Tying new knots in synthetic biology

David K. Karig and Michael L. Simpson

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

(Received 19 March 2008; published online 17 April 2008)

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]

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 proteins, 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-
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” (Andriananantoandro 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 bio-like 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 open the hairpin, exposing new toehold regions that were previously inside the hairpin loop. This can trigger additional reactions, as different hairpins or “initiators” 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.
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; Lar tigue et al., 2007).

REFERENCES


