Expert Opinion





Ashley Publications
www.ashley-pub.com

General

# Synthetic biology

A Michael Sismour & Steven A Benner†
†*University of Florida, Departments of Chemistry and Anatomy and Cell Biology, Gainesville, FL 32611-7200, USA*

Chemistry is a broadly powerful discipline in contemporary science because it has the ability to create new forms of the matter that it studies. By doing so, chemistry can test models that connect molecular structure to behaviour without having to rely on what nature has provided. This creation, known as 'synthesis', began to be applied to living systems in the 1980s as recombinant DNA technologies allowed biologists to deliberately change the molecular structure of the microbes that they studied, and automated chemical synthesis of DNA became widely available to support these activities. The impact of the information that has emerged has made biologists aware of a truism that has long been known in chemistry: synthesis drives discovery and understanding in ways that analysis cannot. Synthetic biology is now setting an ambitious goal: to recreate in artificial systems the emergent properties found in natural biology. By doing so, it is advancing our understanding of the molecular basis of genetics in ways that analysis alone cannot. More practically, it has yielded artificial genetic systems that improve the healthcare of some 400,000 Americans annually. Synthetic biology is now set to take the next step, to create artificial Darwinian systems by direct construction. Supported by the National Science Foundation as part of its Chemical Bonding program, this work cannot help but generate clarity in our understanding of how biological systems work.

**Keywords:** artificial genetic systems, biological networks, Darwinian systems, emergent properties, logic circuit, ribozyme, synthetic biology

*Expert Opin. Biol. Ther. (2005) 5(11):1409-1414*

## 1. Introduction

Animate and inanimate matter have been thought to be fundamentally different from the start of humankind's inquiry into the natural world. The boundary between the two became blurred in the 20th century, however, as researchers became able to associate chemical structures to many phenomena that were historically called 'biological'. These included phenomena as complex as human genetics (the human genome sequence is nothing more and nothing less than a description of how carbon, oxygen, nitrogen, phosphorus and hydrogen atoms are placed in inherited molecules) [1] and evolution itself, which is increasingly being described by its molecular signatures [2].

Today, those seeking a distinction between animate and inanimate matter often look to 'emergent properties' for this purpose. The concept of emergence embraces the notion that the behaviour of a collection of molecules is more than the sum of the behaviours of its parts [3]. Terms such as 'biocomplexity' and 'systems biology' attempt to capture this concept as well. Those working under these rubrics hope to move from the simple analysis of biological systems to complex analysis, measuring many parameters within the system rather than a few. By doing so, they hope that understanding (or failing that, predictive power) will emerge if mathematical formalism reproduces the measured phenomenology well enough [4].

As might be expected from past experiences in the mathematical treatment of chemical complexity, such an analytical approach, even one as extensive as that contemplated by many systems biologists, is expected not to deliver 'understanding', however that may be defined. This is not just because increasingly parameterised mathematical models often offer no more than an increasingly precise reproduction of system noise. Rather, the problem is more fundamental.

First, the mathematical model cannot capture, except as a crude approximation, the underlying reality. From quantum mechanics to macroeconomics, the necessary approximations end up missing elements of the system that are critical. The analogy from chemistry (and biology is, at this level, chemistry) would be to have a collection of time-dependent measurements arising from a mixture of benzene, acetic acid, hexane, formamide, ammonia and water, and then trying to understand the system by fitting curves to its properties. There are good reasons to expect that this will not produce understanding, yet this is exactly what we are trying to do when we analyse databases of protein sequences (replace benzene by 'phenylalanine', acetic acid by 'glutamic acid', hexane by 'leucine', formamide by 'asparagine' and ammonia by 'lysine') using mathematics.

## 2. Synthesis as an alternative to analysis

Synthesis offers a different strategy [5]. Instead of a 'probe and model' paradigm, synthesis uses a symmetrical double paradigm: If you understand it, then you can make it; if you can make it, then you can say that you understand it.

In biological chemistry, this paradigm first emerged in the 1950s with the construction of enzyme models in a field then called 'bioorganic chemistry' [6]. Under this rubric, chemists synthesised artificial molecules in the hope of reproducing the catalytic activity and specificity of natural protein enzymes [7]. If they did, then understanding was advanced. If they did not, a new synthetic effort was attempted.

Nobel prizes (Cram, Lehn) and awards from the American Chemical Society (the Breslow prize) recognised the accomplishments in this field [8,9]. Although few artificial enzymes have ever come close to reproducing in magnitude the catalytic power of natural enzymes, many reproduced the mechanisms of natural enzymes [10]. As a consequence, our understanding of virtually every enzymatic reaction is today grounded in the chemical models synthesised by bioorganic chemists in an effort to reproduce the emergent properties of enzymes, where the whole is greater than the sum of the amino acid parts [11].

It was a logical consequence for synthetic biologists to then move to larger emergent properties of biological systems. Self-reproduction, reproduction with errors, and reproduction with errors where the errors themselves are reproducible, are the hallmarks of biological systems [12]. These are also the minimum combination of chemical properties necessary for Darwinian processes to be operative. Darwinian processes, in turn, are the only way that emergent properties are generated in animate systems, at least those known on contemporary Earth.

Approximately 15 years ago, synthetic biologists set out to recreate these emergent properties [13-15]. While artificial Darwinian systems are not yet in hand, considerable progress has been made towards getting them [16]. At the same time, as spin-offs, tools and technologies that benefit humankind have emerged, just as they did over the past century in chemistry.

### 2.1 Artificial genetic systems

One simple synthetic biological system relates to the familiar Watson–Crick Model for the double helix. The Watson–Crick nucleobase pairing follows two rules of chemical complementarity. The first, size complementarity, pairs large purines with small pyrimidines. The second, hydrogen bonding complementarity, pairs hydrogen bond donors from one nucleobase with hydrogen bond acceptors from the other (**Figure 1**).

If nucleobase pairing were indeed so simple, that is, if such simple concepts indeed confer understanding, the synthetic biology double paradigm requires that we be able to synthesise new nucleobase pairs. These nucleobase pairs should be able to support genetic-like behaviours, including sequence-specific binding, encoding and template-directed replication. These are the same emergent behaviours that are found in genetics, of course; the synthetic biologist seeks to gain them using different chemical structures.

In this case, the synthetic biologists were successful. By shuffling hydrogen bond donating and accepting groups, as well as engineering a few side chains and changing the positions of a few nitrogen atoms in the purine and pyrimidine rings, a new genetic system was synthesised. Eight additional synthetic nucleobases, forming four additional base pairs, were created (**Figure 1**). The synthetic genetic alphabet was independently replicatable following an extended set of Watson–Crick rules [17].

Since this early artificial genetic system was created, synthetic biologists have worked to create a synthetic molecular biology to handle these. With only minor modifications of existing DNA polymerases and reverse transcriptases, it proved possible to get DNA-like molecules incorporating non-standard nucleobases to direct their own replication [16]. Such replication, as with natural DNA, is subject to mutation, with the mutant forms themselves able to direct their own replication. This is, of course, an artificial Darwinian system, albeit one that relies heavily on natural biopolymers and help from the synthetic biologist to survive.

As it provides rule-based molecular recognition that is orthogonal to recognition provided by natural DNA, this synthetic genetic system is found in the clinic today. As part of Bayer's VERSANT branched DNA diagnostic assay [18,19], synthetic biology helps manage the care of ~ 400,000 patients infected with HIV and hepatitis viruses each year. This assay (**Figure 2**) assembles a branched nanostructure using a series of binding events.

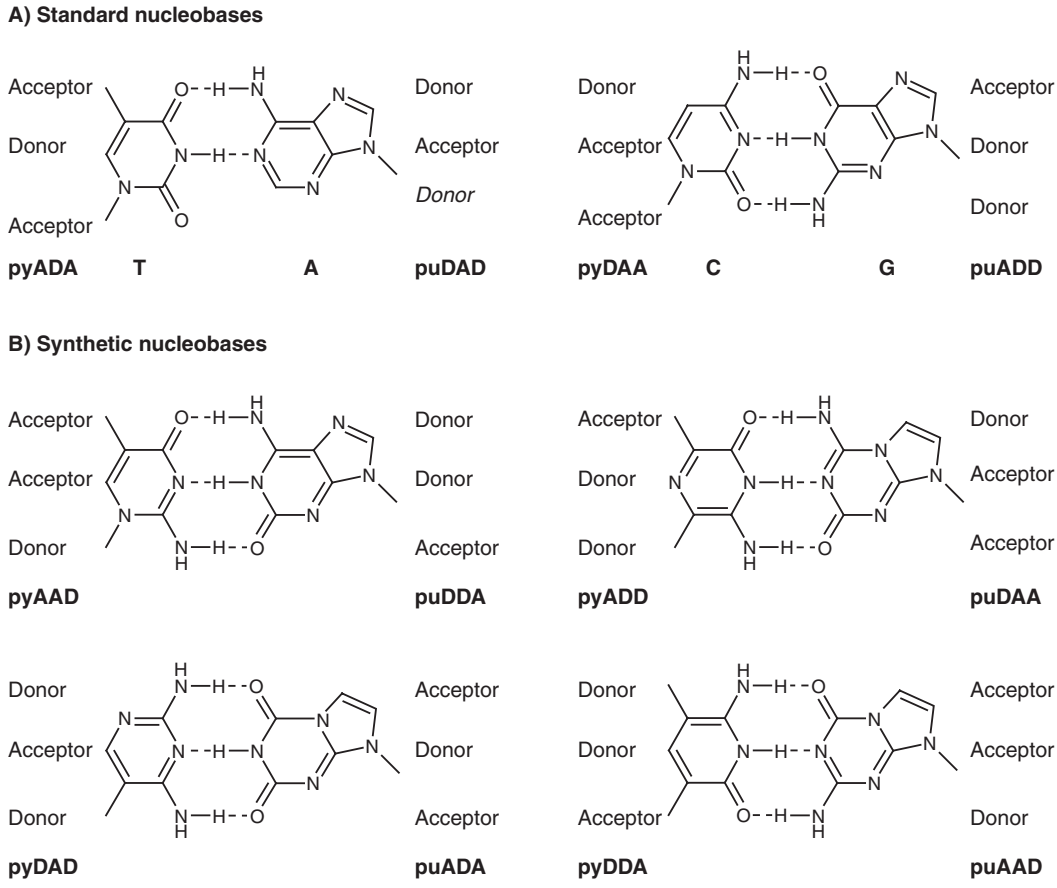


**Figure 1. An example of a synthetic genetic system.** Shown are **A)** the natural nucleobases and **B)** the artificial nucleobases with shuffled hydrogen bonding patterns. The small pyrimidines (left) pair with large purines (right) to achieve size complementarity. Hydrogen bonding complementarity is achieved by pairing hydrogen bond donors (D) with acceptors (A). Shuffling these donor and acceptor groups creates new nucleobase pairs.
A: Adenine; C: Cytosine; G: Guanine; pu: Purine; py: Pyrimidine; T: Thymidine.

In this case, synthesis showed that we do understand nucleobase pairing, as we can synthesise something that behaves the same, but in a different form. Furthermore, the synthetic genetic system is able to encode proteins that contain extra amino acids [20], support the construction of dendrimeric and other artificial genetic architectures, and have novel, semipredictable, biophysical properties [21,22].

### 2.2 Another meaning for 'synthetic biology': reassembling biological parts

Synthetic biology now has three meanings. The first, dating back to the 1980s, is the synthesis of artificial forms of life, and is exemplified by the artificial genetic systems discussed above [23]. The second, introduced by Kool in 2000 [24], concerns the synthesis of unnatural molecules that function in natural systems. The last, emerging more recently, concerns the assembly of natural biological parts in unnatural ways [25].

Engineers working under the last definition have shown how natural biological parts can be recombined to produce interesting properties [26,27]. An early example is the development of an *Escherichia coli* that can act as an IMPLIES logic gate, a basic computing component [28]. The resynthesised bacterium contains a combination of genetic elements obtained from a variety of sources. These act together to generate a simple genetic circuit based on the transcription regulation of several genetic elements.

The circuit, shown in **Figure 3**, relies on the fluorescent EYFP protein to generate the output signal. This protein is expressed under the control of a λPRO12 promoter. On a separate plasmid, the λ repressor protein, cl, is under the transcriptional control of the lac promoter, which is controlled by isopropyl-β-D-thiogalactopyranoside (IPTG) and the lac repressor. The outcome of these elements working together is a system that generates fluorescence in the absence of the input, IPTG, but not in the presence of the input.

A more recent success in this area includes the development of bacteria able to act as a timing device [29]. Here, synthetic biologists placed three transcriptional repressor systems from

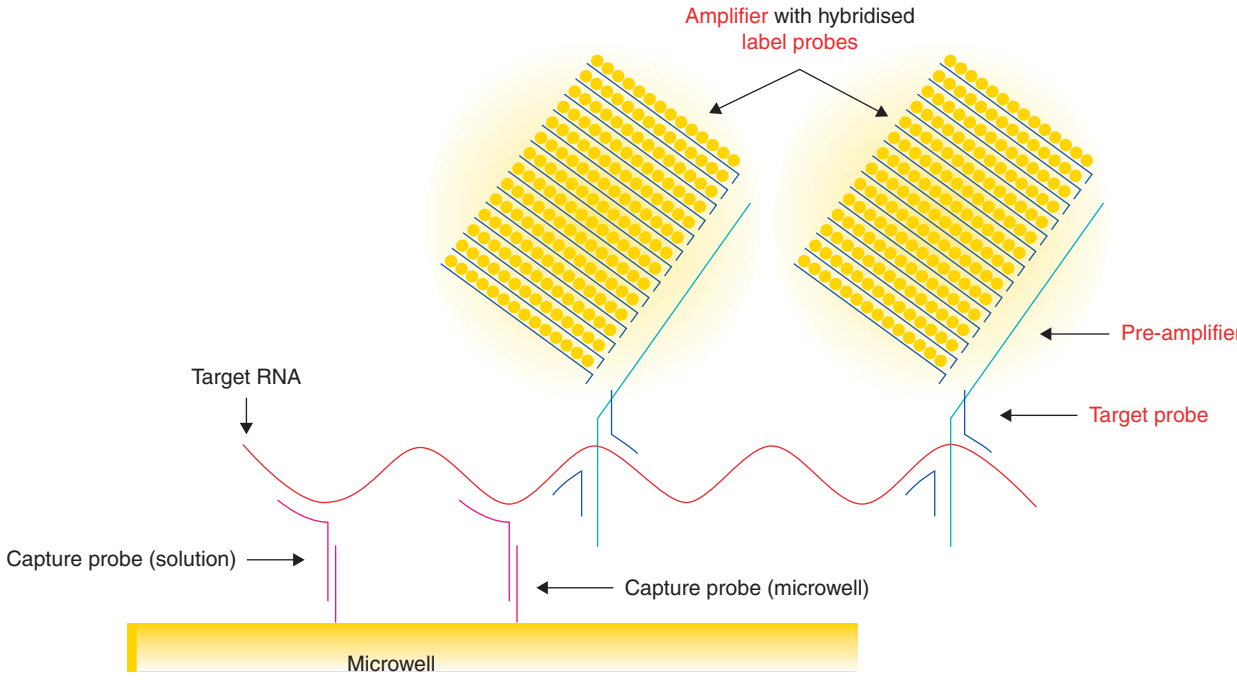


**Figure 2. Bayer's VERSANT assay.** The target RNA is bound to the microwell via hybridisation using a series of capture probes. The target RNA then complexes with the fluorescent label probe through a series of hybrisation events involving the target probe, preamplifier and amplifier. The use of non-canonical nucleosides in the amplifier, preamplifier and label probe increases the specificity of the assay, as non-target DNA present in the sample cannot nonspecifically hybridise with the artificial DNA.

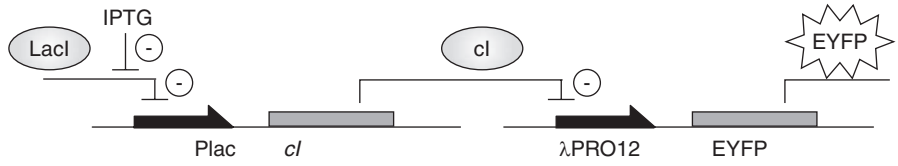


**Figure 3. The genetic circuit acting as an IMPLIES logic gate.** The logic gate, using IPTG and the LacI repressor as input molecules, consists of the fluorescent reporter molecule, EYFP, under transcriptional control of the λPRO12 promoter, which in turn is negatively controlled by the cI repressor protein. As the cI repressor is under transcriptional control of the lac promoter, the circuit generates fluorescent output in the absence of IPTG and no output in the presence of IPTG.
EYFP: Enhanced yellow fluorescent protein; λPRO12: Synthetic λ right promoter; IPTG: Isopropyl-β-D-thiogalactopyranoside.

independent sources into *E. coli*. The system was designed so that each promoter controls the expression of a protein acting as a repressor for another promoter in the system. The result is a time-dependent oscillation of fluorescent output from the green fluorescent protein. As the output cycle time was on the order of hours, longer than the 20 min division time of *E. coli*, the signal was passed to the progeny of the parent cells that began the cycle.

There seems to be no limit to the innovation of synthetic biologists generating these systems [30-34]. Logic circuits, for example, are being developed *in vitro* with ribozymes, catalytic RNA molecules obtained via *in vitro* evolution. A recent example is the development of a molecular automaton able to play tic-tac-toe with a human opponent [35]. The molecular automaton is nothing more than a series of ribozyme-based logic gates engineered to play a childhood game.

## 3. Expert opinion and conclusion

To celebrate the new millennium, the editors of *Chemical and Engineering News* (December 6, 1999) asked prominent chemists of their views about the major advances in the current century. Rita Colwell, Director of the National Science Foundation, commented that *'chemists … will develop self-replicating molecular systems to provide insights into the molecular origins of life'*. Shohei Inoue, President of the Chemical Society of Japan, elaborated further: *'One of the most*

*important unsolved scientific problems where chemistry should play a central role in is the origin of life',* he wrote. *'Recently, a system with order, one of the remarkable characteristics of life, has been demonstrated to appear on computer display from a disordered system by selecting an appropriate program. However, such a 'virtual' system has no relation to the substances that constitute life existing on Earth. There is … the possibility of the creation of a new ordered 'living' system independent of existing life'.*

These visions are well on their way to being realised, with the benefit coming from an enhanced understanding of the complexity of the living system that analysis cannot itself generate. While the potential for application of synthetic biology in human health and commerce has been demonstrated by its actually doing so, even the toys mentioned above have impact. Setting a difficult goal drives scientists and engineers across uncharted territory, forces them to solve unscripted problems, and requires successful solutions of these before the goal can be achieved. Thus, biology is not only acquiring its language from chemistry; it is also obtaining a new paradigm from chemistry.

## Bibliography


1. PENNISI E: Human genome – reaching their goal early, sequencing labs celebrate. *Science* (2003) **300**(5618):409-409.
2. BENNER SA, CARACO MD, THOMSON JM, GAUCHER EA: Evolution – planetary biology – paleontological, geological, and molecular histories of life. *Science* (2002) **296**(5569):864-868.
3. WOESE CR: A new biology for a new century. *Microbiol. Mol. Biol. Rev.* (2004) **68**(2):173-186.
4. VOIGT CA, MAYO SL, ARNOLD FH, WANG ZG: Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci. USA* (2001) **98**(7):3778-3783.
5. BENNER SA, SISMOUR AM: Synthetic biology. *Nat. Rev. Genet.* (2005) **6**(7):533-543.
6. BRESLOW R: The mechanism of thiamine action.2. Rapid deuterium exchange in thiazolium salts. *J. Am. Chem. Soc.* (1957) **79**(7):1762-1763.
7. BRESLOW R: Centenary Lecture – Biomimetic chemistry. *Chem. Soc. Rev.* (1972) **1**(4):553-580.
8. LEHN JM: Perspectives in supramolecular chemistry – from molecular recognition towards molecular information-processing and self-organization. *Angew. Chem. Int. Edit. Engl.* (1990) **29**(11):1304-1319.
9. CRAM DJ: The design of molecular hosts, guests, and their complexes. *Science* (1988) **240**(4853):760-767.
10. BRESLOW R, YANG J, YAN JM: Biomimetic hydroxylation of saturated carbons with artificial cytochrome P450 enzymes – liberating chemistry from the tyranny of functional groups. *Tetrahedron* (2002) **58**(4):653-659.
11. MILDVAN AS: Inverse thinking about double mutants of enzymes. *Biochemistry* (2004) **43**(46):14517-14520.
12. BENNER SA, RICARDO A, CARRIGAN MA: Is there a common chemical model for life in the universe? *Curr. Opin. Chem. Biol.* (2004) **8**(6):672-689.
13. JOYCE GF, INOUE T, ORGEL LE: Non-enzymatic template-directed synthesis on RNA random copolymers – poly(C,U) templates. *J. Mol. Biol.* (1984) **176**(2):279-306.
14. SWITZER CY, MORONEY SE, BENNER SA: Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* (1989) **111**(21):8322-8323.
15. DOUDNA JA, SZOSTAK JW: RNA-catalyzed synthesis of complementary-strand RNA. *Nature* (1989) **339**(6225):519-522.
16. SISMOUR AM, BENNER SA: The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res.* (2005) **33**(17):5640-5646.
17. GEYER CR, BATTERSBY TR, BENNER SA: Nucleobase pairing in Watson-Crick-like genetic expanded information systems. *Structure* (2003) **11**(12):1485-1498.
18. ELBEIK T, MARKOWITZ N, NASSOS P *et al.*: Simultaneous runs of the Bayer VERSANT HIV-1 version 3.0 and HCV bDNA version 3.0 quantitative assays on the system 340 platform provide reliable quantitation and improved work flow. *J. Clin. Microbiol.* (2004) **42**(7):3120-3127.
19. ELBEIK T, SURTIHADI J, DESTREE M *et al.*: Multicenter evaluation of the performance characteristics of the Bayer VERSANT HCV RNA 3.0 assay (bDNA). *J. Clin. Microbiol.* (2004) **42**(2):563-569.
20. BAIN JD, SWITZER C, CHAMBERLIN AR, BENNER SA: Ribosome-mediated incorporation of a nonstandard amino-acid into a peptide through expansion of the genetic-code. *Nature* (1992) **356**(6369):537-539.
21. HUTTER D, BLAETTLER MO, BENNER SA: From phosphate to bis(methylene) sulfone: non-ionic backbone linkers in DNA. *Helv. Chim. Acta* (2002) **85**(9):2777-2806.
22. LIU HB, GAO JM, LYNCH SR *et al.*: A four-base paired genetic helix with expanded size. *Science* (2003) **302**(5646):868-871.
23. HOBOM B: Surgery of genes – at the doorstep of synthetic biology. *Med. Klin.* (1980) **75**(24):14-21.
24. RAWLS RL: 'Synthetic biology' makes its debut. *Chem. Eng. News* (2000) **78**(17):49-53.
25. BALL P: Synthetic biology: starting from scratch. *Nature* (2004) **431**(7009):624-626.
26. MARTIN VJ, PITERA DJ, WITHERS ST, NEWMAN JD, KEASLING JD: Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* (2003) **21**(7):796-802.
27. HOWARD PL, CHIA MC, DEL RIZZO S, LIU FF, PAWSON T: Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins. *Proc. Natl. Acad. Sci. USA* (2003) **100**(20):11267-11272.
28. YOKOBAYASHI Y, WEISS R, ARNOLD FH: Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA* (2002) **99**(26):16587-16591.
29. ELOWITZ MB, LEIBLER S: A synthetic oscillatory network of transcriptional regulators. *Nature* (2000) **403**(6767):335-338.

30. SAGHATELIAN A, YOKOBAYASHI Y, SOLTANI K, GHADIRI MR: A chiroselective peptide replicator. *Nature* (2001) **409**(6822):797-801.

31. HASTY J, MCMILLEN D, COLLINS JJ: Engineered gene circuits. *Nature* (2002) **420**(6912):224-230.

32. SHIH WM, QUISPE JD, JOYCE GF: A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* (2004) **427**(6975):618-621.

33. CHEN IA, ROBERTS RW, SZOSTAK JW: The emergence of competition between model protocells. *Science* (2004) **305**(5689):1474-1476.

34. GAO JM, STRASSLER C, TAHMASSEBI D, KOOL ET: Libraries of composite polyfluors built from fluorescent deoxyribosides. *J. Am. Chem. Soc.* (2002) **124**(39):11590-11591.

35. STOJANOVIC MN, STEFANOVIC D: A deoxyribozyme-based molecular automaton. *Nat. Biotechnol.* (2003) **21**(9):1069-1074.

## Affiliation

A Michael Sismour[1] & Steven A Benner[†2]
[†]Author for correspondence
[1]University of Florida, Chemistry Department, P.O. Box 117200, Gainesville, Florida 32611-7200, USA
Tel: +1 352 846 1856;
E-mail: msismour@chem.ufl.edu
[2]University of Florida, Departments of Chemistry and Anatomy and Cell Biology, Gainesville, FL 32611-7200, USA
Tel: +1 352 392 7773;
E-mail: benner@chem.ufl.edu