Vol 451|17 January 2008|doi:10.1038/nature06451

LETTERS

Programming biomolecular self-assembly pathways

Peng Yin[1,2], Harry M. T. Choi[1], Colby R. Calvert[1] & Niles A. Pierce[1,3]

In nature, self-assembling and disassembling complexes of proteins and nucleic acids bound to a variety of ligands perform intricate and diverse dynamic functions. In contrast, attempts to rationally encode structure and function into synthetic amino acid and nucleic acid sequences have largely focused on engineering molecules that self-assemble into prescribed target structures, rather than on engineering transient system dynamics[1,2]. To design systems that perform dynamic functions without human intervention, it is necessary to encode within the biopolymer sequences the reaction pathways by which self-assembly occurs. Nucleic acids show promise as a design medium for engineering dynamic functions, including catalytic hybridization[3–6], triggered self-assembly[7] and molecular computation[8,9]. Here, we program diverse molecular self-assembly and disassembly pathways using a 'reaction graph' abstraction to specify complementarity relationships between modular domains in a versatile DNA hairpin motif. Molecular programs are executed for a variety of dynamic functions: catalytic formation of branched junctions, autocatalytic duplex formation by a cross-catalytic circuit, nucleated dendritic growth of a binary molecular 'tree', and autonomous locomotion of a bipedal walker.

The hairpin motif (A in Fig. 1a) comprises three concatenated domains, a, b and c. Each domain contains a special nucleation site called a toehold[10], denoted a_t, b_t and c_t. Two basic reactions can be programmed using this motif, as illustrated for the example of catalytic duplex formation in Fig. 1b. First, an assembly reaction (1) occurs when a single-stranded initiator I, containing an exposed toehold a_t*, nucleates at the exposed toehold a_t of hairpin A, initiating a branch migration that opens the hairpin. Hairpin domains b and c, with newly exposed toeholds b_t and c_t, can then serve as assembly initiators for other suitably defined hairpins, permitting cascading (for example, in reaction (2), domain b of hairpin A assembles with domain b* of hairpin B, opening the hairpin). Second, a disassembly reaction (3) occurs when a single-stranded domain (a* of B) initiates a branch migration that displaces the initiator I from A. In this example, I catalyses the formation of duplex A•B through a prescribed reaction pathway.

To assist in programming more complex reaction pathways, we abstract the motif of Fig. 1a as a node with three ports (Fig. 1c): a triangular input port and two circular output ports. The state of each port is either accessible (open triangle/circle) or inaccessible (solid triangle/circle), depending on whether the toehold of the corresponding motif domain is exposed or sequestered. Functional relationships between ports within a node are implicit in the definition of the nodal abstraction corresponding to a particular motif (for example, for the node of Fig. 1c, the output ports flip to accessible states if the input port is flipped to an inaccessible state through an interaction with a complementary upstream output port). By depicting assembly reactions by solid arrows and disassembly reactions by dashed arrows (each directed from an output port to a complementary input port of a different node), reaction pathways can be specified abstractly in the form of a reaction graph, representing a program to be executed by nucleic acid molecules.

The reactions depicted in the secondary structure mechanism of Fig. 1b are specified using a reaction graph in Fig. 1d. The initial conditions for this program are described via the state of each port in the reaction graph. Figure 1e depicts the execution of this reaction graph through cascaded assembly and disassembly reactions. An assembly reaction is executed when ports connected by a solid arrow are simultaneously accessible. For the initial conditions depicted in Fig. 1d, the program must start with the execution of reaction (1).

Reaction 1 (assembly): in an assembly reaction (executed here by the accessible output port of I and the complementary accessible input port of A), a bond is made between the ports and they are flipped to inaccessible states; the two output ports of A are flipped

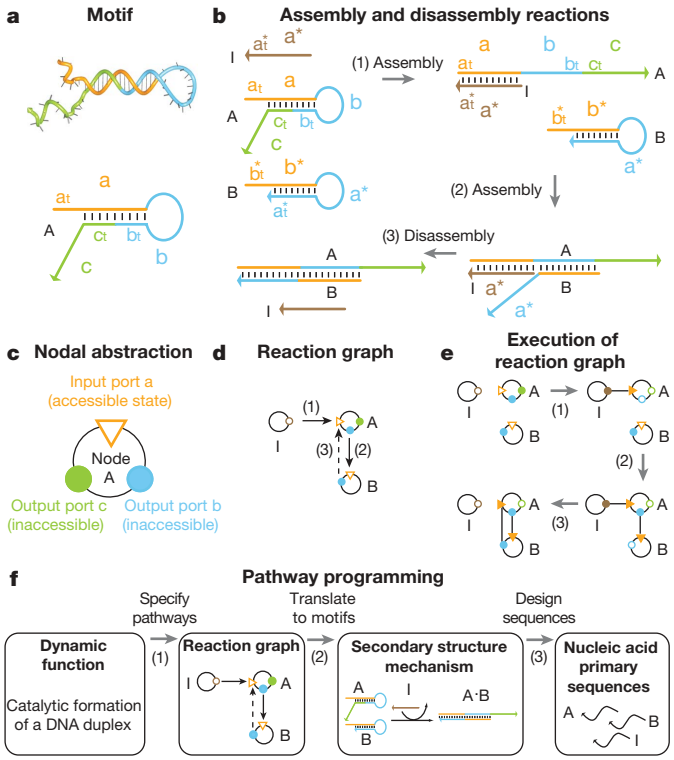

Figure 1 | Programming biomolecular self-assembly pathways. **a**, Secondary structure of the hairpin motif. Coloured lines represent strand domains; short black lines represent base pairs; arrowheads indicate 3′ ends. Domain c is optional. **b**, Secondary structure mechanism illustrating assembly and disassembly reactions during catalytic duplex formation. Asterisks denote complementarity. **c**, Abstraction of the motif A as a node with three ports (colour use is consistent with **a**). **d**, A reaction graph representing a molecular program executed schematically in **b** and **e**. **e**, Execution of the reaction graph of **d**. **f**, Hierarchical design process.

[1]Department of Bioengineering, [2]Department of Computer Science, [3]Department of Applied & Computational Mathematics, California Institute of Technology, Pasadena, California 91125, USA.

to accessible states (based on the internal logic of node A). Reaction 2 (assembly): a bond is made between the newly accessible blue output port of A and the complementary accessible input port of B and both ports are flipped to inaccessible states; the output port of B is flipped to the accessible state (based on the internal logic of node B). Reaction 3 (disassembly): in a disassembly reaction (executed here by the newly accessible output port of B, the inaccessible input port of A, and the inaccessible output port of I), the bond between the output port of I and the input port of A is displaced by a bond between the output port of B and the input port of A; the states of the two output ports are flipped (see Supplementary Information 2 for additional details).

The reaction graph provides a simple representation of assembly (and disassembly) pathways that can be translated directly into molecular executables: nodes represent motifs, ports represent domains, states describe accessibility, arrows represent assembly and disassembly reactions between complementary ports. Starting from a conceptual dynamic function, a molecular implementation is realized in three steps (Fig. 1f): (1) pathway specification via a reaction graph; (2) translation into secondary structure motifs; (3) computational design of motif primary sequences (see Methods for details). We demonstrate the utility of this hierarchical design process by experimentally executing molecular programs encoding four distinct dynamic functions.

Program 1: Catalytic geometry. Current protocols for self-assembling synthetic DNA nanostructures often rely on annealing procedures to bring interacting DNA strands to equilibrium on the free-energy landscape[11–13]. By contrast, self-assembly in biology proceeds isothermally and assembly kinetics are often controlled by catalysts. Until now, synthetic DNA catalysts[3–6] have been used to control the kinetics of the formation of DNA duplex structures. The next challenge is to catalyse the formation of branched DNA structures, the basic building blocks for DNA structural nanotechnology[14,15].

First, we demonstrate the catalytic formation of a three-arm DNA junction. The assembly and disassembly pathways specified in the reaction graph of Fig. 2a are translated into the motif-based molecular implementation of Fig. 2b (see Supplementary Information 3.1 for details). The complementarity relationships between the segments of hairpins A, B, and C are specified (Fig. 2b, top) so that in the absence of initiator strand I, the hairpins are kinetically impeded from forming the three-arm junction that is predicted to dominate at equilibrium. In the reaction graph, this property is programmed by the absence of a starting point if node I is removed from the graph (that is, no pair of accessible ports connected by an assembly arrow). The introduction of I into the system (Fig. 2b, bottom) activates a cascade of assembly steps with A, B and C, followed by a disassembly step in which C displaces I from the complex, freeing I to catalyse the self-assembly of additional branched junctions.

Gel electrophoresis confirms that the hairpins assemble slowly in the absence of initiator and that assembly is markedly accelerated by the addition of initiator (Fig. 2c). Disassembly of the initiator leads to catalytic turnover, as indicated by the nearly complete consumption of hairpins even at substoichiometric initiator concentrations. Interestingly, only minimal assembly is achieved by annealing the hairpin mixture, illustrating the utility of pathway programming for traversing free-energy landscapes with kinetic traps that cannot be overcome by traditional annealing approaches.

Direct imaging of the catalysed self-assembly product A•B•C by atomic force microscopy (AFM) reveals the expected three-arm junction morphology (Fig. 2d). In principle, the reaction pathway can be extended to the catalytic self-assembly of k-arm junctions (Supplementary Information 3.5). We illustrate $k = 4$ with the reaction graph and AFM image of Fig. 2e and f.

Program 2: Catalytic circuitry. By programming cross-catalytic self-assembly pathways in the reaction graph of Fig. 3a, we obtain an autocatalytic system with exponential kinetics. In the corresponding molecular implementation, four hairpin species, A, B, C and D, coexist metastably in the absence of initiator I (Fig. 3b, top). The initiator catalyses the assembly of hairpins A and B to form duplex A•B (steps 1–2, Fig. 3b, bottom), bringing the system to an exponential amplification stage powered by a cross-catalytic circuit: the duplex A•B has a single-stranded region that catalyses the assembly of C and D to form C•D (steps 3–4); duplex C•D in turn has a single-stranded region that is identical to I and can thus catalyse A and B to form A•B (steps 5–6). Hence, A•B and C•D form an autocatalytic set capable of catalysing its own production. Disassembly (steps 2b, 4b and 6b) is fundamental to the implementation of autocatalysis and sterically uninhibited exponential growth.

Each step in the reaction is examined using native polyacrylamide gel electrophoresis (Supplementary Fig. 12), showing the expected assembly and disassembly behaviour. System kinetics are examined in a fluorescence quenching experiment (Fig. 3c). Spontaneous initiation in the absence of initiator reflects the finite timescale associated with the metastability of the hairpins and yields a sigmoidal time course characteristic of an autocatalytic system[16]. As expected, the curve shifts to the left as the concentration of initiator is increased. A plot of 10% completion time against the logarithm of the concentration shows a linear regime, consistent with exponential kinetics and analytical modelling (Fig. 3c, inset). The minimal leakage of a system containing only A and B (labelled A + B in Fig. 3c) emphasizes that the sigmoidal kinetics of spontaneous initiation for the full system (A + B + C + D) are due to cross-catalysis.

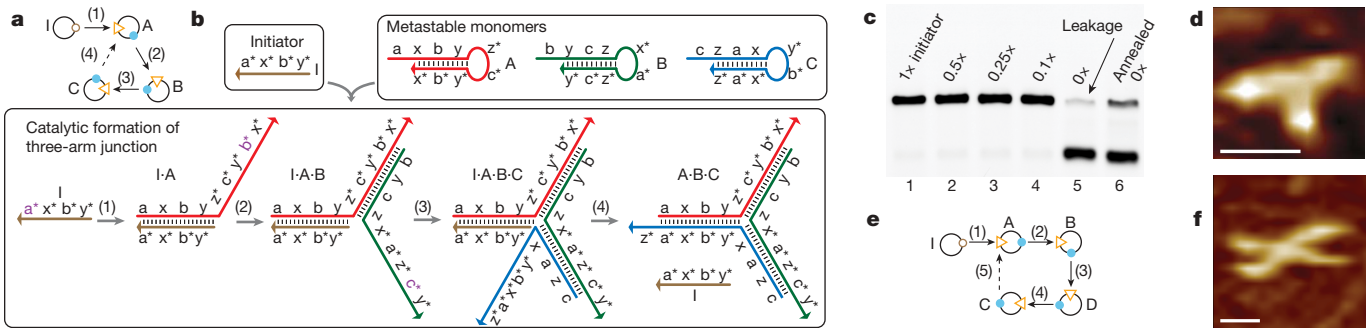

Figure 2 | Programming catalytic geometry: catalytic self-assembly of three-arm and four-arm branched junctions. See Supplementary Information 3 for details. **a**, Reaction graph for three-arm junctions. **b**, Secondary structure mechanism. Each letter-labelled segment is six nucleotides in length. The initially accessible (a* for step 1) or newly exposed (b* for Step 2, c* for step 3) toeholds that mediate assembly reactions are labelled with purple letters. **c**, Agarose gel electrophoresis demonstrating catalytic self-assembly for the three-arm system with 750-nM hairpins. Nearly complete conversion of hairpins to reaction products using stoichiometric or substoichiometric initiator I (lanes 1–4). Minimal conversion in the absence of initiator (lane 5), even with annealing (lane 6). **d**, AFM image of a three-arm junction. Scale bar: 10 nm. **e**, Reaction graph and **f**, AFM image for a four-arm junction. Scale bar: 10 nm.

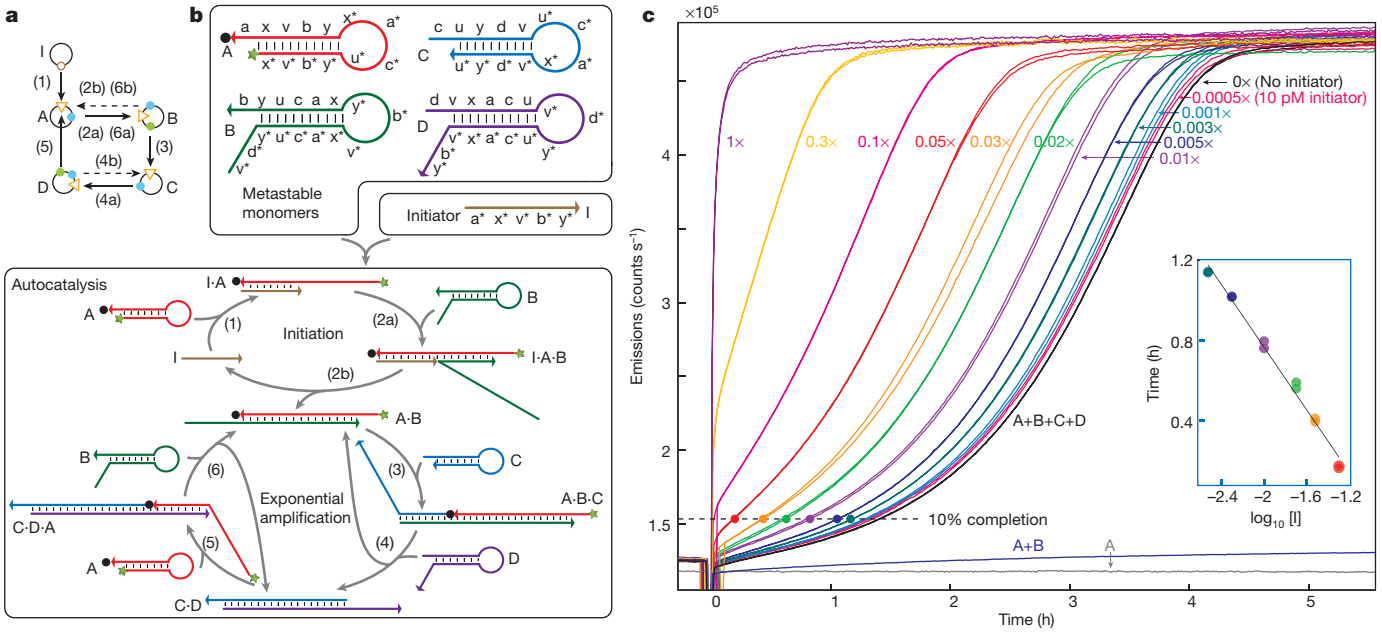

Figure 3 | Programming catalytic circuitry: autocatalytic duplex formation by a cross-catalytic circuit with exponential kinetics. See Supplementary Information 4 for details. **a**, Reaction graph. Multiple assembly arrows entering the same input port depict parallel processes on separate copies of the nodal species. **b**, Secondary structure mechanism. **c**, System kinetics examined by fluorescence quenching. Formation of A•B is monitored by the increase in fluorescence resulting from increased spatial separation between the fluorophore (green star in **b**) and the quencher (black dot in **b**) at either end of A. Raw data for two independent reactions are displayed for each initiator concentration (20-nM hairpins). Single traces are shown for the controls containing only A and B or only A. Inset: linear fit of the 10% completion time against the logarithm of the relative concentration of I ($0.003\times \leq [\mathrm{I}] \leq 0.05\times$). High-concentration end points ($[\mathrm{I}] \geq 0.1\times$) are excluded based on theoretical analysis; low-concentration end points ($[\mathrm{I}] \leq 0.001\times$) are excluded because of signal poisoning by leakage. See Supplementary Information 4.4 for a detailed treatment.

This system demonstrates synthetic biomolecular autocatalysis[17–20] driven by the free energy of base-pair formation. Autocatalysis and exponential system kinetics can also be achieved through entropy-driven hybridization mechanisms[21]. For sensing applications, the triggered exponential growth of these systems suggest the possibility of engineering enzyme-free isothermal detection methods.

Program 3: Nucleated dendritic growth. The molecular program in Fig. 4a depicts the triggered self-assembly of a binary molecular tree of a prescribed size. The reaction starts with the assembly of an initiator node I with a root node A1. Each assembled node subsequently assembles with two child nodes during the next generation of growth, requiring two new node species per generation. In the absence of steric effects, a G-generation dendrimer requires $2G-1$ node species and yields a binary tree containing 2^{G-1} monomers, that is, a linear increase in the number of node species yields an exponential increase in the size of the dendrimer product. Figure 4b depicts the motif based implementation of the program depicted in Fig. 4a: hairpins are metastable in the absence of initiator; the

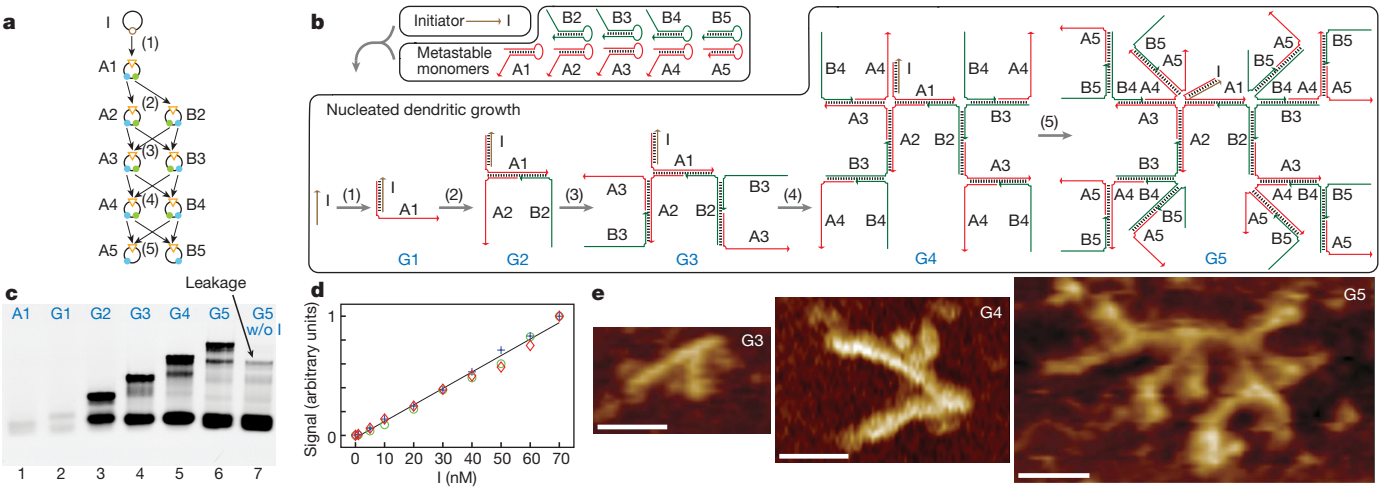

Figure 4 | Programming nucleated dendritic growth: triggered assembly of quantized binary molecular trees. See Supplementary Information 5 for details. **a**, Reaction graph. Multiple assembly arrows entering the same input port depict parallel processes on separate copies of the nodal species. **b**, Secondary structure mechanism. **c**, Agarose gel electrophoresis demonstrating triggered self-assembly. Lanes 1–6: the dominant reaction band shifts with the addition of each generation of hairpins. Subdominant bands are presumed to represent imperfect dendrimers. Lane 7: minimal conversion to reaction products in the absence of initiator. Hairpins A1, A2, B2 at 62.5 nM; the concentration doubles for each subsequent generation of hairpins. Initiator I at 50 nM. **d**, Linear relationship between amplification signal (putative G5 reaction product) and initiator for three independent experiments (cross, diamond, circle). See Supplementary Fig. 17 for details. **e**, AFM images of G3, G4 and G5 dendrimers. Scale bars: 30 nm.

initiator I triggers the growth of a dendrimer with five generations of branching (G5).

We constructed trees with $G = 1$, 2, 3, 4 and 5. The nucleated growth of the tree is examined using native agarose gel electrophoresis. Band shifting demonstrates increasing dendrimer size with each generation of growth (Fig. 4c). Figure 4d demonstrates that the concentration of dendrimer depends linearly on the concentration of the initiator in the system. Finally, AFM imaging of dendrimers for $G = 3$, 4 and 5 reveals the expected morphologies (Fig. 4e). Measurements of the dendrimer segment lengths agree well with the design (Supplementary Information 5.4).

In contrast to previous work in which DNA dendrimer target structures were synthesized by sequential ligation of structural subunits[22], here we program self-assembly pathways so that DNA monomers form dendrimers only on detection of a target nucleation molecule. By growing to a prescribed size, these dendrimers provide quantitative signal amplification with strength exponential in the number of constituent species.

Program 4: Autonomous locomotion. The challenge of engineering molecular machines capable of nanoscale autonomous locomotion has attracted much interest in recent years[23–27]. Inspired by the bipedal motor protein, kinesin, which hauls intracellular cargo by striding along microtubules[28], we have developed an autonomous enzyme-free bipedal DNA walker capable of stochastic locomotion along a DNA track.

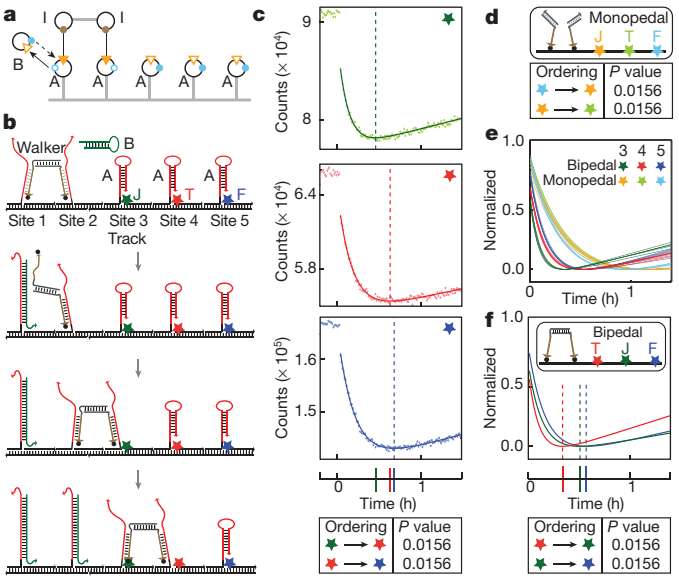

Figure 5 | Programming autonomous locomotion: stochastic movement of a bipedal walker. See Supplementary Information 6 for details. **a**, Reaction graph. Bonds between output ports on I and input ports on A represent initial conditions. Static structural elements are depicted by grey line segments. **b**, Secondary structure mechanism depicting processive locomotion. See Supplementary Information 6.1 and 6.3 for non-processive trajectories. **c–f**, Fluorescence quenching experiments measuring the proximity of the quenchers (black dots) on the walker feet to the fluorophores (coloured stars) decorating the track. Fitted curves (solid) are used to determine the time at which the minimum fluorescence (maximum quenching) was observed (dashed vertical line) for each fluorophore. **c**, Bipedal walker with track labelled by fluorophores JOE (green star) → TAMRA (red) → FAM (blue) as in **b**. For each pair of consecutive minima (JOE → TAMRA and TAMRA → FAM), we test the null hypothesis that the median time difference between the minima is zero against the alternative hypothesis that the time difference is positive. Based on a statistical analysis of six independent experiments (see Supplementary Information 6.6, 6.7), the null hypothesis can be rejected for both time differences with the same *P*-value of 0.0156, supporting the interpretation that the observed minima are sampled from a distribution in which the ordering of the minima matches the physical ordering of the fluorophores along the track. Similar interpretations apply to the ordering of minima for **d** and **f**. **d**, Monopedal walkers on the same track (JOE (orange star) → TAMRA (pale green) → FAM (pale blue)). **e**, Comparison of time scales for bipedal and monopedal walkers (eighteen traces per walker type: three fluorophores, six experiments). **f**, Bipedal walker with track labelled TAMRA (red star) → JOE (green) → FAM (blue).

Joined by a duplex torso, each of two identical walker legs, I, is capable of catalysing the formation of waste duplex A•B from metastable fuel hairpins A and B through a reaction pathway in which I assembles with A, which assembles with B, which subsequently disassembles I from the complex (see Fig. 5a and b for the reaction graph and corresponding molecular implementation). The track consists of five A hairpins arranged linearly at regular intervals along a nicked DNA duplex. In the presence of hairpin B, a subpopulation of walkers is expected to move unidirectionally along the track by sequentially catalysing the formation of A•B. Because of the one-dimensional arrangement of anchor sites, this processive motion occurs only for those walkers that use a foot-over-foot gait by stochastically lifting the back foot at each step.

We investigate walker locomotion using a bulk fluorescence assay that tests whether there is a subpopulation of walkers that moves processively through positions 3, 4 and 5, starting from an initial condition with legs anchored at positions 1 and 2. Quenchers are attached to the walker's legs and spectrally distinct fluorophores are positioned proximal to anchorages 3, 4 and 5. Consistent with processivity, the anticipated sequential transient quenching of the fluorophores at positions 3, 4 and 5 is observed (Fig. 5c). To rule out the possibility that this signal arises from non-processive walker diffusion through the bulk solution from one position to the next, we repeated the experiments using monopedal walkers that lack a mechanism for achieving processivity. In this case, the sequential transient quenching no longer matches the ordering of the fluorophores along the track (Fig. 5d) and the timescale for visiting any one of the three anchorages is longer than the timescale to visit all three anchorages for the bipedal system (Fig. 5e). Additional control experiments (Supplementary Information 6.9) show that this difference in timescales cannot be explained by the relative rates with which freely diffusing bipedal and monopedal walkers land on the track. As a further test of processivity for the bipedal walker, reordering the fluorophores along the track leads to the expected change in the ordering of the transient quenching (Fig. 5f).

The experimental execution of these four molecular programs demonstrates that the hairpin motif functions as a modular programmable kinetic trap, and that rewiring the connections between nodes in the reaction graph corresponds to rewiring the connections between kinetic traps in the underlying free-energy landscape. In the physical systems, metastable hairpins are initially caught in engineered kinetic traps; the introduction of initiator molecules begins a chain reaction of kinetic escapes in which the hairpin species interact through programmed assembly and disassembly steps to implement dynamic functions. It is important that the timescale of metastability for kinetically trapped molecules is longer than the timescale relevant for the execution of the program. We found it helpful to incorporate clamping segments at the ends of helices to discourage the initiation of non-toehold-mediated branch migrations (see Supplementary Information 3.1). We also found that impure strand syntheses artificially reduce the strength of metastable traps and increase leakage rates. System fidelity was improved by ligating hairpins out of two shorter segments to increase strand purity (Supplementary Information 7.1).

Reaction graphs can be extended beyond the present versatile motif by defining new nodal species that abstract the functional relationships between domains in other motifs. The present hierarchical approach to encoding dynamic function in nucleic acid sequences represents a promising step towards the goal of constructing a compiler for biomolecular function—an automated design process that requires as input a modular conceptual system design, and provides as output a set of biopolymer sequences that encode the

desired dynamic system behaviour (Supplementary Information 7.2).

METHODS SUMMARY

Starting from a conceptual dynamic function, a molecular implementation is realized in three steps summarized in Fig. 1f. See Supplementary Information 3.1 for an example illustrating the design of the catalytic three-arm junction system. Step (1): pathway specification. We specify the pathway that implements a target dynamic function using a reaction graph. Step (2): translation to motifs. The reaction graph is directly translated to motif secondary structures. First, the basic complementarity requirements are defined and then clamping/padding segments are added (as in Supplementary Information 3.1). Initial dimensioning of the number of nucleotides in each segment is performed using the NUPACK server (www.nupack.org), which models the behaviour of strand species in the context of a dilute solution (including unintended species of complexes)[29]. Step (3): sequence design. Sequences are designed by considering a suite of structures that punctuate the intended reaction pathway or that explicitly preclude undesired off-pathway interactions (for example, structures specifying the absence of an interaction between two strands that should not pair). The sequences are optimized computationally (J. N. Zadeh and R. M. Dirks, personal communication) to maximize affinity and specificity for this suite of structures by minimizing the average number of incorrectly paired bases at equilibrium[30]. We then synthesize and verify the system using gel electrophoresis, bulk fluorescence quenching, or single-molecule AFM.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 July; accepted 31 October 2007.

1. Butterfoss, G. L. & Kuhlman, B. Computer-based design of novel protein structures. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 49–65 (2006).
2. Seeman, N. C. DNA in a material world. *Nature* **421,** 427–431 (2003).
3. Turberfield, A. J. *et al.* DNA fuel for free-running nanomachines. *Phys. Rev. Lett.* **90,** 118102 (2003).
4. Bois, J. S. *et al.* Topological constraints in nucleic acid hybridization kinetics. *Nucleic Acids Res.* **33,** 4090–4095 (2005).
5. Green, S. J., Lubrich, D. & Turberfield, A. J. DNA hairpins: Fuel for autonomous DNA devices. *Biophys. J.* **91,** 2966–2975 (2006).
6. Seelig, G., Yurke, B. & Winfree, E. Catalyzed relaxation of a metastable DNA fuel. *J. Am. Chem. Soc.* **128,** 12211–12220 (2006).
7. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proc. Natl Acad. Sci. USA* **101,** 15275–15278 (2004).
8. Rothemund, P. W. K., Papadakis, N. & Winfree, E. Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biol.* **2,** 2041–2053 (2004).
9. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314,** 1585–1588 (2006).
10. Yurke, B., Turberfield, A. J., Mills, J. A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406,** 605–608 (2000).
11. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394,** 539–544 (1998).
12. Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **427,** 618–621 (2004).
13. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).
14. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
15. Feldkamp, U. & Niemeyer, C. M. Rational design of DNA nanoarchitectures. *Angew. Chem. Int. Edn Engl.* **45,** 1856–1876 (2006).
16. Robertson, A., Sinclair, A. J. & Philp, D. Minimal self-replicating systems. *Chem. Soc. Rev.* **29,** 141–152 (2000).
17. von Kiedrowski, G. A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Edn Engl.* **25,** 932–935 (1986).
18. Paul, N. & Joyce, G. F. A self-replicating ligase ribozyme. *Proc. Natl Acad. Sci. USA* **99,** 12733–12740 (2002).
19. Levy, M. & Ellington, A. D. Exponential growth by cross-catalytic cleavage of deoxyribozymogens. *Proc. Natl Acad. Sci. USA* **100,** 6416–6421 (2003).
20. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature* **382,** 525–528 (1996).
21. Zhang, D. Y., Turberfield, A. J., Yurke, B. & Winfree, E. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318,** 1121–1125 (2007).
22. Li, Y. *et al.* Controlled assembly of dendrimer-like DNA. *Nature Mater.* **3,** 38–42 (2004).
23. Yin, P., Yan, H., Daniell, X. G., Turberfield, A. J. & Reif, J. H. A unidirectional DNA walker that moves autonomously along a track. *Angew. Chem. Int. Edn Engl.* **43,** 4906–4911 (2004).
24. 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. Edn Engl.* **44,** 4355–4358 (2005).
25. Bath, J., Green, S. J. & Turberfield, A. J. A free-running DNA motor powered by a nicking enzyme. *Angew. Chem. Int. Edn Engl.* **44,** 4358–4361 (2005).
26. Pei, R. *et al.* Behavior of polycatalytic assemblies in a substrate-displaying matrix. *J. Am. Chem. Soc.* **128,** 12693–12699 (2006).
27. Venkataraman, S., Dirks, R. M., Rothemund, P. W. K., Winfree, E. & Pierce, N. A. An autonomous polymerization motor powered by DNA hybridization. *Nature Nanotechnol.* **2,** 490–494 (2007).
28. Asbury, C. L. Kinesin: world's tiniest biped. *Curr. Opin. Cell Biol.* **17,** 89–97 (2005).
29. Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E. & Pierce, N. A. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.* **49,** 65–88 (2007).
30. Dirks, R. M., Lin, M., Winfree, E. & Pierce, N. A. Paradigms for computational nucleic acid design. *Nucleic Acids Res.* **32,** 1392–1403 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the following for discussions: J. S. Bois, R. M. Dirks, M. Grazier G'Sell, R. F. Hariadi, J. A. Othmer, J. E. Padilla, P. W. K. Rothemund, T. Schneider, R. Schulman, M. Schwarzkopf, G. Seelig, D. Sprinzak, S. Venkataraman, E. Winfree, J. N. Zadeh and D. Y. Zhang. We also thank J. N. Zadeh, R. M. Dirks and J. M. Schaeffer for the use of unpublished software, and R. F. Hariadi and S. H. Park for advice on AFM imaging. This work is funded by the NIH, the NSF, the Caltech Center for Biological Circuit Design, the Beckman Institute at Caltech, and the Gates Grubstake Fund at Caltech.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on *Nature*'s website (http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.A.P. (niles@caltech.edu).

METHODS

System design. A molecular implementation is realized in three steps summarized in Fig. 1f and illustrated in Supplementary Information 3.1. Step (1): pathway specification. Step (2): translation to motifs. Following initial dimensioning using the NUPACK server, the segment dimensions are sometimes further optimized based on subsequent experimental testing. Step (3): sequence design. After computational optimization, occasional further manual optimization was performed using the same design metric on a subset of crucial target structures. We then further analysed the thermodynamic behaviour of the sequences using the NUPACK server. For some systems, stochastic kinetic simulations[31] (J. M. Schaeffer, personal communication) were carried out to confirm the absence of significant kinetic traps along the target reaction pathways. The sequences are shown in Supplementary Information 8.

System synthesis. DNA was synthesized and purified by Integrated DNA Technologies. The purified DNA strands were reconstituted in ultrapure water (resistance of 18 MΩ cm). We determined the concentrations of the DNA solutions by measuring ultraviolet light absorption at 260 nm.

Hairpins were synthesized as two pieces which were then ligated to produce the full hairpin (see Supplementary Information 7.1 for details). We performed the ligation using T4 DNA ligase (New England Biolabs) at either room temperature or 16 °C for a minimum of 2 h. We further purified ligated strands using denaturing polyacrylamide gel electrophoresis. The bands corresponding to the DNA strands of expected sizes were visualized by ultraviolet shadowing and excised from the gel. The DNA strands were then eluted and recovered by ethanol precipitation.

For monomer preparation, we diluted the concentrated DNA strands to reaction conditions: 50 mM Na_2HPO_4, 0.5 M NaCl, pH = 6.8 for species in Fig. 2 and Supplementary Fig. 4; and 20 mM Tris, pH = 7.6, 2 mM EDTA, 12.5 mM Mg^{2+} (1 × TAE/Mg^{2+} buffer) for species in Fig. 3, Supplementary Fig. 12, and Fig. 4. We then annealed the hairpins by heating for 5 min at 90 °C, and then turning off the heating block to allow the system to cool to room temperature (requiring at least 2 h). For walker system assembly, see Supplementary Information 6.4.

Gel electrophoresis. For the gel in Fig. 2c, 12 μl of each 3-μM hairpin species were mixed by pipetting. Portions of this master mix were aliquoted into five separate tubes (6 μl per tube). To these tubes we added 2 μl of either 3 μM I (lane 1), 1.5 μM I (lane 2), 0.75 μM I (lane 3), 0.3 μM I (lane 4), or 1× reaction buffer (50 mM Na_2HPO_4, 0.5 M NaCl, pH = 6.8) (lane 5) to reach a total reaction volume of 8 μl. The samples were then mixed by pipetting and allowed to react for 2.5 h at room temperature. The annealed reaction (lane 6), prepared 0.5 h in advance, was made by mixing 2 μl of each hairpin with 2 μl of the 1× reaction buffer, and then annealing as described in monomer preparation. A 2% native agarose gel was prepared for use in 1× LB buffer (Faster Better Media, LLC). We then mixed 1 μl of each sample with 1 μl of 5× SYBR Gold loading buffer: 50% glycerol/50% H_2O/SYBR Gold (Invitrogen) and loaded this into the gel. The gel was run at 350 V for 10 min at room temperature and imaged using an FLA-5100 imaging system (Fuji Photo Film).

For the gel in Fig. 4c, we annealed the hairpins at the following concentrations: A1, A2, B2, A3 and B3 at 1 μM; A4 and B4 at 2 μM; A5 and B5 at 4 μM. The initiator I was prepared at 800 nM. The following sample mixtures were prepared: lane 1, A1; lane 2, I + A1; lane 3, I + A1 + A2 + B2; lane 4, I + A1 + A2 + B2 + A3 + B3; lane 5, I + A1 + A2 + B2 + A3 + B3 + A4 + B4; lane 6, I + A1 + A2 + B2 + A3 + B3 + A4 + B4 + A5 + B5; lane 7, A1 + A2 + B2 + A3 + B3 + A4 + B4 + A5 + B5. Here, I, A1, A2 and B2 were added at 1 μl; A3, B3, A4, B4, A5 and B5 at 2 μl. We added 1× reaction buffer (20 mM Tris, pH = 7.6, 2 mM EDTA, 12.5 mM Mg^{2+}) to bring the total volume of each sample to 16 μl. We mixed the samples by pipetting and allowed them to react for 2 h at room temperature. A 1% native agarose gel was prepared in 1× LB buffer. We added 8 μl of each sample to 2 μl 5× SYBR Gold loading buffer and loaded 8 μl of this sample/loading-buffer mix into the gel. The gel was run at 350 V for 10 min at room temperature and then imaged using an FLA-5100 imaging system. For the reactions in Fig. 4d, the hairpins were mixed to reach the following final concentration: A1-Cy5 (see Supplementary Information 8.4), A2, B2, 100 nM; A3, B3, 200 nM; A4, B4, 400 nM; A5, B5, 800 nM. We then aliquoted portions of this mix into 10 separate tubes (9 μl per tube). To these tubes we added either 1× TAE/Mg^{2+} reaction buffer or the initiator I to give the indicated final concentration of I and a final volume of 11 μl. The samples were mixed by pipetting and allowed to react for 1 h at room temperature. We then mixed the sample with 5× LB loading buffer (Faster Better Media, LLC) to reach 1× loading buffer concentration (8 μl sample, 2 μl loading buffer). We loaded the sample/loading buffer mix into a 1% native agarose gel prepared in 1× LB buffer. The gel was run at 350 V for 10 min at room temperature and then imaged and quantified using an FLA-5100 imaging system. The experiments were performed with 10 μM inert 25-nt poly-T carrier strands[21] in the reaction solution.

AFM imaging. We obtained AFM images using a multimode scanning probe microscope (Veeco Instruments), equipped with a Q-Control module for analogue AFM systems (Atomic Force F&E). The images were obtained in liquid phase under tapping mode using DNP-S oxide sharpened silicon nitride cantilevers (Veeco). We first diluted samples in 1× TAE/Mg^{2+} buffer to achieve the desired imaging density. We applied a 20 μl drop of 1× TAE/Mg^{2+} and a 5 μl drop of sample to the surface of freshly cleaved mica and allowed them to bind for approximately 2 min. We added supplemental Ni^{2+} (15–30 mM) to increase the strength of DNA–mica binding[32]. Before placing the fluid cell on top of the mica puck, we added an additional 15–20 μl of 1× TAE/Mg^{2+} buffer to the cavity between the fluid cell and the AFM cantilever chip to avoid bubbles.

Fluorescence experiments. For catalytic circuitry experiments, we obtained fluorescence data using a QM-6/2005 steady state spectrofluorometer (Photon Technology International), equipped with a Turret 400^{TM} four-position cuvette holder (Quantum Northwest) and 3.5-ml QS quartz cuvettes (Hellma). The temperature was set to 25 °C. We set the excitation and emission wavelengths to 520 nm (2-nm bandwith) and 540 nm (4-nm bandwidth), respectively. For the experiments in Fig. 3c, we prepared hairpin monomers, A, B, C and D, and initiator, I, separately as described above. We added 40 μl 1-μM A to 1.8 ml 1× TAE/Mg^{2+} buffer and mixed it by rapid pipetting eight times using a 1-ml tip. We recorded the baseline signal for ~16 min. Then we added 40 μl of 1-μM B, C and D and the appropriate concentration of I (or 1× TAE/Mg^{2+} buffer in the case of 0× I) to the cuvette (to reach the target concentrations described in Fig. 3c) and mixed by rapid pipetting eight times using a 1-ml tip. The control with 20-nM A alone was monitored continuously. The final volume was 2 ml for all experiments. We carried out the experiments with 10-μM inert 25-nt poly-T carrier strand[21] in the individual hairpin and initiator stock solutions and ~1-μM inert 25-nt poly-T carrier strands in the final reaction solution.

For autonomous locomotion experiments, we used the same spectrofluorometer as above with the temperature controller set to 21 °C. We used two 3.5-ml QS quartz cuvettes (Hellma) in each set of experiments. Excitation and emission wavelengths were set to 492 nm and 517 nm (for FAM), 527 nm and 551 nm (for JOE), and 558 nm and 578 nm (for TAMRA), respectively, with 4-nm bandwidths. The assembly of the walker system is described in Supplementary Information 6.4. We snap-cooled hairpin B in the reaction buffer (4 mM $MgCl_2$, 15 mM KCl and 10 mM Tris-HCl, pH = 8.0): heating at 95 °C for 90 s, rapid cooling at room temperature, sitting at room temperature for 30 min before use. The system was assembled using 4 nM track and 3.5 nM bipedal walker. We used a substochiometric amount of walker to ensure that no free-floating walker would bind to hairpin A on the track. For the same reason, we used substoichiometric monopedal walker (7 nM) in the diffusion experiments. The final concentration of hairpin B was 20 nM, which was equimolar with the five A hairpins on the track (5 × 4 nM = 20 nM). The assembled track was first introduced to record the fluorescence baselines for FAM, JOE and TAMRA. We then introduced hairpin B and mixed 100 times by rapid pipetting to start walker locomotion.

31. Flamm, C., Fontana, W., Hofacker, I. L. & Schuster, P. RNA folding at elementary step resolution. *RNA* **6**, 325–338 (2000).
32. Hansma, H. G. & Laney, D. E. DNA binding to mica correlates with cationic radius: assay by atomic force microscopy. *Biophys. J.* **70**, 1933–1939 (1996).