Experimental tools to reduce the burden of bacterial synthetic biology

Short title: Synthetic Biology tools for reduced burden

Grob Alice1,2*, Di Blasi Roberto1,2*, Ceroni Francesca1,2*

1. Department of Chemical Engineering, Imperial College London, South Kensington Campus, London (UK)
2. Imperial College Centre for Synthetic Biology, South Kensington Campus, London (UK)

* these authors contributed equally to this work
* correspondence: f.ceroni@imperial.ac.uk

Abstract

Cellular burden limits the applications of bacterial synthetic biology. Experimental approaches for burden minimisation have recently become available. Tools to identify construct design with low footprint on the host include capacity monitors that quantify cellular capacity, high-throughput approaches and cell-free systems for construct prototyping. Orthogonal ribosomes and feedback controllers are instead useful to seek control of resource allocation and lower burden. Other approaches include genome reduction to increase the available resource pool and synthetic addiction to couple cell fitness and product accumulation. However, controlling the cellular response to exogenous expression is still a challenge, and more tools are needed in order to widen the applications of synthetic biology. Further effort that combines novel evolutionary data with burden-aware tools can set the foundation to increase the stability and robustness of future genetic systems.

Introduction

The existence of context dependencies and their impact on the behaviour and predictability of synthetic systems has become clear from the early years of synthetic biology[1-4]. Reports describing the finiteness of cellular resources in bacteria have drawn attention to the strategies that cells use to allocate these resources to different tasks and how heterologous expression can disrupt such regulation to favour synthetic construct expression[5,6]. The impact of resource competition and burden on the functionality of synthetic constructs has been extensively reported and characterised in the field[7-9]. Strategies to ease burden and its impact have focused on rewiring the native cellular machinery[10], the design of orthogonal systems for controlled allocation of cellular resources between endogenous and exogenous genes[11-13], identification of low-burden designs[14], adoption of biocontrollers to balance cellular fitness and exogenous expression[15-17] as well as less complex systems where gene expression resources are separated from the native cellular context[18-
While also a plethora of computational approaches is now available for host-aware bacterial engineering [21-26], in this review we aim at providing a schematic overview of some more recent experimental strategies adopted to reduce burden in bacteria, describing their advantages and limitations. From these examples it will become clear that we have improved in our ability to take burden into account at the design stage, even if we are still facing uncertainties in controlling the response of engineered cells. Deeper knowledge on how cells evolve in response to burden together with an expanded repertoire of tools for a growing number of organisms are awaited to guide the possible future directions of the field.

**Design and characterisation of parts with lower footprint on the host.** In considering experimental tools developed to design burden-aware synthetic systems, we would like to start by looking at methods that support part characterisation and improvement, based on their impact on the host and its resources. Ceroni et al [14] presented a fluorescent-based, genome integrated *capacity monitor* that functions as a proxy for the gene expression capacity (e.g. transcription and translation resources) in *Escherichia coli* (Figure 1A). The authors demonstrated that the *capacity monitor* is an effective tool for tracking the decrease in cellular capacity and gene expression resources caused by exogenous expression. The monitor also aided the identification of best-performing designs that support high production and low burden, and can therefore be very informative at the construct design stage. More recently, Gorochowski et al [27] were able to quantify translation regulation as well as burden by comparing ribosome distribution between synthetic and endogenous genes combining ribosome sequencing with quantitative RNA-seq. This approach allows high-throughput measurements of genetic parts behaviour and their effects on the host cell, which is valuable to elucidate the source of a circuit failure and thus reduces the need to engineer multiple genetic circuit designs to identify the most efficient ones.

In an effort to identify construct designs that excel for performance and impact on the host, changing one parameter at a time can be time consuming. Design of Experiment (DoE) can thus prove helpful in highlighting design rules predictive of reduced burden within a more efficient process [28]. Cambray et al [29] used a high-throughput DoE method to design and evaluate the impact of nucleotide, secondary structure and amino acid properties of 244,000 DNA sequence variations on translation efficiency (Figure 1B). The library was systematically evaluated by reporter transcripts whose abundance, decay, ribosome profiling and protein production were closely monitored together with impact on cellular growth rate. The authors hence highlighted the key role of mRNA secondary structures in efficient translation initiation and burden as a result of ribosome sequestration. However, such a comprehensive approach could only account for 53% of the total variance in protein production, highlighting that fundamental mechanisms in that matter are yet to be discovered. A
powerful platform to avoid native cellular complexity in the characterisation of genetic systems is represented by cell-free environments. Borkowski et al.[30] developed a cell free-based computational approach to predict the burden imposed by synthetic constructs when expressed in vivo in E. coli (Figure 1C). First, they found correlation between in vivo and in vitro cost for protein production. They then used a previously developed model of translation to predict the resource cost for the production of different constructs and verified that model predictions based on lysate measurements correlated well with in vivo results. Finally, they demonstrated that the approach is also valid for two-gene operons as well as for a metabolic pathway such as the biosynthesis of beta-carotene. The study combines the advantages of cell-free systems in fast and high-throughput testing with mathematical modelling, opening up a way for easy benchmarking of construct designs based on the load they impose in vitro. The work also highlights how gene expression burden only accounts for a part of the total burden caused by engineered pathways. The cost of engineered enzymatic pathways is well known by metabolic engineers and it has been shown to arise through depletion of essential precursors as well as via metabolite-induced toxicity effects not necessarily limited to gene expression overload [31-35]. Indeed, to efficiently decouple the expression burden caused by production of the beta-carotene biosynthesis enzymes, the authors introduced an inactivating mutation within the active site of the first enzyme of the pathway. This led to a stronger correlation between model-based predictions and in vivo burden measurements, also pointing to the need for new modelling strategies able to capture these different contributions to burden.

**Orthogonal systems and resource allocation.** Cellular functions are interconnected, and exogenous constructs can be affected by retroactivity, where changes in a downstream process impact the functionality of upstream modules, leading to difficult-to-predict effects[36,37]. For this reason, attention has been dedicated to the development of orthogonal components that enable the insulation of exogenous processes from the native gene expression machinery[38,39]. Among these, orthogonal polymerases[12], transcription factor-promoter pairs[40], proteases[41] and ribosomes[13,42-45] have been designed and adopted. We here focus on orthogonal ribosomes and their potential in controlled resource allocation to minimise burden (Figure 2A). Recently, Aleksashin et al.[13] proposed OSYRIS, an orthogonal translation system with isolated ribosome subunits and an improved version of the previously developed tethered Ribo-T system[42]. Authors showed that it is possible to allocate Ribo-T to endogenous proteome expression and OSYRIS to production of exogenous genes, achieving fully independent and parallel translation of selected genes. However, since Ribo-T is not efficient in translation, the developed strains displayed slow growth and translation rates. These limitations were addressed in a parallel work by Carlson et al.[43] resulting in a second-generation Ribo-T with more efficient functionality and ability to incorporate non canonical amino
acids, expanding the potential future applications. The benefits of using orthogonal ribosomes to reduce burden were directly exemplified by Darlington et al[17]. The authors exploited a previously developed orthogonal 16S rRNAs-RBS pair[46] and partitioned translation between an endogenous host pool (h-ribosome) and an orthogonal pool (o-ribosome) to alleviate the overall resource competition. Improvements of this orthogonal translation machinery were generated by tethered o-ribosomes, as well as by the elaboration of a feedback controller of o-ribosomes levels. While this feedback controller is translated by o-ribosomes, it also represses their transcription, thus generating a dynamic resource allocator. These same tools were adopted by the authors to allocate the orthogonal machinery to different, co-expressed, synthetic constructs, providing a route taking into account the inter-connection between co-expressed genes[47]. Efficient design of such resource allocators can be informed by computational design [48]. However, it must be said that translational allocators only account for competition at the ribosome level and that integrated control of multiple resources is desirable[38].

**Feedback-based controllers.** Feedback regulation is commonly found in natural systems where robust control of native processes in response to perturbations is needed[49,50]. In synthetic biology, negative feedback controllers have been adopted to uncouple native and exogenous gene expression. Shopera et al[51] considered the competition between two circuits placed on two different plasmids. While circuit 1 bore an inducible fluorescent reporter, circuit 2 was considered in three different topologies with increasing complexity (i.e. constitutive reporter, ExsA-inducible and InvF/SicA-inducible reporters). Authors demonstrated that the non-linearity of resource-coupled interference between circuits varies based on circuit complexity. They then adopted negative feedback regulation of output in circuit 2 based on the ExsA-ExsD activator/anti-activator system of *P. aeruginosa*, where ExsD can bind and sequester the transcriptional activator ExsA and prevent circuit activation. Feedback-based control of circuit 2 led to decreased coupling between circuit outputs. The authors nicely demonstrated that feedback that was too strong can lead to more coupling between circuits than the absence of feedback, pointing to the need for fine-tuning of feedback controllers. Integral feedback controllers can be a powerful tool to design expression systems more robust to resource competition [52,53]. Huang et al[16] used such a strategy to overcome resource sharing limitation in free ribosomes availability. The authors developed a negative feedback controller consisting of a synthetic siRNA silencing the gene of interest (GOI) mRNA according to cellular translational resources. This quasi-integral controller successfully maintained the GOI expression level upon activation of a resource competitor, thus displaying robustness to resource loading by competing modules. In order to limit the impact on the host, burden itself can also be used as input for circuit regulation. Ceroni et al developed a burden-responsive negative feedback controller to mitigate burden in *E. coli*[15]
The burden-responsive htpG1 promoter, identified by RNAseq of engineered E. coli cells, was used to drive the expression of a sgRNA which, together with a constitutively expressed dCas9, was customised to repress exogenous expression to a non-burdensome level. The controller was able to rescue both cellular capacity and growth while maximising product yields. Different sgRNAs were also adopted to prove tunability in repression levels. While similar feedback controllers were previously developed, this system possesses the advantage of portable, quick and tuneable RNA-based regulation, compared to slower and more burdensome transcription factor-based or genome integration-based systems. However, generally feedback-based approaches suffer from a reduction of the output expression and the need for the controller itself to be optimised. Thus, their application implies a trade-off between the burden imposed and the desired output levels.

**Other approaches.** When output maximisation is sought other approaches can be adopted. These include genome minimisation to increase the pool of available resources as well as synthetic addiction to couple cellular growth and synthetic production. Chassis modification is a well-established approach used in metabolic engineering to maximise the yield of a bioproduct. Although many studies have focused on redirecting metabolic fluxes to favour certain biochemical reactions, chassis modification can also be employed to free up cellular resources that can then be used to increase heterologous expression. Lastiri-Pancardo et al developed a computational-based method, named ReProMin, that identifies the minimum number of genomic deletions needed to free up cellular resources and increase exogenous expression (Figure 2C). They focused their attention on transcription factors, integrating transcription factor interaction networks and proteomics data to identify non-essential targets, and generated two knockout strains with maximised resource release and proteome reduction. They ultimately demonstrated that they can achieve 18% overall increase in recombinant violacein production compared to the wild-type strain. One advantage of the approach is that it is pathway independent and authors maximised the yield of different products together with that of violacein. However, the limitation lays in the need for completely new analysis and engineering in case another host is adopted. Approaches that couple cellular growth with synthetic production stand out as valuable alternatives to reduce evolutionary drift of engineered populations without compromising synthetic product expression (Figure 2D) [57]. Rugbjerg et al used synthetic addiction to increase mevalonic acid production in engineered E. coli cells (Figure 2D). Faced with emergence of non-producing cell populations in long-term cultivations, the authors placed the genes for key growth intermediates (folP-glmM) under the control of a promoter responsive to the end-product. The approach led to higher mevalonic acid titers via enrichment of producing cells in the population. Similar systems where allosteric transcription factors (aTFs) drive synthetic addiction have also been developed for yeast [59]. The advantage of these approaches is their independence from
external stimuli and maximisation of output expression, while their main disadvantage is their reliance on the availability of product-responsive sensors (i.e. promoter and transcription factors) which make them not portable to different contexts.

**Conclusions.** Our understanding of cellular burden in bacteria has considerably improved in recent years. However, we are still far from being able to control the cellular response to burden, especially in settings outside the laboratory where the control of environmental conditions is difficult. In particular, large-scale bioproduction set ups and the space of bacterial therapeutics, where growth conditions vary considerably and heterogeneity in cell population can lead to genetic drift and loss of expression, are especially challenging areas [60,61]. Deep sequencing-based evolutionary data have so far proven useful to highlight failure modes in expression and production [62]. In this direction, insertion sequence elements and error-prone DNA polymerases have been deleted, leading to increased genome stability and reduced escape rates[63,64]. Further work on strain improvement is also foreseeable, along with the possibility of pinpointing mutational hotspots for a given expression construct that can be mitigated by intelligent design, but this comes with the limitation of an *ad hoc* approach [10]. Similarly, while our ability to design burden-aware circuits with reduced resource demand has considerably improved, more still needs to be done to apply this to a wider number of organisms in order to make it widely applicable [65-67]. Ultimately, there is still much to do in the field and a unique solution to burden may be very difficult to find. Going forward efforts should be directed toward integrating strain-based approaches with tools for low-burden construct engineering. This can lead to novel cellular backgrounds with reduced mutational rate and increased resource budget where constructs with minimal impact can be engineered. Combining these aspects is highly challenging but it would hugely benefit the implementation of genetic circuits with predictable outputs as well as of high-producing pathways for longer-term applications.
Declarations of interest: none.

Acknowledgements. This work was supported by the Biotechnology and Biological Sciences Research Council (grant BB/V00882X/1 to A.G.).

Figures

Figure 1. Synthetic biology approaches to identify low-burden genetic designs. (A) Fluorescent-based integrated capacity monitor quantifies the resource uptake by different synthetic designs aiding high-throughput screening of low burden candidates[14]. (B) Design of experiments followed by high-throughput library construction and in vivo measurements highlight design rules predictive of reduced burden[29]. (C) TX-TL experiments coupled to mathematical modelling inform on the impact that construct designs have in vivo[30].
Figure 2. Recent strategies to mitigate burden in bacteria. (A) Orthogonal ribosomes (o-ribosomes) were used to build resource allocation controllers[17]. Assembly of o-ribosomes and host ribosomes (h-ribosomes) is a competitive process. To ease this competition, production of o-ribosomes is placed under the control of a repressor (orange circles) translated by o-ribosomes, generating feedback on o-ribosome synthesis. In the presence of a synthetic circuit, the synthetic product (yellow pentagon) and the repressor compete for translation by o-ribosomes, leading to a decrease in repressor and an increase in o-ribosome production to be allocated to exogenous expression. (B) Burden-driven feedback regulation was adopted to decrease protein production and rescue cellular fitness mitigating cellular burden[15]. (C) Genome minimization proved to free up cellular resources and increase heterologous protein production[10]. (D) Synthetic product expression was coupled to cellular growth preventing population heterogeneity in long-term cultivation[58].
References


