Synthetic Biology: Putting Engineering into Bioengineering

2 authors, including:

Matthias Heinemann
University of Groningen
210 PUBLICATIONS **8,619** CITATIONS

Some of the authors of this publication are also working on these related projects:

Development of RNA-based biosensors for intracellular metabolites View project

Synthetic biology – putting engineering into biology

Matthias Heinemann[1][$], Sven Panke[*][1]

[1]ETH Zurich, Bioprocess Laboratory, Institute of Process Engineering, Universitätsstrasse 6, 8092 Zurich, Switzerland,
Phone: +41 44 6320413, Email: panke@ipe.mavt.ethz.ch

Associate Editor: Jonathan Wren

ABSTRACT

Synthetic biology is interpreted as the engineering-driven building of increasingly complex biological entities for novel applications. Encouraged by progress in the design of artificial gene networks, *de novo* DNA synthesis and protein engineering, we review the case for this emerging discipline. Key aspects of an engineering approach are purpose-orientation, deep insight into the underlying scientific principles, a hierarchy of abstraction including suitable interfaces between and within the levels of the hierarchy, standardization, and the separation of design and fabrication. Synthetic biology investigates possibilities to implement these requirements into the process of engineering biological systems. This is illustrated on the DNA level by the implementation of engineering-inspired artificial operations such as toggle switching, oscillating, or production of spatial patterns. On the protein level, the functionally self-contained domain structure of a number of proteins suggests possibilities for essentially Lego-like recombination which can be exploited for reprogramming DNA binding domain specificities or signaling pathways. Alternatively, computational design emerges to rationally reprogram enzyme function. Finally, the increasing facility of *de novo* DNA synthesis - synthetic biology's system fabrication process - supplies the possibility to implement novel designs for ever more complex systems. Some of these elements have merged to realize the first tangible synthetic biology applications in the area of manufacturing of pharmaceutical compounds.

1   INTRODUCTION

The advent of systems biology, the steady development of foundational technologies such as *de novo* DNA synthesis (Tian, et al., 2004), milestone experiments such as the computational re-design of enzymes (Dwyer, et al., 2004), the opportunity to widely recombine zinc fingers to re-program DNA-binding site specificity (Dreier, 2001) and the availability of well-studied model regulatory systems for the design of engineering-inspired molecular devices provide a very powerful knowledge and technology basis for building novel biological entities. Encouraging applications come from such diverse areas as the design of artificial gene networks (Sprinzak and Elowitz, 2005), the refactoring of small genomes (Chan, et al., 2005), the reprogramming of signaling pathways (Dueber, et al., 2003), or metabolic engineering (Martin, et al., 2003) and have been referred to as "synthetic biology". These examples entertain the notion that adaptation and assembly of functionally self-contained parts such as promoters, ribosome bind-ing sites, coding sequences, terminators, or protein domains is a promising way to re-constitute existing or to produce novel, biological entities. To extend such an approach to the many opportunities typically associated with the application of biological systems, we need to establish a number of rules. Building novel parts, devices, and in particular complex systems will require a systematic approach that relies on modularity and abstraction at various cellular levels in order to be useful to a broad group of biotechnologists. It will also require a technology to fabricate the designs, more precisely the DNA encoding the system. This is synonymous to following an engineering approach, even if biology is in many aspects not understood well enough to consider it a sufficient knowledge base for an engineering discipline.

This review will discuss prominent works that have helped launch this engineering vision and will attempt – in one review – to touch upon all the fields that we perceive as the currently most relevant to synthetic biology, spanning the entire trajectory from fabrication via design to applications on protein, gene networks, and systems level. Several separate aspects of these fields have recently been reviewed (e.g. (Andrianantoandro, et al., 2006; Benner and Sismour, 2005; Endy, 2005; Forster and Church, 2006; Hasty, et al., 2002; Kærn, et al., 2003; McDaniel and Weiss, 2005; Sprinzak and Elowitz, 2005). However, we reason that it is important to cover all major aspects of the engineering vision in one review to give emphasis to synthetic biology's rather comprehensive claim of re-organizing the bioengineering endeavor. After a brief discussion on the fundamental characteristics of an engineering approach, we bring together prominent examples from these aspects and finally discuss the requirements and implications for bioinformatics.

2   TOWARDS ENGINEERING OF BIOLOGY

Engineers' efforts in the field of engineering biology are furnished with only a few success stories. This reflects the fact that the ability to engineer biology in a directed and successful manner is still rather limited today and as a consequence, the complexity of things we can efficiently make is still quite small. Synthetic biology with its engineering vision aims to overcome the existing fundamental inabilities in system design and system fabrication, by developing foundational principles and technologies to ultimately enable a systematic forward-engineering of (parts of) biological systems for improved and novel applications (Fig. 1). Conceptually, this needs to encompass several characteristic features of other established engineering disciplines (Endy, 2005):

[i] In non-biology related areas, engineers can usually draw on a sound knowledge base. In-depth understanding (sometimes down to the first principles) permits computer-based design of new sys-

[*]To whom correspondence should be addressed.

[$]New address: ETH Zurich, Institute of Molecular Systems Biology, Wolfgang-Pauli-Str. 16, 8093 Zurich, Switzerland
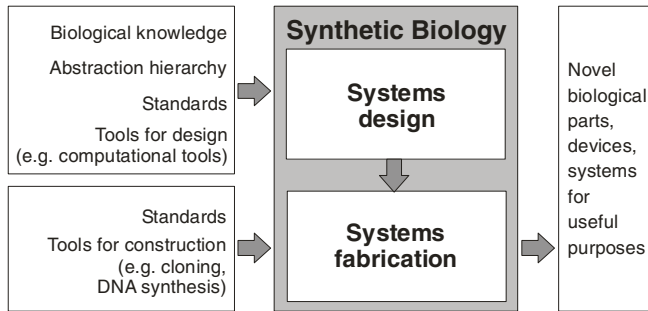
Biological knowledge	**Synthetic Biology**	Novel
Abstraction hierarchy		biological
Standards	**Systems**	parts,
	design	devices,
Tools for design		systems
(e.g. computational tools)		for
		useful
Standards	**Systems**	purposes
Tools for construction	**fabrication**	
(e.g. cloning,		
DNA synthesis)		

Fig. 1. Synthetic Biology encompasses systems design and fabrication. Each part has its specific prerequisites and inputs. Ultimately, Synthetic Biology will deliver novel biological entities with improved functionality.

tems by going through iterations between computer models and simulations. This allows an extensive *in silico* testing of new design variants.

[ii] The application of abstraction hierarchies is a common characteristic in almost every engineering endeavor and results in several practical advantages. First, the introduction of system boundaries basically hides information and is thus a way to manage complexity. In other words, abstraction is useful as it allows individuals to work independently at each level of the hierarchy. It is an organizational prerequisite for combining parts into complex systems.

[iii] In order to guarantee "plug-and-play" compatibility of different components in an abstraction hierarchy, the connections between the different parts need to be defined, i.e. standardization is required. Only broadly implemented standards for the components themselves and their in- and output criteria would ensure that interfaces between components fit, even when they are designed and fabricated by different laboratories or, ultimately, different companies.

[iv] Finally, another characteristic feature of "true" engineering is the decoupling of the design process from the actual fabrication of components or systems. Both tasks require a distinct set of skills and expertise, which is typically not provided by the same individuals. However, as a consequence this requires that the design engineers also need to have a sound knowledge about how things are actually produced and how parts are assembled. In other words, the design for an object is useless if it has been designed in such a way that no possibility exists to fabricate it.

We argue that design in biology along these engineering characteristics is a realistic goal, to a sufficiently promising extent, for the following reasons:

[i] There is a ubiquitous tendency to move from qualitative to quantitative analyses in biological sciences. This starts to provide on a large scale the data sets that are needed to understand and structure biological complexity, for example by means of mathematical models. Even if the complexity we can manage does not currently go much beyond small artificial gene networks, whose DNA-basis can be designed relatively independently from the remaining cellular environment, our ever increasing knowledge and system understanding will allow us to extend our realm of design further and further.

[ii] The introduction of system boundaries in living systems is of course difficult as information flow relies on the diffusion of mole-

cules through the cytoplasm. This entails all sorts of cross-activities and complicates the design of functionally self-contained elements. Furthermore, many cellular elements have more than one function. On the other hand, we are able to engineer excellent specificities in small molecule-protein, DNA-protein, and protein-protein interactions that help to functionally separate different elements. Furthermore, work on orthogonal (i.e. insulated) translational systems in general is only about to start, and first results are encouraging (Isaacs, et al., 2004; Rackham and Chin, 2005). Finally, there are strong ongoing efforts towards minimal (bacterial) systems and it can be expected that such systems – owing to their reduced complexity – have a much smaller number of cross-reactions, so that implementation of novel elements stands a much better chance of remaining functionally isolated. In summary, we reason that even though proper system boundaries might be difficult to implement in current biological systems, there is good indication that these systems can be sufficiently adapted to allow this in the near future.

[iii] The wide-spread implementation of standards in biological design has never been a serious issue in biotechnology. This fact resulted in a large diversity in bioengineering tools (strains, plasmids, expression systems, etc) which hampers the wide distribution of biological parts across laboratories. Current attempts to implement standards (see below) have the potential to fundamentally change this and have already helped to enable the operation of large projects such as the iGEM2006 competition (http://parts.mit.edu), where more than 30 projects all over the world rely on and provide parts for one parts warehouse (see below).

[iv] Finally, the biological equivalent to fabrication is genomic scale *de novo* DNA synthesis, which - as we will discuss below - is increasingly possible already today.

We have to point out that there is a fundamental difference between engineering biology and engineering in other natural sciences such as chemistry or physics: biological systems have the capacity to replicate and to evolve. This fundamental characteristic will interfere at least with the long-term stability of a number of designed systems and will require constant monitoring of the integrity (of crucial parts) of the systems. As the ultimate goal will be to build complex systems into specific hosts, it is likely that interference from mutations becomes a serious issue. However, it is important to point out that the problem as such is not new to biotechnology – rather, it has accompanied every major strain development effort in industrial fermentation and has contributed to the development of appropriate selection programs for sufficiently stable strains and suitable strain storage routines.

In summary, we see synthetic biology as a novel, true engineering approach to conducting biotechnology. In synthetic biology, we would include endeavors that exploit at least some of the important elements of an engineering approach, as discussed, in their methodology. For example, *de novo* DNA synthesizing an entire pathway, which is composed of genes from various species and includes optimized codon usage, adapted secondary mRNA structure, tailored regulatory elements, and a restriction site strategy that allows the modular replacement of specific genes by improved versions, can be considered a synthetic biology approach to a metabolic engineering problem. Alternatively, the forward-engineering design of genetic modules that can be freely combined

to program gene regulation can be considered a synthetic biology approach to genetic engineering.

3 TOOLS AND ENABLING TECHNOLOGIES FOR SYNTHETIC BIOLOGY - FABRICATION

Synthetic biology aims to follow the standard engineering protocol of design and fabrication. We will discuss first, the status of a number of items related to the issue of fabrication, including standardized cloning, *de novo* DNA synthesis and work on minimizing genomes.

3.1 Standardized cloning

Today, traditional cloning represents the most important fabrication tool in synthetic biology. Using standard, typically PCR-based technologies, an existing DNA fragment is adapted for a novel purpose and inserted into one of a myriad of different cloning vectors according to the specifications of the multi-cloning site. Insertion of alternative elements from vectors with different restriction site structures (for example from different laboratories), introduction of point mutations for novel restriction sites or adaptation of codon usage, etc, all require several rounds of manipulations. In other words, the current system fabrication process is laborious and inefficient.

An important initial step to improve the situation could be the introduction of a standardized vector format which lends itself easily to assembly and allows interoperability of the assembled sequences, such as the set of "idempotent vectors" (https://dspace.mit.edu/bitstream/1721.1/21168/1/biobricks.pdf) or the NOMAD technology (Rebatchouk, et al., 1996). Here, vectors are designed in such a way that insertion of a DNA fragment into a vector recreates exactly the same restriction site architecture. This way, multiple rounds of insertion on either side of an insert are easily possible.

Through the implementation of such standards, rapid exchange of parts is made possible and in fact realized by the "Registry of Standard Biological Parts" of the Massachusetts Institute of Technology working group on synthetic biology (http://parts.mit.edu). The registry intends to serve as a source of well characterized parts that have been formatted according to specific rules, to make assembly easy. Furthermore, the organization of the registry into basic DNA parts such as promoters, ribosome binding sites (RBS), and coding sequences sparks the hope that it will become much easier to obtain suitable genetic elements for a specific design purpose, so that ultimately desired experimental outcomes can be achieved much faster.

Of course, the ideas for standardization need to remain open for optimization and will hopefully include in the future standardization of strains as well as vectors and directions on which suitable cellular parameters should be quantified by which standardized protocol. Here, it should be pointed out, that in many cases the means to rapidly quantify the important connections and interactions between parts hardly exist, not even for such a simple concept as promoter strength. In other words, to be successful as an engineering discipline, synthetic biology will need to repeat the corresponding developments of its sister engineering disciplines that have lead to highly organized fields such as mechanical and electrical engineering.

3.2 *De novo* DNA synthesis

Even after the potential improvements in standardization and organization discussed above, parts and especially system assembly will remain a major issue in any synthetic biology project. The question of how to produce physical instances of systems requiring tens (existing gene networks) or even hundreds (genome reprogramming) of kilobases of novel DNA sequence is basically, still, unsolved. It will be imperative to provide access to *de novo* synthesized DNA sequences. This technology will allow assembly of all the desired changes from promoter strength to codon usage directly into a novel sequence. This sequence should then be available within weeks or even days after its design.

DNA *de novo* synthesis is currently performed by assembling overlapping short (25 to 70 bp long) and chemically synthesized oligonucleotides into longer DNA fragments in a PCR-based assembly process (Stemmer, et al., 1995). This technology has already led to the complete reconstruction of some smaller phage genomes such as the polio virus (Cello, et al., 2002). However, it suffers from two major cost entries: the costs of synthesizing the DNA oligonucleotides by standard phosphoamidite chemistry and the limitations in the accuracy of the chemical synthesis (Young and Dong, 2004). For example, the *de novo* reconstruction of the phage ΦX174 genome relied on a two-fold selection process: [i] oligonucleotides were first gel-purified to ensure the correct length of the set and [ii] correctly reassembled DNA was recovered from plaques after transformation, effectively providing a positive selection strategy for biological function (Smith, et al., 2003). Cost and accuracy associated with the current method will have to be addressed before *de novo* DNA synthesis can become the routine technology that will be required to fabricate more complex biological systems.

Research to optimize the current procedure follows three paths: [i] miniaturizing oligonucleotide production; [ii] optimizing sequence verification costs; and [iii] eliminating "false" oligonucleotides by enzymatic and/or hybridization methods.

[i] Regarding miniaturization, a promising, though not yet fully developed, option is the utilization of microfluidics-based arrays for synthesis. Exploiting the small scale should lead not only to a reduction in the materials costs but also allow optimized reaction conditions and thus reduce error frequencies (Zhou, et al., 2004). Alternatively, oligonucleotide synthesis has been miniaturized on photo-programmable chips (Tian, et al., 2004).

[ii] Due to the inherently error-prone chemical DNA synthesis process, sequence verification of the novel fragment is important. As the error-frequency is increasing with oligonucleotide length but oligo-assembly is simpler with longer molecules, there should be an optimal oligonucleotide length from which to start assembly. This length is estimated to be 40 bp. From there, intermediary 500 bp "synthons" are assembled, which in turn are assembled into 5 kb fragments (Kodumal, et al., 2004).

[ii] Finally, different techniques can be used to remove the remaining errors: for example, it is highly unlikely that two synthesized complementary oligonucleotides will have mutations in complementary positions, so there will be mismatches. These can serve as entry points for digestion of different enzyme systems (Greger and Kemper, 1998; Smith and Modrich, 1997; Young and Dong, 2004). Alternatively, such erroneous oligonucleotides can be sorted out by having all the oligonucleotides designed to hybridize at one specific temperature and then very carefully control this temperature during hybridization. Oligonucleotides with mis-

matches can then be washed away (Tian, et al., 2004). Combinations of these approaches put the current accuracy at error rates of only 1 in 1400 bp (Tian, et al., 2004).

In summary, if synthesis costs can be reduced by a factor ten from the current prices in the order of 1 US$/bp, it is clear that DNA *de novo* synthesis will become an overwhelming force in systems fabrication. Coupled to community-wide standards on the availability, documentation, characterization, and standardization of parts, there is reasonable perspective to eliminate system fabrication as a prohibitive part in future synthetic biology endeavors.

3.3   Providing engineering chassis

Designed and synthesized DNA segments that encode novel functions need to be implemented into a suitable organism, for the time being, by one of the many available genome engineering techniques (see, for example, (Kolisnychenko, et al., 2002)), or in the future by novel mega-size cloning strategies (Itaya, et al., 2005). As the complexity of existing biological systems is the major problem in implementing synthetic biology's engineering vision, it is desirable to reduce this complexity. One option is to reduce the genome of the host - the chassis - into which the new sequence is implemented, which would eliminate many possibilities for interference. For an intended chemical production with the designed system, this genome reduction will probably be limited to reducing the metabolic capabilities of a strain. A true "minimal genome" - the minimum set of genes that is necessary for a cell to propagate under specific environmental conditions – would be a useful point to start when trying to engage in re-building more complex systems (Forster and Church, 2006).

According to theoretical considerations, growth in the presence of a rich but synthetic and defined medium requires as few as 206 genes, basically comprising the DNA replication, transcriptional, and translational machinery, rudimentary DNA repair functions, protein processing and degradation, cell division, and rudimentary metabolic and energy functions (Gil, et al., 2004). Towards this goal, one can either substantially reduce the relatively large genomes of established model systems and exploit the abundance of molecular biology tools for these model organisms, or work on the already very small genome of other organisms in exchange for the requirement to develop novel molecular biology tools.

Regarding the latter, non-pathogenic *Mesoplasma florum* with very attractive cultivation properties and a genome size of 793 kb, is currently being established as such a chassis. Its genomic sequence has recently become available and molecular biology methods are developed. A similar approach is followed with *Mycoplasma genitalium*, for which extensive data on non-essential genes is available (Glass, et al., 2006).

Regarding the former, a prominent example is *E. coli* whose genome has been reduced in various projects by 6% (Yu, et al., 2002), 8% (Kolisnychenko, et al., 2002), or 15% (Posfai, et al, 2006), without any noticeable effect on the investigated physiological properties and by 30% resulting in defects in cell replication (Hashimoto, et al., 2005). *Bacillus subtilis*' genome has been reduced by 8%, again with only minor effects on physiology (Westers, et al., 2003), confirming the hypothesis that under controlled laboratory conditions a substantial part of a bacterium's genome is indeed dispensable.

4   EXAMPLES FOR SYNTHETIC BIOLOGY ENDEAVORS

The ultimate goal of synthetic biology is the efficient design of biological systems. In a few areas, this has already advanced quite substantially. Here, we will outline examples - at various levels of complexity - of protein, circuit, and system engineering.

4.1   Engineering of proteins

To facilitate the design process, it would be highly desirable to have protein modules that are functionally self-contained and can be freely combined to new functionalities. The domain architecture of many regulatory proteins plays very much in favor of such approaches.

One specific example how novel functionality can be obtained from recombining protein modules is the design of polydactyl zinc finger DNA-binding proteins (Blancafort, et al., 2004). Here, combinations of zinc finger domains provide the sequence specificity of the DNA-binding domain (DBD) by essentially recognizing a subset of three or four nucleotides per zinc finger domain. Such proteins display modularity in two ways: typically, the DBD is functionally independent from the effector domain and zinc finger domains are functionally, relatively independent from each other. Therefore by selecting a set of specific zinc fingers *in silico* one can specify arbitrary sequence specificity for a novel DBD, which can then be coupled to a novel effector domain.

The success rates in this process can be quite attractive. The zinc-finger design can for example be based on a comprehensive table of zinc fingers covering all possible binding DNA-sequences for a single zinc-finger. This table has been compiled by extrapolations from X-ray structures and 5 out of 10 combinations of three zinc-fingers selected from this table showed indeed excellent DNA-binding properties for decamer DNA-sequences (Sera and Uranga, 2002). Even better rates can be achieved when the functional modularity between adjacent zinc-fingers is improved by combinatorial and rational approaches. Currently, the corresponding designs are available for all possible 5'-GNN (Segal, et al., 1999) and 5'-ANN (Dreier, et al., 2001) DNA-target sequences and have been successfully exploited for specific binding of up to octadecamer DNA sequences (Dreier, et al., 2001).

Similar functional reprogramming could also be achieved on the level of protein-protein interactions with signaling proteins. For example, the eukaryotic neuronal N-WASP protein accommodates the substitution of its receiving domains or changes in their affinities, insertion of alternative linker structures between domains, different N-WASP output domains, and even different domain architectures. These changes produced proteins with completely new logical behavior (Dueber, et al., 2003).

Other exploitations of functional modularity in signaling pathways stem from scaffold proteins that recruit a kinase and the kinase's substrate, assemble them in close proximity for phosphorylation, and thus provide specificity (Park, et al., 2003). The domain structure of theses scaffolds allows to recombine scaffolds that share the same "nodes" (a common kinase that interacts with both scaffolds) and thus to redirect signals to novel outputs, such as making osmotolerance a function of the induction of the mating signaling pathway (Park, et al., 2003). In human cell lines, mitogenic signals can be rewired to apoptotic behavior by the same strategy (Howard, et al., 2003).
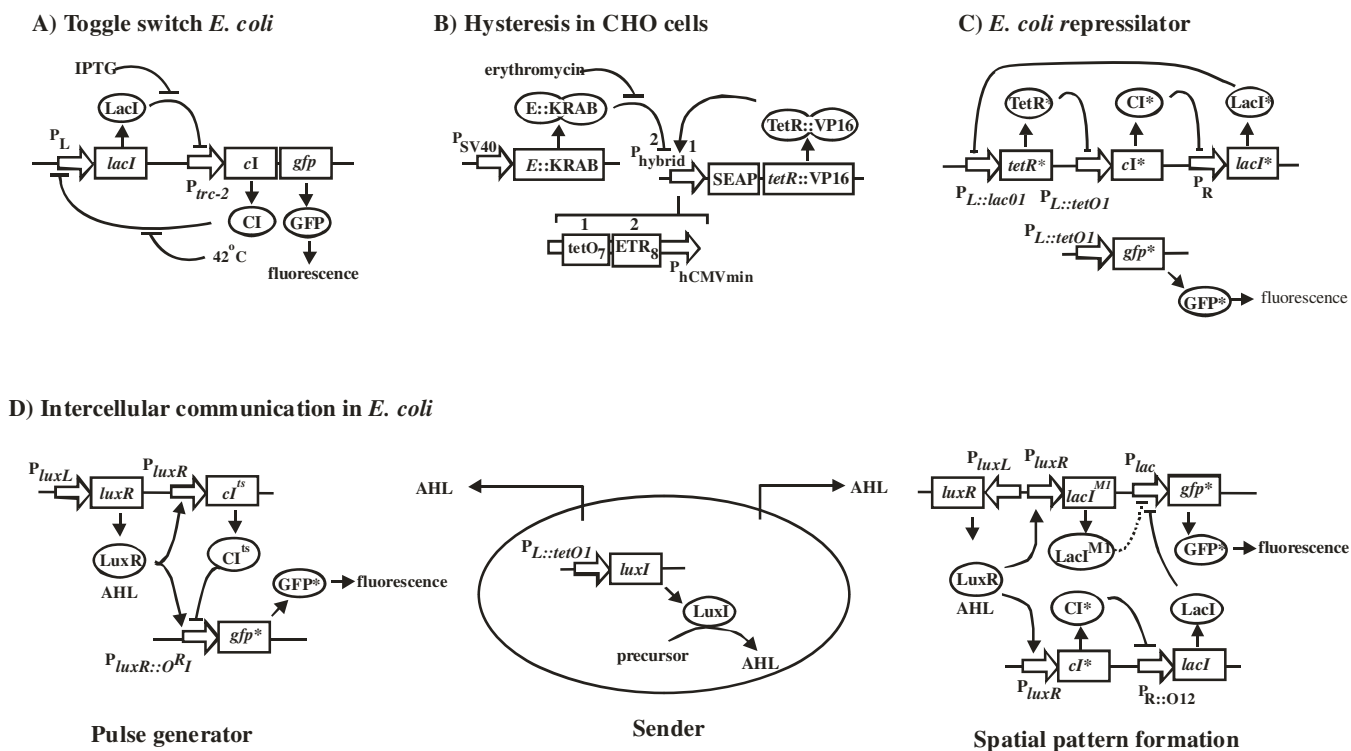
A) Toggle switch *E. coli*

B) Hysteresis in CHO cells

C) *E. coli* repressilator

D) Intercellular communication in *E. coli*

Pulse generator

Sender

Spatial pattern formation

Fig. 2. Artificial genetic networks in *E. coli* and mammalian cells. Boxes represent genes, arrows promoters, boxes on arrows operators. Lines connecting promoters and genes represent connecting DNA that is not further specified here. Different lines represent different replicons. Terminators and ribosome binding sites have not been included. Ovals represent proteins. Lines ending in arrows symbolize activation, lines ending in orthogonal lines represent blocking. Stars indicate proteins (and their corresponding genes) that carry a tag that targets them for accelerated degradation. A) Toggle switch in *E. coli*. IPTG: isopropylthiogalactoside, inactivates LacI; P_L: Left promoter from phage λ, CI-repressible; P_{trc-2}: fusion promoter of promoters from tryptophane- and *lac* operon, LacI-repressible; *gfp*: gene encoding green fluorescent protein. LacI: *E. coli* Lac-repressor. CI^{ts}: temperature sensitive variant of phage λ CI repressor. B) Hysteretic network behaviour in chinese hamster ovary cells: P_{SV40}: simian virus 40 early promoter; $P_{hCMVmin}$: truncated version of the human cytomegalovirus promoter that can be made responsive to the *tet*-repressor; E: *E. coli*'s erythromycin-repressible repressor MphR(A); KRAB (Kruppel-associated box): human transrepressor domain; SEAP: human placental secreted alkaline phosphatase, can be assayed by its color-inducing dephosphorylation activity for p-nitrophenylphophate; TetR: here: DNA binding domain of the *tet*-repressor, interacting with *tet*-operator; VP16: transactivation domain of protein VP16 of *Herpes simplex* virus; tetO7: seven copies of *tet*-operator; ETR8: eight MphR(A)-binding operators. C) *E. coli* repressilator: $P_{L::tetO1}$: fusion of the P_L promoter and the operator site 2 of the Tn10 tetracycline resistance operon, makes the P_L promoter repressible by bacterial TetR; $P_{L::lacO1}$: fusion of the P_L promoter to operator sequences of the *lac* operon, makes the promoter repressible by LacI; P_R: right promoter from phage λ, CI-repressible. D) AHL: acly-homoserinelactone, activates LuxR; P_{luxL}: left promoter for autoregulated *luxR* expression; P_{luxR}: right promoter for *luxI* expression. Requires LuxR activated by AHL for expression; LuxR: regulator for activating transcription from P_{luxR}; $P_{luxR::ORI}$: fusion of P_{luxR} and operator site from P_R, making the promoter repressible by CI; LuxI: enzyme catalyzing formation of AHL; $P_{R::O12}$: truncated version of the λ P_R promoter that lacks the third operator site. LacI^{M1}: Protein identical to LacI but expressed from a codon-modified *lacI^{M1}* to prevent recombination with the second *lacI*. For further details and references, see text.

The above results have to be seen in the light of recent improvements in the algorithms for computational design of proteins that allow for example the rational engineering of novel substrate binding specificities (Looger, et al., 2003) or even the conversion of binding proteins into functional enzymes (Dwyer, et al., 2004). The novel enzyme possessed a catalytic constant in the order of 0.1 s^{-1}, which is at least one to two orders of magnitudes higher than previous comparable efforts (Bolon, et al., 2002). Taken together, the recent successes in modular protein engineering and computational design suggest that rational protein design for many applications will be feasible in the near future.

4.2 Engineering of artificial gene networks

Significant efforts were recently undertaken in the design of artificial genetic networks in prokaryotic and eukaryotic systems. Here,

different genetic elements or "parts" are (ultimately) rationally combined to "devices" to realize specific cellular behaviors that have frequently analogies to elements from electric circuits such as switches and oscillators. We will outline recent efforts in the development of networks with complex behaviors and for cell-cell-communication.

4.2.1 Switches. A switch lets the cell adopt one of two possible states, depending either on the presence or absence of a chemical inducer, or on two separate external stimuli (toggle switch) (Becskei, et al., 2001).

The latter behavior can be easily designed from any two repressors that reciprocally inhibit the transcription of their genes (Fig. 2A) (Gardner, et al., 2000; Kramer, et al., 2004). Switching between states can be achieved by intermittently inactivating the repressor that maintains the current state (such as adding a chemi-

cal inducer or increasing the temperature). Essentially, this property conveys a cell with a memory of its previous cultivation history and thus represents an epigenetic toggle switch.

The former behavior requires positive feedback in the regulatory processes, such as [i] the positive autoregulation of a positive regulator's gene transcription or [ii] the concomitant upregulation of an operon by external inducer and of the gene that encodes the transporter protein for entrance of the inducer. Besides the artificial design of such systems, this behavior is rather common in a number of well-characterized bacterial expression systems such as the bacterial lactose and arabinose systems, (Atkinson, et al., 2003; Khlebnikov, et al., 2001; Ozbudak, et al., 2004; Vilar, et al., 2003).

In addition, the artificial gene networks can be engineered with a hysteretic character, so that the system switches into the "ON" state at a higher concentration of external signal than is required to switch back to the "OFF" state. This requires that the concentration of activator or active repressor can be made a function of the history of the cell, e. g. by adding another regulatory layer on top of the positive feedback element. This can be a concentration-dependent inactivation of a repressor that competes with an activator (Fig. 2B). Depending on the previous state of the cell, a given concentration of active repressor interacts with either high or low concentrations of activators, leading to a differentiation in response depending on the history (Atkinson, et al., 2003; Kramer and Fussenegger, 2005).

4.2.2 Complex networks. In an oscillator, an artificial network is used to produce regular fluctuations in network elements such as reporter proteins. Oscillators have been realized in two ways: as ring-oscillators ("repressilators" (Fig. 2C)) or as a combination of activation and repression elements. The ring oscillator consists of three repressor genes that are coupled to three corresponding promoters in such a way, that each repressor protein can turn off the synthesis of one other repressor protein. This design worked on single cell level, but not on culture-level, which probably has to do with the noise involved on gene expression level (see below) (Elowitz and Leibler, 2000). However, by combining positive and negative regulation, it is possible to reduce the noise to such a degree that population-synchronized oscillation behavior over three periods can be observed in a turbidostat (Atkinson, et al., 2003). Interestingly, such oscillating systems can be extended to include metabolite concentrations (Fung, et al., 2005).

In order to execute ever more complex logical behavior, it will be important to be able to "integrate" more and more signals into determining one or more cellular functions. This is facilitated by the high level of modularity in the regulatory elements of eukaryotic systems. This modularity makes them particularly amenable to design and can be used to implement a wide variety of logical behaviors for two and three signal inputs while exploiting only a limited number of genetic elements (Kramer, et al., 2004).

4.2.3 Networks for intercellular communications. Creating macroscopically observable artificial functional behavior in a cell population requires some kind of synchronization. Such synchronization can be enforced by adding chemical inducers or by letting the cells themselves produce a signal in response to a change in a culture property. One example for such a property is cell density which can be communicated by quorum sensing, for example via the *luxR/luxI* system of *Vibrio fischeri* or via artificially engineered

systems such as acetate accumulation in *E. coli* cultures (Bulter, et al., 2004).

The *luxR/luxI* system has been used to trigger a variety of population-density dependent responses, such as flipping of a toggle switch (Kobayashi, et al., 2004) or programmed population control (You, et al., 2004). The system has also been exploited to design spatial patterns of behavior that re-build aspects of multicellular systems (Fig. 2D): When producer cells send the autoinducer signal of the *lux* system via diffusion through a plate, cells at different distances from the senders experience differently steep gradients once the autoinducer reaches them. Alternatively, cells can be used to detect the differences in inducer concentration in resulting (quasi-)steady state. Networks can be designed which are able to detect these rather subtle differences in environmental conditions and which translate them into adequate cellular responses such as different pulses of reporter proteins or stable colorimetric patterns (Basu, et al., 2005; Basu, et al., 2004), introducing space as an additional design-parameter into the synthetic biology realm.

4.2.4 Issues related to the design of genetic circuits. For the design of genetic networks, the availability of functional elements with specific properties (such as binding constants and degradation rates) that fit the design purpose is crucial. So far, we are only at the beginning of being able to easily measure, let alone program kinetic parameters, cooperativities, or binding constants (see above). Consequently, the design process remains - for the time being - an iterative process that still contains considerable elements of trial and error. Nevertheless, some workaround tools are available today in order to – at least crudely – shift certain characteristics from wild-type values to values that allow a desired behavior to be implemented. These include variations in gene dosage via changes in plasmid replicon (e. g. (Atkinson, et al., 2003)), the increase of protein degradation rates by fusion to suitable protease sensitive tag-sequences (Elowitz and Leibler, 2000), variations in the strength of RBSs (e. g. (Yokobayashi, et al., 2002)), or drawing on the large number of mutants that are available for a number of model systems (such as phage λ, the *lac* system, or the *tet* system). Alternatively, parameters can be adapted to the desired behavior by directed evolution, if a suitable assay is available (Yokobayashi, et al., 2002). However, it is not really clear how such directed evolution assays can be easily tailored to screen for relatively subtle differences in properties important for optimizing design. In summary, a primary task for the immediate future is to gain access to complete system parameter sets, which can then serve as the starting point to produce parts with parameter values that span suitable ranges.

4.3   Engineering of systems

Synthetic biology is a very young discipline that follows a powerful technological vision. However, there are no examples available where the whole approach (as sketched above) has been implemented. Still, in some cases specific aspects of synthetic biology have been of critical importance. We will discuss the following examples: the design of an *E. coli* capable of image processing, refactoring of the phage T7, the design of novel polyketide antibiotics, and the manufacturing of precursors for the anti-malaria drug artemisinin.

An original example for new applications that derives from the interface of engineering and life sciences, which came out of the

iGEM student competition, is the image-processing *E. coli*. By designing proteins that couple light-detection to well-known *E. coli* regulatory circuits, first steps towards light-detecting pixel sizes of micrometer dimensions are possible (Levskaya, et al., 2005).

A more fundamental aspect is covered by the work on the phage T7, which tries to help to answer the question whether it is indeed possible to refactor significant portions of small genomes. In other words, can we indeed modify those genomes according to the requirements of "engineerability" such as monofunctionality of a part of the sequence and organization of the DNA into functional segments. Refactoring 10 kb of the T7 genome, representing about a quarter of the total genome, still produces functional phages, though their efficiency in propagation is reduced (Chan, et al., 2005). This is an important validation of the synthetic biology approach, even though on a small scale. It remains to be seen whether the same concepts can be applied to more complex systems such as microbes.

Two examples for application of synthetic biology concepts come from the area of pharmaceutical production and involve primarily the opportunities offered by *de novo* DNA synthesis, such as the direct adaptation of codon usage, implementation of suitable regulatory circuitry, and the possibility to modularize the DNA sequences by restriction sites to facilitate iterative optimizations. The first example involves the adaptation of polyketide synthesis to well studied *E. coli* production strains (Kodumal, et al., 2004) and the subsequent design of novel polyketides by semi-randomized recombination of polyketide synthase genes. These recombinations were easily enforced along the interfaces of the different functional modules that make up a synthase and resulted in a rather high success rate of detecting novel polyketides (Menzella, et al., 2005).

Along similar lines, another project that very much catches the spirit of synthetic biology is the construction, from scratch, of a cheap terpenoid production pathway in *E. coli* leading to artemisinic acid, a precursor to the anti-malaria drug artemisinin. This goal essentially requires the design of an entirely new pathway in a suitable production organism. The corresponding pathway elements can be recruited from bacteria (*E. coli*), yeast (*Saccharomyces cerevisiae*), and plant (*Artemisa annua*), redesigned and functionally expressed in bacteria or yeast, effectively paving the road to a low-cost production route to effective malaria treatment (Martin, et al., 2003; Ro, et al., 2006).

Although the design of novel biological systems is only beginning, all ingredients of the engineering approach are visible: the role of *de novo* DNA synthesis/fabrication, the design of well-behaved parts on the DNA and protein level, the organization of parts into the next functional level of devices and the corresponding abstractions, and the attempt to introduce standardization, even though for the time being only on a parts level. With the design of ever more complex systems, the need to emphasize these elements will undoubtedly increase.

5 REQUIREMENTS FOR COMPUTATIONAL AND INFORMATIC EXPERTISE

Synthetic biology encompasses the building of novel biological entities for useful purposes and the corresponding endeavors can be subdivided into two distinct types of tasks: systems design and systems fabrication. Here, we will discuss the essential elements of these two tasks with a special focus on the computational and informatics requirements.

Fabrication deals with the transformation of design plans into actual physical instances. Today, this still involves a significant amount of cloning work, which should decrease in the future due to *de novo* DNA synthesis (see above). The fabrication as such, is not expected to create a great demand for novel informatics tools.

In contrast, systems design consisting of forward-engineering of biological parts, devices or systems strongly relies on computing and informatics tools that assist the design process. Ultimately, it would be desirable to have computer aided design tools – CAD tools for biological engineering - in analogy to the respective software tools in the areas of mechanical or civil engineering. Using such software, the synthetic biology design engineer would try to improve the behavior of a biological system *in silico* by optimizing design parameters targeting a selected objective function. Design variants would be tested computationally by means of simulations.

Such design tools will be based on quantitative mechanistic models that reproduce biological behavior, and - in order to be useful for forward-engineering design - would also have predictive power. In biology, we have not yet reached a level of understanding where such models can be developed on a large scale, and consequently, true biological engineering is hardly possible until now (Endy and Brent, 2001). In fact, in most cases today, we are faced with highly uncertain or even unknown model topologies, mechanisms and parameters. The recent advances in the post-genomic research and especially in systems biology, however, provide hope that sooner or later we will be able to draw on a body of knowledge that allows for the envisioned directed engineering of biology (Endy and Brent, 2001). Ultimately, mathematical models developed for research purposes (e.g. in systems biology) will be employed as design models in synthetic biology. In contrast to the current lack of predictive models, tools for modelling and simulation exist in large numbers (cf. (Lemerle, et al., 2005)).

We envision that in the long run we will require models and design software for the following tasks: [i] sequence-based (*ab initio*) prediction of structure, function and interactions of macromolecules, in particular proteins and mRNA, [ii] prediction of the dynamics of signaling and regulatory networks; and [iii] prediction of the dynamics of metabolic networks. For each of these areas, we will shortly sketch the current status of development and also elaborate on future tasks.

5.1 Design of functions and interactions of macromolecules

We would like to predict - starting from a linear sequence of nucleotides or amino acids – 2D (mRNA) and 3D structures of the respective macromolecules (RNA, proteins), as well as their function and their interaction parameters with other cellular components (DNA, metabolites, etc.). In other words, as outlined above we would like to have the possibility to modify sequences in a targeted manner to obtain, for example, novel transcription factors (i.e. with altered binding constants or kinetics) or proteins with novel functions introduced.

However today, as an example, *de novo* protein structure prediction from a linear amino acids sequence can only be achieved for small protein domains at significant computational costs (Bradley, et al., 2005; Misura, et al., 2006). Nevertheless, starting from

known structures of 'scaffold' proteins, design methods are available, which can be used to rationally modify the proteins' structure and function, i.e. to build completely new active sites into proteins (Dwyer, et al., 2004) or to redesign binding specificities of proteins (Looger, et al., 2003). However, such design processes still go through several cycles of iterative improvement involving design, analysis, redesign, etc., where computational tools such as FoldX (Schymkowitz, et al., 2005) are typically employed. In other words, the design of tailored catalytic activities on artificial proteins seems to be within reach, while quantitative prediction of enzymatic activity and selectivity from 3D protein structures in general is not yet feasible. For further information on the current status in modeling of protein structures and interactions, the reader is referred to a recent review (Schueler-Furman, et al., 2005).

Based on structure models, molecular dynamics simulation have shown to be a versatile tool to investigate the dynamic behavior of complexes between DNA binding sites and respective DNA target sites (Marco, et al., 2003; Obika, et al., 2003). These tools can also be employed to predict the effect of structural modulations on protein-ligand interactions in a way that would allow forward-engineering design of, for example, DNA-binding specificity of transcription factors.

5.2   Design of signaling and regulatory networks

Artificial signaling and regulatory gene networks will need to be assembled for synthetic biology. Today, such circuits are still frequently assembled by intuition and optimized through several rounds of trial-and-error (Kærn, et al., 2003) and the mathematical models are only developed once proper *in vivo* function has been demonstrated. Deterministic or stochastic models (or a combination of both) are then used to describe the observed dynamic behavior of the circuit (Fung, et al., 2005; Gardner, et al., 2000; Kobayashi, et al., 2004; You, et al., 2004).

Ideal, however, would be models that allow deriving *in silico* suggestions for optimal design strategies or debugging, prior to implementation of the circuit *in vivo* (Gardner, et al., 2000; Sprinzak and Elowitz, 2005). Such models should be able to capture the dynamic behavior of the gene networks. In cases where only small molecule numbers are involved (as in gene transcription or translation, where transcription factors and mRNA molecules only occur in low copy numbers), the models would also need to be able to reproduce the inherent stochasticity of such processes. This is imperative as it was shown that stochasticity in combination with certain system architectures can result in different system states (Kærn, et al., 2005; Pedraza and van Oudenaarden, 2005). A robust design of new devices and systems must exclude such eventualities.

In summary, to enable the envisioned forward-engineering (model-based) design of signaling and regulatory circuits, improvements are required in the following areas: It is necessary [i] to obtain an improved quantitative understanding of regulatory and signaling processes (Sprinzak and Elowitz, 2005); [ii] to develop effective rules (cf. (Wall, et al., 2003; Wall, et al., 2004)) and standards for characterizing modules (Kærn, et al., 2003; Sprinzak and Elowitz, 2005), and [iii] to improve multiscale simulation algorithms as the existing ones are limited in a way that the participating reactions have to occur on a comparative time scale and the participating reaction species have to fulfill certain population size requirements.

5.3   Design of metabolic networks

Besides the design and construction of signaling and regulatory networks, we would also like to engineer metabolic networks. Models currently available in this area are [i] stoichiometric models which display – in most cases on a genome-wide level – an organism's metabolic (stoichiometric) capabilities and [ii] enzymatic reaction network models, which describe the kinetics of a network but mostly cover only small and well studied areas of metabolism.

For design of metabolic networks based on stoichiometric models, the most prominent tools to analyze these networks are flux balance analysis (FBA) (Edwards, et al., 1999; Varma and Palsson, 1994) and elementary flux mode (Schuster, et al., 1999) or extreme pathways analysis (Schilling, et al., 2000). Extended tools - conceptually based on FBA - were developed that allow assisting the targeted modification of existing or the design of new, metabolic networks (Patil, et al., 2005; Pharkya, et al., 2004).

On the side of the kinetic models (consisting of ordinary differential equations), a large set of models is available from model repositories (cf. www.siliconcell.net). However, due to largely unknown reaction mechanisms and parameters, these models in most cases only cover small parts of metabolism. Tools for modeling and simulation are widely available ranging from general simulation platforms, such as SmartCell (http://smartcell.embl.de/, (Ander, et al., 2004)), via object-oriented software suites (www.e-cell.org) to web-based suites for modeling, simulation, and analysis of biological cells (http://www.webcell.org/) (Lee, et al., 2006). An overview about available tools can be obtained from http://sbml.org.

6   CONCLUSION

The move from describing biology to exploiting it for our requirements has always been a part of the biological enterprise – and thus always reflected the current main lines of biological research. So, as molecular biology has for a long time attempted to unravel the molecular mechanisms that are important in cellular function, biotechnology has exploited this knowledge and adopted some of these mechanisms to produce chemicals, enzymes, and biopharmaceuticals. Now, synthetic biology is adopting a very ambitious agenda in building novel biological entities on an ever more complex level for novel applications.

Currently, biology is characterized by the shift from the study of single aspects of systems to the appreciation of the system as a whole. With this goes the expectation that many of the failures that biotechnology suffered can be understood from this new perspective. At the same time, to advance exploitation of biological systems (instead of single aspects), we adopt an approach that is sufficiently organized and robust. We adapt our technologies – for example in fabrication – and in methodologies – such as design – to the new system-scale of the task through adopting the crucial elements of classical, non-life science related engineering disciplines. The aspects that have been discussed here, from *de novo* DNA synthesis via modular protein design, the design of novel gene networks, to the reconstruction of artificial pathways, address - in our view - this transition in an exemplary way.

REFERENCES

Ander, M., Beltrao, P., Di Ventura, B., Ferkinghoff-Borg, J., Foglierini, M., Kaplan, A., Lemerle, C., Tomas-Oliveira, I. and Serrano, L. (2004) SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks, *Systems Biology*, **1**, 129-138.

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

Atkinson, M.R., Savageau, M.A., Myers, J.T. and Ninfa, A.J. (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*, *Cell*, **113**, 597-607.

Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. and Weiss, R. (2005) A synthetic multicellular system for programmed pattern formation, *Nature*, **434**, 1130-1134.

Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. and Weiss, R. (2004) Spatiotemporal control of gene expression with pulse-generating networks, *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6355-6360.

Becskei, A., Séraphin, B. and Serrano, L. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion, *EMBO Journal* **20**, 2528-2535.

Benner, S.A. and Sismour, A.M. (2005) Synthetic Biology, *Nature Reviews Genetics*, **6**, 533-543.

Blancafort, P., Segal, D.J. and Barbas III, C.F. (2004) Designing transcription factor architectures for drug discovery, *Molecular Pharmacology*, **66**, 1361-1371.

Bolon, D.N., Voigt, C.A. and Mayo, S.L. (2002) *De novo* design of bioctalysts, *Current Opinion in Chemical Biology*, **6**, 125-129.

Bradley, P., Misura, K.M.S. and Baker, D. (2005) Toward high-resolution *de novo* structure prediction for small proteins, *Science*, **309**, 1868-1871.

Bulter, T., Lee, S.-G., Wong, W.W., Fung, E., Connor, M.R. and Liao, J.C. (2004) Design of artificial cell-cell communication using gene and metabolic networks, *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 2299-2304.

Cello, J., Paul, A.V. and Wimmer, E. (2002) Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template, *Science*, **297**, 1016-1018.

Chan, L.Y., Kosuri, S. and Endy, D. (2005) Refactoring bacteriophage T7, *Molecular Systems Biology*, **1**, msb4100025-E4100021-msb4100025-E4100010.

Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and III, C.F.B. (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors, *Journal of Biological Chemistry*, **276**, 29466-29478.

Dueber, J.E., Yeh, B.J., Chak, K. and Lim, W.A. (2003) Reprogramming control of an allosteric signaling switch through modular recombination, *Science* **301**, 1904-1908.

Dwyer, M.A., Looger, L.L. and Hellinga, H.W. (2004) Computational design of a biologically active enzyme, *Science*, **304**, 1967-1971.

Edwards, J.S., Ramakrishna, R., Schilling, C.H. and Palsson, B.O. (1999) Metabolic flux balance analysis. In Lee, S.Y. and Papoutsakis, E.T. (eds), *Metabolic engineering*. Marcel Dekker, New York, 13-57.

Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators, *Nature*, **403**, 335-338.

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

Endy, D. and Brent, R. (2001) Modelling cellular behaviour, *Nature*, **409**, 391-395.

Forster, A.C. and Church, G.M. (2006) Towards synthesis of a minimal cell, *Mol Syst Biol*, **2**.

Fung, E., Wong, W.W., Suen, J.K., Bulter, T., Lee, S.-g. and Liao, J.C. (2005) A synthetic gene-metabolic oscillator, *Nature*, **435**, 118-122.

Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*, *Nature*, **403**, 339-342.

Gil, R., Silva, F.J., Pereto, J. and Moya, A. (2004) Determination of the core of a minimal bacterial gene set, *Microbiology and Molecular Biology Reviews*, **68**, 518-537.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison III, C.A., Smith, H.O. and Venter, J.C. (2006) Essential genes of a minimal bacterium, *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 425-430.

Greger, B. and Kemper, B. (1998) An apyrimidinic site kinks DNA and triggers incision by endonuclease VII of phage T4, *Nucleic Acids Research*, **26**, 4432-4438.

Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., Ote, T., Yamawaki, T., Yamazaki, Y., Mori, H., Katayama, T. and Kato, J.-i. (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome, *Molecular Microbiology*, **55**, 137-149.

Hasty, J., McMillen, D. and Collins, J.J. (2002) Engineered gene circuits, *Nature*, **420**, 224-230.

Howard, P.L., Chia, M.C., Rizzo, S.D., Liu, F.F. and Pawson, T. (2003) Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins, *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 11267-11272.

Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R. and Collins, J.J. (2004) Engineered riboregulators enable post-transcriptional control of gene expression, **22**, 841-847.

Itaya, M., Tsuge, K., Koizumi, M. and Fujita, K. (2005) Combining two genomes in one cell: Stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome, *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 15971-15976.

Kærn, M., Blake, W.J. and Collins, J.J. (2003) The engineering of gene regulatory networks, *Annual Review of Biomedical Engineering*, **5**, 179-206.

Kærn, M., Elston, T.C., Blake, W.J. and Collins, J.J. (2005) Stochasticity in gene expression: from theories to phenotypes, *Nature Reviews Genetics*, **6**, 451-464.

Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L. and Keasling, J.D. (2001) Homogeneous expression of the P_{BAD} promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter, *Microbiology*, **147**, 3241-3247.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R. and Collins, J.J. (2004) Programmable cells: interfacing natural and engineered gene networks, *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 8414-8419.

Kodumal, S.J., Patel, K.G., Reid, R., Menzella, H.G., Welch, M. and Santi, D.V. (2004) Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster, *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15573-15578.

Kolisnychenko, V., Plunkett, G., Herring, C.D., Feher, T., Posfai, J., Blattner, F.R. and Posfai, G. (2002) Engineering a reduced *Escherichia coli* genome, *Genome Research*, **12**, 640-647.

Kramer, B.P., Fischer, C. and Fussenegger, M. (2004) BioLogic gates enable logical transcription control in mammalian cells, *Biotechnology and Bioengineering*, **87**, 478-484.

Kramer, B.P. and Fussenegger, M. (2005) Hysteresis in a synthetic mammalian gene network, *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 9517-9522.

Kramer, B.P., Viretta, A.U., Baba, M.D.-E., Aubel, D., Weber, W. and Fussenegger, M. (2004) An engineered epigenetic transgene switch in mammalian cells, *Nature Biotechnology*, **22**, 867-870.

Lee, D.-Y., Yun, C., Cho, A., Hou, B.K., Park, S. and Lee, S.Y. (2006) WebCell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks, *Bioinformatics*, **22**, 1150-1151.

Lemerle, C., Di Ventura, B. and Serrano, L. (2005) Space as the final frontier in stochastic simulations of biological systems, *FEBS Letters*, **579**, 1789-1794.

Levskaya, A., Chevalier, A.A., Tabor, J.J., Simpson, Z.B., Lavery, L.A., Levy, M., Davidson, E.A., Scouras, A., Ellington, A.D., Marcotte, E.M. and Voigt, C.A. (2005) Synthetic biology: engineering *Escherichia coli* to see light, *Nature*, **438**, 441-442.

Looger, L.L., Dwyer, M.A., Smith, J.J. and Hellinga, H.W. (2003) Computational design of receptor and sensor proteins with novel functions, *Nature*, **423**, 185-190.

Marco, E., Garcia-Nieto, R. and Gago, F. (2003) Assessment by molecular dynamics simulations of the structural determinants of DNA-binding specificity for transcription factor Sp1, *Journal of Molecular Biology*, **328**, 9-32.

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D. and Keasling, J.D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids, *Nature Biotechnology*, **21**, 796-802.

McDaniel, R. and Weiss, R. (2005) Advances in synthetic biology: on the path from prototypes to applications, *Current Opinion in Biotechnology*, **16**, 476-483.

Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A. and Santi, D.V. (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes, *Nature Biotechnology*, **23**, 1171-1176.

Misura, K.M.S., Chivian, D., Rohl, C.A., Kim, D.E. and Baker, D. (2006) Physically realistic homology models built with ROSETTA can be more accurate than their templates, *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5361-5366.

Obika, S., Reddy, S.Y. and Bruice, T.C. (2003) Sequence specific DNA binding of Ets-1 transcription factor: Molecular dynamics study on the Ets domain-DNA complexes, *Journal of Molecular Biology*, **331**, 345-359.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I. and van Oudenaarden, A. (2004) Multistability in the lactose utilization network of *Escherichia coli*, *Nature*, **427**, 737-740.

Park, S.-Y., Zarrinpar, A. and Lim, W.A. (2003) Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms, *Science*, **299**, 1061-1064.

Patil, K., Rocha, I., Forster, J. and Nielsen, J. (2005) Evolutionary programming as a platform for in silico metabolic engineering, *BMC Bioinformatics*, **6**, 308.

Pedraza, J.M. and van Oudenaarden, A. (2005) Noise propagation in gene networks, *Science*, **307**, 1965-1969.

Pharkya, P., Burgard, A.P. and Maranas, C.D. (2004) OptStrain: A computational framework for redesign of microbial production systems, *Genome Research*, **14**, 2367-2376.

Rackham, O. and Chin, J.W. (2005) A network of orthogonal ribosome-mRNA pairs, *Nature Chemical Biology*, **1**, 159-166.

Rebatchouk, D., Daraselia, N. and Narita, J.O. (1996) NOMAD: A versatile strategy for *in vitro* DNA manipulation applied to promoter analysis and vector design, *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 10891-10896.

Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R. and Keasling, J.D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast, *Nature*, **440**, 940-943.

Schilling, C.H., Letscher, D. and Palsson, B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective, *Journal of Theoretical Biology*, **203**, 229-248.

Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. and Baker, D. (2005) Progress in modeling of protein structures and interactions, *Science*, **310**, 638-642.

Schuster, S., Dandekar, T. and Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering, *Trends in Biotechnology*, **17**, 53-60.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field, *Nucleic Acids Research*, **33**, W382-388.

Segal, D.J., Dreier, B., Beerli, R.R. and III, C.F.B. (1999) Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5 '-GNN-3 ' DNA target sequences, *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 2758-2763.

Sera, T. and Uranga, C. (2002) Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table, *Biochemistry*, **41**, 7074-7081.

Smith, H.O., Hutchison, C.A., 3rd, Pfannkoch, C. and Venter, J.C. (2003) Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides, *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 15440-15445.

Smith, J. and Modrich, P. (1997) Removal of polymerase-produced mutant sequences from PCR products, *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 6847-6850.

Sprinzak, D. and Elowitz, M.B. (2005) Reconstruction of genetic circuits, *Nature*, **438**, 443-448.

Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides, *Gene*, **164**, 49-53.

Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X. and Church, G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips, *Nature*, **432**, 1050-1054.

Varma, A. and Palsson, B.O. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110, *Applied and Environmental Microbiology*, **60**, 3724-3731.

Vilar, J.M.G., Guet, C.C. and Leibler, S. (2003) Modeling network dynamics: the lac operon, a case study, *Journal of Cell Biology*, **161**, 471-476.

Wall, M.E., Hlavacek, W.S. and Savageau, M.A. (2003) Design principles for regulator gene expression in a repressible gene circuit, *Journal of Molecular Biology*, **332**, 861-876.

Wall, M.E., Hlavacek, W.S. and Savageau, M.A. (2004) Design of gene circuits: lessons from bacteria, *Nature Reviews Genetics*, **5**, 34-42.

Westers, H., Dorenbos, R., van Dijl, J.M., Kabel, J., Flanagan, T., Devine, K.M., Jude, F., Seror, S.J., Beekman, A.C., Darmon, E., Eschevins, C., de Jong, A., Bron, S., Kuipers, O.P., Albertini, A.M., Antelmann, H., Hecker, M., Zamboni, N., Sauer, U., Bruand, C., Ehrlich, D.S., Alonso, J.C., Salas, M. and Quax, W.J. (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*, *Molecular Biology and Evolution*, **20**, 2076-2090.

Yokobayashi, Y., Weiss, R. and Arnold, F.H. (2002) Directed evolution of a genetic circuit, *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16587-16591.

You, L., Cox, R.S., 3rd, Weiss, R. and Arnold, F.H. (2004) Programmed population control by cell-cell communication and regulated killing, *Nature*, **428**, 868-871.

Young, L. and Dong, Q.H. (2004) Two-step total gene synthesis method, *Nucleic Acids Research*, **32**, e59.

Yu, B.J., Sung, B.H., Koob, M.D., Lee, C.H., Lee, J.H., Lee, W.S., Kim, M.S. and Kim, S.C. (2002) Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/*loxP* excision system, *Nature Biotechnology*, **20**, 1018-1023.

Zhou, X., Cai, S., Hong, A., You, Q., Yu, P., Sheng, N., Srivannavit, O., Muranjan, S., Rouillard, J.M., Xia, Y., Zhang, X., Xiang, Q., Ganesh, R., Zhu, Q., Matejko, A., Gulari, E. and Gao, X. (2004) Microfluidic picoarray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences, *Nucleic Acids Research*, **32**, 5409-5417.