ACRE: Absolute concentration robustness exploration in module-based combinatorial networks

Hiroyuki Kuwahara,[1,2,*] Ramzan Umarov,[1,2] Islam Almasri,[2] and Xin Gao[1,2,*]

[1]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia and [2]Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia

*Corresponding author: E-mail: hiro.kuwahara@kaust.edu.sa; xin.gao@kaust.edu.sa

Abstract

To engineer cells for industrial-scale application, a deep understanding of how to design molecular control mechanisms to tightly maintain functional stability under various fluctuations is crucial. Absolute concentration robustness (ACR) is a category of robustness in reaction network models in which the steady-state concentration of a molecular species is guaranteed to be invariant even with perturbations in the other molecular species in the network. Here, we introduce a software tool, absolute concentration robustness explorer (ACRE), which efficiently explores combinatorial biochemical networks for the ACR property. ACRE has a user-friendly interface, and it can facilitate efficient analysis of key structural features that guarantee the presence and the absence of the ACR property from combinatorial networks. Such analysis is expected to be useful in synthetic biology as it can increase our understanding of how to design molecular mechanisms to tightly control the concentration of molecular species. ACRE is freely available at https://github.com/ramzan1990/ACRE.

Key words: biological design principle; combinatorial network analysis; concentration robustness; chemical reaction network theory.

Introduction

A fundamental feature of living systems is their robustness to various internal and external stimuli at multiple organizational levels (Kitano 2004). At the cellular level, the molecular mechanisms to tightly control the steady-state concentration of certain molecular species under a range of perturbations are an important feature that is required for homeostasis and adaptive responses to environmental stimuli (Barkai and Leibler 1997). Such biological robustness features to maintain functional stability under environmental stress are essential to the engineering of microorganisms for industrial-scale applications (Abdel-Banat et al. 2010; Zhu et al. 2012; Jia et al. 2016). Thus, a better understanding of how to design molecular control mechanisms

to achieve high-concentration robustness to various fluctuations is key to the success of synthetic biology.

While a traditional approach to investigating such molecular control mechanisms is based on genetic and biochemical perturbations, synthetic biology offers a complementary approach from the perspective of 'understanding by building' (Sprinzak and Elowitz 2005). In this approach, by exploiting the modular nature of biological components (Hartwell et al. 1999; Khosla and Harbury 2001; Ihmels et al. 2002; Ravasz et al. 2002; Alon 2006, 2007), synthetic biologists can construct combinatorial circuits through rewiring of biological parts and explore the design search space for a given functional specification (Guet et al. 2002; Sprinzak and Elowitz 2005; Lim 2010; Peisajovich et al. 2010; Wang and Buck 2012). Unlike traditional approaches that

focus on one specific biological network instance, thus, this combinatorial approach allows researchers to explore many different networks and interrogate the core rules for the design of molecular systems for a given function. Importantly, the integration of theoretical analysis through the use of kinetic modeling tools in synthetic biology experiments can guide and streamline this exploratory approach by efficiently enumerating combinatorial networks and providing essential information about the structure-function mapping (Ma et al. 2009; Soyer et al. 2009; Gunawardena 2010; Chau et al. 2012; Nielsen et al. 2016; Roquet et al. 2016). Such integrative approaches can also help improve the current theoretical understanding of biological systems when there are discrepancies between the results from experiments and those from a corresponding model.

A strong class of concentration robustness is called *absolute concentration robustness* (ACR) where the steady-state concentration of a molecular species in a kinetic model remains invariant to perturbations in the other species. Recently, Shinar and Feinberg proposed a theorem that gives a sufficient structural condition to determining ACR in a broad class of mass-action reaction networks (Shinar and Feinberg 2010). This theorem provides a powerful statement since it treats ACR as a structural property, which emerges independent of the underlying kinetic parameter settings. Hence, unlike simulation-based analysis—in which ACR can be checked only for kinetic models with limited subsets of parameter combinations—this theorem enables formulation of a stronger class of ACR whereby the network structure itself guarantees the presence of strong concentration robustness.

In this article, we introduce a software tool, absolute concentration robustness explorer (ACRE), that applies Shinar and Feinberg's theorem to combinatorial biological networks for systems-level analysis of the concentration robustness. ACRE has a user-friendly interface and can automatically and efficiently explore a large number of mass-action reaction systems to search for network structures with the ACR property. By facilitating a platform to conveniently analyze essential topological features that underpin the machinery of the concentration robustness, ACRE is expected to be a useful tool to uncover general principles to design molecular control mechanisms for robust biological systems.

2. Shinar and Feinberg's theorem

To illustrate the ACR property, we consider the following simple two-species mass-action reaction system (Shinar and Feinberg 2010):

$$A + B \xrightarrow{k_1} 2B,$$

$$B \xrightarrow{k_2} A,$$
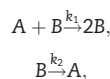
where the time evolution of the system is governed by

$$\frac{dx_A}{dt} = k_2 x_B - k_1 x_A x_B$$

$$\frac{dx_B}{dt} = -\frac{dx_A}{dt}$$

with x_A and x_B being the concentration of A and B, respectively. By letting x_T be the total concentration of A and B, that is,

$x_T = x_A(0) + x_B(0)$, the positive steady-state solution of this system becomes:

$$x_A^{ss} = \frac{k_2}{k_1}$$

$$x_B^{ss} = x_T - \frac{k_2}{k_1},$$

where the superscript *ss* indicates the steady-state solution. Here, while the positive steady-state solution of B depends on the total concentration of A and B, that of A only depends on parameters k_1 and k_2 and is invariant of the changes in the concentration of A and B. Thus, this system has ACR in species A.

Shinar and Feinberg showed that the question of finding this ACR property in some mass-action network models can be formulated within the framework of chemical reaction network theory (Shinar and Feinberg, 2010). The chemical reaction network theory has been used to connect the structure of mass-action networks to dynamical properties such as multistability (Siegal-Gaskins et al. 2011). Similarly, Shinar and Feinberg's theory can examine structural conditions that can guarantee the presence of the ACR property. Here, we provide a very brief introduction of the theory. Interested readers are referred to Shinar and Feinberg (2010) for further information about the theory. Shinar and Feinberg's theory states that, for a given mass-action network model that has positive steady state with its network deficiency being one, if there are two non-terminals that differ only in molecular species S, then this network has ACR in S. Here, non-terminals are a certain class of linear combinations of species based on participation in reactions, while the deficiency of a network can be defined as a non-negative integer which represents the dimension of a certain vector space (Gunawardena 2003).

3. ACRE

3.1 Function of ACRE

ACRE allows the user to conveniently construct combinatorial reaction networks and determine which of the networks have the ACR property. To generate combinatorial reaction networks using ACRE, the user specifies network building units and how they are connected. With this information, ACRE enumerates all combinatorial biochemical reaction networks and determines which of the networks have structural features for the ACR property by applying Shinar and Feinberg's theorem. It can also identify some closed systems that are not able to possess the ACR property by checking the deficiency of those networks (Shinar et al. 2009).

3.3 Workflow of ACRE

Figure 1 shows the workflow of ACRE. To use ACRE for the analysis of ACR property in combinatorial networks, the user first loads a set of reactions, which we call a *module*, in the System Biology Markup Language (SBML) format, a standardized format for encoding biological models (Hucka et al. 2003). The user specifies each module by defining its input ports, output ports, and reaction structure. User-specified modules can then be used as network building blocks to compose combinatorial networks. To have the structure of a module, ACRE requires each reaction in an SBML file to have information about its reactants,
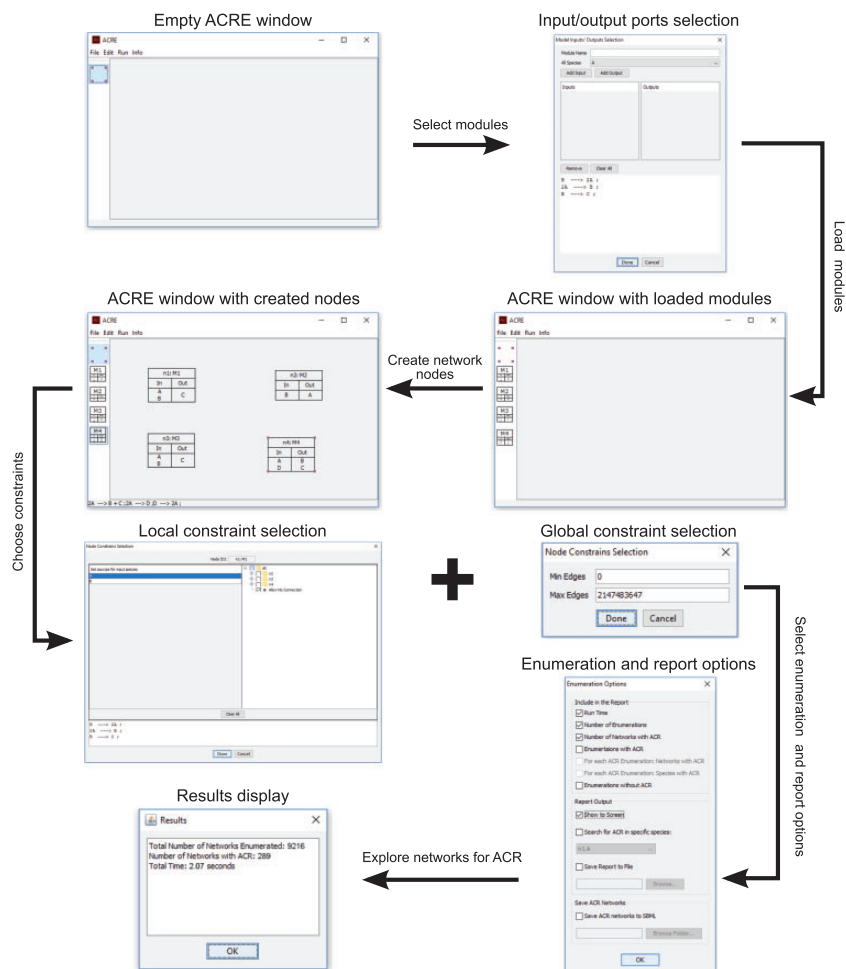
Figure 1. An illustration of the workflow of ACRE.

products, and reaction stoichiometry. The user can find in the 'Modules' directory under the ACRE installation home a number of SBML files that can be used as modules, including those used in the user manual and in this article.

After a module is successfully loaded into the ACRE tool, the tool notifies the user if its structure has the ACR property. After this check, the user specifies the input and output ports of the module. An input port of a module specifies a molecular species that can be replaced, while an output port of a module specifies a molecular species that can replace another species. After the selection of the ports, a rectangular icon representing the module is shown on the left pane. If the user wishes to specify different input and output ports for the same reaction structure, the user can load the same SBML file multiple times and construct different modules with different sets of input and output ports.

To construct combinatorial networks, the user creates instances of these modules, which we call *nodes*. To create a node of a given module, the user first highlights the module on the left pane by clicking on it, and then the user clicks any space in the workspace, which displays a rectangular node showing its name, module, and input and output ports. After creating nodes, the user can specifies two types of constraints, *local constraints* and *global constraints*. To display the window to specify the local constraint of a node, the user can double-click its rectangular icon. ACRE enumerates all possible networks with the created nodes under these user-specified constraints. The local

constraint is defined at the node level, and it is used to constrain mapping by defining which output ports can replace which input ports. By default, the local constraint sets each input port to have 'no connection', which means that the molecular species specified in an input port is not replaced. The global constraint is, on the other hand, defined at the network level. To specify the global constraints, the user can select it under the 'Edit' menu. The user can specify the minimum and maximum number of *edges* that each generated network can possess. Here, edges are defined to be mapping connections from output ports to input ports. The global constraint can prevent the construction of over-sparse or over-dense networks, and it can also limit the search space of all possible networks in order to achieve higher computational efficiency. For convenience, ACRE has a feature to save and load workspace, allowing the user to reuse nodes and constraints easily.

For each of the generated networks, ACRE analyzes whether or not it has ACR. For the ACR analysis, the user can select which pieces of the information to be included in a report. The selection includes information about runtime, the number of constructed networks, and the number of networks with ACR. The report can also show, for each network with ACR, which species has the ACR property. Furthermore, the tool allows the user to save combinatorial networks with ACR in SBML. After choosing these options, the user can start the ACR analysis of the combinatorial networks by clicking the 'Run' button.
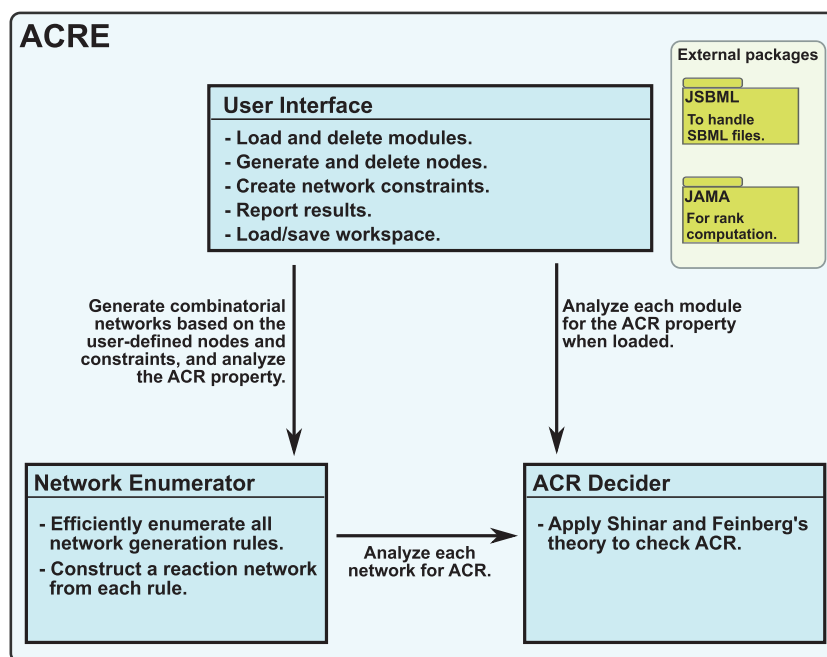
Figure 2. A high-level illustration of the software architecture of ACRE.
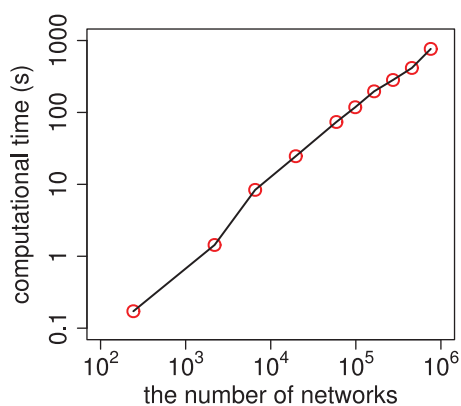

Figure 3. Computational performance of ACRE. Computational time (in seconds) to check the presence of ACR is shown as a function of the cardinality of combinatorial networks.

3.3 Software design

ACRE is a Java-based system (JDK 1.7) that has been tested on various runtime environments, including Windows-, Mac-, and Ubuntu-based machines. Figure 2 illustrates a high-level architecture of ACRE. From the software design point of view, ACRE can be divided into three components: user interface; network enumerator; and ACR decider. User interface allows the user to input necessary information about network building blocks (i.e., nodes) and network connection rules (i.e., constraints) and to run ACR analysis on all combinatorial reaction networks. Network enumerator uses the nodes and network constraints to generate combinatorial reaction networks, each of which is analyzed for the ACR property. ACR decider takes a reaction network as input and checks if its network structure alone is sufficient for the ACR property. In ACRE, we utilized two external Java packages: JSBML (Dräger et al. 2011) and JAMA (http://math.nist.gov/javanumerics/jama/).

In user interface, we used JSBML to parse and generate SBML files, while in ACR decider, we used JAMA to perform matrix computations.

3.4 Computational performance

The scalability is crucial for software tools that analyze properties of combinatorial networks. To evaluate the computational performance of ACRE, we measured computation time required for the ACR analysis of combinatorial reaction networks with given cardinality. To this end, we changed the network constraints to have 10 different numbers of combinatorial networks. We found that, on our testing computational environment, it took about 8 seconds, 118 seconds, and 765 seconds to analyze combinatorial networks of the size 6,561, 98,415, and 759,375, respectively (Fig. 3). While the actual computation time depends on a specific machine used for the runtime environment of ACRE, our analysis shows that the computation time increases only linearly with respect to the cardinality of the combinatorial networks.

3.5 Combinatorial circuit design example

To illustrate the use of ACRE, we consider combinatorial signaling circuits. In signaling circuits, network modules can be allosteric enzyme reactions based on a combination of well-characterized catalytic and regulatory domains (Khosla and Harbury 2001; Kiel et al. 2010; Peisajovich et al. 2010; Mayer 2015). By recombining catalytic and regulatory domains of signaling proteins in the yeast pheromone-response pathway, e.g., synthetic biologists have generated combinatorial signaling circuits with diverse mating response dynamics (Lim 2010; Peisajovich et al. 2010). A recent theoretical study has used a small number of such signaling circuits with allosteric proteins to analyze structural and parametric constraints that can give rise to different types of robustness (Dexter et al. 2015). In this study, we used allosteric regulation reaction for signaling
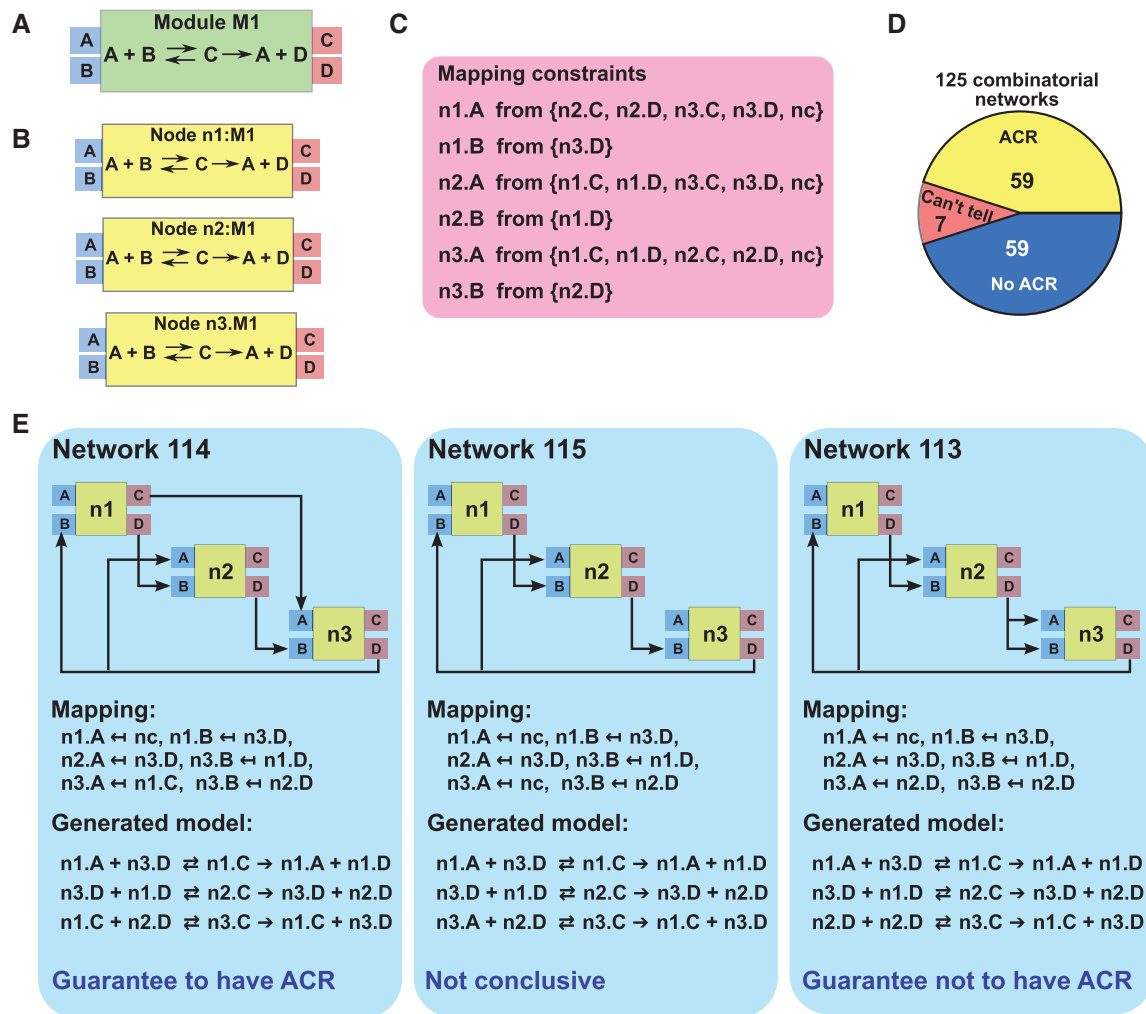
Figure 4. An example of ACR analysis for combinatorial signaling circuits. (A) Network module. (B) Network nodes used to generate combinatorial reaction networks. (C) The local constraints. (D) Allocation of different types of outcomes from the ACR analysis of the 125 circuits. (E) Sample networks for the three different types of the outcome.

proteins as the network module to generate combinatorial signaling circuits for the ACR analysis.

Figure 4A shows our network module M1, whose reaction structure can be loaded from an SBML file *signaling_module.xml* in the 'Modules' subdirectory under the ACRE installation home. In this module, molecular species A and B are defined as the input ports, while molecular species C and D are defined as the output ports. In M1, species A acts as an enzyme to catalyze the functional state transition of a signaling species from B to D, while species C is an intermediate complex form of species A and B.

From M1, we created three nodes, n1, n2, and n3 (Fig. 4B), and we chose to have local constraints to create feedback loop in all of the combinatorial circuits. To this end, we set the input port B to be connected with the output port D sequentially (e.g., input port B of n2 is connected to output port D of n1). We allowed the input port A of each node to be connected with any output ports of the other nodes and to have the 'no connection' option (Fig. 4C). By using the default values for the global constraints, we generated, in total, 125 signaling circuits from these three nodes and the network constraints. By running the ACR analysis, we found that out of 125 networks, 59 had structural features for the presence of ACR, 59 had structural features for the absence of ACR, and seven were inconclusive based solely

on their structures (Fig. 4D). Remarkably, this shows that over 94% of the combinatorial signaling circuits had structural features that can determine the presence or absence of the ACR property. Figure 4E shows a sample circuit of each of these three outcomes. The computation time of this ACR analysis was only 0.17 seconds. The workspace of this example was saved as *signaling_feedback_ws*, and it can be found in the 'Workspaces' subdirectory under the ACRE installation home.

4. Conclusion

In this article, we have introduced our new software tool, called ACRE, which explores combinatorial biochemical reaction networks and searches for those with structural features to guarantee the ACR property by applying Shinar and Feinberg's (2010) theorem. This tool allows an exploratory approach to connect many reaction network structures to the ACR property, facilitating the analysis of general principles to design molecular control mechanisms for higher functional stability. In this line of theoretical studies, bottom-up approaches to build combinatorial networks have indeed been used to explore structural features for various dynamical properties such as adaptation and stochastic effects (Ma et al. 2009; Soyer et al. 2009; Kuwahara

and Gao 2013). However, due to overwhelming analytical complexity, such approaches are often confined to analysis based on simulations that only consider a relatively small subset of parameter spaces. Compared with such simulation-based approaches, thus, ACRE can give a stronger statement about the connection between reaction networks and their dynamical properties. Given its capacity to streamline the analysis of general design principles for concentration robustness, ACRE is expected to be a valuable tool for synthetic biology. In addition, since the robustness to internal and external fluctuations is an essential feature of living systems, analysis of general principles for concentration robustness enabled by ACRE may facilitate new insights into the evolutionary constraints on a given biological system that drove the selection of one specific network structure over other possible structures capable of exhibiting strong concentration robustness. ACRE uses the Apache 2.0 license, and it can be downloaded at https://github.com/ramzan1990/ACRE.

References

Abdel-Banat, B. M. A. et al. (2010) 'High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast?' *Applied Microbiology and Biotechnology*, 85: 861–7.

Alon, U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits*, vol. 10. Chapman & Hall/CRC.

—— (2007) 'Network motifs: theory and experimental approaches', *Nature Reviews Genetics*, 8: 450–61.

Barkai, N, and Leibler S. (1997) 'Robustness in simple biochemical networks', *Nature*, 387: 913–7.

Chau, A. H. et al. (2012) 'Designing synthetic regulatory networks capable of self-organizing cell polarization', *Cell*, 151: 320–32.

Dexter, J. P., Dasgupta T., and Gunawardena J. (2015) 'Invariants reveal multiple forms of robustness in bifunctional enzyme systems', *Integrative Biology: Quantitative Biosciences from Nano to Macro*, 7: 883–94.

Dräger, A. et al. (2011) 'JSBML: a flexible java library for working with SBML', *Bioinformatics (Oxford, England)*, 27: 2167–8.

Guet, C. C. et al. (2002) 'Combinatorial synthesis of genetic networks', *Science (New York, NY)*, 296: 1466–70.

Gunawardena, J. (2003) *Chemical Reaction Network Theory for In-Silico Biologists* <http://vcp.med.harvard.edu/papers/crnt.pdf > accessed 25 May 2016.

—— (2010) 'Biological systems theory', *Science (New York, NY)*, 328: 581–2.

Hartwell, L. H. et al. (1999) 'From molecular to modular cell biology', *Nature*, 402: C47–52.

Hucka, M. et al. (2003) 'The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models', *Bioinformatics*, 19: 524–31.

Ihmels, J. et al. (2002) 'Revealing modular organization in the yeast transcriptional network', *Nature Genetics*, 31: 370–7.

Jia, H. et al. (2016) 'Intelligent microbial heat-regulating engine (IMHeRE) for improved thermo-robustness and efficiency of bioconversion', *ACS Synthetic Biology*, 5: 312–20.

Khosla, C., and Harbury P. B. (2001) 'Modular enzymes', *Nature*, 409: 247–52.

Kiel, C., Yus E., and Serrano L. (2010) 'Engineering signal transduction pathways', *Cell*, 140: 33–47.

Kitano, H. (2004) 'Biological robustness', *Nature Reviews Genetics*, 5: 826–37.

Kuwahara, H., and Gao X. (2013) 'Stochastic effects as a force to increase the complexity of signaling networks', *Scientific Reports*, 3: 2297.

Lim, W. A. (2010) 'Designing customized cell signalling circuits', *Nature Reviews Molecular Cell Biology*, 11: 393–03.

Ma, W. et al. (2009) 'Defining network topologies that can achieve biochemical adaptation', *Cell*, 138: 760–73.

Mayer, B. J. (2015) 'The discovery of modular binding domains: building blocks of cell signalling', *Nature Reviews Molecular Cell Biology*, 16: 691–8.

Nielsen, A. A. K. et al. (2016) 'Genetic circuit design automation', *Science (New York, NY)*, 352: aac7341.

Peisajovich, S. G. et al. (2010) 'Rapid diversification of cell signaling phenotypes by modular domain recombination', *Science*, 328: 368–72.

Ravasz, E. et al. (2002) 'Hierarchical organization of modularity in metabolic networks', *Science*, 297: 1551–5.

Roquet, N. et al. (2016) 'Synthetic recombinase-based state machines in living cells', *Science (New York, NY)*, 353: aad8559.

Shinar, G., Alon U., and Feinberg M. (2009) 'Sensitivity and robustness in chemical reaction networks', *SIAM Journal on Applied Mathematics*, 69: 977–98.

——, ——, and —— (2010) 'Structural sources of robustness in biochemical reaction networks', *Science* 327: 1389–91.

Siegal-Gaskins, D. et al. (2011) 'Emergence of switch-like behavior in a large family of simple biochemical networks', *PLoS Computational Biology*, 7: e1002039.

Soyer, O. S., Kuwahara H., and Csiksz-Nagy A. (2009) 'Regulating the total level of a signaling protein can vary its dynamics in a range from switch like ultrasensitivity to adaptive responses', *FEBS Journal*, 276: 3290–8.

Sprinzak, D., and Elowitz M. B. (2005) 'Reconstruction of genetic circuits', *Nature*, 438: 443–8.

Wang, B., and Buck M. (2012) 'Customizing cell signaling using engineered genetic logic circuits', *Trends in Microbiology*, 20: 376–84.

Zhu, L. et al. (2012) 'Engineering the robustness of industrial microbes through synthetic biology'. *Trends in Microbiology*, 20: 94–101.