



**This is a pre- or post-print of an article published in**

**Lam, C.M.C., Suárez Diez, M., Godinho, M., Martins Dos Santos, V.A.P.**

**Programmable bacterial catalysis - Designing cells for biosynthesis of value-added compounds**

**(2012) FEBS Letters, 586 (15), pp. 2184-2190.**

**Title:** Programmable Bacterial Catalysis — Designing Cells for Biosynthesis of Value-added Compounds

**Authors and full addresses:** Carolyn M.C. Lam<sup>a,b,1</sup>, María Suárez Diez<sup>b</sup>, Miguel Godinho<sup>a,c</sup>, Vítor A.P. Martins dos Santos<sup>b,\*</sup>

<sup>a</sup>Systems and Synthetic Biology Group, Helmholtz Centre for Infection Research, Inhoffenstraße 7, D-38124 Braunschweig, Germany.

<sup>b</sup>Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, Building number 316, 6703 HB Wageningen, the Netherlands.

<sup>c</sup>Lifewizz Lda, Rua Pero de Alenquer 123 7 E, 4150-616 Porto, Portugal.

<sup>1</sup> Current address: Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, Building number 316, 6703 HB Wageningen, The Netherlands.

\* Corresponding author. E-mail address: vitor.martinsdossantos@wur.nl (V.A.P. Martins dos Santos).

**Abstract:**

Bacteria have long been used for the synthesis of a wide range of useful proteins and compounds. The developments of new bioprocesses and improvements of existing strategies for syntheses of valuable products in various bacterial cell hosts have their own challenges and limitations. The field of synthetic biology has combined knowledge from different science and engineering disciplines and facilitated the advancement of novel biological components which has inspired the design of targeted biosynthesis. Here we discuss recent advances in synthetic biology with relevance to biosynthesis in bacteria and the applications of computational algorithms and tools for manipulation of cellular components. Continuous improvements are necessary to keep up with increasing demands in terms of complexity, scale, and predictability of biosynthesis products.

**Keywords:** gene circuits; chassis; genome-scale modelling; *Pseudomonas putida*; synthetic biology

## 1. Introduction

The natural abilities of many bacteria are highly exploitable for production processes to make food, medicines, and plenty of other useful compounds. As technology advanced, new applications of micro-organisms have been developed to synthesize enzymes for various catalysis, biofuels from alternative raw materials, artificial molecules such as drugs and biodegradable polymers, and to metabolize harmful environmental pollutants. Traditional cell engineering is a species-dependent task since the microbiological and genetic tools applicable to one type of cells are often dysfunctional or sub-optimal in another cell host. Therefore, it is desirable to have standardized methodologies to forward engineer suitable metabolic responses and capabilities across a wide range of species. With an increasing amount of knowledge of bacteria stemming from systems biology as well as other disciplines including molecular biology, genetics, bioinformatics, biochemistry, engineering, and computational simulation, the field synthetic biology has assembled all relevant expertise and facilitated the development of artificial biological parts and functions to serve numerous purposes. There is much potential in the solutions that can be found to solve some of the current challenges in cell culture processes regarding productivity, lack of suitable enzymes or metabolic pathways, and limitations from existing cell hosts. But much research is still needed to further understand the interactions among natural cellular components and artificial elements as well as develop appropriate computational methods for integration of the characterization, analysis, and design of biological parts to increase efficiency.

## 2. Cell hosts for biosynthesis

Owing to their outstanding versatility, bacteria are common and preferred hosts for synthesis of a wide range of chemicals and biological compounds. Enzymatic catalysis for production of target compounds is an attractive method especially when the structure of the product molecule is complex. Despite a long history of application in the industry with continuous process improvement to enhance product yields and qualities, there are still many challenges to be overcome. For example, there are general issues with low product yield, slow growth, and missing pathways for specific value-added compounds. For *Escherichia coli*, one of the commonly used bacterial strains, the post-transcriptional modification capability and metabolic capacity is limited that only a narrow range of natural proteins and compounds can be synthesized [1]. The protein products in *E. coli* sometimes are secreted into the periplasm or form inclusion bodies, which increase the difficulty for product extraction, instead of being excreted into the medium [2,3]; and the bacterium contains an endotoxin lipopolysaccharide (LPS) which, if not completely removed in purification of proteins for clinical purposes, causes fever in humans [4]. *Bacillus subtilis* is another popular bacterial strain for protein syntheses which is endotoxin-free and has high capacity of protein secretion into extracellular medium [5] but there are certain bottlenecks such as insufficient suitable expression vectors, plasmid instability, misassembled proteins, intracellular insoluble aggregates of heterologous proteins, and protein degradation by proteases that need to be addressed [6-8]. An illustration of these limitations is shown in Fig. 1. Some of the challenges are tackled by over-expression of molecular chaperones to reduce intracellular heterologous protein aggregates [6], protease gene knock-out to reduce product degradation [8], engineering of metabolic pathways [9] or membrane-bound enzymes [10] that affect product synthesis, or selecting a different cell host such as the use of *Thermus thermophilus* for production of thermophilic proteins which cannot be synthesized in *E. coli* [11], and the use of *Pseudomonas putida* for its resistance to toxic aromatic compounds [12-14]. But new cell host candidates are relatively less

characterized and require adaptation of cell engineering techniques from other micro-organisms. As the synthesis bottlenecks are also pathway-dependent, there is no single optimal strategy for all bacterial systems. Cell engineering for biosynthesis would benefit from a more systematic approach of designing cellular responses, developing suitable cell hosts, creating new metabolic pathways, and adding novel artificial functions to the biological system.

**Figure 1:** Illustration of limitations for synthesis of foreign proteins or molecules due to low product yield, slow growth, missing pathways for specific value-added compounds, limited post-transcriptional modification capability, protein products being secreted into periplasm or form inclusion bodies, endotoxin lipopolysaccharide (LPS), insufficient suitable expression vectors, plasmid instability, misassembled proteins, intracellular insoluble aggregates of heterologous proteins, and protein degradation by proteases in common bacterial hosts.

### 3. An ideal chassis for bioprocesses

Genes with unknown functions are commonplace in the sequenced genomes of many bacteria. Redundancy in genomes often poses additional challenge in predicting the behaviors of synthetic parts in the cellular system since unexpected interactions can arise. An ideal candidate for biosynthesis programming would be a well-characterized micro-environment containing only the necessary functions to synthesize the product of interest. Such a chassis, that can satisfy the production requirements of a bioprocess, may take the form of a simplified living cell, or alternatively an artificial environment capable of biosynthesis but unable to grow. The search for a standardized host in synthetic biology has been divided into the bottom-up versus top-down approaches. The bottom-up approach focuses on creating a protocell using defined components. Lipids have been used to form bilayer vesicles [15] encapsulating genetic information and transcriptional/translational machinery for protein synthesis [16]; and lipid vesicles retaining genetic materials have been shown to be thermostable at various temperature [17]. After about a decade's investigation of RNA and DNA replication and protein expression in protocells, lately more research have dedicated to the *in silico* understanding of the properties of lipid vesicles, laying down further necessary knowledge for engineering protocells. Stochastic analysis of a minimal cell model, known as a chemoton, resembling a protocell suggested convergence of the cell division time to a stable value [18]; and competing template replications can coexist when there is coupling between their reactions [19]. On the other hand, existing organisms already possess all elements necessary for life. Thus, in the top-down approach redundant genes in micro-organisms are identified and removed to generate mutants with smaller genomes. Several experimental genome minimization studies have reduced the genomes of a few bacteria by different extent, such as a recent study of the human and amoeba pathogen *Legionella pneumophila* Philadelphia 1 which have led to a strain with 18.5% of the original 3.4 Mbp genome removed, but nonetheless it is able to grow on bacteriological culture [20]; an improvement in expression of recombinant enzymes in a mutant of *B. subtilis* strain 168 with 20.7% of the 4.2 Mbp genome removed [21]; and a reduction of the 6.2 Mbp genome of the versatile soil bacterium *P. putida* by 7.4% [22]. Progresses in gene deletion tools, including recent examples of a vector platform developed by Martínez-García and de Lorenzo which allows precise and accumulative gene deletion in a wide range of Gram-negative bacteria with successful application to study the multi-resistant antibiotic profile of *P. putida* [23] and a method based on the Flp-*FRT* recombination system and minitransposon which enables rapid random large-scale genomic deletion [24], have significantly increased the efficiency in

genome minimization. From the trend that protocells are gaining capability in handling biochemical processes with increasing complexity, and genomes of bacteria are being reduced to simplify the metabolic system, it is conceivable that both approaches can provide candidates with high applicability in synthetic biology (Fig.2). A construct, or a chassis, which can provide an optimal and robust environment to host synthetic circuitries to support any desirable biosynthetic processes will be a very valuable tool for implementing synthetic cellular functions. The bottom-up approach delivers protocells that are becoming more usable for production of proteins, enzymes, and complex molecules as the number of basic functions related to self-maintenance and synthesis/transportation of the target molecules is increased (Fig.2). Protocells have the advantage of simplicity in structure, which is attractive for biosynthesis of simple molecules. But the addition of functions to protocell is likely to reach a plateau eventually as a result of high marginal costs over further increase in functionality. On the other hand, as the genome of a living cell is continuously reduced, the degree of minimization is limited by the cardinality of the smallest subset, or subsets, of genes which are required to maintain a minimal living and reproducible cell under different nutrient availability (Fig.2). The relationship between genome size and nutrient dependency can be seen from the reliance on amino acids, purines, and pyrimidines from the surroundings for the symbiotic *Mycoplasma pneumoniae* with a 0.8 Mbp genome but not for the cyanobacterium *Prochlorococcus marinus* which has a genome of 1.75 Mbp [25]. The minimum number of genes estimated to be essential for cell growth is so far dependent on the organisms studied. Comparison between *Mycoplasma genitalium* and *Haemophilus influenzae* suggested that 256 genes are close to the minimal to support a living cell [26] but 271 indispensable genes were identified for *B. subtilis* [27]. It has also been proposed that 151 genes are sufficient to build a minimum cell [28]. Since analyses of sequenced genomes revealed a lack of ubiquitous set of genes for core functions, it is likely that there are orthologous genes sharing the same cellular functions [29,30]. Overall, the chassis obtained from each approach is likely to be suitable for specific types of bioprocess. Since high-throughput techniques are starting to provide quantitative data for the construction and validation of whole-system models, such as the measurements of the transcriptome [31], proteome [32], and metabolome [33] of *M. pneumoniae*, the development of standardized chassis implies not only the availability of suitable biological constructs to host the introduced circuits but also the *in silico* construction of standard quantitative models which are able to reproduce biological behaviors and predict new properties for each of the chosen chassis.

**Figure 2:** Conceptual projection of the applicability of artificial cellular constructs based on the bottom-up and top-down approaches as suitable chassis for biosynthetic processes. Protocells and streamlined bacterial genomes are becoming more useful for production of value-added compounds when the functionality of protocells continues to grow and redundant genes in existing and new bacterial hosts are gradually being removed. It is not clear what the maximum applicability that can be reached by each of the approaches will be. In the figure both approaches are shown to reach a similar maximum applicability to reflect the fact that most likely each of the approaches will produce optimal chassis for different types of process.

#### 4. Designing biological circuits and metabolic pathways

Gene networks play an important role in determining the behavior of a cell. The notions of logic gates and circuits have gained popularity in describing gene networks since the former allow engineering methodologies for systematic network design to be conceptually transferred

into biological systems. Most of the biological circuits developed earlier in synthetic biology involve DNA transcriptional control where activating or repressing relationships are designed among transcription factors, inhibitors, promoter, and upstream operator, producing transcription responses resembling Boolean logics and logic gates with AND, OR, NOT, NAND, NOR, or more complex input-output relationships. It has been demonstrated recently that the normal quorum sensing mechanism in bacteria can be linked to a synthetic gene circuit to stimulate over-expression of recombinant proteins under high cell density [34]. *E. coli* continues to serve as a popular host for new circuit applications, such as a synthetic modular electron transfer circuit which facilitated hydrogen production [35], and a synthetic reporter protein specific for copper ions has transformed the bacterium into an automatic copper remover [36]. Other bacterial hosts have gradually gained applicability in circuit design, including a hybrid genetic circuit for investigating protein secretion in *Salmonella typhimurium* SL1344 [37]; and the TOL circuit in *P. putida* mt2 which has been modelled as a set of logic gates to understand the growth patterns in mixed carbon sources [38,39] and a metabolic amplifier motif has been revealed from logic modelling of the circuit [40]. Circuits involving RNAs are also becoming popular in synthetic bacterial circuits [41,42]. In recent years, designs of genetic circuits have become automated using libraries of genetic elements [43] and rational approach for specified circuit output properties without involving complex optimization algorithms [44]. The design, analysis, and DNA sequence inference of gene circuits have gained efficiency from a computer-aided circuit design tool [45,46] and an algorithm generating DNA sequence from circuit connection [47]. As a large amount of artificial biological circuits have been developed which can serve to construct a wide range of input-output relationship, the need of standardization to promote cross-applicability of the circuit parts is becoming stronger. It is also important that the biological circuits are well-characterized, particularly since there has not been a single standard for the development of biological parts being globally adopted by most designed circuits. Unknown interaction between synthetic biological circuit and host [48] and influence from intricate parameters which are usually not considered in circuit design and modelling often hamper the intended feature of the artificial circuit. For example, a suicide circuit for *E. coli* induced unexpected population oscillation due to unknown feedback between cell density and plasmid amplification [49]. In some occasions, an exhaustive analysis of the properties of the underlying cellular system is required to explain, for instance, the fluctuation of intracellular protein concentration which can have subtle effect on exogenous circuit. An example can be found from the interactions between cell cycle and oscillations of cell volume and protein concentration in yeast [50]. It is necessary to have the characteristics of synthetic gene circuits structurally organised to facilitate a wide scope of application.

Among the possible biological circuits that can be designed and introduced into a cell host, those leading to the productions of valuable molecules using the cellular machinery of the hosts are highly relevant for improvement of existing bioprocesses. Natural organisms evolve to adapt to the environment but their goals can be far from optimizing metabolic pathways of industrial interest. The ability to artificially design new enzymes and metabolic networks in a systematic way has revolutionized the manipulation of metabolism in cells. Extensive computational analysis is gradually adopted into the design process of novel proteins. For example, a multi-objective computational design approach is able to extend protein promiscuity and to endow the thioredoxin from *E. coli* with a promiscuous esterase function while maintaining the native oxidoreductase activity [51]. For the production of compounds for which no natural pathways have been elucidated, feasible solutions can be predicted through a retro-biosynthetic approach similar to the retro-synthesis method developed in organic chemistry that the metabolic pathway leading to the synthesis of a target compound is

specified by considering the biotransformation of functional groups rather than the entire structure, assuming the availability of enzymes for the desired transformation. The retro-biosynthetic approach may result in numerous candidate pathways which need to be further selected to identify the most promising option. Carbonell et al. and Cho et al. have proposed different prioritizing strategies and tested them with the prediction of novel biosynthesis pathways for isobutanol, 3-hydroxypropionate, butyryl-CoA [52], penicillin G, and cephalosporin [53] in *E. coli*. The substrate promiscuity property of already known enzymes may present a source for the required catalysts of retro-biosynthesis. Advances in protein design also make this an increasingly likely proposition [51,54]. Alternatively, utilizing enzymes directly from other organisms offers operable results, such as the construction of a synthetic pathway for the conversion of glucuronic acid to glucaric acid from glucose in *E. coli* by assembling genes originally from *Saccharomyces cerevisiae*, mice, and *Pseudomonas syringae* [55]. Constructions of missing pathways in micro-organisms have also been sped up by computational selection of enzyme-coding genes from other species [56]. Computational prediction of new protein interactions [57], and algorithmic creation of new enzymes either by using pure computational tools [58] or by combining these tools with *in vitro* directed evolution techniques [59] have considerably increased the efficacy of extracting correlations from experimental data and screening feasible candidates for new metabolic pathway and protein structures to reduce the time and cost of experimental trials. Traditional optimization of biological components often involve directed evolution where experimental random alterations create a library of variants which are then screened for the required properties or further evolved under specific conditions to achieve target functions. But the accumulation of experimental data has gradually enabled development of computational models and algorithms which represent known biological interactions and can be used to discover hidden relationships and novel system properties. Computational selection of the designs of proteins, enzymes, and metabolic pathways is becoming more indispensable and will continue to reduce experimentation time as new experimental data are fed backwards into models and algorithms to improve prediction accuracy (Fig. 3). Although bioinformatics has already delivered a broad range of tools essential for the engineering of metabolic pathways, the collection of computational tools that synthetic biology has at its disposal is still fairly fragmented and lacks an integrative solution. As synthetic biology matures, those computational tools need to be shaped and aggregated in ways that permit a smooth workflow and, consequently, higher efficiencies in bioprocess development.

**Figure 3:** Trend of increasing involvement of computational analysis and optimization in the designs of synthetic biological circuit/metabolic pathway connectivity, protein structure, DNA sequence, and other cellular parts which are involved in biosynthesis.

Bacteria have been increasingly used to synthesize foreign compounds and proteins, such as drug molecules, biofuel, and proteins containing unnatural amino acids. Various terpenoid precursors have been successfully synthesized in *E. coli* by, for example, engineering the taxadiene metabolic pathway [60] or the deoxy-xylulose 5-phosphate pathway [61], providing general platforms to produce terpenoids with higher specific productivity than the original plant sources. Bacteria are also used to synthesize other valuable metabolites such as benzyloquinoline alkaloids (BIAs) which can be used for anti-cancer, anti-oxidative, and anti-HIV therapeutics [62], and polyketides with antibiotic functions [63]. Much effort has also been put into synthesizing alkanols, biodiesel, and hydrocarbons in *E. coli* [64-68]. An alkane synthesis pathway in cyanobacteria has been shown to be transferrable into *E. coli* [69], illustrating the possibility of expressing useful biofuel pathways in common cell hosts. Genome-scale metabolic models of the cyanobacteria *Arthrospira (Spirulina) maxima* [70]

and *Synechocystis* [71,72] have been developed to find conditions optimizing hydrogen photoproduction. Comparison between genome-scale metabolic models of various strains are often hindered by differences inherited from their distinct reconstruction methods, but recently a metabolic network reconciliation process has been proposed to identify true biological differences among metabolic models of different strains [73]. Unnatural amino acids (UAAs) have enabled anchoring additional functions onto protein, controlling substrate specificity of an enzyme [74], and significant enhancement of enzymatic activity [75]. Pairs of aminoacyl-tRNA synthetase/tRNA from archaea were found to be orthogonal in *E. coli* and increased the specificity of UAAs incorporation [76]. Borrowing enzymes from other bacteria is an alternative for synthesis of UAA, as in the example of L-homoalanine produced in *E. coli* by incorporating enzymes from *Bacillus* and *Streptomyces* species [77]. A similar strategy has also been adopted to synthesize an effective biosurfactant rhamnolipid in *P. putida* using enzymes originally from pathogenic *P. aeruginosa*, resulting in high production yield and better growth resistance against the antimicrobial effect of the rhamnolipid than other bacteria such as *E. coli* or *B. subtilis* [78]. But the design of effective metabolic pathways to produce foreign compounds, including the utilization of artificial amino acids to generate novel enzymes for new pathways, is relatively less systematic than the development of biological circuit components. Each type of cell host has its unique limitations and the product structures also affect the synthesis bottlenecks, thus the optimal metabolic adjustment for a new compound or a different host is hard to be predicted in advance. The complexity, incomplete understanding of potential cell hosts, and limitation in the natural cell machinery are hurdles to be overcome for the future of bacterial biosynthesis. The demand of biosynthesis of valuable materials and metabolites using bacteria, particularly for simpler cell hosts to produce natural biological products or for robust hosts to synthesize compounds and macromolecules, is growing ever faster. Successful pathway designs have brought considerable reduction in the costs and increase in the supply of many raw materials, drugs, additives, and catalysts. Further improvement in the prediction of cellular metabolic performance will speed up the time required for searching new optimal pathways and enhancing the property of novel molecules which are constantly being designed or discovered to improve living standards.

## 5. Future perspectives

Bacterial systems are promising environments to deliver biosynthesis of a large variety of products, from (semi-)bulk compounds such as organic acids or precursors of bioplastics to fine chemicals and pharmaceuticals. Some applications have already found their ways into the production processes of valuable compounds; and the control as well as detection of specific biological or biochemical reactions have been demonstrated to be highly tunable using synthetic circuits. From the current development of logic gates in bacterial cells, it is foreseeable that gene circuits will be able to selectively adjust or activate metabolic pathways in a robust manner to enable flexible manipulation of biosynthesis. Well-defined and stable chassis in the form of streamlined bacteria or protocells will provide more options for expression of novel or foreign enzymatic routes, which can be pre-selected and optimized through *in silico* analysis. Computational involvement in designing synthetic biological processes will take a more significant role when the accuracy of predictions continues to be improved by refined knowledge of the biological system. A deeper understanding of various bacterial hosts or other suitable chassis and a greater capability to rationally design biological functions empowered by synthetic biology will help to improve bio-industrial processes, such as increasing enzyme activity, changing substrate specificities of enzymes, using alternative

hosts for biosynthesis, creating new metabolic or transportation pathways, faster selection of enzymes for tailor-made functions etc. Further integration between experimentation and computational prediction and analysis is important in order to harvest the synergy to efficiently gain accurate and detailed knowledge of the cellular complexity which is still not completely understood. Synthetic biology is ever more starkly influencing the evolution of bioproduction processes. There are still many challenges ahead, in particular for complex networks or new cell hosts regarding the fact that a synthetic unit may not function fully as predicted in the biological system. Further investigations are necessary to search for appropriate solutions to current bottlenecks and explore for more applications with the synthetic biosynthesis capability of bacteria.

## Acknowledgements

Miguel Godinho acknowledges funding from the Portuguese Fundação para a Ciência e Tecnologia (Project METAGUT; Grant No. ERA-PTG/SAU/0003/2008). Vítor Martins dos Santos acknowledges the financial support from the European Union under the project Microme (FP7-Grant Agreement Number 222886-2).

## References

- [1] Zhang, H., Wang, Y. and Pfeifer, B.A. (2008). Bacterial hosts for natural product production. *Molecular Pharmaceutics* 5, 212-225.
- [2] Arié, J.-P., Miot, M., Sassoon, N. and Betton, J.-M. (2006). Formation of active inclusion bodies in the periplasm of *Escherichia coli*. *Molecular Microbiology* 62, 427-437.
- [3] Idicula-Thomas, S., Kulkarni, A.J., Kulkarni, B.D., Jayaraman, V.K. and Balaji, P.V. (2006). A support vector machine-based method for predicting the propensity of a protein to be soluble or to form inclusion body on overexpression in *Escherichia coli*. *Bioinformatics* 22, 278-284.
- [4] Svensson, M., Han, L., Silfversparre, G., Häggström, L. and Enfors, S.-O. (2005). Control of endotoxin release in *Escherichia coli* fed-batch cultures. *Bioprocess and Biosystems Engineering* 27, 91-97.
- [5] Schallmeyer, M., Singh, A. and Ward, O.P. (2004). Developments in the use of *Bacillus* species for industrial production. *Canadian Journal of Microbiology* 50, 1-17.
- [6] Westers, L., Westers, H. and Quax, W.J. (2004). *Bacillus subtilis* as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochimica et Biophysica Acta - Molecular Cell Research* 1694, 299-310.
- [7] Zweers, J.C. et al. (2008). Towards the development of *Bacillus subtilis* as a cell factory for membrane proteins and protein complexes. *Microbial Cell Factories* 7, 10.
- [8] Li, W., Zhou, X. and Lu, P. (2004). Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Research in Microbiology* 155, 605-610.
- [9] Leonard, E., Yan, Y., Fowler, Z.L., Li, Z., Lim, C.-G., Lim, K.-H. and Koffas, M.A.G. (2008). Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Molecular Pharmaceutics* 5, 257-265.
- [10] Leonard, E. and Koffas, M.A.G. (2007). Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Applied and Environmental Microbiology* 73, 7246-7251.
- [11] Hidalgo, A., Betancor, L., Moreno, R., Zafra, O., Cava, F., Fernandez-Lafuente, R., Guisan, J.M. and Berenguer, J. (2004). *Thermus thermophilus* as a cell factory for the production of a thermophilic Mn-dependent catalase which fails to be synthesized in an active form in *Escherichia coli*. *Applied and Environmental Microbiology* 70, 3839-3844.
- [12] Martins dos Santos, V.A.P., Heim, S., Moore, E.R.B., Strätz, M. and Timmis, K.N. (2004). Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environmental Microbiology* 6, 1264-1286.

- [13] Ramos, J.L., Krell, T., Daniels, C., Segura, A. and Duque, E. (2009). Responses of *Pseudomonas* to small toxic molecules by a mosaic of domains. *Current Opinion in Microbiology* 12, 215-220.
- [14] Domínguez-Cuevas, P., González-Pastor, J.E., Marqués, S., Ramos, J.L. and De Lorenzo, V. (2006). Transcriptional tradeoff between metabolic and stress-response programs in *Pseudomonas putida* KT2440 cells exposed to toluene. *Journal of Biological Chemistry* 281, 11981-11991.
- [15] Mansy, S.S., Schrum, J.P., Krishnamurthy, M., Tobe, S., Treco, D.A. and Szostak, J.W. (2008). Template-directed synthesis of a genetic polymer in a model protocell. *Nature* 454, 122-125.
- [16] Doktycz, M.J. and Simpson, M.L. (2007). Nano-enabled synthetic biology. *Molecular Systems Biology* 3, 125.
- [17] Mansy, S.S. and Szostak, J.W. (2008). Thermostability of model protocell membranes. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13351-13355.
- [18] Van Segbroeck, S., Nowé, A. and Lenaerts, T. (2009). Stochastic simulation of the chemoton. *Artificial Life* 15, 213-226.
- [19] Zachar, I., Fedor, A. and Szathmáry, E. (2011). Two different template replicators coexisting in the same protocell: Stochastic simulation of an extended chemoton model. *PLoS One* 6, e21380.
- [20] O'Connor, T.J., Adepoju, Y., Boyd, D. and Isberg, R.R. (2011). Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. *Proceedings of the National Academy of Sciences of the United States of America* 108, 14733 -14740.
- [21] Manabe, K. et al. (2011). Combined effect of improved cell yield and increased specific productivity enhances recombinant enzyme production in genome-reduced *Bacillus subtilis* strain MGB874. *Applied and Environmental Microbiology* 77, 8370-8381.
- [22] Leprince, A., de Lorenzo, V., Völler, P., van Passel, M.W. and Martins dos Santos, V.A. (2012). Random and cyclical deletion of large DNA segments in the genome of *Pseudomonas putida*. *Environmental Microbiology* (accepted).
- [23] Martínez-García, E. and de Lorenzo, V. (2011). Engineering multiple genomic deletions in Gram-negative bacteria: Analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environmental Microbiology* 13, 2702–2716.
- [24] Leprince, A., Janus, D., de Lorenzo, V. and Martins dos Santos, V. (2012). Streamlining of a *Pseudomonas putida* genome using a combinatorial deletion method based on minitransposon insertion and the Flp-*FRT* recombination system. *Methods in Molecular Biology* 813, 249-266.
- [25] DeWall, M.T. and Cheng, D.W. (2011). The minimal genome - A metabolic and environmental comparison. *Briefings in Functional Genomics* 10, 312-315.
- [26] Mushegian, A.R. and Koonin, E.V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proceedings of the National Academy of Sciences of the United States of America* 93, 10268-10273.
- [27] Kobayashi, K. et al. (2003). Essential *Bacillus subtilis* genes. *Proceedings of the National Academy of Sciences of the United States of America* 100, 4678-4683.
- [28] Forster, A.C. and Church, G.M. (2006). Towards synthesis of a minimal cell. *Molecular Systems Biology* 2, 45.
- [29] Koonin, E.V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Reviews Microbiology* 1, 127-136.

- [30] Carbone, A. (2006). Computational prediction of genomic functional cores specific to different microbes. *Journal of Molecular Evolution* 63, 733-746.
- [31] Guell, M. et al. (2009). Transcriptome complexity in a genome-reduced bacterium. *Science* 326, 1268-1271.
- [32] Kuhner, S. et al. (2009). Proteome organization in a genome-reduced bacterium. *Science* 326, 1235-1240.
- [33] Yus, E. et al. (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326, 1263-1268.
- [34] Tsao, C.Y., Hooshangi, S., Wu, H.C., Valdes, J.J. and Bentley, W.E. (2010). Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of *E. coli*. *Metabolic Engineering* 12, 291-297.
- [35] Agapakis, C.M. and Silver, P.A. (2010). Modular electron transfer circuits for synthetic biology Insulation of an engineered biohydrogen pathway. *Bioengineered Bugs* 1, 413-418.
- [36] Ravikumar, S., Yoo, I.-K., Lee, S. and Hong, S. (2011). Construction of copper removing bacteria through the integration of two-component system and cell surface display. *Applied Biochemistry and Biotechnology*, (Online First, 23 Sept. 2011).
- [37] Widmaier, D.M. and Voigt, C.A. (2010). Quantification of the physiochemical constraints on the export of spider silk proteins by *Salmonella* type III secretion. *Microbial Cell Factories* 9, 78.
- [38] Koutinas, M. et al. (2011). Improving the prediction of *Pseudomonas putida* mt-2 growth kinetics with the use of a gene expression regulation model of the TOL plasmid. *Biochemical Engineering Journal* 55, 108-118.
- [39] Koutinas, M., Kiparissides, A., Silva-Rocha, R., Lam, M.C., Martins dos Santos, V.A.P., de Lorenzo, V., Pistikopoulos, E.N. and Mantalaris, A. (2011). Linking genes to microbial growth kinetics — An integrated biochemical systems engineering approach. *Metabolic Engineering* 13, 401-413.
- [40] Silva-Rocha, R., de Jong, H., Tamames, J. and de Lorenzo, V. (2011). The logic layout of the TOL network of *Pseudomonas putida* pWW0 plasmid stems from a metabolic amplifier motif (MAM) that optimizes biodegradation of *m*-xylene. *BMC Systems Biology* 5, 191.
- [41] Saito, H. and Inoue, T. (2009). Synthetic biology with RNA motifs. *International Journal of Biochemistry & Cell Biology* 41, 398-404.
- [42] Win, M.N., Liang, J.C. and Smolke, C.D. (2009). Frameworks for programming biological function through RNA parts and devices. *Chemistry and Biology* 16, 298-310.
- [43] Rodrigo, G., Carrera, J. and Jaramillo, A. (2011). Computational design of synthetic regulatory networks from a genetic library to characterize the designability of dynamical behaviors. *Nucleic Acids Research* 39, e138.
- [44] Marchisio, M.A. and Stelling, J. (2011). Automatic design of digital synthetic gene circuits. *PLoS Computational Biology* 7, e1001083.
- [45] Chandran, D., Bergmann, F.T. and Sauro, H.M. (2009). TinkerCell: Modular CAD tool for synthetic biology. *Journal of Biological Engineering* 3, 19.
- [46] Chandran, D., Bergmann, F.T. and Sauro, H.M. (2010). Computer-aided design of biological circuits using TinkerCell. *Bioengineered Bugs* 1, 274-281.
- [47] Misirli, G., Hallinan, J.S., Yu, T., Lawson, J.R., Wimalaratne, S.M., Cooling, M.T. and Wipat, A. (2011). Model annotation for synthetic biology: Automating model to nucleotide sequence conversion. *Bioinformatics* 27, 973-979.
- [48] Nandagopal, N. and Elowitz, M.B. (2011). Synthetic biology: Integrated gene circuits. *Science* 333, 1244-1248.

- [49] Marguet, P., Tanouchi, Y., Spitz, E., Smith, C. and You, L. (2010). Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology. *PLoS One* 5, e11909.
- [50] Cookson, N.A., Cookson, S.W., Tsimring, L.S. and Hasty, J. (2010). Cell cycle-dependent variations in protein concentration. *Nucleic Acids Research* 38, 2676-2681.
- [51] Suárez, M., Tortosa, P., Garcia-Mira, M.M., Rodríguez-Larrea, D., Godoy-Ruiz, R., Ibarra-Molero, B., Sanchez-Ruiz, J.M. and Jaramillo, A. (2010). Using multi-objective computational design to extend protein promiscuity. *Biophysical Chemistry* 147, 13-19.
- [52] Cho, A., Yun, H., Park, J.H., Lee, S.Y. and Park, S. (2010). Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Systems Biology* 4, 35.
- [53] Carbonell, P., Planson, A.G., Fichera, D. and Faulon, J.L. (2011). A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Systems Biology* 5, 122.
- [54] Lippow, S.M. et al. (2010). Engineering enzyme specificity using computational design of a defined-sequence library. *Chemistry and Biology* 17, 1306-1315.
- [55] Moon, T.S., Yoon, S.H., Lanza, A.M., Roy-Mayhew, J.D. and Prather, K.L.J. (2009). Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*. *Appl. Environ. Microbiol.* 75, 589-595.
- [56] Chou, C.H., Chang, W.C., Chiu, C.M., Huang, C.C. and Huang, H.D. (2009). FMM: A web server for metabolic pathway reconstruction and comparative analysis. *Nucleic Acids Research* 37, W129-W134.
- [57] Kiel, C., Beltrao, P. and Serrano, L. (2008). Analyzing protein interaction networks using structural information. *Annual Review of Biochemistry* 77, 415-441.
- [58] Jiang, L. et al. (2008). *De novo* computational design of retro-aldol enzymes. *Science* 319, 1387-1391.
- [59] Khersonsky, O. et al. (2011). Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *Journal of Molecular Biology* 407, 391-412.
- [60] Ajikumar, P.K. et al. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70-74.
- [61] Meng, H., Wang, Y., Hua, Q., Zhang, S. and Wang, X. (2011). *In silico* analysis and experimental improvement of taxadiene heterologous biosynthesis in *Escherichia coli*. *Biotechnology and Bioprocess Engineering* 16, 205-215.
- [62] Nakagawa, A., Minami, H., Kim, J.S., Koyanagi, T., Katayama, T., Sato, F. and Kumagai, H. (2011). A bacterial platform for fermentative production of plant alkaloids. *Nature Communications* 2, 326.
- [63] Menzella, H.G., Carney, J.R., Yong, L. and Santi, D.V. (2010). Using chemobiosynthesis and synthetic mini-polyketide synthases to produce pharmaceutical intermediates in *Escherichia coli*. *Applied and Environmental Microbiology* 76, 5221-5227.
- [64] Hoover, S.W. et al. (2011). Bacterial production of free fatty acids from freshwater macroalgal cellulose. *Applied Microbiology and Biotechnology* 91, 435-446.
- [65] Smith, K.M. and Liao, J.C. (2011). An evolutionary strategy for isobutanol production strain development in *Escherichia coli*. *Metabolic Engineering* 13, 674-681.
- [66] Trinh, C.T., Li, J., Blanch, H.W. and Clark, D.S. (2011). Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. *Applied and Environmental Microbiology* 77, 4894-4904.

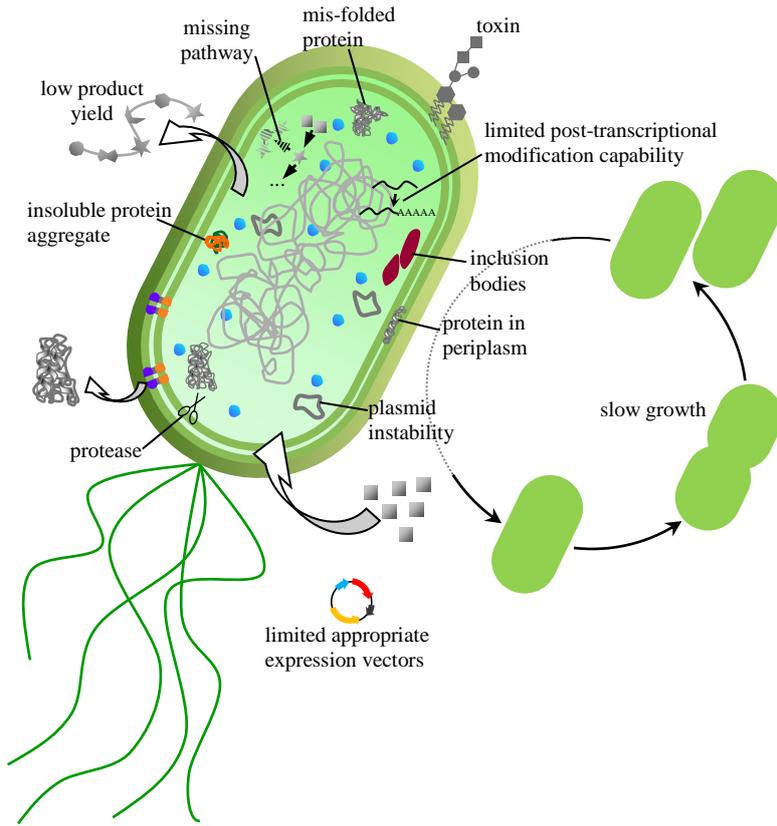
- [67] Dellomonaco, C., Clomburg, J.M., Miller, E.N. and Gonzalez, R. (2011). Engineered reversal of the  $\beta$ -oxidation cycle for the synthesis of fuels and chemicals. *Nature* 476, 355-359.
- [68] Duan, Y., Zhu, Z., Cai, K., Tan, X. and Lu, X. (2011). *De novo* biosynthesis of biodiesel by *Escherichia coli* in optimized fed-batch cultivation. *PLoS ONE* 6, e20265.
- [69] Schirmer, A., Rude, M.A., Li, X., Popova, E. and Del Cardayre, S.B. (2010). Microbial biosynthesis of alkanes. *Science* 329, 559-562.
- [70] Ananyev, G., Carrieri, D. and Dismukes, G.C. (2008). Optimization of metabolic capacity and flux through environmental cues to maximize hydrogen production by the cyanobacterium '*Arthrospira (Spirulina) maxima*' *Applied and Environmental Microbiology* 74, 6102-6113.
- [71] Montagud, A., Navarro, E., Fernández de Córdoba, P., Urchueguía, J.F. and Patil, K.R. (2010). Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology* 4, 156.
- [72] Montagud, A., Zelezniak, A., Navarro, E., de Córdoba, P.F., Urchueguía, J.F. and Patil, K.R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnology Journal* 6, 330-342.
- [73] Oberhardt, M.A., Puchalka, J., Martins dos Santos, V.A.P. and Papin, J.A. (2011). Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. *PLoS Computational Biology* 7, e1001116.
- [74] Mehta, K.R., Yang, C.Y. and Montclare, J.K. (2011). Modulating substrate specificity of histone acetyltransferase with unnatural amino acids. *Molecular BioSystems* 7, 3050-3055.
- [75] Ugwumba, I.N. et al. (2011). Improving a natural enzyme activity through incorporation of unnatural amino acids. *Journal of the American Chemical Society* 133, 326-333.
- [76] Armen, R.S., Schiller, S.M. and Brooks Iii, C.L. (2010). Steric and thermodynamic limits of design for the incorporation of large unnatural amino acids in aminoacyl-tRNA synthetase enzymes. *Proteins: Structure, Function and Bioinformatics* 78, 1926-1938.
- [77] Zhang, K., Li, H., Cho, K.M. and Liao, J.C. (2010). Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proceedings of the National Academy of Sciences of the United States of America* 107, 6234-6239.
- [78] Wittgens, A. et al. (2011). Growth independent rhamnolipid production from glucose using the non-pathogenic *Pseudomonas putida* KT2440. *Microbial Cell Factories* 10, 80.

### **Figure captions and figures**

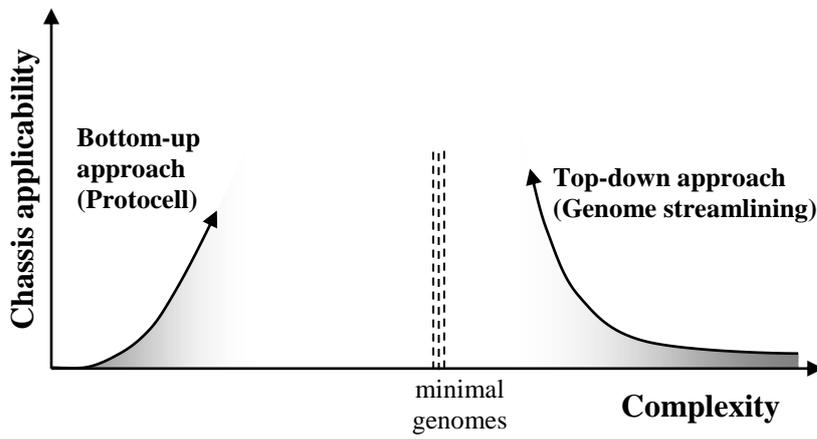
**Figure 1:** Illustration of limitations for synthesis of foreign proteins or molecules due to low product yield, slow growth, missing pathways for specific value-added compounds, limited post-transcriptional modification capability, protein products being secreted into periplasm or form inclusion bodies, endotoxin lipopolysaccharide (LPS), insufficient suitable expression vectors, plasmid instability, misassembled proteins, intracellular insoluble aggregates of heterologous proteins, and protein degradation by proteases in common bacterial hosts.

**Figure 2:** Conceptual projection of the applicability of artificial cellular constructs based on the bottom-up and top-down approaches as suitable chassis for biosynthetic processes. Protocells and streamlined bacterial genomes are becoming more useful for production of value-added compounds when the functionality of protocells continues to grow and redundant genes in existing and new bacterial hosts are gradually being removed. It is not clear what the maximum applicability that can be reached by the each of the approaches will be. In the figure both approaches are shown to reach a similar maximum applicability to reflect the fact that most likely each of the approaches will produce optimal chassis for different types of process.

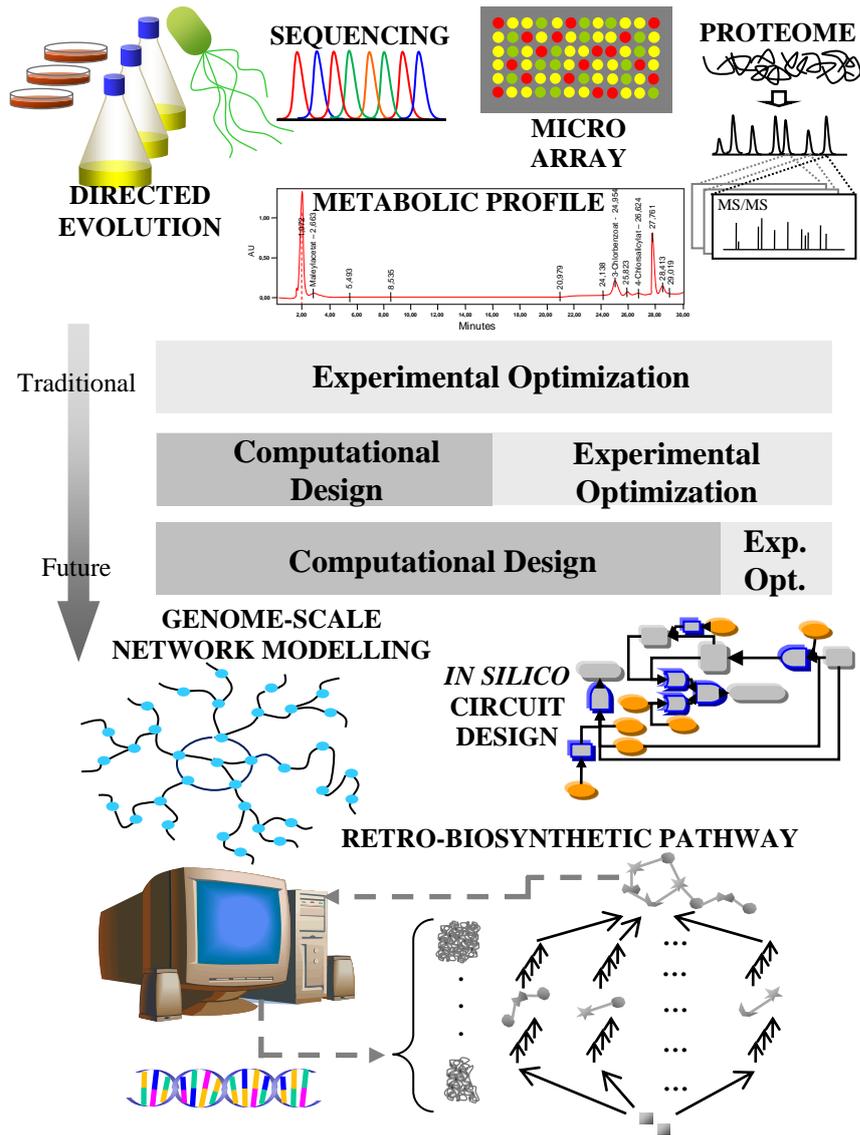
**Figure 3:** Trend of increasing involvement of computational analysis and optimization in the designs of synthetic biological circuit/metabolic pathway connectivity, protein structure, DNA sequence, and other cellular parts which are involved in biosynthesis.



**Figure 1**



**Figure 2**



**Figure 3**