DYNAMIC SUPRAMOLECULAR SYSTEMS BUILT BY NUCLEIC ACID SELF-ASSEMBLY: QUICK, FLEXIBLE, PROGRAMMABLE

ANDREW J. TURBERFIELD
Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, United Kingdom
Kavli Institute for Nanoscience Discovery, University of Oxford, Dorothy Crowfoot Hodgkin Building, South Parks Rd, Oxford OX1 3QU, United Kingdom

Introduction

The capacity of nucleic acids to store and transmit information plays a central role in the dynamic chemistry of life and must therefore have great potential in the creation and control of synthetic supramolecular systems. Oligonucleotides store information digitally using an alphabet of four bases: what distinguishes them from other information-rich molecules, such as proteins, is the ease with which interactions mediated by base-pairing can be manipulated. Control of intra- and intermolecular bonding through rational design of base sequence underpins research into the reprogramming of natural systems (synthetic biology), devices and systems built from oligonucleotide components (DNA and RNA nanotechnology), archival information storage and molecular computation. There is natural precedent for each of these applications, even molecular construction: nucleic acids are the material of genes, they play key roles in the regulation of gene expression and constitute the molecular machinery for ribosomal protein production.
This article provides a broad-brush, personal perspective of a powerful but idiosyncratic technology that is still discovering its place in the physical and life sciences.

Building dynamic systems with nucleic acids

Nucleic acid nanofabrication [1], using information encoded in base sequence to program assembly, is rapid and versatile. Nanostructures made from DNA make use of a small set of structural motifs: the double helix provides rigidity and the opportunity to link complementary domains of two molecules; interhelical strand crossovers serve as vertices in mesh-like structures [2,3] or bind together parallel helices to fill space and increase rigidity [4,5]. Assembly, typically by annealing multiple oligonucleotide components [6], is programmed through control of patterns of base pairing through sequence design. It is usual to assign a unique sequence to each double-helical segment, even in otherwise symmetrical objects, such that the target assembly is unambiguously the most stable product. The simplicity and modularity of this construction principle means that computational design tools can be used to automate sequence design [7,8]. It also means that DNA nanofabrication is extremely versatile – the same set of structural motifs can be tweaked to improve a design or reconfigured to make something completely different. Construction with RNA adds a richer set of tertiary structural

interactions [9] and, for intracellular applications, poses the challenge of design for assembly by out-of-equilibrium co-transcriptional folding of a single strand of RNA [10]. The field relies on the quality, speed and rapidly decreasing price of commercial DNA synthesis: a DNA origami nanostructure (typically 4.5MDa) can be designed, ordered, and assembled within a few weeks with a materials cost of the order of €1000.

Most synthetic nucleic acid systems rely on a limited repertoire of chemical reactions, the base-pairing and other non-covalent interactions that determine DNA and RNA secondary and tertiary structure, sometimes supplemented by enzyme-catalyzed covalent reactions (restriction, ligation, templated polymerization). DNA [11] and RNA [12,13] have catalytic potential which can be used in dynamic devices [14,15]. Physical and chemical diversity can be added through use of covalently modified oligonucleotides to add lipids, fluorophores, backbone inserts, multivalent vertices, reactive linker groups and thereby peptides, proteins, nanoparticles etc. [16]. Because each part of a nucleic acid nanostructure can usually be uniquely identified by a local sequence of base pairs, the stoichiometry and position of each modification can be precisely controlled. An interesting application is DNA-templated synthesis, in which dynamic systems of chemically modified oligonucleotides are used to control a sequence of covalent reactions by bringing reactive building blocks into proximity in a programmed sequence [17].

A significant advantage of nucleic acids in the creation of supramolecular systems is that the sequence-specific base-pairing that controls nanostructure assembly can be used to mediate dynamic interactions. A simple example is a strand-displacement reaction, in which one oligonucleotide is displaced from a duplex by another with a domain of similar sequence [18,19]. Strand displacement can be used to change nanostructure conformation – for example, to open and close tweezer- or cage-like structures [18,20] to move a "walker" along a track [21] or to initiate a templated coupling reaction [22]. The rates and equilibria of strand-displacement reactions can be widely tuned by breaking symmetry between the initial and final duplexes by adding or subtracting terminal base pairs [23], introducing or healing base-pairing defects [24,25] or by using transient interactions to co-localize reactants [26,27]. Kinetic control of strand-displacement reactions can be used to create molecular machinery that operates autonomously [28,29] and systems for molecular computation [30,31]. Molecular computation based on strand-displacement reactions can be surprisingly powerful in practice because it lends itself to composable design: circuit components replicated with different sequences can operate more-or-less orthogonally and oligonucleotides embodying information-carrying signals can act as both inputs and outputs of elementary operations.

An important class of dynamic system makes use of natural DNA- and RNA-modifying enzymes [32]. Operations include ligation (which can be conditional on hybridization to a complementary "splint"), extension by a polymerase of a primer hybridized to a template, restriction (nicking or complete cutting) of a duplex incorporating a precisely defined sequence motif, and degradation (hydrolysis of the backbone). More closely biomimetic genetic circuits make use of, for example, riboswitches or ribosomal production of transcription factors to control gene expression in cell-free systems, including geometrically controlled artificial environments [33], or in cells [34].

Dynamic systems of nucleic acids can also be designed to respond to chemical stimuli, typically through competition between a ligand-binding motif, such as an aptamer, G-quadruplex or i-motif, and an alternative base-paired secondary structure, and to light. Such systems can form complex adaptive reaction networks [35].

There is a formidable strand of application-focused research into the use of DNA for archival information storage [36]. This is motivated by its chemical stability, information storage density and certain protection from obsolescence of technologies for manipulating and reading DNA. At its most straightforward this technology involves only base-by-base solid-support synthesis and sequencing technologies – but there is scope for more complex synthesis schemes and context-dependent information retrieval that build on techniques of DNA computation [37].

Some advantages and frustrations of building with nucleic acids

Compared to conventional top-down nanofabrication by lithography, DNA and RNA self-assembly offers considerably greater resolution, enabling construction with isotropic sub-nanometre precision [38] of micrometer-scale objects [39]. However, assembly defects are ubiquitous and increase dramatically when patterning at larger length scales.
Compared to covalent synthesis and synthetic protein production, construction by DNA and RNA self-assembly is quicker and more flexible, capable of producing larger and more complex systems (in dimension and number of distinct components), and has an embedded capacity for embedded information and reprogrammable interaction that is difficult to emulate. It is not well adapted to provide the fine spatial control (on sub-nanometre length scales) required to create a synthetic enzyme, offers much less chemical functionality, and currently operates at very low synthesis scales (typically in the picomole to nanomole range).

Perspective

Applications of nucleic acids in nanostructure assembly, synthetic molecular machinery, dynamic reaction networks and computation are all programmed – and can be reprogrammed – through information encoded in base sequence. For this reason, nucleic acids provide a uniquely flexible system for the exploration of the fundamental science of dynamic molecular systems. For the creation and testing of model systems the limitations of dynamic systems of nucleic acids are outweighed by their flexibility and speed.
In developing more practical applications of synthetic systems of nucleic acids, it is important to recognize their limitations. DNA synthesis costs are falling dramatically and it is already reasonable to contemplate nanostructure fabrication on gramme scales [40]. However, it is likely that complex DNA and RNA nanostructures will be limited to high-value applications. One promising (but static) example is the use of DNA templates to lay out the components of three-dimensional molecular electronic circuits. This would be a technological revolution – but the drive to discover a radically new way to continue to miniaturize electronics in the spirit of Moore's Law will soon become irresistible.
A second class of application where economic drive has the potential to overcome economic drawback is in medicine. Synthetic nucleic acid nanostructures can probe biological systems [41] and have the potential to combine sensing, computation and actuation to and to create theranostic devices, that is, autonomous systems that combine local diagnosis and therapy. Nucleic acids provide a natural interface to natural genetic control systems and to RNA-directed gene editing: potential applications include tissue- and cell- specific treatment for genetic disease or senescence. Pursuit of applications within living systems suggest use of RNA nanostructures folded within the cell from RNA transcribed *in situ*. This creates new challenges and opportunities to develop

dynamic assembly techniques: co-transcriptional folding from a single RNA strand is a non-equilibrium process in which kinetic traps must be overcome or exploited [10].

DNA-templated chemistry can be applied to enable sequence-controlled oligomer synthesis and could form the basis of a discovery technology to explore new chemical spaces [17,42]. Synthetic molecular machines, made from DNA and programmed by synthetic "genes", would generate libraries of oligomers from natural and non-natural building blocks in parallel, autonomously, in a one-pot reaction. Retention of the connection between programming gene and product would enable amplification and readout of the genes attached to the few products selected (for example, for receptor binding or catalytic activity), resynthesis and even mutation and evolution. Selected products would then be synthesized conventionally for characterization and application – programmed synthesis by molecular machinery would only be used for discovery, for which current synthesis scales are already appropriate.

A related potential technology builds on Drexler's vision for additive manufacture with molecular or atomic control [43], using DNA-programmed machinery built from DNA to position a "write head" with nanometer precision. A two-dimensional molecular printing system has been demonstrated [44]; extension to three dimensions and to more useful chemistry is work in progress. The question of whether such a manufacturing technology could be useful given limitations related to DNA synthesis scales is unresolved: it is straightforward now to operate 10^{12} molecular devices in parallel, ten or more orders of magnitude more than a typical number of parallel conventional production lines but fewer by at least the same factor than the number of parallel reactions in bulk chemical synthesis. An appealing possibility is that DNA machinery could be used to bootstrap a second generation of molecular printers using materials with less intrinsic programmability but more desirable physical characteristics.

Elements of DNA computation are embedded in all of the applications discussed above: design for effective nanostructure assembly involves control of assembly pathways that are closely connected to computational strand-displacement cascades; dynamic reaction networks map onto electronic circuits; a theranostic device must compute on its sensor inputs to determine its mode and level of output; and context-specific information retrieval from a DNA archive is a form of computation. DNA strand-displacement networks are capable of executing much higher-level programs using quantities of material so small that they could be distributed throughout a structure or material: future uses of this extraordinary capability are difficult to predict.

References

1 N. C. Seeman, H. F. Sleiman, *Nat Rev Mater* **3**, 17068 (2018).
2 J. H. Chen, N. C. Seeman, *Nature* **350**, 631 (1991).
3 R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben *et al.*, *Science* **310**, 1661 (2005).
4 P. W. K. Rothemund, *Nature* **440**, 297 (2006).
5 S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg *et al.*, *Nature* **459**, 414 (2009).
6 K. E. Dunn, F. Dannenberg, T. E. Ouldridge, M. Kwiatkowska *et al.*, *Nature* **525** (2015).
7 S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez *et al.*, *Nucleic Acids Res* **37**, 5001 (2009).

8 E. Benson, A. Mohammed, J. Gardell, S. Masich *et al.*, *Nature* **523**, 441 (2015).
9 A. Chworos, I. Severcan, A. Y. Koyfman, P. Weinkam *et al.*, *Science* **306**, 2068 (2004).
10 C. Geary, G. Grossi, E. K. S. McRae, P. W. K. Rothemund *et al.*, *Nat Chem* **13**, 549 (2021).
11 P. Nissen, J. Hansen, N. Ban, P. B. Moore *et al.*, *Science* **289**, 920 (2000).
12 D. S. Wilson, J. W. Szostak, *Annu Rev Biochem* **68**, 611 (1999).
13 S. W. Santoro, G. F. Joyce, *Proc Natl Acad Sci USA* **94**, 4262 (1997).
14 M. N. Stojanovic, D. Stefanovic, *Nat Biotechnol* **21**, 1069 (2003).
15 K. Lund, A. J. Manzo, N. Dabby, N. Michelotti *et al.*, *Nature* **465**, 206 (2010).
16 C. K. McLaughlin, G. D. Hamblin, H. F. Sleiman, *Chem Soc Rev* **40**, 5647 (2011).
17 X. Y. Li, D. R. Liu, *Angew Chem Int Ed* **43**, 4848 (2004).
18 B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel *et al.*, *Nature* **406**, 605 (2000).
19 D. Y. Zhang, G. Seelig, *Nat Chem* **3**, 103 (2011).
20 R. P. Goodman, M. Heilemann, S. Doose, C. M. Erben *et al.*, *Nat Nanotechnol* **3**, 93 (2008).
21 J. S. Shin, N. A. Pierce, *J Am Chem Soc* **126**, 10834 (2004).
22 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes *et al.*, *Nat Chem* **8**, 542 (2016).
23 B. Yurke, A. P. Mills, *Genet Program Evol M* **4**, 111 (2003).
24 R. R. F. Machinek, T. E. Ouldridge, N. E. C. Haley, J. Bath *et al.*, *Nat Commun* **5**, 5324 (2014).
25 N. E. C. Haley, T. E. Ouldridge, I. M. Ruiz, A. Geraldini *et al.*, *Nat Commun* **11**, 2562 (2020).
26 A. J. Genot, D. Y. Zhang, J. Bath, A. J. Turberfield, *J Am Chem Soc* **133**, 2177 (2011).
27 J. Cabello-Garcia, W. Bae, G. B. V. Stan, T. E. Ouldridge, *ACS Nano* **15**, 3272 (2021).
28 A. J. Turberfield, J. C. Mitchell, B. Yurke, A. P. Mills *et al.*, *Phys Rev Lett* **90**, 118102 (2003).
29 S. J. Green, J. Bath, A. J. Turberfield, *Phys Rev Lett* **101**, 238101 (2008).
30 D. Soloveichik, G. Seelig, E. Winfree, *Proc Natl Acad Sci USA* **107**, 5393 (2010).
31 L. Qian, E. Winfree, *Science* **332**, 1196 (2011).
32 K. Montagne, R. Plasson, Y. Sakai, T. Fujii *et al.*, *Mol Syst Biol* **7**, 466 (2011).
33 E. Karzbrun, A. M. Tayar, V. Noireaux, R. H. Bar-Ziv, *Science* **345**, 829 (2014).
34 M. N. Win, C. D. Smolke, *Science* **322**, 456 (2008).
35 Z. Z. Li, J. B. Wang, I. Willner, *Adv Funct Mater*, 2200799.
36 L. Ceze, J. Nivala, K. Strauss, *Nat Rev Genet* **20**, 456 (2019).
37 C. Bee, Y. J. Chen, M. Queen, D. Ward *et al.*, *Nat Commun* **12**, 4764 (2021).
38 M. Kube, F. Kohler, E. Feigl, B. Nagel-Yuksel *et al.*, *Nat Commun* **11**, 6229 (2020).
39 K. F. Wagenbauer, C. Sigl, H. Dietz, *Nature* **552**, 78 (2017).
40 F. Praetorius, B. Kick, K. L. Behler, M. N. Honemann *et al.*, *Nature* **552**, 84 (2017).
41 E. Silvester, B. Vollmer, V. Prazak, D. Vasishtan *et al.*, *Cell* **184**, 1110 (2021).
42 R. K. O'Reilly, A. J. Turberfield, T. R. Wilks, *Acc Chem Res* **50**, 2496 (2017).
43 K. E. Drexler, *Proc Natl Acad Sci USA* **78**, 5275 (1981).
44 E. Benson, R. C. Marzo, J. Bath, A. J. Turberfield, *Sci Robot* **7**, eabn5459 (2022).