ISSN 1330-9862
(FTB-3220)

minireview

Synthetic Biology: A Novel Approach for the Construction of Industrial Microorganisms

Daslav Hranueli[1], Antonio Starcevic[1], Jurica Zucko[1], Juan Diego Rojas[2], Janko Diminic[1], Damir Baranasic[1], Ranko Gacesa[1], Gabriel Padilla[2], Paul F. Long[3] and John Cullum[4]*

[1]Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, HR-10000 Zagreb, Croatia

[2]Institute of Biomedical Sciences, University of São Paulo, São Paulo SP-CEP 05508-900, Brasil

[3]King's College, University of London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK

[4]Department of Genetics, University of Kaiserslautern, Postfach 3049, DE-67653 Kaiserslautern, Germany

Received: October 3, 2012
Accepted: January 18, 2013

Summary

The recent achievement of synthesising a functioning bacterial chromosome marks a coming of age for engineering living organisms. In the future this should allow the construction of novel organisms to help solve the problems facing the human race, including health care, food, energy and environmental protection. In this minireview, the current state of the field is described and the role of synthetic biology in biotechnology in the short and medium term is discussed. It is particularly aimed at the needs of food technologists, nutritionists and other biotechnologists, who might not be aware of the potential significance of synthetic biology to the research and development in their fields. The potential of synthetic biology to produce interesting new polyketide compounds is discussed in detail.

Key words: molecular biotechnology, industrial microorganisms, DNA synthesis, homologous recombination, recombinant gene clusters

Traditional and Molecular Biotechnology

Traditional biotechnology uses microorganisms for a variety of purposes. For thousands of years, fermentation has been used in the food and drinks industry. In the twentieth century many other microorganism processes became important, including the production of antibiotics, industrial enzymes and vitamins. With the advent of genetic engineering, there has been an explosion in biotechnology in which, not only microorganisms, but also animals and plants have been used.

Biotechnology is often divided into red, green, white and blue subdivisions (*1*). The products and technologies of red biotechnology are useful in medicine for the production of drugs, particularly therapeutic proteins. Red biotechnology also includes the manufacturing processes that have evolved out of pharmacogenomics, which can directly be attributed as benefits coming from sequencing the human genome, *i.e.* personalised medicines, gene therapies and molecular-based diagnostics (*2*). The products and technologies of green biotechnology are useful to agriculture and animal husbandry. For example, breeding of agricultural plants and domestic animals to increase their yields and their nutritive content, as well as for the improvement of texture, flavour and complexion of raw materials for the food industry. The products and technologies of white biotechnology are useful to the chemical and related industries for the production of fine and bulk chemicals, plastics, fabric, and what is emerging as an especially important biotechno-

*Corresponding author; Phone: ++385 1 4605 013; Fax: ++385 1 4836 083; E-mail: dhranueli@pbf.hr

logy – production of biofuels as alternative energy sources to dwindling fossil fuel reserves. Products and technologies of blue biotechnology are applicable in environmental protection, particularly for the bioremediation of specific pollutants such as oil after environmental disasters.

Much of biotechnology utilises microorganisms. Many microorganisms that are used have been refined to improve the most desirable traits. For example, when penicillin was discovered some 80 years ago (*3*), its industrial production started with the *Penicillium chrysogenum* wild-type isolate that could naturally produce a little over 2 μg/mL. Today's industrial strains can produce more than 100 mg/mL of penicillin, some 50 000 times more than the wild-type. To achieve this, scientists used traditional genetic methods of strain improvement, which were heavily dependent on repeated cycles of mutagenesis and screening. More modern molecular genetic techniques encompassing recombinant DNA technologies were easily incorporated into this process. A typical industrial strain improvement pipeline would start with the screening of mutants for higher product yield following random mutagenesis. The mutations in these high producing mutants would then be mapped using techniques such as conjugation, transformation, transduction (for the historical overview see references in: *4*) or protoplast fusion (*5*,*6*). Usually, the molecular basis for high production would cover many genetic loci and would never be discovered. An alternative approach using recombinant DNA technology is to clone the desirable gene or gene cluster, followed by DNA sequencing and bioinformatics analysis of the sequence in order to recognise structural and regulatory regions. Increased product yield could therefore be achieved either by specific mutation of these sequences or replacement by more efficient sequences such as promoters.

Functional genomics has developed as a new broad field of science for analysing the interactions of biological information objects. It includes a number of 'omes': the most important being genome, transcriptome, proteome, cellome, tissueome and organome (Fig. 1) (*7*). The main focus is to discover the properties of the system being studied rather than concentrating on single parts. It is aimed to characterize all of the information objects of the system (such as genes, proteins and ligands) and find the interactions between the objects. This will often include engineering networks and objects to understand and manipulate the regulatory mechanisms and it is usually necessary to integrate various omes and omics. Modern research in food science and nutrition is also moving from classical methodologies to advanced functional genomics in which synthetic biology is the most recent development. In this context, nutrigenomics (*8*) and foodomics (*9*) have recently been defined as new disciplines that study food and nutrition domains through

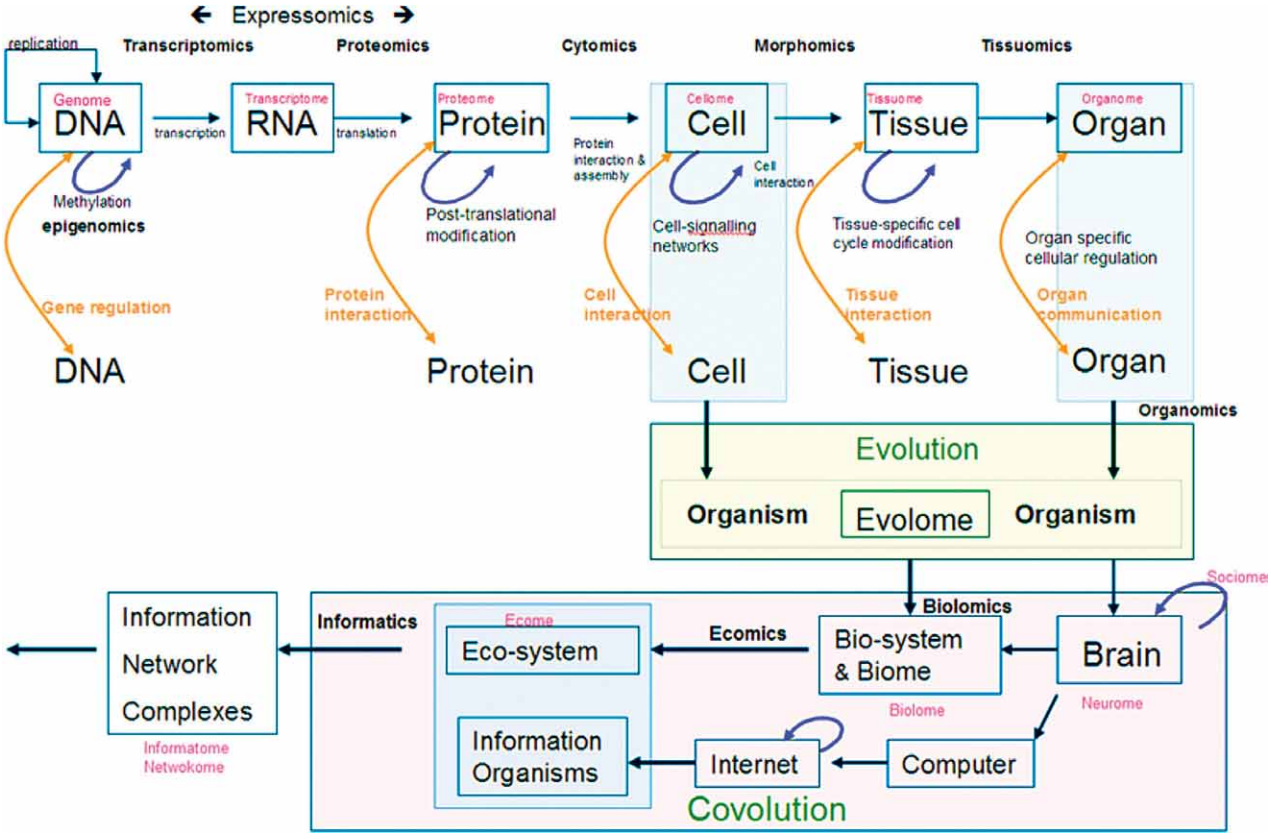

Fig. 1. Omics pathway. The traditional 'omes' are shown in the top part of the figure. The genome, transcriptome and proteome are in principle catalogues of the DNA, RNA and protein molecules present in an organism. The cellome, tissuome and organome combine catalogues of molecules with information about regulatory interactions. The black arrows indicate information flow between the 'omes' and the resulting hierarchy. The blue arrows indicate modification pathways and the yellow arrows interaction pathways. The 'omes' in the lower part of the figure are included for completeness, but are less relevant for the topic of this review (reprinted from ref. *7* with the permission of the Web page owner)

the application of advanced omics technologies. Applications of nutrigenomics and foodomics include the genomic, transcriptomic, proteomic, and/or metabolomic study of foods. This includes compound profiling to establish authenticity, and/or biomarker detection related to food quality or safety. There are also omics-based studies of food bioactivity and the effects of food on human health. These are all relevant to establishing the safety of new transgenic foods as well as novel food processing technologies.

Synthetic Biology

Genetic engineering has been very successful for both commercial purposes and for fundamental biological studies. This has usually involved the manipulation of a small number of genes in an organism. The dream of synthetic biology was to make designer organisms in which the whole genome sequence is determined by the experimenter. Recently Craig Venter's group has successfully synthesized a complete bacterial chromosome and transplanted it into a cell to obtain a viable living organism (for details see reference: *10*). In this minireview, we will discuss the importance of synthetic biology for applications in food technology and biotechnology and try to pinpoint areas in which progress is possible in the near future. At present synthetic biology is a rapidly growing field with numerous potential applications. Our intention is not to achieve exhaustive coverage of the entire field, but, rather, to select examples that will alert scientists from fields such as food technology and nutrition to developments, which will become essential for them in the near future.

Routine oligonucleotide synthesis using phosphoramidite chemistry has been possible for many years and can, in principle, produce oligonucleotides of over 200 b in length. However, in practice, it is most efficient to produce oligonucleotides of about 50 b. The main innovations that reduced the costs of synthesis were the development of techniques allowing the synthesis of many oligonucleotides in parallel, thus reducing the cost per oligonucleotide. For synthetic biology applications, the yield of each oligonucleotide is not a problem. The first application of oligonucleotides to produce DNA molecules, which were replicated, was to construct linkers to insert new restriction sites into molecules. A good example is the construction of multiple cloning sites for pUC and related vectors (*11*). Here the DNA sequences at the 5'-end of the *lacZ*α gene fragment were altered to introduce many restriction sites, while still preserving a viable protein-coding gene. Somewhat later synthetic genes were constructed by hybridizing and ligating a series of oligonucleotides. Some care is needed in designing the individual oligonucleotides to avoid problems with secondary structure and mispairing. However, this technology has become routine and companies offer gene synthesis as a routine service. This works well for sequences of around 1–5 kb. However, considerably larger sequences cannot be effectively synthesized in this way, because significant error frequencies occur.

In 1995 scientists, led by Craig Venter, sequenced the chromosome of *Mycoplasma genitalium*, a bacterium with the smallest known genome of any living organism, only 583 kb in size and encoding 517 genes (*12*). Comparison with the genome sequence of another human pathogen, *Haemophilus influenza* (*13*), that encodes around 1700 genes revealed about 250 genes common to both bacteria that by inference must be absolutely essential for life. They called these 250 genes the 'minimal genome'. The essential requirement of each gene was confirmed by directed mutagenesis. An effective experimental approach to show whether the 'minimal genome' is sufficient to support life requires the synthesis of bacterial chromosomes and their installation in a cell. It was first necessary to show that it was technically possible to achieve synthesis and activation of a genome known to be functional. In this case, the entire *M. genitalium* chromosome, so the discipline of synthetic biology was born. Craig Venter's team started construction of the entire *M. genitalium* chromosome from hundreds of pieces of DNA between 5 and 6 kb in size (*10,14*), which could be obtained from routine well-understood assembly of oligonucleotides. The construction of the entire chromosome was achieved in two steps. In the first step, fragments of DNA, with sticky 5' and 3' ends were ligated into larger pieces of DNA using an *E. coli* host. In the second step, these large pieces were transformed into the yeast *Saccharomyces cerevisiae* and assembled into a single chromosome by the homologous recombination system found in *S. cerevisiae*, which is unusual in allowing recombination of short identical sequences of 50 b or larger (see references in: *15*) (Fig. 2; *14,16*). These experiments showed that it was technically feasible to synthesize a complete bacterial chromosome and assemble it in yeast. However, there was still another major hurdle to overcome: that of introducing the newly synthesized chromosome into a bacterial cell in a way that it would function. Although the *M. genitalium* chromosome had the advantage of small size, the organism is very slow growing, so that any transplantation experiments would need weeks of waiting before it was known whether they were successful. Therefore, *M. mycoides*, a close relative of *M. genitalium*, which grows much faster, was used. The chromosome of *M. mycoides* is considerably larger (approx. 1 Mb, *17*) than that of *M. genitalium* (583 kb). After cloning the chromosome of *M. mycoides* into yeast and introducing some genetic changes, it proved possible to transplant it to *M. capricolum* and produce a viable *M. mycoides* strain (*18*).

Venter's team then turned to synthesize the chromosome of *M. mycoides*. The researchers started building their new synthetic chromosome by going DNA shopping. They bought from a commercial source more than 1000 of 1-kb sequences that covered the whole *M. mycoides* chromosome. To facilitate their assembly in the correct order, the ends of each sequence had 80 bp that overlapped with its neighbours. In order to make sure that the assembled chromosome would be recognizable as synthetic, six of the ordered DNA sequences contained strings of bases that spelt out the names of some of the people involved in the project (*19*). However, when they introduced that synthetic chromosome into *M. capricolum* cells nothing happened. The researchers had to correct mistakes in the chemically synthesized DNA one by one in the same way informaticians debug newly developed software. This process took many months of unsuccess-

fully transplanting these various new chromosomes until the correct combinations were found (*16*). Proofreading and correcting errors introduced during total synthesis of large pieces of DNA in a timely manner will be one of the major challenges for synthetic biology in the future.

The success of Venter's team was achieved at great expense, an estimated 40 million dollars, and an effort of 20 people working for more than a decade. Despite this success, creating heavily customized genomes, such as ones that make fuels or pharmaceuticals, and getting them to boot up the same way in a cell is not yet a reality. Professor Paul Keim, a molecular geneticist at Northern Arizona University said that 'There are great challenges ahead before genetic engineers can mix, match, and fully design an organism's genome from scratch'. However, before they succeed governmental regulative organisations, such as Food and Drug Administration in the US, will need to establish proper regulations for such cases. Professor Eckard Wimmer from Stony Brook, University in New York, who led a team that in 2002 created the first synthetic virus (see references in: *20*) believes that 'The possibility of misuse unfortunately exists' (*21*). In the meantime Craig Venter has founded a new company, Synthetic Genomics Inc. (*22*) which applied for several patents covering the work. A technology consulting agency, ETC Group in Ottawa, has argued that these actions could result in a monopoly on synthesized life (*23*) but others are not worried. Given the current climate for granting and upholding patents of this type, it is unlikely that the company Synthetic Genomics Inc. will ever become the Microsoft of synthetic biology. However, one thing is sure, interesting creatures will be bubbling out of the Craig Venter Institute laboratories. Because of its importance, it is not surprising that the synthetic biology immediately got its scientific, the SynBio (*24*), and industrial, the Syndustry (*25*) (Fig. 3), acronyms. By the synthetic biology living beings can be changed on the molecular level, on the level of the biosynthetic pathway, on the cellular level and on the level of the entire multicellular organism (*26*).

The chemical synthesis of a bacterial chromosome is an impressive achievement, but the *de novo* design of bacteria is still a distant prospect, because the fundamental knowledge needed is still lacking. Work by George M. Church and his group from Harvard Medical School in Boston and others has tried to adopt an 'engineering' approach, *i.e.* to standardise approaches and develop useful components (*27*). The International Genetically Engineered Machine (iGEM) Foundation (*28*) is an organisation dedicated to the advancement of Synthetic Biology with open exchange of information. The main concept is to develop standard parts to allow the construction of organisms, which can perform a particular task. Each year it sponsors a competition for student teams to achieve this aim (there were 190 teams for the iGEM 2012 competition). There is a catalogue of standard parts (*28*) and the aim is to make Synthetic Biology an engineering discipline: most of a construction should use robust stan-

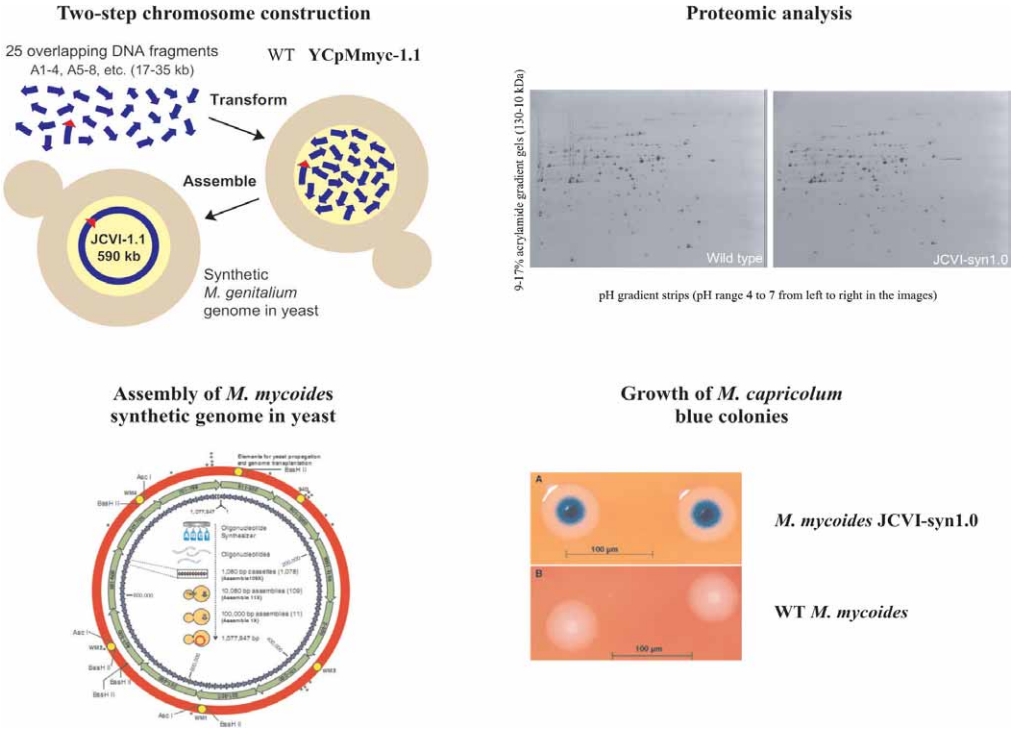

Fig. 2. Illustration of some important steps in the construction of a synthetic bacterial chromosome. The two-step chromosome construction shows in cartoon form how the whole *Mycoplasma genitalium* chromosome was assembled in yeast from shorter DNA fragments assembled in *E. coli*. The same strategy was used to assemble the *M. mycoides* chromosome and the assembly steps are shown in more detail. After successful introduction of the synthetic *M. mycoides* chromosome into a bacterial cell, the proteome was compared to that of wild-type *M. mycoides* using 2-D gel electrophoresis to confirm that a normal protein complement was present. The synthetic chromosome contained a β-galactosidase gene allowing colonies containing the synthetic chromosome to be recognised by their blue colour on media containing X-gal (modified from refs. *14* and *16* with the permission of publishers)

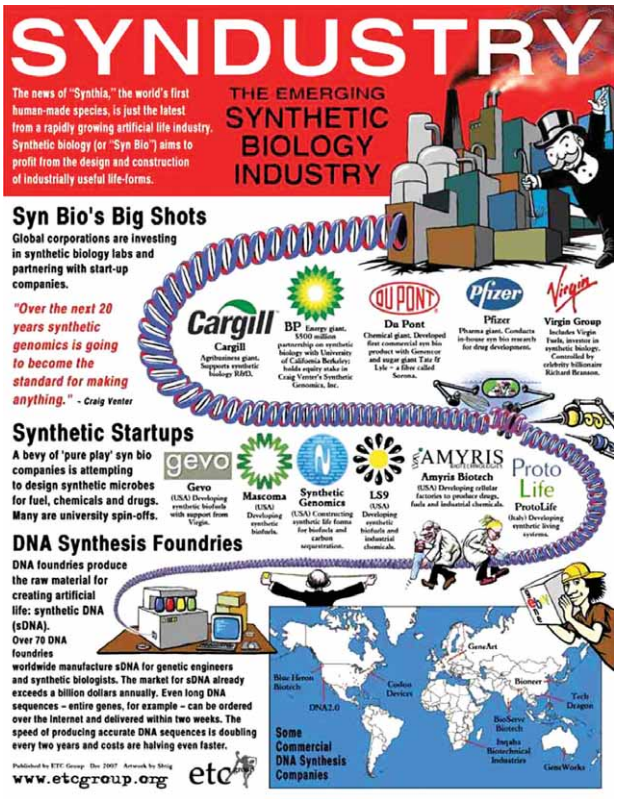

Fig. 3. The cartoon illustrating possible applications of synthetic biology (reprinted from ref. *25* with the permission of the Web page owner)

dard parts with well understood properties. The BioBricks Foundation (*29*) supports similar aims and has established standards for parts ('BioBricks'). Many of the parts are designed for microorganisms, but there are also parts for animals and plants. One proof of principle was to use a BioBrick approach to introduce genes for the sweet protein brazzein and the flavour-inverter protein miraculin, which makes sour foods taste sweet, into the plant *Arabidopsis thaliana* (*30*). This sort of development will move the emphasis of genetic engineering from today's situation in which genetic engineering experts drive the field to one where the engineer sits down and designs the desired modification using a standard toolkit. Food technologists will be able to decide favourable components for food and design plants and animals with the correct properties, without knowing all the subtle details of vector design and cloning strategies.

In many cases, it will probably be more effective to redesign existing organisms rather than building a new organism *de novo*. For many years, it has been possible to use directed mutagenesis to change one gene in an organism. Recently methods have been developed for introducing many mutations simultaneously (*31*). These depend on the properties of the recombination protein β of bacteriophage λ, which allows recombination between short regions of homology. If an oligonucleotide is introduced into an *E. coli* cell whilst it is replicating, protein β can promote recombination of the oligonucleotide with the chromosome at the replication fork. The authors designed an automatic system for carrying out repeated cycles of transformation with a mixture of oligonucleotides so as to achieve a high frequency of recombinants in multiple genes without selection; this approach was called multiplex automated genome engineering (MAGE). The system was tested by targeting 24 genes simultaneously to try to improve the production of the industrially important isoprenoid lycopene. Four of the genes were targeted for inactivation by introducing stop codons, whereas the other 20 genes were targeted for higher expression, trying to improve translational efficiency by modifying the ribosome binding sites (RBS). As the optimal ribosome binding sites were not known, a mixture of oligonucleotides was synthesized for each RBS using a choice of nucleotide at several sites. This approach resulted in a five-fold increase in lycopene production within three days of MAGE cycling. A more ambitious aim is to reprogram the genetic code of an organism. This recoding could prevent expression of foreign genetic material giving protection against viruses and providing genetic containment for recombinant genes. If it were possible by recoding to introduce novel amino acids in protein synthesis, there would be many potential commercial applications. A first step to this aim has been taken by eliminating the use of the amber stop codon (UAG) in *E. coli*, thus providing a free codon for reprogramming (*32*). There are 314 *E. coli* genes, which use a UAG stop codon and a strain was produced with all 314 replaced by UAA stop codons. The approach used MAGE to construct strains with 10 changes in the same region of the chromosome. The mutations in these strains were then assembled into a single strain using conjugative assembly genome engineering (CAGE). This impressive feat shows that large scale reengineering of existing strains is possible. However, the methods used to achieve a high efficiency of MAGE also resulted in other random mutations occurring in the genome, which could be a considerable drawback for engineering strains. However, it may be possible to improve the methodology to reduce this problem.

It is claimed that the use of synthetic biology, or Syndustry, will allow the construction of superior industrial microorganisms suitable for the production of novel drugs, raw materials for the food industry, chemicals and plastics (Fig. 3). There is also considerable interest in the production of alternative fuels, as well as for the bioremediation of pollution. However, most of the concrete examples of synthetic biology do not involve systems of industrial interest. Synthetic biology does not employ novel methodologies, but rather extends the scope of existing methods of genetic manipulation. These facts raise the question as to whether synthetic biology is a passing fad or is an important field, which biotechnologists should take seriously. We believe the latter, because synthetic biology introduces a fundamental difference in how biotechnologists should operate. Classical biotechnology selects a microorganism and the design of the process is constrained by the physiological properties of the organism. Thus, in a secondary metabolite fermentation using a *Streptomyces* strain, fermentation will occur over many days usually using a complex fed batch process with problems arising from the rheological properties of the fermentor contents. Although strain improvement programmes improve yield, they do not alter

the fact that the process engineering in dominated by the physiological properties of the organisms used. Synthetic biology makes the very ambitious claim that the biotechnologist can design the organism using standard procedures in a similar way to the engineering design of other components of the process. At present, the knowledge base only allows limited design possibilities. However, the demonstration that it is possible to synthesize bacterial chromosomes and the continuing development of standard parts for genetic manipulation will lead to rapid improvements. Synthetic biology is already being integrated into pipelines for the incremental improvement of existing processes, but will really come of age in industry, when novel processes are introduced, which were not feasible with more traditional methods. In a few years, every biotechnologist will have to understand the fundamentals of synthetic biology.

Synthetic Biology for Large Gene Clusters for the Biosynthesis of Natural Products

Progress in synthetic biology will be predicated on the development of better and cheaper technology to synthesize long DNA sequences. This will be stimulated by the demand for DNA synthesis from different projects. It seems unlikely that, in the near future, many groups will be synthesizing complete bacterial chromosomes. An interesting target for synthetic biology is natural product biosynthetic clusters. These are frequently large (50–150 kb in size), so they need comparable technology to that for bacterial chromosomes. However, they are much better understood, so there will be an immediate pay-off in terms of functioning clusters. Our laboratories have been working on pharmaceutically important natural products that are synthesized by Type I and Type II polyketide synthases (PKS) (*e.g. 33–35*). There has been much work on the *in vivo* combinatorial biosynthesis of polyketides for the past two decades, but the field remains in its infancy (see references in: *36*), mainly because of the problem that most constructs have very low product yields. In our laboratories we have developed an expert system for the *in silico* drug design and discovery, *3DIS* (*37*). The expert system consists of two generic program packages, *ClustScan* (*38*,*39*) and *CompGen* (*40*,*41*), and two custom databases *CSDB* and *r-CSDB* (see databases at: *42*). These two suites of programs allow rapid mining of large genomic and metagenomic data sets for modular polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS) and hybrid polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) biosynthetic gene clusters, collectively called Thiotemplate Modular Systems (TMS) (*43*). The annotation is performed with hierarchical structuring from DNA sequence into polypeptides and then catalytic enzyme domains, as well as storage and graphical presentations of the data. A prediction of the most likely chemical structures produced by these enzyme-catalysed reactions is also given, thereby providing comprehensive annotation 'from genes to compounds' (*39*,*44*,*45*). As mentioned above, these annotations and likely chemical structures have been used to build two custom databases, one of which is a repository of known compounds, *CSDB*, and the other a repository of novel recombinant products, *r-CSDB*, that can be used for *in silico* screening with computer aided drug design technology (*46*). The program packages use profiles derived from hidden Markov models (HMMs) (*47*) to locate catalytic domains in proteins and then, through consideration of the collected set of domain structures, predict the likely enzyme products. It is possible to use standard profiles (Pfam) (*48*), proprietary profiles supplied with the program and also for the user to supply their own proprietary profiles (*e.g.* built with profile HMMs for protein sequence analysis). Chemical structures of metabolites are stored and output as extended isomeric SMILES (Simplified Molecular Input Line Entry System, *49*) strings. One of the important advantages of *ClustScan* and *CompGen* is that they make it easy to identify potential sequencing and assembly errors and interesting features of gene clusters (*38*,*41*).

The *r-CSDB* database contains predicted entirely novel recombinant products. The recombination sites are chosen on the basis of modelling homologous recombination between clusters, because it seems likely that this will alleviate the problems associated with junctions encountered during attempts at combinatorial biosynthesis. At present there are 47 parental PKS gene clusters, 777 cluster pairs and 20187 recombinant gene clusters in *r-CSDB* database (Fig. 4) that generates 11 796 unique compounds. Like *CSDB*, *r-CSDB* also contains all data starting with gene cluster recombinant DNA sequence, the DNA and protein sequences of genes, modules and domains and of the recombinant gene clusters present in FASTA formats. It also contains all known polyketide and peptide building blocks in the form of isomeric SMILES (*49*), along with the programmed logic that allows total prediction of linear and partial prediction cyclic polyketide and peptide chains and aglycons in the 2-D or 3-D forms suitable for further computer processing. Parental and recombinant linear chain and agycons can be also visualised using Jmol (*50*) or ChemAxon (*51*). The *r-CSDB* database is also fully searchable using *CompGen* suit of programs of TMS gene cluster annotations as well as recombinant compound structures. As *CSDB*, the *r-CSDB* data can also be manipulated using a number of conventional bioinformatic tools.

In *r-CSDB* we have DNA sequences of all recombinant gene clusters generated so far. They could now be synthetized, cloned in an appropriate vector and expressed in a suitable host. If the DNA sequences of several parent clusters were synthesized in suitable segments, a large number of different recombinants could be constructed from the same material. In general, the number of possible recombinants scales with the square of the number of parent clusters available. A synthetic biology approach needs suitable vectors. One such vector, pKW-201, is described by Starcevic *et al.* (*40*) and would be useful for its ability to transfer gene clusters between different plasmid vectors and/or chromosomes of different species using the *rpsL*-based constructs (*52*). A major issue for the pharmaceutical industry is maintaining a continuous supply of promising new leads for drug development. We propose that recombinatorial biosynthesis and synthetic biology offers a new and exciting strategy whereby large and chemically diverse libraries of polyketides and non-ribosomal peptides can first be screened *in silico*

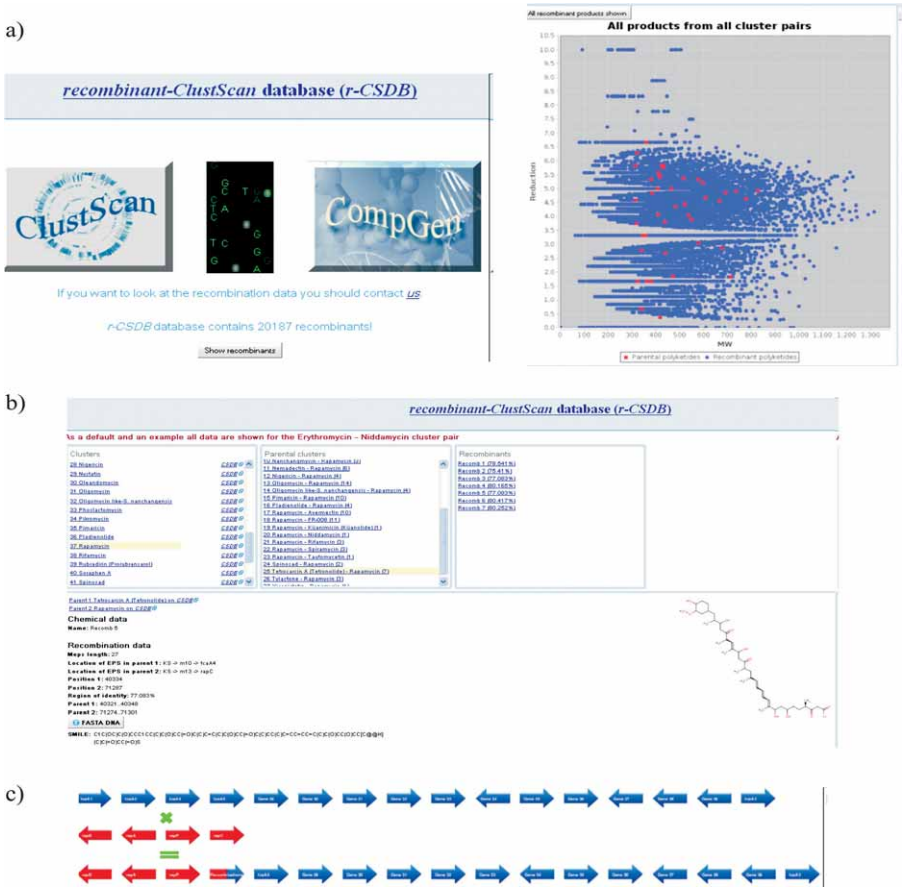

Fig. 4. Screenshots from *r-CSDB*: a) *r-CSDB* homepage (left) and all recombinant products in the database (right), b) the data of 5th recombinant between hybrid PKS/NRPS *Rap* and PKS *Tca* gene clusters and c) the graphical presentation of homeologous recombination (see database at: *42*)

and then generated in the laboratory for further new lead development.

When a commercially interesting cluster has been identified, synthetic biology can be employed to generate an industrial microorganism. The ideal organism would probably be a fast-growing unicellular microorganism rather than a mycelial slow-growing *Streptomyces* strain. As synthetic biology matures, it will become possible to manipulate a suitable organism to provide the necessary precursors and switch on the biosynthesis genes at the correct time. At present it should be possible to express the cluster in a suitable *Streptomyces* host using known regulation systems to induce the genes at the correct time. For instance, it would be possible to use *Streptomyces rimosus*, which has disperse growth and a proven track record for production of high titres of a polyketide (*53*). Given that many polyketides and non-ribosomal peptides are used clinically (*54*), this new expert system (*37*) and synthetic biology come at an important time for hit and lead identification.

Conclusions and Future Prospects

Apart from the construction of industrial microorganisms reviewed by Chen *et al*. (*55*) and by Krivoruchko *et al*. (*56*) in which structural synthetic biotechnology as a new field in biotechnology was proposed for the improved production of industrial enzymes (*57*), natural product metabolites (*58*), other pharmaceutical products (*59*) and for biofuel production (*60*), there are numerous other applications of synthetic biology. For example, it was proposed that the synthetic biology can be applied in evolutionary and toxicology studies (*61,62*) and that it will even move into clinical studies (*63,64*) for the treatment of infectious diseases and cancer; as well as for the vaccine development, microbiome engineering, cell therapy, and regenerative medicine. The chemical synthesis of single genes up to about 5 kb in length is now routine and not very expensive. The synthesis of longer DNA sequences (*e.g*. longer than 100 kb) is still very difficult and expensive; a very important constraint is the presence of mutations, which are time-consuming to eliminate. The speed and costs of such DNA synthesis will only drop significantly if there is commercial demand, analogous to the drop in DNA sequencing costs fuelled by the Human Genome Project. It seems unlikely that there will be enough demand for the synthesis of complete bacterial chromosomes in advance of technology improvement, because the short term commercial advantages are unclear. A more realistic target is clusters for the biosynthesis of natural products, which range in size from about 10–200 kb. The products of these clusters are of proven commercial values and, as pointed out in this review, fragments of the clusters can be used for

various combinatorial processes, reducing the synthesis costs of each novel cluster. Rapid and cheap synthesis of longer DNA sequences would open up possibilities for novel applications in higher risk fields and make the routine synthesis of whole bacterial chromosomes attractive.

It also seems likely that the movement to establish an engineering approach to synthetic biology (*28,29*) will lead to a rapid increase in applications once the library of standard parts reaches a critical size. This will mean that most of the components needed for any application already exist in a standard tested form and most of the work is selecting appropriate components and building them together. This approach generates an increasing need for computational tools that can support synthetic biology. A range of algorithms has been developed that can be used for different applications. The key existing tools and suggestions how informatics can help to shape the future of synthetic microbiology is reviewed in a number of articles (*65–68*). It seems that the synthetic biology software will soon drive the wet lab implementation of DNA sequences.

Acknowledgements

This work was supported by grant 09/5 (to D.H.) from the Croatian Science Foundation, Republic of Croatia, and by a cooperation grant of the German Academic Exchange Service (DAAD) and the Ministry of Science, Education and Sports, Republic of Croatia (to D.H. and J.C.). We would also like to thank CNPq-TWAS (Conselho Nacional de Pesquisa do Brazil Third World Academy of Sciences, Brazil) and FAPES (Fundação de Apoio a Pesquisa do Estado de São Paulo, Brazil) for grants to JDR and FAPESP-King's College, London, UK for grant to GP and PFL.

References

1. E.J. DaSilva, The colours of biotechnology: Science, development and humankind, *Electron. J. Biotechnol. 7* (2004) Article no. 3.
2. J.P. Evans, D.C. Dale, C. Fomous, Preparing for a consumer-driven genomic age, *N. Engl. J. Med. 363* (2010) 1099–1103.
3. A. Fleming, On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae*, *Br. J. Exp. Pathol. 10* (1929) 226–236.
4. W. Hayes: *The Genetics of Bacteria and Their Viruses*, Blackwell UK Ltd., Oxford, UK (1964).
5. J. Pigac, D. Hranueli, T. Smokvina, M. Alačević, Optimal cultural and physiological conditions for handling *Streptomyces rimosus* protoplasts, *Appl. Environ. Microbiol. 44* (1982) 1178–1186.
6. D. Hranueli, J. Pigac, T. Smokvina, M. Alačević, Genetic interactions in *Streptomyces rimosus* mediated by conjugation and by protoplast fusion, *J. Gen. Microbiol. 129* (1983) 1415–1422.
7. Omics.org, La Jolla, CA, USA (*http://omics.org/index.php/Main_Page*).
8. A.P. Simopoulos, J.A. Milner: *Personalized Nutrition: Translating Nutrigenetic/Nutrigenomic Research into Dietary Guidelines*, S. Karger AG, Basel, Switzerland (2009).
9. M. Herrero, C. Simó, V. García-Cañas, E. Ibáñez, A. Cifuentes, Foodomics: MS-based strategies in modern food science and nutrition, *Mass Spectrom. Rev. 31* (2012) 49–69.
10. D.G. Gibson, G.A. Benders, C. Andrews-Pfannkoch, E.A. Denisova, H. Baden-Tillson, J. Zaveri *et al.*, Complete chemical synthesis, assembly, and cloning of *Mycoplasma genitalium* genome, *Science, 319* (2008) 1215–1220.
11. J. Messing, Cloning single-stranded DNA, *Mol. Biotechnol. 5* (1996) 39–47.
12. C.M. Fraser, J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann *et al.*, The minimal gene complement of *Mycoplasma genitalium*, *Science, 270* (1995) 397–404.
13. R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage *et al.*, Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd., *Science, 269* (1995) 496–512.
14. D.G. Gibson, G.A. Benders, K.C. Axelrod, J. Zaveri, M.A. Algire, M. Moodie *et al.*, One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome, *Proc. Natl. Acad. Sci. USA, 105* (2008) 20404–20409.
15. S. Fogel, R.K. Mortimer, Recombination in yeast, *Annu. Rev. Genet. 5* (1971) 219–236.
16. D.G. Gibson, J.I. Glass, C. Lartigue, V.N. Noskov, R.Y. Chuang, M.A. Algire *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome, *Science, 329* (2010) 52–56.
17. J. Westberg, A. Persson, A. Holmberg, A. Goesmann, J. Lundeberg, K.E. Johansson *et al.*, The genome sequence of *Mycoplasma mycoides* subsp. *mycoides* SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP), *Genome Res. 14* (2004) 221–227.
18. C. Lartigue, S. Vashee, M.A. Algire, R.Y. Chuanq, G.A. Benders, L. Ma *et al.*, Creating bacterial strains from genomes that have been cloned and engineered in yeast, *Science, 325* (2009) 1693–1696.
19. M. Bittner, D. Vapnek, Versatile cloning vectors derived from the runaway-replication plasmid pKN402, *Gene, 15* (1981) 319–329.
20. E. Wimmer, A.V. Paul, Synthetic poliovirus and other designer viruses: What have we learned from them?, *Annu. Rev. Microbiol. 65* (2011) 583–609.
21. E. Pennisi, Two steps forward for synthetic biology, *Science, 21* (2009) 928–929.
22. Synthetic Genomics Inc., La Jolla, CA, USA (*http://www.syntheticgenomics.com/*).
23. J. Kaiser, Attempt to patent artificial organism draws a protest, *Science, 15* (2007) 1557.
24. A.J. Newson, Current ethical issues in synthetic biology: Where should we go from here?, *Account. Res. 18* (2011) 181–193.
25. Converging New Technologies, ETC Group, Ottawa, ON, Canada (*http://www.etcgroup.org/*).
26. J. Liang, Y. Luo, H. Zhao, Synthetic biology: Putting synthesis into biology, *Wiley Interdiscip. Rev. Syst. Biol. Med. 3* (2011) 7–20.
27. H.W. Harris, G.M. Church, Multiplexed genome engineering and genotyping methods: Applications for synthetic biology and metabolic engineering. *Methods Enzymol. 498* (2011) 409–426.
28. Registry of Standard Biological Parts, iGEM – Synthetic Biology Based on Standard Parts, Cambridge, MA, USA (*http://partsregistry.org/Main_Page*).
29. The BioBricks Foundation, Cambridge, MA, USA (*http://biobricks.org/*).
30. P.M. Boyle, D.R. Burrill, M.C. Inniss, C.M. Agapakis, A. Deardon, J.G. DeWerd *et al.*, A BioBrick compatible strategy for genetic modification of plants, *J. Biol. Eng. 6* (2012) Article No. 8.
31. H.H. Wang, F.J. Isaacs, P.A. Carr, Z.Z. Sun, G. Xu, C.R. Forest *et al.*, Programming cells by multiplex genome engi-

neering and accelerated evolution, *Nature, 460* (2009) 894–898.

32. F.J. Isaacs, P.A. Carr, H.H. Wang, M.J. Lajoie, B. Sterling, L. Kraal *et al.*, Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement, *Science, 333* (2011) 348–353.
33. D. Hranueli, N. Perić, B. Borovička, S. Bogdan, J. Cullum, P.G. Waterman *et al.*, Molecular biology of polyketide biosynthesis, *Food Technol. Biotechnol. 39* (2001) 203–213.
34. J. Zucko, N. Skunca, T. Curk, B. Zupan, P.F. Long, J. Cullum *et al.*, Polyketide synthase genes and the natural products potential of *Dictyostelium discoideum*, *Bioinformatics, 23* (2007) 2543–2549.
35. G. Castaldo, J. Zucko, S. Heidelberger, D. Vujaklija, D. Hranueli, J. Cullum *et al.*, Proposed arrangement of proteins forming a bacterial type II polyketide synthase, *Chem. Biol. 15* (2008) 1156–1165.
36. F.T. Wong, C. Khosla, Combinatorial biosynthesis of polyketides – A perspective, *Curr. Opin. Chem. Biol. 16* (2012) 117–123.
37. A. Starcevic, J. Diminic, J. Zucko, P.F. Long, J. Cullum, D. Hranueli, *3DIS*: An expert system for *in silico* drug design and discovery, *Proceedings of the 7th International Congress of Food Technologists, Biotechnologists and Nutritionists*, H. Medić (Ed.), Opatija, Croatia (2011) pp. 10–15.
38. A. Starcevic, J. Zucko, J. Simunkovic, P.F. Long, J. Cullum, D. Hranueli, *ClustScan*: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and *in silico* prediction of novel chemical structures, *Nucleic Acids Res. 36* (2008) 6882–6892.
39. J. Zucko, A. Starcevic, J. Diminic, M. Elbekali, M. Lisfi, P.F. Long *et al.*, From DNA sequences to chemical structures – Methods for mining microbial genomic and metagenomic datasets for new natural products, *Food Technol. Biotechnol. 48* (2010) 234–242.
40. A. Starcevic, J. Diminic, J. Zucko, M. Elbekali, T. Schlosser, M. Lisfi *et al.*, A novel docking domain interface model that can predict recombination between homoeologous modular biosynthetic gene clusters, *J. Ind. Microbiol. Biotechnol. 38* (2011) 1295–1304.
41. A. Starcevic, K. Wolf, J. Diminic, J. Zucko, I. Trninic Ruzic, P.F. Long *et al.*, Recombinatorial biosynthesis of polyketides, *J. Ind. Microbiol. Biotechnol. 39* (2012) 503–511.
42. Thiotemplate Modular Systems Studies, Faculty of Food Technology and Biotechnology, Zagreb, Croatia (*http://bioserv.pbf.hr/cms/*).
43. S. Donadio, P. Monciardini, M. Sosio, Polyketide synthases and nonribosomal peptide synthetases: The emerging view from bacterial genomics, *Nat. Prod. Rep. 24* (2007) 1073–1109.
44. J.D. Rojas, L.D. Sette, W.L. Araújo, M.S.G. Lopes, L.F. Silva, R.L.A. Furlan, G. Padilla, The diversity of polyketide synthase genes from sugarcane-derived fungi, *Microb. Ecol. 63* (2012) 565–577.
45. L.M. Garrido, F. Lombo, I. Braig, M. Nur-e-Alam, R.L.A. Furlan, C.C. Borda *et al.*, Insights in the glycosilation steps during biosynthesis of the antitumour anthracycline cosmomycin: Characterization of two glycosyltransferase genes, *Appl. Microbiol. Biotechnol. 73* (2006) 122–131.
46. S. Kalyaanamoorthy, Y.P. Chen, Structure-based drug design to augment hit discovery, *Drug Discov. Today, 16* (2011) 831–839.
47. S.R. Eddy, Profile hidden Markov models, *Bioinformatics, 14* (1998) 755–763.
48. A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S.R. Eddy *et al.*, The Pfam protein families database, *Nucleic Acids Res. 30* (2002) 276–280.
49. D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci. 28* (1988) 31–36.
50. Jmol, An Open-Source Java Viewer for Chemical Structures in 3D, Urbandale, IA, USA (*http://www.jmol.org/*).
51. F. Csizmadia, JChem: Java applets and modules supporting chemical database handling from web browsers, *J. Chem. Inf. Comput. Sci. 40* (2000) 323–324.
52. T.J. Hosted, R.H. Baltz, Use of *rpsL* for dominance selection and gene replacement in *Streptomyces roseosporus*, *J. Bacteriol. 179* (1997) 180–186.
53. H. Petković, J. Cullum, D. Hranueli, I.S. Hunter, N. Perić-Concha, J. Pigac *et al.*, Genetics of *Streptomyces rimosus*, the oxytetracycline producer, *Microbiol. Mol. Biol. Rev. 70* (2006) 704–728.
54. D. Hranueli, J. Cullum, B. Basrak, P. Goldstein, P.F. Long, Plasticity of the *Streptomyces genome* – Evolution and engineering of new antibiotics, *Curr. Med. Chem. 12* (2005) 1697–1704.
55. Z. Chen, M. Wilmanns, A.P. Zeng, Structural synthetic biotechnology: From molecular structure to predictable design for industrial strain development, *Trends Biotechnol. 28* (2010) 534–542.
56. A. Krivoruchko, V. Siewers, J. Nielsen, Opportunities for yeast metabolic engineering: Lessons from synthetic biology, *Biotechnol. J. 6* (2011) 262–276.
57. J. Pleiss, Protein design in metabolic engineering and synthetic biology, *Curr. Opin. Biotechnol. 22* (2011) 611–617.
58. H. Gao, Y. Zhuo, E. Ashforth, L. Zhang, Engineering of a genome-reduced host: Practical application of synthetic biology in the overproduction of desired secondary metabolites, *Protein Cell, 1* (2010) 621–626.
59. H. Neumann, P. Neumann-Staubitz, Synthetic biology approaches in drug discovery and pharmaceutical biotechnology, *Appl. Microbiol. Biotechnol. 87* (2010) 75–86.
60. V.L. Colin, A. Rodríguez, H.A. Cristóbal, The role of synthetic biology in the design of microbial cell factories for biofuel production, *J. Biomed. Biotechnol. 2011* (2011) Article ID 601834.
61. T.S. Bayer, Using synthetic biology to understand the evolution of gene expression, *Curr. Biol. 20* (2010) R772–R779.
62. M. Schmidt, L. Pei, Synthetic toxicology: Where engineering meets biology and toxicology, *Toxicol. Sci. 120* (2011) S204–224.
63. W.C. Ruder, T. Lu, J.J. Collins, Synthetic biology moving into the clinic, *Science, 333* (2011) 1248–1252.
64. S. Shankar, M.R. Pillai, Translating cancer research by synthetic biology, *Mol. Biosyst. 7* (2011) 1802–1810.
65. J.T. MacDonald, C. Barnes, R.I. Kitney, P.S. Freemont, G.B. Stan, Computational design approaches and tools for synthetic biology, *Integr. Biol. Biol. 3* (2011) 97–108.
66. T. Ellis, T. Adie, G.S. Baldwin, DNA assembly for synthetic biology: From parts to pathways and beyond, *Integr. Biol. 3* (2011) 109–118.
67. M.H. Medema, R. van Raaphorst, E. Takano, R. Breitling, Computational tools for the synthetic design of biochemical pathways, *Nat. Rev. Microbiol. 10* (2012) 191–202.
68. M.A. Marchisio, *In silico* implementation of synthetic gene networks, *Methods Mol. Biol. 813* (2012) 3–21.