# Tying new knots in synthetic biology

## David K. Karig[1] and Michael L. Simpson[1]

[1]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

**Recent years have seen the emergence of synthetic biology, which encompasses the engineering of living organisms as well as the implementation of biological behavior in non-living substrates. Many of these engineered systems have harnessed the diverse toolkit of proteins, genes, and cellular processes that nature offers. While these efforts have been fruitful, they have also illustrated the difficulty associated with programming highly complex functions by tapping into cellular processes. Another set of efforts has focused on building circuits, performing computation, and constructing nanoscale machines using nucleic acids. Zhang *et al.*, 2007, *Science* 318, 1121–1125 and Yin *et al.*, 2008, *Nature* 451, 318–322 recently demonstrated flexible approaches for the modular construction of such biochemical devices exclusively using DNA. These approaches have exciting implications both for engineering living cells and for mimicking life-like behavior at the nanoscale. [DOI: 10.2976/1.2907240]**

CORRESPONDENCE
Michael L. Simpson: simpsonmll@ornl.gov

Living organisms thrive in environments ranging from arctic glaciers to hot springs and ocean floors (Finsinger *et al.*, 2008; Pawlowski *et al.*, 2007; Sonjak *et al.*, 2007), harness a wide variety of energy sources (Kristjánsson and Hreggvidsson, 1995), and theoretically possess a computational density far greater than existing electronic technologies (Simpson *et al.*, 2004). These remarkable properties have motivated the forward engineering of living biological systems as well as the creation of non-living systems with bio-inspired functionality (Andrianantoandro *et al.*, 2006; Doktycz and Simpson, 2007; Endy, 2005; Sprinzak and Elowitz, 2005). Such engineering efforts have manifested themselves in the emerging field of synthetic biology. To understand the broad picture of the current state of this field, it is helpful to consider the source of biological diversity and robustness. Csete and Doyle (2004) point to a "bowtie" architecture that emerges both in biological and complex technological systems, where a diversity of input signals are processed through a highly-regulated central core (knot of the bowtie) before fanning out to an equally diverse set of outcomes. For example, a myriad of different genes are fed into the same "knot" of transcription and translation processes to produce a tremendous variety of different pro-

teins, as illustrated in Fig. 1(A). The past few years have seen great progress in the area of developing synthetic systems by rewiring and remodeling components at the edges of the bowtie (e.g., utilizing genes, proteins, signal molecules, metabolic processes), while leaving the knot of the bowtie (e.g., transcription and translation processes) intact. In contrast, recent works by Zhang *et al.* (2007) and Yin *et al.* (2008) mark great strides towards tying new "knots" by using nucleic acids alone to implement a wide range of different functions.

By engineering networks of DNA, RNA, protein, and signal molecule interactions, synthetic biologists have built a variety of different systems. These include efficient production of the malaria drug artemisinin (Ro *et al.*, 2006), detection of TNT (Looger *et al.*, 2003), development of model systems to explore natural phenomena (Austin *et al.*, 2006; Elowitz *et al.*, 2002; Rosenfeld *et al.*, 2002), and construction of oscillators (Atkinson *et al.*, 2003; Elowitz and Leibler, 2000), switches (Gardner *et al.*, 2000), and pattern formation systems (Basu *et al.*, 2005). Many of these efforts have focused on engineering whole cells, thus tapping into natural cellular resources and taking advantage of cellular capabilities such as autonomous reproduction and functionality in varying envi-
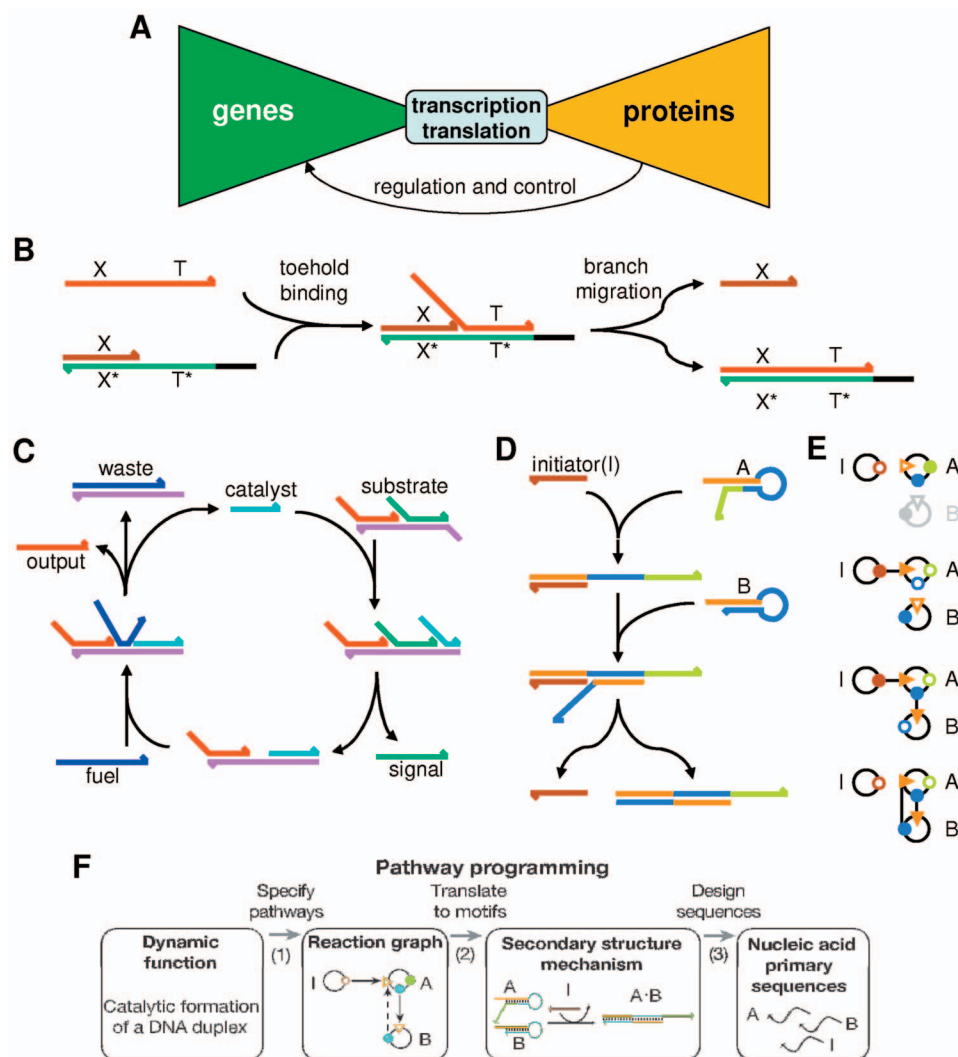
**Figure 1. (A) Many complex, robust systems exhibit "bowtie" architectures**. For example, in biology, a diverse set of genes is fed through the "knot" of transcription and translation processes before fanning out to an equally diverse set of proteins. (B) Both the Zhang *et al.* (2007) and Yin *et al.* (2008) methods make use of "toehold" binding and consequential branch migration. A single stranded toehold domain (T) can recognize and bind an exposed region of another DNA strand or complex that is complementary to the toehold (T·). If the remaining dangling region (X) is also complementary, it can displace an originally bound strand (brown X) through the process of branch migration. (C) Simplified catalytic cycle that serves as the building block for Zhang *et al.* (2007) systems. (D) In the Yin *et al.* (2008) approach, reactions are triggered by a single stranded "initiator" and are fueled by metastable hairpins which can open when a strand binds an exposed toehold and migrates towards the loop. (E) Each initiator and hairpin can be represented by a node (right) with a port corresponding to each toehold. A port is open if the corresponding toehold is exposed and closed if the toehold is bound. (F) Hierarchical design procedure for the Yin *et al.* (2008) approach. (Reprinted by permission from Macmillan Publishers Ltd: *Nature* **451**, 318–322, copyright 2008.)

ronments (Brenner *et al.*, 2007). One major challenge that this approach faces is that natural components taken from different organisms are not naturally designed to work together and are, thus, far from being "plug and play" (Andrianantoandro *et al.*, 2006). Rather, considerable modification and tuning is necessary to properly interface the components (Weiss *et al.*, 2003).

Another major obstacle when introducing a complex circuit consisting of components that span the bowtie is that some of the components may engage in spurious interactions with the host cell's machinery. Thus, the underlying set of

reactions which can impact the synthetic circuit, either directly or via competition for cell resources, is large, complex, and difficult to discern (Hakes *et al.*, 2008). This can tremendously complicate the debugging process. Furthermore, even *in vitro* systems that use high-level components still depend on complex processes of transcription, translation, and degradation (Kim *et al.*, 2006; Simpson, 2006).

A different approach to creating new biological or biolike systems is to focus on re-engineering the synthetic "knot". The past decade has seen the use of DNA to solve difficult computational problems (Adleman, 1994), perform

digital logic operations (Hagiya *et al.*, 2006; Macdonald *et al.*, 2006), and mimic dynamic biological processes such as the walking of kinesin down microtubules (Sherman and Seeman, 2004). These efforts strip away many of the unknown interactions and complexities associated with operation in a living cell and utilization of transcriptional and translational machinery. In addition, DNA-based systems generally perform logic and computation on a faster timescale than transcription/translation systems which involve thousands of reactions to produce a single transcript or protein. Nevertheless, many of these DNA-based approaches still rely on the use of enzymes such as ligase or restriction endonucleases and are specific to the particular problem approached (Ezziane, 2006). In contrast, the methodologies recently presented by Zhang *et al.* (2007) and Yin *et al.* (2008) avoid the use of proteins and guide the construction of various different types of DNA systems and circuits. Thus, these flexible methodologies embody a major step towards engineering new synthetic "knots" consisting of simplified, modular components that interact in well-defined ways.

## EXPLANATION OF APPROACHES

Like most natural biological systems, reactions in the Zhang *et al.* (2007) and Yin *et al.* (2008) approaches proceed isothermally and effectively utilize catalysts to drive reactions. In both strategies, reactions are set in motion when a specific "toehold" region of one strand binds the exposed complementary region of another strand or complex, as shown in Fig. 1(B). Both strategies also harness branch migration, which allows one strand to effectively displace another. The Zhang *et al.* (2007) systems are driven by the entropic gain of molecules that are liberated by the system reactions. On the other hand, Yin *et al.* (2008) drive reactions through the free energy of base formation. In both cases, energy can be added to the system by the addition of DNA "fuel" strands (Benenson *et al.*, 2003; Turberfield *et al.*, 2003).

Specifically, the Zhang *et al.* (2007) approach is based on constructing and interfacing different catalytic cycles. Figure 1(C) provides a simplified illustration of such a catalytic cycle. First, the single stranded input "catalyst" binds a "toehold" overhang region on the "substrate" DNA complex. Branch migration allows the input to displace the "signal" strand from the "substrate" complex, which in turn frees up a spot for the "fuel" strand to bind. The "fuel" strand binds and ultimately displaces the "output signal" as well as the input "catalyst," thus completing the catalytic cycle.

By contrast, the main components of the Yin *et al.* (2008) systems are single stranded "initiators" and metastable hairpins with different overhanging "toehold" regions. As depicted in Fig. 1(D), an "initiator" strand can bind the compatible toehold of a metastable hairpin. Branch migration of the "initiator" then opens the hairpin, exposing new toehold regions that were previously inside the hairpin loop. This can trigger additional reactions, as different hairpins or "initia-

tors" can bind the newly exposed toeholds and possibly dislodge the original "initiator" through branch migration.

## DESIGN METHODOLOGY

One key feature of both strategies is the capacity for organized and modular design. In the case of Zhang *et al.*, (2007), a "catalyst," "signal," "fuel," or "output" strand of one catalytic cycle can serve as either the "catalyst," "signal," "fuel," or "output" in another cycle. This modularity hinges upon the independence of the input and output sequences. Careful design of "specificity domains" in the input and output strands helps to ensure this independence. Strands are divided into a series of specificity and toehold domains. Initially, these domains are designed to contain only A, C, and T bases. Problematic subsequences are altered, and the domains are then concatenated to form the overall strands. This is followed by an iterative procedure of checking the strands for dimer and hairpin formation tendencies, changing problematic subsequences, and repeating.

Yin *et al.* (2008) present a hierarchical approach to programming reaction pathways. Design begins by representing a desired set of reactions with a reaction graph. In a reaction graph, each type of "initiator" and hairpin is represented by a node. A node has ports corresponding to each toehold domain. A port is open if the toehold is exposed and closed if the toehold is inaccessible. Two complementary ports can interact if both are accessible, and this interaction can affect the accessibility of other ports, thus setting up other possible reactions. As an example, Fig. 1(E) shows the nodes and port states for each reaction step to the left in Figure 1(D).

The next step in the pathway programming procedure [Fig. 1(F)] is to translate the reaction graph nodes into DNA motifs. Segments are added to help prevent undesired leakage reactions, and segment lengths are then adjusted to further reduce leakage. Finally, the actual sequences of these motifs are selected such that the expected number of incorrectly paired bases at equilibrium is minimized.

## SIGNIFICANCE OF ACCOMPLISHMENTS

The modularity and flexibility of the two approaches embody an important step towards the development of complex DNA systems and circuits. The authors have demonstrated this flexibility by creating and testing a variety of example systems. Zhang *et al.* particularly focus on issues confronting the modular construction of complex circuits. Previous studies have shown functional cascades of logic gates in the nucleic acid substrate but have also revealed the need for amplification to combat signal loss in deep cascades (Seelig *et al.*, 2006). Toward this end, Zhang *et al.* (2007) present a feed forward cascade capable of 900-fold amplification, as shown in Fig. 2(A). Also, both Zhang *et al.* (2007) and Yin *et al.* (2008) have constructed autocatalytic loops that exhibit exponential kinetics. These could be useful not only for amplifying weak outputs of long cascades but also for detection
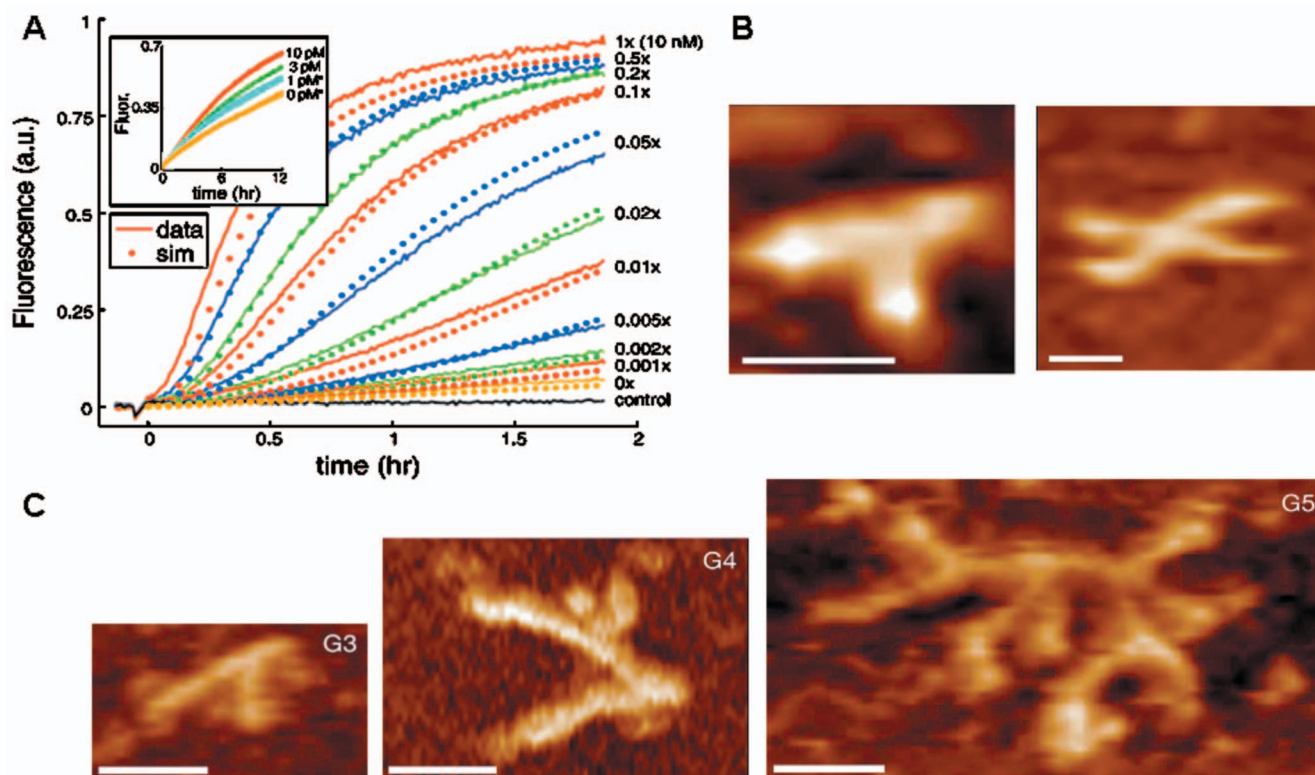
**Figure 2.** (**A**) **The feed-forward cascade constructed by Zhang et al. (2007) exhibits 900-fold amplification**. Reprinted by permission from AAAS: *Science* **318**, 1121–1125, copyright 2008. (B) AFM images of three arm and four arm branch structures formed by Yin *et al.* (10 nm scale bar). (Reprinted by permission from Macmillan Publishers Ltd: *Nature* **451**, 318–322, copyright 2008.) (C) AFM images of different steps in the progression of a dendritic growth system (30 nm scale bar). (Reprinted by permission from Macmillan Publishers Ltd: *Nature* **451**, 318–322, copyright 2008.)

applications. Yin *et al.* (2008) focus on controlling the formation of self-assembly structures and on mimicking dynamic biological pathways. For example, they demonstrate the programmed formation of three and four armed branch structures [Fig. 2(B)] as well as dendritic growth [Fig. 2(C)]. In addition, they present a DNA system that imitates the walking of kinesin molecules down microtubules.

**CHALLENGES**

A few key challenges confront the scalability of the Zhang *et al.* (2007) and Yin *et al.* (2008) approaches. As systems grow larger in size, crosstalk must be minimized to mitigate undesired alternative reaction pathways. Also, in large systems, the timescales of operation of different components and modules must be taken into consideration. For example, in the case of Yin *et al.* (2008), the timescale of hairpin metastability must be greater than the overall program timescale. One possible solution would be to program checkpoints to effectively synchronize and validate certain operations before proceeding to the next stage. Such checkpoints are seen in natural biological processes such as the cell cycle (Nyberg *et al.*, 2002). Yin *et al.* (2008) specifically mention the development of a "compiler." Such a tool for systematic, automated design could considerably help to address the above

challenges when tackling the implementation of complex systems. Another practical consideration is that circuit function is not necessarily immune to reactions with the external environment. Thus, the context of operation must be carefully considered in applications. Towards demonstrating the plausibility of their circuits in biomedical applications, Zhang *et al.* (2007) tested the function of a basic catalytic reaction in mouse liver total RNA and found that performance was not dramatically degraded. They also demonstrated robust performance of the catalytic circuit under a broad range of salt and temperature conditions.

**CONCLUSION**

Zhang *et al.* (2007) and Yin *et al.* (2008) have demonstrated robust and flexible methodologies for designing DNA circuits and systems, thus opening the door to a host of applications. One exciting possibility is the convergence between the approach of engineering basic components and the approach of engineering the whole cell. The faster execution, smaller size, simpler set of reactions, and potentially greater isolation from natural cellular components make DNA circuits an appealing choice for programming complex functionality. DNA circuitry can serve as a synthetic "knot" which can be interfaced to the higher level circuitry that is

necessary for probing natural systems or for applications such as drug fabrication. Such an interface may be provided by interactions with RNA, aptamers, and specific helix-turn-helix motifs of proteins. This merging of approaches to engineering biological functionality is analogous to the design of mixed analog and digital electronic devices. Specifically, although analog design and layout is largely an art, analog circuitry is often necessary for providing an interface to signals that are fundamentally analog in nature. For example, a cell phone must handle radio frequency transmission and reception, microphone input, and speaker output. However, the analog signals are converted to digital signals and passed to the digital circuitry, which executes the bulk of complex computation (Carey, 2006). Likewise, circuits such as the ones proposed by Zhang *et al.* (2007) and Yin *et al.* (2008) can handle the bulk of computation, while the more difficult to design circuitry consisting of higher level components can provide an interface to the whole cell. Yet another exciting front involves efforts to ultimately redefine the meaning of "whole cell" by minimizing the genome, reprogramming the genome, or entirely replacing the cell with similarly functioning synthetic components (Doktycz and Simpson, 2007; Gibson *et al.*, 2008; Lartigue *et al.*, 2007).

## REFERENCES

Adleman, LM (1994). "Molecular computation of solutions to combinatorial problems." *Science* **266**, 1021–1024.

Andrianantoandro, E, Basu, S, Karig, DK, and Weiss, R (2006). "Synthetic biology: new engineering rules for an emerging discipline." *Mol Syst Biol* **2**, 2006 0028.

Atkinson, MR, Savageau, MA, Myers, JT, and Ninfa, AJ (2003). "Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli." *Cell* **113**, 597–607.

Austin, DW, Allen, MS, McCollum, JM, Dar, RD, Wilgus, JR, Sayler, GS, Samatova, NF, Cox, CD, and Simpson, ML (2006). "Gene network shaping of inherent noise spectra." *Nature* **439**, 608–611.

Basu, S, Gerchman, Y, Collins, CH, Arnold, FH, and Weiss, R (2005). "A synthetic multicellular system for programmed pattern formation." *Nature* **434**, 1130–1134.

Benenson, Y, Adar, R, Paz-Elizur, T, Livneh, Z, and Shapiro, E (2003). "DNA molecule provides a computing machine with both data and fuel." *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2191–2196.

Brenner, K, Karig, DK, Weiss, R, and Arnold, FH (2007). "Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium." *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17300–17304.

Carey, D (2006). "En route to one-chip cell phone." *Electon. Eng. Times*.

Csete, M, and Doyle, J (2004). "Bow ties, metabolism and disease." *Trends Biotechnol.* **22**, 446–450.

Doktycz, MJ, and Simpson, ML (2007). "Nano-enabled synthetic biology." *Mol Syst Biol* **3**, 125.

Elowitz, MB, and Leibler, S (2000). "A synthetic oscillatory network of transcriptional regulators." *Nature* **403**, 335–338.

Elowitz, MB, Levine, AJ, Siggia, ED, and Swain, PS (2002). "Stochastic gene expression in a single cell." *Science* **297**, 1183–1186.

Endy, D (2005). "Foundations for engineering biology." *Nature* **438**, 449–453.

Ezziane, Z (2006). "DNA computing: applications and challenges." *Nanotechnology* **17**, R27–R39.

Finsinger, K, Scholz, I, Serrano, A, Morales, S, Uribe-Lorio, L, Mora, M, Sittenfeld, A, Weckesser, J, and Hess, WR (2008). "Characterization of true-branching cyanobacteria from geothermal sites and hot springs of Costa Rica." *Environ. Microbiol.* **10**, 460–473.

Gardner, TS, Cantor, CR, and Collins, JJ (2000). "Construction of a genetic toggle switch in *Escherichia coli*." *Nature* **403**, 339–342.

Gibson, DG, Benders, GA, Andrews-Pfannkoch, C, Denisova, EA, Baden-Tillson, H, Zaveri, J, Stockwell, TB, Brownley, A, Thomas, DW, Algire, MA, Merryman, C, Young, L, Noskov, VN, Glass, JI, Venter, JC, Hutchison, CA, 3rd, and Smith, HO (2008). "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome." *Science* **319**, 1215–1220.

Hagiya, M, Yaegashi, S, and Takahashi, K (2006). "Computing with Hairpins and Secondary Structures of DNA." In Nanotechnology, Science and Computation pp. 293–308.

Hakes, L, Pinney, JW, Robertson, DL, and Lovell, SC (2008). "Protein-protein interaction networks and biology—what's the connection?." *Nat. Biotechnol.* **26**, 69–72.

Kim, J, White, KS, and Winfree, E (2006). "Construction of an *in vitro* bistable circuit from synthetic transcriptional switches." *Mol Syst Biol* **2**, 68.

Kristjánsson, JK, and Hreggvidsson, GO (1995). "Ecology and habitats of extremophiles." *World J. Microbiol. Biotechnol.* **11**, 17–25.

Lartigue, C, Glass, JI, Alperovich, N, Pieper, R, Parmar, PP, Hutchison, CA, 3rd, Smith, HO, and Venter, JC (2007). "Genome transplantation in bacteria: changing one species to another." *Science* **317**, 632–638.

Looger, LL, Dwyer, MA, Smith, JJ, and Hellinga, HW (2003). "Computational design of receptor and sensor proteins with novel functions." *Nature* **423**, 185–190.

Macdonald, J, Li, Y, Sutovic, M, Lederman, H, Pendri, K, Lu, W, Andrews, BL, Stefanovic, D, and Stojanovic, MN (2006). "Medium scale integration of molecular logic gates in an automaton." *Nano Lett.* **6**, 2598–2603.

Nyberg, KA, Michelson, RJ, Putnam, CW, and Weinert, TA (2002). "Toward maintaining the genome, DNA damage and replication checkpoints." *Annu. Rev. Genet.* **36**, 617–656.

Pawlowski, J, Fahrni, J, Lecroq, B, Longet, D, Cornelius, N, Excoffier, L, Cedhagen, T, and Gooday, AJ (2007). "Bipolar gene flow in deep-sea benthic foraminifera." *Mol. Ecol.* **16**, 4089–4096.

Ro, DK, Paradise, EM, Ouellet, M, Fisher, KJ, Newman, KL, Ndungu, JM, Ho, KA, Eachus, RA, Ham, TS, Kirby, J, Chang, MC, Withers, ST, Shiba, Y, Sarpong, R, and Keasling, JD (2006). "Production of the antimalarial drug precursor artemisinic acid in engineered yeast." *Nature* **440**, 940–943.

Rosenfeld, N, Elowitz, MB, and Alon, U (2002). "Negative autoregulation speeds the response times of transcription networks." *J. Mol. Biol.* **323**, 785–793.

Seelig, G, Soloveichik, D, Zhang, DY, and Winfree, E (2006). "Enzyme-free nucleic acid logic circuits." *Science* **314**, 1585–1588.

Sherman, WB, and Seeman, NC (2004). "A precisely controlled DNA biped walking device." *Nano Lett.* **4**, 1203–1207.

Simpson, ML (2006). "Cell-free synthetic biology: a bottom-up approach to discovery by design." *Mol Syst Biol* **2**, 69.

Simpson, ML, Cox, CD, Peterson, GD, and Sayler, GS, (2004). "Engineering in the biological substrate: information processing in genetic circuits." *Proc. IEEE* **92**, 848–863.

Sonjak, S, Ursic, V, Frisvad, JC, and Gunde-Cimerman, N (2007). "Penicillium svalbardense, a new species from Arctic glacial ice." *Antonie van Leeuwenhoek* **92**, 43–51.

Sprinzak, D, and Elowitz, MB (2005). "Reconstruction of genetic circuits." *Nature* **438**, 443–448.

Turberfield, AJ, Mitchell, JC, Yurke, B, Mills, AP, Jr., Blakey, MI, and Simmel, FC (2003). "DNA fuel for free-running nanomachines." *Phys. Rev. Lett.* **90**, 118102.

Weiss, R, Basu, S, Hooshangi, S, Kalmbach, A, Karig, D, Mehreja, R, and Netravali, I (2003). "Genetic circuit building blocks for cellular computation, communications, and signal processing." ⟨http://dx.doi.org/10.1023/A:1023307812034⟩ *Natural Computing: an international journal* **2**, 47–84.

Yin, P, Choi, HM, Calvert, CR, and Pierce, NA (2008). "Programming biomolecular self-assembly pathways." *Nature* **451**, 318–322.

Zhang, DY, Turberfield, AJ, Yurke, B, and Winfree, E (2007). "Engineering entropy-driven reactions and networks catalyzed by DNA." *Science* **318**, 1121–1125.