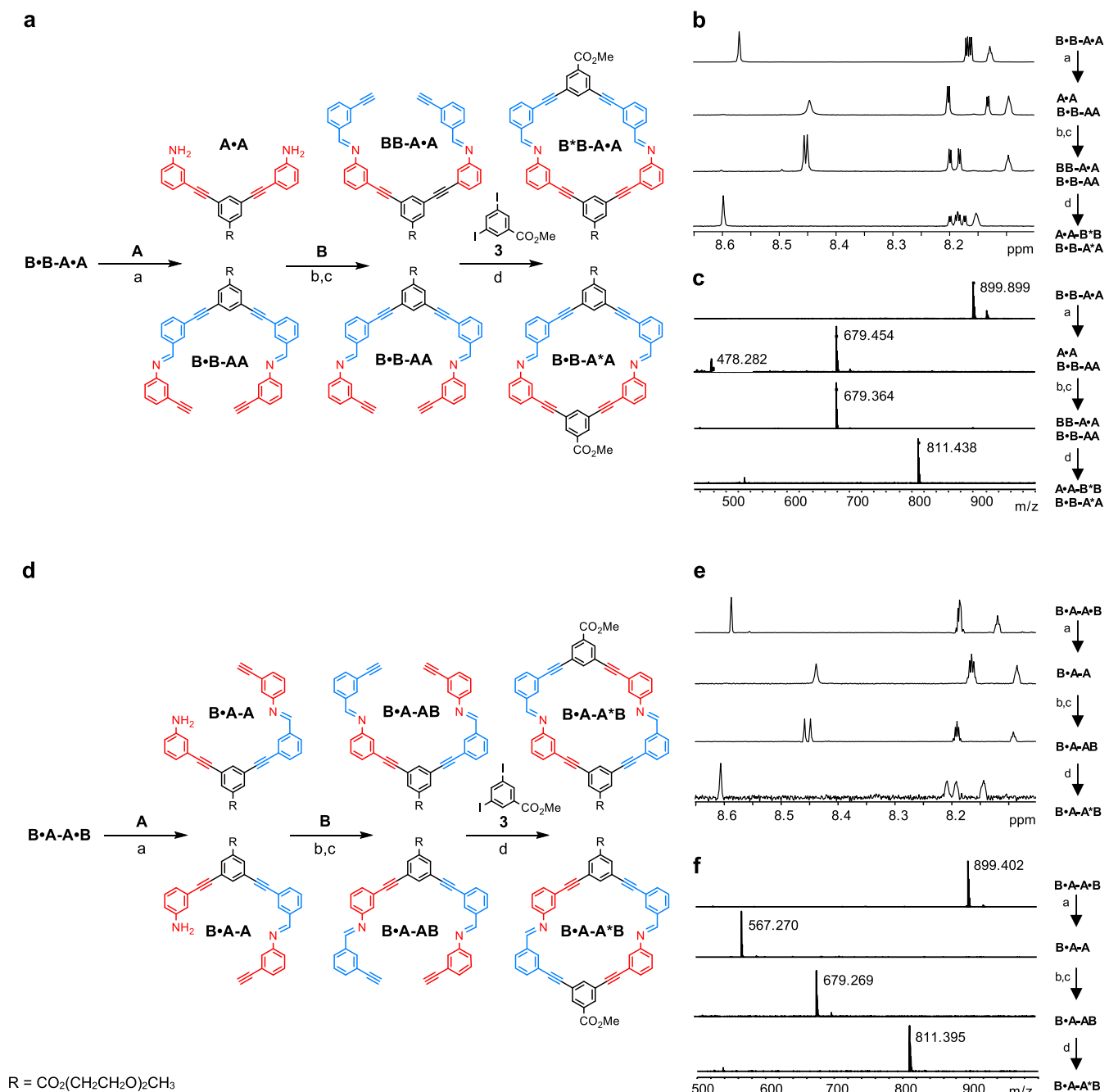
Figure 3. Replication of **B·B—A·A** and **B·A—A·B**: (a) 30 equiv of 3-ethynylaniline **A**, CDCl$_3$ 0.02% TFA; (b) 60 equiv of 3-ethynylbenzaldehyde **B** and 4 Å molecular sieves; (c) sublimation, 0.02 mbar, 120 °C, neat; (d) 5 equiv of methyl 3,5 diiodobenzoate **2**, 1 equiv of Pd(PPh$_3$)$_4$, 0.5 equiv of CuI, 6 μM template complex in degassed DMF, 60 °C. (a) Scheme showing the four-step replication sequence of **B·B—A·A**. (b) NMR expansion of downfield aromatic and imine protons for replication of **B·B—A·A**, steps a—d, from left to right, imine protons, *meta* ester protons, *meta* imine (carbon side) singlet. (c) MALDI-TOF spectra for replication of **B·B—A·A**, steps a—d. (d) Scheme showing the four-step replication sequence of **B·A—A·B**. (e) NMR expansion of downfield aromatic and imine protons for replication of **B·A—A·B**, steps a—d, from left to right, imine protons, *meta* ester protons, *meta* imine (carbon side) singlet. (f) MALDI-TOF spectra for replication of **B·A—A·B**, steps a—d.

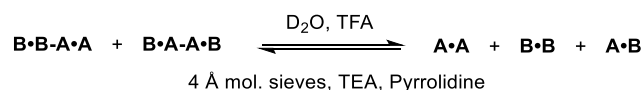Figure 4) in addition to NMR and MALDI-MS (Figure 3b,c top spectra).

The synthesis of heterodimer **A·B** was carried out via two sequential Sonogashira reactions (Figure 2). First, 3-ethynylaniline **A** and a 5-fold excess of diglymyl 3,5-diiodobenzoate **1** were coupled to produce aryl iodide **2** (58%). A second Sonogashira cross coupling of **2** with 3-ethynylbenzaldehyde **B** produced a mixture of amino aldehyde dimer **A·B** and duplex **B·A—A·B**. Given the self-complemen-

tary nature of **A·B**, the spontaneous formation of the **B·A—A·B** duplex is not surprising. Mild heating followed by gentle removal of the solvent under vacuum afforded the desired duplex **B·A—A·B**, which was obtained as a pure white solid after precipitation from a mixture of chloroform and benzene. The structure of **B·A—A·B** was confirmed by NMR and MALDI-MS (Figure 3e,f top spectra).

In the absence of water, the starting duplexes **A·A—B·B** and **B·A—A·B** were stable in CDCl$_3$ for months at 4 °C; however,

they could be completely hydrolyzed back to the single-stranded dimers by the addition of D_2O and trace TFA. Using pyrrolidine as a catalyst,[30] TEA as a base, and 4 Å molecular sieves to remove water, the duplexes were reformed nearly quantitatively (Supplementary Figure 3). Additionally, a $CDCl_3$ solution containing both duplexes, **B·B−A·A** and **B·A−A·B**, gave all the three dimers, **A·A**, **B·B**, and **A·B**, on hydrolysis with D_2O and trace acid. This mixture of dimers was then reannealed in the same NMR tube by the addition of TEA, 4 Å molecular sieves, and pyrrolidine, reforming the duplexes corresponding to the sequence complementary pairs (Scheme 1).

Scheme 1. Hydrolysis and Annealing of B·B, A·A, and A·B Dimers

$$\text{B·B-A·A} + \text{B·A-A·B} \quad \underset{\text{4 Å mol. sieves, TEA, Pyrrolidine}}{\overset{D_2O,\ \text{TFA}}{\rightleftharpoons}} \quad \text{A·A} + \text{B·B} + \text{A·B}$$

Replication of both duplexes **B·B−A·A** and **B·A−A·B** was carried out under identical conditions in four steps, a−d (Figure 3). All steps were conducted sequentially without chromatography. To monitor the progress of the reactions, the first two steps were conducted in an NMR tube in $CDCl_3$.

In step a, the duplexes were separated by treatment with an excess of 3-ethynylaniline **A** and catalytic TFA. This prompted the transimination of the duplex imines with the large excess driving the separation of the duplex to completion and quantitatively condensing 3-ethynylaniline **A** to the template strands. For the **B·B−A·A** duplex, this produced a solution with the **B·B** half of the starting duplex templating two unlinked **A** monomers (**B·B−AA**), while the **A·A** half of the starting duplex was left free in solution (Figure 3a, step a). The NMR spectrum of this solution showed **A·A** and a new upfield-shifted resonance consistent with the two imines of **B·B−AA** (Figure 3b, step a). The MALDI-TOF spectrum showed two peaks, m/z 478.282 and 679.454, consistent with m + H/1 for **A·A** and **B·B−AA**, respectively (Figure 3c, step a). For the complex **B·A−A·B**, monomer **A** was attached to the **B** subunit of each half of the duplex giving 2 equiv of iminoaniline **B·A−A** (Figure 3d, step a). The NMR spectrum of this solution showed an imine proton and set of aromatic resonances consistent with **B·A−A** (Figure 3e, step a). The MALDI-TOF spectrum showed a single peak, m/z 567.270, consistent with m + H/1 for **B·A−A** (Figure 3f, step a).

In step b, an excess of 3-ethynylbenzaldehyde **B** was added along with 4 Å molecular sieves. Just as in step a, the excess of monomer drove the equilibrium of imine formation to fully associate the free **A** subunits of the template strands with the **B** monomer. For **B·B−A·A**, the **A·A** half of the starting duplex was templated with two **B** monomers (**A·A−BB**), and **B·B−AA** produced in step a was left unchanged (Figure 3a, step b). The NMR spectrum of this solution showed unchanged **B·B−AA**, the disappearance of **A·A**, and the appearance of a new imine resonance and set of aromatic resonances consistent with **A·A−BB** (Figure 3b, step b). The MALDI-TOF spectrum showed a single peak, m/z 679.361, consistent with m+H/1 for both **A·A−BB** and **B·B−AA**, as they have the same mass (Figure 3c, step b). For **B·A−A·B**, the untemplated **A** subunit of **B·A−A** was templated to give **B·A−AB** (Figure 3d, step b). The NMR spectrum of this solution showed two imine protons and a set of aromatic resonances consistent with **B·A−AB**

(Figure 3e, step b). The MALDI-TOF spectrum showed a single peak, m/z 679.269, consistent with m + H/1 for **B·A−AB** (Figure 3f, step b).

The larger excess of **B** required in step b compared to step a (60 equiv compared to 30 equiv) was necessary to fully condense the residual untemplated aniline monomer **A** from the 30-fold excess used in step a, and the 4 Å molecular sieves were necessary to remove the water generated by this condensation (Scheme 2).

Scheme 2. Condensation of A and B Monomers
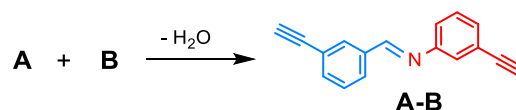
The order of monomer addition, **A** followed by **B**, was also essential for successful replication. Without complete conversion of the nucleophilic anilines to imines, transimination reactions would persist during the subsequent steps. With no obvious pathway for hydrolysis mediated imine exchange, transimination, or imine metathesis,[31] the equilibrium established by the large excess of **A** and **B** was essentially "fixed" by the removal of water and aniline.

At this stage in replication, the parent duplexes had been separated and the desired monomer−template strands were present in solution. However, an excess of monomer **B**, and the condensation product **A−B**, was also present. Removal of these undesired monomers was necessary to prevent them from reacting during the final replication step. Given the susceptibility to hydrolysis of the template bound monomers, standard purification techniques such as column chromatography were not successful. Surprisingly, given that **A−B** contains 17 carbon atoms, vacuum sublimation overnight at 120 °C removed the superfluous monomers completely (step c). The sublimate consisted of **B** and **A−B** and the residue consisted exclusively of the template strands with the associated monomers bound. In the absence of water or anilines, the solutions of template-bound monomers were stable, and no imine exchange was observed in chloroform over several weeks.

Finally, the monomers bound to the parent template were linked via Sonogashira coupling with 3,5-diiodobenzoate ester **3** using $Pd(PPh_3)_4$ and CuI catalysts in dilute DMF (step d). To distinguish the starting parent duplexes **B·B−A·A** and **B·A−A·B** from the daughter duplexes, methyl 3,5-diiodobenzoate **3** was used instead of the diglymyl 3,5-diiodobenzoate **1** used to link the parent dimers. This provided a convenient mass difference between the parent and daughter duplexes, as the starting duplexes were linked by two diglymyl benzoates (denoted by a dot) and the daughters were expected to have one methyl benzoate (donated with an asterisk) and one diglymyl benzoate. The use of methyl benzoate **3** as a linker also provided a means to distinguish off-template from on-template coupling of monomers. The only aryl-iodide available in step d was methyl benzoate **3**, so off-template coupling and annealing would give products with methyl benzoates on both sides of the duplex (e.g., **B*B−A*A**).

Replication that began with duplex **B·B−A·A** gave exclusively the expected daughter duplexes **B*B−A·A** and **B·B−A*A** (Figure 3a, step d). The NMR spectrum showed two overlapping sets of imine protons with a similar downfield shift

(δ 8.60) to the starting duplex and a set of aromatic resonances consistent with **B*B−A·A** and **B·B−A*A** in a 1:1 ratio (Figure 3b, step d). The MALDI-TOF spectrum showed a single peak, m/z 811.438, consistent with the mass of the expected daughters containing one methyl benzoate backbone and one diglymyl benzoate backbone (Figure 3c, step d).

Replication that began with duplex **B·A−A·B** also gave the expected daughter duplex **B·A−A*B** (Figure 3d, step d). Multiple rounds of precipitation from CHCl₃ with C₆H₆ failed to remove the PPh₃ impurity. The NMR spectrum showed two overlapping imine protons with a similar downfield shift (δ 8.61) to the starting duplex and a set of aromatic resonances consistent with **B·A−A*B** (Figure 3e, step d). The MALDI-TOF spectrum showed a single peak, m/z 811.395, consistent with the mass of the expected daughter containing one methyl benzoate backbone and one diglymyl benzoate backbone (Figure 3f, step d).

Successful replication of parent−template complexes **B·B−AA**, **BB−A·A**, and **B·A−AB** required intermolecular coupling to diiodide **3**, followed by intramolecular macrocyclization (Figure 4). Two alternative Sonogashira coupling pathways
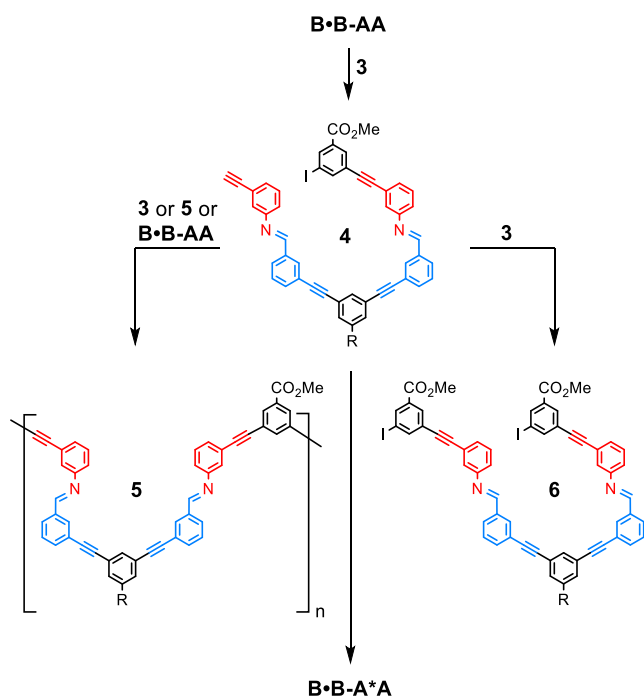

Figure 4. Stepwise replication of **B·B−AA** and potential side products.

were anticipated to divert replication: polymerization, to products such as **5**, caused by intermolecular reactions between template strands and capping, to products such as **6**, caused by the intermolecular reaction with a second diiodide. Polymerization reactions were prevented by using a 5-fold excess of the diiodobenzoate **3**. Capping reactions were prevented by running the reaction under dilute conditions (6 μM template). Reactions at concentrations greater than 50 μM gave complex mixtures presumed to be mixtures of polymers and capped products. Given the high dilution, a full equivalent of Pd(PPh₃)₄ and 0.5 equiv of CuI were used. Likely because of the large amount of the CuI catalyst, the coupling was very sensitive to dissolved O₂, and rigorous degassing was required to prevent unwanted Glaser coupling products.

CONCLUSIONS

Two duplexes encoding two different sequences were synthesized and replicated under the same set of conditions. Parent duplexes **B·B−A·A** and **B·A−A·B** functioned as templates for the synthesis of two daughter duplexes, and the sequence information contained in the parent was transferred to the daughters. Although these dimers only encoded a single bit of information, this demonstrates that molecules and chemistry unrelated to biopolymers can be used to form base-pairing interactions and facilitate the template-directed transfer of sequence information.

The differences between this system and RNA and DNA are notable. These dimers replicated in the absence of water as a solvent, with all the steps of replication occurring in aprotic organic solvent. Instead of hydrogen bonds, the base-pairing interactions were covalent bonds. While the phosphate backbone of DNA and RNA confers water solubility, the backbone of these dimers used a diglymyl moiety to confer organic solubility. Replication in DNA and RNA, both nonenzymatic and enzymatic, occurs by a substitution reaction; the hydroxyl of a sugar on the end of the polymers attacks a phosphate of the monomer to be added. The chemistry employed here was a double palladium-/copper-catalyzed Sonogashira reaction, and two new bonds were formed as the backbone was inserted between adjacent monomers.

While these dimers are small and do not yet approximate the abilities of DNA and RNA to transfer information, they hint that replicating polymers with a great diversity of structures and chemistries may be possible. In the absence of enzymes, the transfer of sequence information is a challenging problem even for nucleotide based oligomers and has only been demonstrated to work well for sequences up to 10 bases long.[32,33] Previous work by Moore and colleagues has shown that polyimine polymers, similar in structure to the dimers constructed here, form duplexes reversibly to at least tetramers.[28] Those achievements lead us to believe that our methods can be extended to construct longer information containing oligomeric materials capable of information transfer, work which is currently underway.

Many of the extraordinary functions of DNA and RNA, such as the ability to evolve function, stem from their being information-containing polymers that can self-replicate. Access to synthetic polymers that can mimic these lifelike feats may have other lifelike properties and help to elucidate the transition from chemistry and biology.

EXPERIMENTAL SECTION

General Methods. The ¹H NMR and ¹³C NMR spectra were recorded at 117.42 kG (¹H 500 MHz, ¹³C 125 MHz) or 93.94 kG (¹H 400 MHz, ¹³C 100 MHz), as noted, at ambient temperature. Hydrogen chemical shifts are expressed in parts per million (ppm) relative to the residual protio-solvent resonance: CDCl₃ δ 7.26. For ¹³C spectra, the centerline of the solvent signal was used as internal reference: CDCl₃ δ 77.00. Unless otherwise noted, each carbon resonance represents a single carbon (relative intensity). All exchangeable −OH and −NH hydrogen resonances were confirmed by D₂O exchange. Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry experiments were carried out as follows: Five 0.5 μL portions of a solution of CHCl₃ saturated with 2,5-dihydroxybenzoic acid matrix (Sigma) were deposited on a MTP 384 polished steel BC target plate (Bruker), allowing the solvent to evaporate between each addition. A single 0.5 μL aliquot of analyte in CDCl₃ was then added on top of the 2,5-

dihydroxybenzoic acid film and the sample analyzed with an autoflex speed LRF MALDI-TOF mass spectrometer (Bruker). High-resolution mass spectrometric data were obtained on a ToF (time-of-flight) Agilent Technologies system by electrospray (ESI) in the positive-ion mode. Mobile phases were water and acetonitrile (1:8) with 0.1% formic acid. W with a flow rate of 0.25 mL/min. The MS settings were as follows: capillary voltage 3.5 kV and desolvation temperature 325 °C.

General Sonogashira Cross Coupling. The aryl alkyne and aryl iodide, at the given ratio of equivalents, were dissolved in tetrahydrofuran (0.1 M alkyne) in a screw-capped vial containing a magnetic stir bar. Five equivalents of TEA, with respect to the aryl iodide, was added followed by 10 mol % of Pd(PPh$_3$)$_4$ and 5 mol % of CuI. The vial was sealed with a screw cap and allowed to stir between 1 and 24 h at room temperature. The reaction was monitored by silica TLC (10% EtOAc in hexanes) until it reached full conversion, whereupon it was diluted in ethyl acetate and mixed with silica. The solvent was removed in vacuo, and the crude mixture was purified by silica gel flash chromatography using a CombiFlash (Teledyne Isco Lincoln, NB).

Transimination of Duplexes B·B−A·A and B·A−A·B with 3-Ethynylaniline A (Step a). B·B−A·A or B·A−A·B (4 mg, 0.004 mmol) was added to an NMR tube, dissolved in CDCl$_3$ and TFA (0.5 mL CDCl$_3$, 0.02% TFA), and placed over 4 Å molecular sieves. To the NMR tube was added 3-ethynylaniline (15 uL, 0.12 mmol), and transimination was monitored by ^{1}H NMR (500 MHz) until the reactants reached equilibrium and the amount of the duplex was reduced to <3% (approximately 2 h). For duplex B·B−A·A, this procedure resulted in a solution containing diimine B·B−AA, dianiline A·A, and and 3-ethynlaniline A in a 1:1:28 ratio, as measured by NMR, respectively. For duplex B·A−A·B this procedure resulted in a solution containing imine−aniline B·A−A and 3-ethynlaniline A in a 2:28 ratio, respectively.

Imine Formation with 3-Ethynylbenzaldehyde B (Step b). Following step a, additional 4 Å molecular sieves (20 mg) and 3-ethynylbenzaldehyde B (31 mg, 0.24 mmol) were added to the NMR tube. Imine formation was monitored by ^{1}H NMR until the amines of both the templates (B·B−A·A or B·A−A·B) and the 3-ethynylaniline A were condensed to imines (2−16 h). For the procedure beginning with duplex B·B−A·A, this step resulted in a solution containing diimine B·B−AA, diimine BB−A·A, imine A−B, and 3-ethynlbenzaldehyde B in a 1:1:28:30 ratio, as measured by NMR, respectively. For duplex B·A−A·B, this step resulted in a solution containing diimine B·A−AB, imine A−B, and 3-ethynlenzaldehyde B in a 2:28:30 ratio, respectively.

Removal of Monoimine A−B and 3-Ethynlbenzaldehyde B (Step c). Following step b, the molecular sieves were pelleted via centrifugation, and the CDCl$_3$ supernatant was decanted into a 5 mL glass vial. The solvent was removed under a stream of dry N$_2$ and the resulting white solid sublimated under reduced pressure (0.02 mbar) at 120 °C for 16 h using a BÜCHI GKR-50 glass tube oven (BÜCHI Labortechnik Flawil, Switzerland). Following hydrolysis, the sublimate was found to contain a 30:28 mixture of 3-ethynylbenzaldehyde B and 3-ethynylaniline A with >99% mass recovery. The unsublimated residue was found to contain nearly pure diimines B·B−AA and BB−A·A in a 1:1 ratio for the replication reaction mixture starting with duplex B·B−A·A and diimine B·A−BA for replication reaction mixture which started with duplex B·A−A·B.

Linking of Template-Bound Monomers of B·B−AA and BB−A·A or B·A−BA (Step d). The residue obtained in step c was dissolved in CDCl$_3$ (500 uL) and placed over 4 Å molecular sieves (20 mg). Methyl 3,5-diiodobenzoate (8 mg, 0.02 mmol) and DABCO (4.5 mg, 0.04 mmol) were added to the NMR tube, and the relative stoichiometry was verified by ^{1}H NMR. A 1/6 portion (approximately 0.6 μmol of B·B−AA + BB−A·A, or B·A−BA) of this solution was aliquoted for the reaction and added to an oven-dried Schlenk flask equipped with a stir bar. The solution was diluted with amine-free, anhydrous DMF (100 mL), placed over 4 Å molecular sieves (50 mg), and then degassed by sparging with N$_2$ through a sintered glass frit for 2 h. Removal of dissolved O$_2$ was imperative to the success of

this reaction as Glaser coupling products were otherwise obtained in significant quantities. While still under a positive N$_2$ pressure, Pd(PPh$_3$)$_4$ (1 mg) and CuI (0.1 mg) were added. The head space of the Schlenk flask was evacuated, backfilled with N$_2$ three times, and then sealed. After the mixture was stirred at 60 °C in an oil bath for 16 h, the DMF was removed in vacuo. The resulting residue was taken up in CHCl$_3$ (10 mL), the 4 Å molecular sieves were pelleted via centrifugation, and the CHCl$_3$ was decanted and then removed in vacuo. The resulting light brown oil was precipitated from a 3:1 mixture of benzene:CHCl$_3$ (50 μL). For replication beginning with duplex B·B−A·A, a mixture of B·B−A*A and B*B−A·A was obtained exclusively. For replication beginning with duplex B·A−A·B, a mixture of B·A−A*B and PPh$_3$ was obtained.

2-(2-Methoxyethoxy)ethyl 3,5-Diiodobenzoate (1). Dry K$_2$CO$_3$ (10 eq, 128 mmol, 17.7 g) and freshly activated 4 Å molecular sieves (17.7 g) were added to a flame-dried round-bottom flask equipped with a stir bar. Then 2-(2-(2-methoxyethoxy)ethoxy)ethan-1-ol (30 eq, 45 mL) was added, and the solution was stirred for 30 min. Next, methyl 3,5-diiodobenzoate (5 g, 12.8 mmol) was added. After being stirred at rt for 1.5 h, the solution was filtered through a glass frit, and the solids were washed with EtOAc (200 mL). The EtOAc was removed in vacuo, leaving a solution of 2-(2-(2-methoxyethoxy)-ethoxy)ethan-1-ol and the desired product 1. Reversed-phase C18 flash chromatography using a CombiFlash (Teledyne Isco Lincoln, NB) gave pure 1 (rt 5−6 min with 100% ACN, 5.05 g, 83%) as a white solid: ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.33 (d, J = 1.6 Hz, 2H), 8.23 (t, J = 1.6 Hz, 1H), 4.52−4.44 (m, 2H), 3.85−3.79 (m, 2H), 3.71−3.66 (m, 2H), 3.60−3.54 (m, 2H), 3.40 (s, 3H); ^{13}C{^{1}H} NMR (125 MHz, CDCl$_3$) δ 162.9, 148.6, 137.2 (2C), 132.7, 94.2 (2C), 71.4, 70.0, 68.4, 64.2, 58.6; HRMS (ESI-TOF) m/z [M + Na]$^+$ calcd for C$_{12}$H$_{14}$I$_2$O$_4$Na 498.8874, found 498.8890.

2-(2-Methoxyethoxy)ethyl 3,5-Bis((3formylphenyl)ethynyl)-benzoate (B·B). Prepared according to the general Sonogashira cross coupling procedure: 3-Ethynylbenzaldehyde B (27 mg, 0.2 mmol) was reacted with 2-(2-methoxyethoxy)ethyl 3,5-diiodoben-zoate 1 (20 mg, 0.04 mmol) in the presence of Pd(PPh$_3$)$_4$ (30 mg, 0.026 mmol), CuI (10 mg, 0.052 mmol), and TEA (1 mL) in THF (10 mL) at rt for 16 h. CombiFlash silica gel purification (gradient of 0−100% EtOAc in hexanes) gave 3 (20 mg, 98%) as an off-white solid: ^{1}H NMR (500 MHz, CDCl$_3$) δ 10.03 (s, 2H), 8.20 (d, J = 1.6 Hz, 2H), 8.05 (dd, J = 1.9, 1.4 Hz, 2H), 7.90−7.86 (m, 3H), 7.79 (ddd, J = 7.7, 1.4, 1.4 Hz, 2H), 7.56 (dd, J = 7.7, 7.7 Hz, 2H), 4.56−4.51 (m, 2H), 3.90−3.85 (m, 2H), 3.73−3.70 (m, 2H), 3.61−3.57 (m, 2H), 3.40 (s, 3H); ^{13}C{^{1}H} NMR (125 MHz, CDCl$_3$) δ 191.6 (2C), 165.3, 138.6, 137.4 (2C), 136.7 (2C), 133.2 (2C), 132.9 (2C), 131.2, 129.6 (2C), 129.5 (2C), 124.0 (2C), 123.9 (2C), 89.6 (2C), 89.1 (2C), 72.1, 70.8, 69.4, 64.8, 59.3; HRMS (ESI-TOF) m/z [M + H]$^+$ calcd for C$_{30}$H$_{25}$O$_6$ 481.1646, found 481.1672.

2-(2-Methoxyethoxy)ethyl 3,5-Bis((3-aminophenyl)ethynyl)-benzoate (A·A). Prepared according to the general Sonogashira cross coupling procedure: 3-Ethynylaniline A (24 μL, 0.2 mmol) was reacted with 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate 1 (20 mg, 0.04 mmol) in the presence of Pd(PPh$_3$)$_4$ (5 mg, 0.004 mmol), CuI (0.5 mg, 0.003 mmol), and TEA (0.1 mL) in tetrahydrofuran (1 mL) at rt for 16 h. CombiFlash silica gel purification (gradient of 0−100% EtOAc in hexanes) gave A·A (16.5 mg, 86%) as an off-white solid: ^{1}H NMR (500 MHz, CDCl$_3$) 8.12 (d, J = 1.6 Hz, 2H), 7.81 (t, J = 1.6 Hz, 1H), 7.14 (dd, J = 8.1, 7.6 Hz, 2H), 6.94 (ddd, J = 7.6, 1.6, 1.0 Hz, 2H), 6.85 (dd, J = 2.5, 1.6, 2H), 6.68 (ddd, J = 8.1, 2.5, 1.0 Hz, 2H), 4.54−4.48 (m, 2H), 3.88−3.82 (m, 2H), 3.74 (br s, 4NH) 3.73−3.68 (m, 2H), 3.60−3.55 (m, 2H), 3.39 (s, 3H); ^{13}C{^{1}H} NMR (125 MHz, CDCl$_3$) δ 165.2, 146.3 (2C), 138.2, 132.0 (2C), 130.6, 129.3 (2C), 124.1 (2C), 123.2 (2C), 122.0 (2C), 117.7 (2C), 115.7 (2C), 91.1 (2C), 86.9 (2C), 71.8, 70.5, 69.1, 64.4, 59.0; HRMS (ESI-TOF) m/z [M + H]$^+$ calcd for C$_{28}$H$_{27}$N$_2$O$_4$ 455.1965, found 455.2007.

2-(2-Methoxyethoxy)ethyl 3-((3-Aminophenyl)ethynyl)-5-iodo-benzoate (2). Prepared according to General Procedure A: 3-Ethynylaniline A (50 μL, 0.44 mmol) was reacted with 2-(2-methoxyethoxy)ethyl 3,5-diiodobenzoate 1 (1.2 g, 5.7 equiv, 2.52

mmol) in the presence of Pd(PPh$_3$)$_4$ (60 mg, 0.06 mmol), CuI (5 mg, 0.026 mmol), and TEA (2 mL) in tetrahydrofuran (4 mL) at rt for 16 h. CombiFlash silica gel purification (gradient of 0−100% EtOAc in hexanes) gave **2** (120 mg, 58%) as a light brown oil: ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.32 (dd, J = 1.6, 1.6 Hz, 1H), 8.14 (dd, J = 1.6, 1.6 Hz, 1H), 8.03 (dd, J = 1.6, 1.6 Hz, 1H), 7.14 (dd, J = 8.1, 7.7 Hz, 1H), 6.93 (ddd, J = 7.7, 1.6, 1.0 Hz, 1H), 6.84 (dd, J = 2.4, 1.6 Hz, 1H), 6.69 (ddd, J = 8.1, 2.4, 1.0 Hz, 1H), 4.53−4.47 (m, 2H), 3.87−3.81 (m, 2H), 3.72 (br s, 2NH), 3.72−3.67 (m, 2H), 3.60−3.55 (m, 2H), 3.40 (s, 3H); ^{13}C{^{1}H} NMR (125 MHz, CDCl$_3$) δ 164.8, 146.6, 144.3, 138.1, 132.2, 132.1, 129.6, 125.9, 123.3, 122.4, 118.0, 116.1, 93.4, 92.2, 86.3, 72.1, 70.8, 69.3, 64.8, 59.4; HRMS (ESI-TOF) m/z [M + H]$^+$ calcd for C$_{20}$H$_{21}$INO$_4$ 466.0510, found 466.0537.

*Duplex **B·B**−**A·A**.* A 0.1 M solution of dianiline **A·A** (2 mg) in benzene was added to a 0.1 M solution of dialdehyde **B·B** (2 mg) in benzene. TFA was then added (0.1 uL), and the solution was placed in a freezer. After the solution was allowed to sit for 3 days, colorless, needle-like crystals formed. The benzene supernatant was carefully removed with a pipet leaving pure white crystals of **B·B**−**A·A** (4 mg, 100%): ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.58 (s, 2H), 8.18 (d, J = 1.6 Hz, 2H), 8.17 (d, J = 1.6 Hz, 2H), 8.13 (dd, J = 1.8, 1.6 Hz, 2H), 7.938 (ddd, J = 7.7, 1.7, 1.4, 2H), 7.933 (t, J = 1.6 Hz, 2H), 7.91 (t, J = 1.6 Hz, 1H), 7.66 (ddd, J = 7.7, 1.6, 1.2 Hz, 2H), 7.52 (dd, J = 2.1, 1.6 Hz, 2H), 7.50 (dd, J = 7.7, 7.7 Hz, 2H), 7.45 (ddd, J = 7.5, 1.6, 1.5 Hz, 2H), 7.42 (dd, J = 7.5, 7.5 Hz, 2H), 7.34 (ddd, J = 7.5, 2.1, 1.5 Hz, 2H), 4.58−4.51 (overlap, 4H), 3.92−3.86 (overlap, 4H), 3.77−3.70 (overlap, 4H), 3.63−3.59 (overlap, 4H), 3.42 (s, 6H); ^{13}C{^{1}H} NMR (125 MHz, CDCl$_3$) δ 165.5, 165.4, 159.3 (2C), 151.0 (2C), 139.4, 139.2, 136.6 (2C), 134.3 (2C), 133.0 (2C), 132.24 (2C), 132.22 (2C), 131.15, 131.09, 129.8 (2C), 129.6 (2C), 129.3 (2C), 128.8 (2C), 124.9 (2C), 124.2 (2C), 124.1 (2C), 123.8 (2C), 123.6 (2C), 122.2 (2C), 90.8 (2C), 90.5 (2C), 88.6 (2C), 88.2 (2C), 72.2 (2C), 70.8 (2C), 69.4 (2C), 64.75, 64.73, 59.4; HRMS (MALDI-TOF) m/z [M + H]$^+$ calcd for C$_{58}$H$_{47}$N$_2$O$_8$ 899.3333, found 899.8992. The structure of **B·B**−**A·A** was further confirmed by X-ray crystallography (see the Supporting Information).

*Duplex **B·A**−**A·B**.* Prepared according to the general Sonogashira cross coupling procedure: Aryl iodide **2** (45 mg, 0.1 mmol) was reacted with 3-ethynylbenzaldehyde **B** (14 mg, 0.11 mmol) in the presence of Pd(PPh$_3$)$_4$ (12 mg, 0.01 mmol), CuI (4 mg, 0.02 mmol), and TEA (0.7 mL) in DMF (1 mL) at rt for 16 h. CombiFlash silica gel purification (gradient of 0−100% EtOAc in hexanes) gave a mixture of **A·B** and **B·A**−**A·B**. The fractions containing **A·B** and **B·A**−**A·B** were combined, heated to 60 °C in an oil bath for 1 h, and then concentrated. The residue was precipitated from a mixture of chloroform and benzene (1:4 respectively) to give pure **B·A**−**A·B** (10 mg, 22%) as a white solid: ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.58 (s, 2H), 8.18 (t, J = 1.6 Hz, 2H), 8.17 (t, J = 1.6 Hz, 2H), 8.11 (t, J = 1.7 Hz, 2H), 7.95 (ddd, J = 7.8, 1.7, 1.2 Hz, 2H), 7.92 (t, J = 1.6 Hz, 2H), 7.66 (ddd, J = 7.6, 1.6, 1.2 Hz, 2H), 7.51 (dd, J = 2.1, 1.6 Hz, 3H), 7.49 (dd, J = 7.8, 7.6 Hz, 2H), 7.45 (ddd, J = 7.6, 1.6, 1.4 Hz, 2H), 7.42 (dd, J = 7.6, 7.6 Hz, 2H), 7.35 (ddd, J = 7.6, 2.1, 1.4 Hz, 2H), 4.57−4.53 (m, 4H), 3.91−3.87 (m, 4H), 3.76−3.71 (m, 4H), 3.63−3.59 (m, 4H), 3.42 (s, 6H); ^{13}C{^{1}H} NMR (125 MHz, CDCl$_3$) δ 165.5 (2C), 159.4 (2C), 151.0 (2C), 139.3 (2C), 136.6 (2C), 134.4 (2C), 133.1 (2C), 132.3 (2C), 132.2 (2C), 131.1 (2C), 129.8 (2C), 129.6 (2C), 129.3 (2C), 128.7 (2C), 124.8 (2C), 124.3 (2C), 124.1 (2C), 123.8 (2C), 123.6 (2C), 122.4 (2C), 90.9 (2C), 90.4 (2C), 88.7 (2C), 88.1 (2C), 72.2 (2C), 70.8 (2C), 69.4 (2C), 64.7 (2C), 59.4 (2C); HRMS (MALDI-TOF) m/z [M + H]$^+$ calcd for C$_{58}$H$_{47}$N$_2$O$_8$ 899.3333, found 899.4021.

ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.9b00095.

NMR spectra and details of the hydrolysis and reannealing procedure, NMR spectra of starting compounds and replication procedure, crystallographic data for **B·B**−**A·A** (PDF)

X-ray data for compound **B·B**−**A·A** (CIF)

AUTHOR INFORMATION

Corresponding Authors
*E-mail: kylerstrom@gmail.com.
*E-mail: noam.prywes@gmail.com.

ORCID ⓞ
Kyle R. Strom: 0000-0002-0567-347X
Jack W. Szostak: 0000-0003-4131-1203
Noam Prywes: 0000-0003-2526-2392

Present Address
†(N.P.) The Innovative Genomics Institute, University of California, Berkeley 2151 Berkeley Way, Berkeley, CA 94704.

Notes
The authors declare no competing financial interest.

REFERENCES

(1) Weimann, B. J.; Lohrmann, R.; Orgel, L. E.; Schneider-Bernloehr, H.; Sulston, J. E. Template-directed synthesis with adenosine-5′-phosphorimidazolide. *Science* **1968**, *161*, 387.

(2) von Kiedrowski, G. A Self-Replicating Hexadeoxynucleotide. *Angew. Chem., Int. Ed. Engl.* **1986**, *25* (10), 932−935.

(3) Jauker, M.; Griesser, H.; Richert, C. Copying of RNA Sequences without Pre-Activation. *Angew. Chem., Int. Ed.* **2015**, *54* (48), 14559−14563.

(4) Szostak, J. W. The Narrow Road to the Deep Past: In Search of the Chemistry of the Origin of Life. *Angew. Chem., Int. Ed.* **2017**, *56* (37), 11037−11043.

(5) Eschenmoser, A. Chemical Etiology of Nucleic Acid Structure. *Science* **1999**, *284* (5423), 2118−2124.

(6) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* **2012**, *336* (6079), 341−344.

(7) Tjivikua, T.; Ballester, P.; Rebek, J. Self-replicating system. *J. Am. Chem. Soc.* **1990**, *112* (3), 1249−1250.

(8) Nowick, J. S.; Feng, Q.; Tjivikua, T.; Ballester, P.; Rebek, J. Kinetic studies and modeling of a self-replicating system. *J. Am. Chem. Soc.* **1991**, *113* (23), 8831−8839.

(9) Lo, P. K.; Sleiman, H. F. Nucleobase-Templated Polymerization: Copying the Chain Length and Polydispersity of Living Polymers into Conjugated Polymers. *J. Am. Chem. Soc.* **2009**, *131* (12), 4182−4183.

(10) Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. Self-Assembling Sequence-Adaptive Peptide Nucleic Acids. *Science* **2009**, *325* (5936), 73−77.

(11) Colquhoun, H.; Lutz, J.-F. Information-containing macromolecules. *Nat. Chem.* **2014**, *6*, 455.

(12) ten Brummelhuis, N. Controlling monomer-sequence using supramolecular templates. *Polym. Chem.* **2015**, *6* (5), 654−667.

(13) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341* (6146), 1238149.

(14) Niu, J.; Hili, R.; Liu, D. R. Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nat. Chem.* **2013**, *5*, 282.

(15) Lutz, J.-F. Coding Macromolecules: Inputting Information in Polymers Using Monomer-Based Alphabets. *Macromolecules* **2015**, *48* (14), 4759−4767.

(16) Solleder, S. C.; Schneider, R. V.; Wetzel, K. S.; Boukis, A. C.; Meier, M. A. R. Recent Progress in the Design of Monodisperse, Sequence-Defined Macromolecules. *Macromol. Rapid Commun.* **2017**, *38* (9), 1600711.

(17) Marquis, A.; Smith, V.; Harrowfield, J.; Lehn, J.-M.; Herschbach, H.; Sanvito, R.; Leize-Wagner, E.; Van Dorsselaer, A. Messages in Molecules: Ligand/Cation Coding and Self-Recognition in a Constitutionally Dynamic System of Heterometallic Double Helicates. *Chem. - Eur. J.* **2006**, *12* (21), 5632−5641.

(18) Gong, B. Molecular Duplexes with Encoded Sequences and Stabilities. *Acc. Chem. Res.* **2012**, *45* (12), 2077−2087.

(19) Yashima, E.; Ousaka, N.; Taura, D.; Shimomura, K.; Ikai, T.; Maeda, K. Supramolecular Helical Systems: Helical Assemblies of Small Molecules, Foldamers, and Polymers with Chiral Amplification and Their Functions. *Chem. Rev.* **2016**, *116* (22), 13752−13990.

(20) Stross, A. E.; Iadevaia, G.; Núñez-Villanueva, D.; Hunter, C. A. Sequence-Selective Formation of Synthetic H-Bonded Duplexes. *J. Am. Chem. Soc.* **2017**, *139* (36), 12655−12663.

(21) Swain, J. A.; Iadevaia, G.; Hunter, C. A. H-Bonded Duplexes based on a Phenylacetylene Backbone. *J. Am. Chem. Soc.* **2018**, *140* (36), 11526−11536.

(22) Lin, N.-T.; Lin, S.-Y.; Lee, S.-L.; Chen, C.-h.; Hsu, C.-H.; Hwang, L. P.; Xie, Z.-Y.; Chen, C.-H.; Huang, S.-L.; Luh, T.-Y. From Polynorbornene to the Complementary Polynorbornene by Replication. *Angew. Chem., Int. Ed.* **2007**, *46* (24), 4481−4485.

(23) Kamonsutthipaijit, N.; Anderson, H. L. Template-directed synthesis of linear porphyrin oligomers: classical, Vernier and mutual Vernier. *Chemical Science* **2017**, *8* (4), 2729−2740.

(24) Kosikova, T.; Philp, D. Exploring the emergence of complexity using synthetic replicators. *Chem. Soc. Rev.* **2017**, *46* (23), 7274−7305.

(25) Carnall, J. M. A.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C. A.; Peyralans, J. J.-P.; Otto, S. Mechanosensitive Self-Replication Driven by Self-Organization. *Science* **2010**, *327* (5972), 1502−1506.

(26) Zhao, D.; Moore, J. S. Synthesis and Self-Association of an Imine-Containing m-Phenylene Ethynylene Macrocycle. *J. Org. Chem.* **2002**, *67* (11), 3548−3554.

(27) Hartley, C. S.; Elliott, E. L.; Moore, J. S. Covalent Assembly of Molecular Ladders. *J. Am. Chem. Soc.* **2007**, *129* (15), 4512−4513.

(28) Elliott, E. L.; Hartley, C. S.; Moore, J. S. Covalent ladder formation becomes kinetically trapped beyond four rungs. *Chem. Commun.* **2011**, *47* (17), 5028−5030.

(29) Sonogashira, K. Development of Pd−Cu catalyzed cross-coupling of terminal acetylenes with sp2-carbon halides. *J. Organomet. Chem.* **2002**, *653* (1), 46−49.

(30) Morales, S.; Guijarro, F. G.; García Ruano, J. L.; Cid, M. B. A General Aminocatalytic Method for the Synthesis of Aldimines. *J. Am. Chem. Soc.* **2014**, *136* (3), 1082−1089.

(31) Ciaccia, M.; Di Stefano, S. Mechanisms of imine exchange reactions in organic solvents. *Org. Biomol. Chem.* **2015**, *13* (3), 646−654.

(32) Hänle, E.; Richert, C. Enzyme-Free Replication with Two or Four Bases. *Angew. Chem.* **2018**, *130* (29), 9049−9053.

(33) Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. Enhanced Nonenzymatic RNA Copying with 2-Aminoimidazole Activated Nucleotides. *J. Am. Chem. Soc.* **2017**, *139* (5), 1810−1813.