

# Complexity from Simple Building Blocks: Engineering Large-scale Information-processing Networks from Molecules

Yaakov Benenson\*

**Abstract:** Complexity in molecular systems can manifest itself either structurally or functionally. One of the more complex functions encountered in the natural world is that of information processing, or computation. Similarly, artificial cells will require this capacity to fully exploit their potential. Here I review the state of the art in the field, describe our contribution to this challenge in the framework of NCCR Molecular Systems Engineering, and propose an outlook for future efforts.

**Keywords:** Complexity · Computation · Engineering · Gene circuits · Information processing

## 1. Introduction

Nothing in the physical world surpasses the complexity of a living cell; the emergence of cells from inanimate matter (aka the ‘Origin of Life’ question) and the principles behind their internal organization are, and will remain, the subject of intense study in the next decades and perhaps centuries. One obvious feature, however, is the emergence of complex properties from (often simple) interactions between multiple relatively simple and well-defined players – genes, proteins, lipids, small molecule chemicals. Any attempt to engineer cell-like systems, or augment cells with new functions, should arguably follow the same path, that is, rely on emergence to perform complex tasks rather than on engineering extremely intricate building blocks. This idea inspired the foundation of ‘systems chemistry’,<sup>[1]</sup> and it is one of the driving forces behind the concept of Molecular Systems Engineering.

## 2. Molecular Computing as a Paradigm of Systems Complexity

One field where these ideas have been intensively investigated experimentally is molecular computing.<sup>[2]</sup> Computation is an intrinsically complex task. Interestingly, universal computation (both in theory and in practice) can be performed on machines comprising simple, repetitive building blocks. For example a Turing machine,<sup>[3]</sup> a universal computer, can be described in a single paragraph; a primary school pupil can understand its principles of operation. Similarly, modern computers employ circuit boards that integrate billions of identical components. Computation, or in-

formation processing, can also take place in chemical reaction networks. Chemical concentrations are interpreted as values, either continuous or binary (‘all’ or ‘none’).<sup>[4]</sup> Some chemicals, usually those that trigger the network, are designated as inputs, and others, usually some of the final products, are the outputs (Fig. 1). The output concentrations depend on the inputs in complex fashion, and sometimes the dependency can be concisely formulated as a mathematical function, such as  $[\text{output}] = [\text{input 1}] + [\text{input 2}]$ .

It turns out that most of the complex cellular sub-systems perform some kind of molecular-level computing using environmental and internal inputs to control and

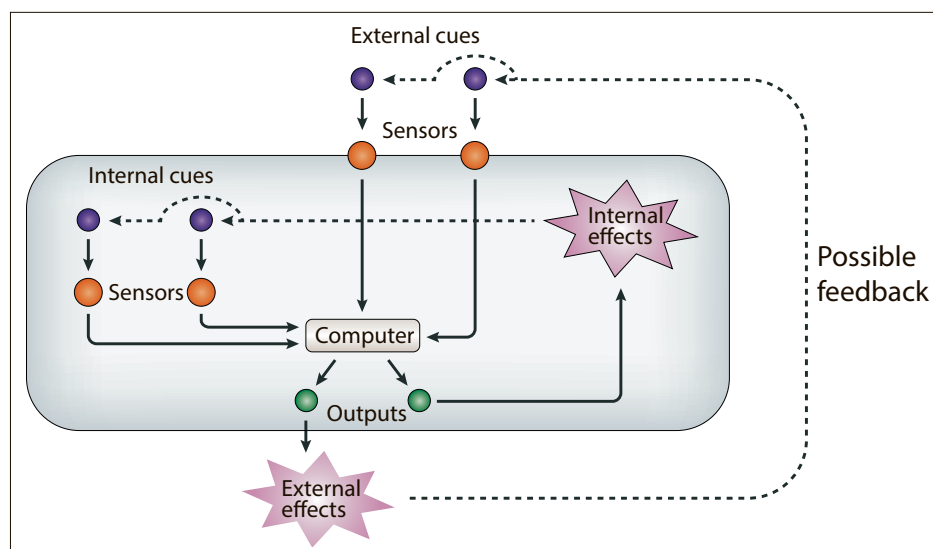


Fig. 1. Schematic description of a molecular computing system (shown here in a context of a live or an artificial cell). The inputs are the cues, external and internal. They are first intercepted by cognate sensors that in turn transduce the information to the computer module where it is integrated. The outcomes of the process – outputs – may trigger internal and external effects, which in turn may alter the inputs via feedback mechanisms. Reproduced from Y. Benenson, *Nature Rev. Genetics*, 2012, 13, 455.<sup>[2a]</sup>

\*Correspondence: Prof. Y. Benenson  
Department of Biosystems Science and Engineering  
ETH Zurich, Mattenstrasse 26  
CH-4058 Basel  
E-mail: kobi.benenson@bsse.ethz.ch

direct their function.<sup>[5]</sup> Therefore, the effort to design an artificial cell must explicitly address the capacity to compute with requisite molecular information. There are two aspects to this challenge: (1) Proof-of-principle level, where the main question is ‘what is possible’; and (2) task-specific challenges where the goal is precisely defined in terms of the task an artificial system is supposed to perform. Here the specifications can be further constrained by the choice of building blocks.

So far, engineered molecular computing systems have been built using a number of different approaches, the choice being dictated most and foremost by the milieu in which the systems operate. Historically, earliest works used organic molecules to implement logic switches;<sup>[6]</sup> typically, the inputs and the outputs of such systems were distinct, preventing cascading of switches into increasingly complex circuits. DNA has become a building block of choice in the last two decades.<sup>[7]</sup> On one hand, synthetic DNA oligonucleotides can be manufactured to precise specifications and at relatively low cost; on the other hand, the very nature of a DNA molecule lends itself to a large variety of information-processing frameworks. Highly-complex computing systems such as state machines,<sup>[9]</sup> logic circuits,<sup>[9]</sup> and neural circuits,<sup>[10]</sup> were successfully made of synthetic DNA. Later on, a number of reports described enzyme-based logic switches and circuits.<sup>[11]</sup> However, both DNA- and enzyme-based systems operate in well-controlled buffer solutions and are chemical, rather than biological, in nature. In parallel, the effort to implement computation in live cells has revolved around genetically-encoded systems.<sup>[12]</sup> The main reason for this choice is technical feasibility: genetic encoding is inherently compatible with a living cell, while any other approach requires extra effort to integrate with cellular processes. To give an example, it took more than a decade of intense effort in a number of research labs to lead to the first report on DNA-based logic gates operating in mammalian cells a few months ago.<sup>[13]</sup>

### 3. Molecular Computing as an Enabling Technology for Drug Discovery

Using genetic encoding allows exploring a variety of engineering approaches without being excessively hindered by technical difficulty of implementing non-native switches in cells. We have so far pursued this strategy in the framework of NCCR Molecular Systems Engineering, and described an approach that uses a complex computing network to facilitate the drug discovery process.<sup>[14]</sup> In particu-

lar, we tackled the challenge of detecting intended (on-target) and unintended (off-target) effects simultaneously using a complex network of sensors and switches. In summary, a set of sensors are linked to potential off-targets; if there is an off-target effect, at least one of these sensors is triggered. The sensors are linked in a logic circuit that signals if at least one of the off-target sensors is affected. In this way, a large number of off-target effects can be probed simultaneously without increasing the number of readouts (Fig. 2). The same network also reports separately on an on-target effect. The proof of concept was implemented using microRNA drug targets, but the technology can be adapted to additional drug target families. In summary, a single experiment can tell whether a drug candidate has a desired effect without unwanted side effects.

### 4. Outlook

Augmenting living cells with complex information-processing capabilities is an important milestone in biotechnology. This is currently being done using genetically-encoded circuits. Engineering efforts toward artificial cells likewise need to address the challenge of creating computing systems in these cells to provide them with sensory and decision-making capa-

bilities. Artificial cells do not necessarily need to mimic nature’s design principles. This opens new venues for engineering information-processing architectures that are based on organic chemistry or some combination of organic chemistry, biochemistry (e.g. purified enzymes), and genetic encoding. State of the art knowledge in molecular switches and gates can form the foundation on which more complex systems will be built. Enzyme-based logic circuits can offer another suitable enabling technology. In addition, gene expression has been reconstituted in a number of artificial cell settings,<sup>[15]</sup> and this can in principle enable combining genetic material and chemical components. The challenges ahead need not be underestimated. For an information-processing system to operate, it must exist outside of equilibrium and consume energy. This is a given in a living cell where the gene products are constantly recycled and equilibrium means death. Implementing an artificial cell that exists in a steady state away from equilibrium for extended periods of time is a necessary prerequisite for moving forward.

Received: April 4, 2016

- [1] J. Stankiewicz, L. H. Eckardt, *Angew. Chem.-Int. Edit.* **2006**, *45*, 342.  
 [2] a) Y. Benenson, *Nat. Rev. Genet.* **2012**, *13*, 455;  
 b) Y. Benenson, *Molecular Biosystems* **2009**, *5*, 675.

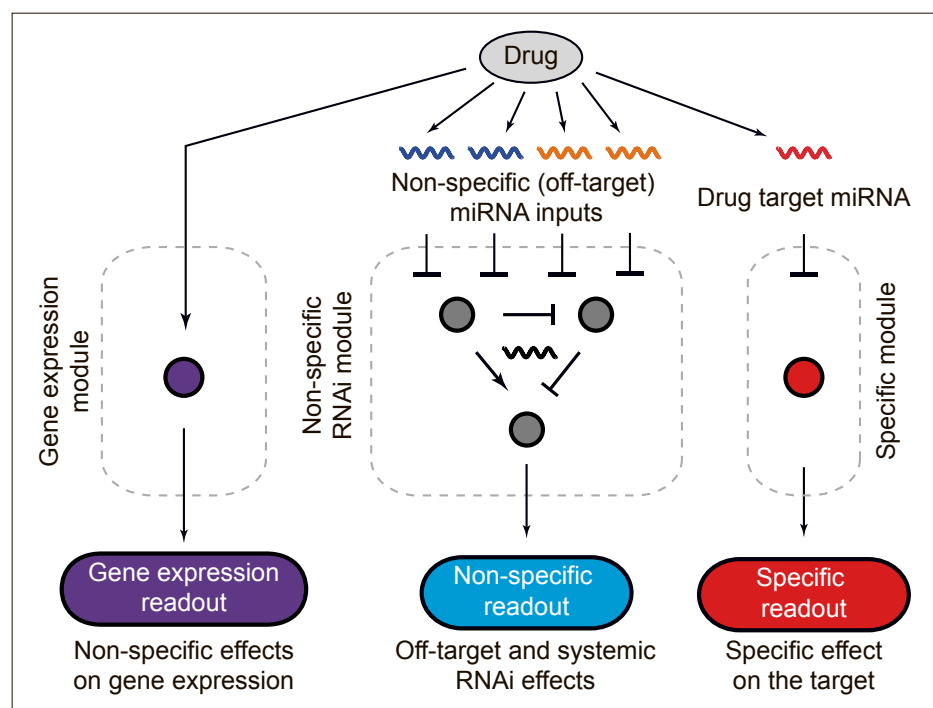


Fig. 2. High-level schematics of a logic circuit for rapid drug discovery. A drug candidate may interact with the intended target (right) or with a number of plausible unintended targets (off-targets, middle). It may also interfere with gene expression in general (left). The unintended targets are integrated in a logic circuit that generates high readout when the off-targets are in their unperturbed state. If at least one of the off-targets is affected, this readout will change, either down- or upward. Reproduced from B. Haefliger, L. Prochazka, B. Angelici, Y. Benenson, *Nat. Commun.* **2016**, *7*, 12.<sup>[14]</sup>

- [3] A. M. Turing, *Proc. London Math. Soc.* **1937**, 42, 230.
- [4] D. Soloveichik, M. Cook, E. Winfree, J. Bruck, *Natural Computing* **2008**, 7, 615.
- [5] A. E. Mayo, Y. Setty, S. Shavit, A. Zaslaver, U. Alon, *PLoS. Biol.* **2006**, 4, 555.
- [6] A. P. De Silva, H. Q. N. Gunaratne, C. P. McCoy, *Nature* **1993**, 364, 42.
- [7] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, E. Shapiro, *Nature* **2001**, 414, 430.
- [8] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, E. Shapiro, *Nature* **2004**, 429, 423.
- [9] M. N. Stojanovic, D. Stefanovic, *Nat. Biotechnol.* **2003**, 21, 1069.
- [10] L. Qian, E. Winfree, J. Bruck, *Nature* **2011**, 475, 368.
- [11] E. Katz, V. Privman, *Chem. Soc. Rev.* **2010**, 39, 1835.
- [12] K. Rinaudo, L. Bleris, R. Maddamsetti, S. Subramanian, R. Weiss, Y. Benenson, *Nat. Biotechnol.* **2007**, 25, 795.
- [13] B. Groves, Y. J. Chen, C. Zurla, S. Pochekailov, J. L. Kirschman, P. J. Santangelo, G. Seelig, *Nat. Nanotechnol.* **2016**, 11, 287.
- [14] B. Haefliger, L. Prochazka, B. Angelici, Y. Benenson, *Nat. Commun.* **2016**, 7, 12.
- [15] S. Nomura, K. Tsumoto, T. Hamada, K. Akiyoshi, Y. Nakatani, K. Yoshikawa, *ChemBioChem* **2003**, 4, 1172.