Confidence, tolerance, and allowance in biological engineering: The nuts and bolts of living things

*Manuel Porcar[1)2)], Antoine Danchin[3)] and Víctor de Lorenzo[4)]**

The emphasis of systems and synthetic biology on quantitative understanding of biological objects and their eventual re-design has raised the question of whether description and construction standards that are commonplace in electric and mechanical engineering are applicable to live systems. The tuning of genetic devices to deliver a given activity is generally context-dependent, thereby undermining the re-usability of parts, and predictability of function, necessary for manufacturing new biological objects. Tolerance (acceptable limits within the unavoidable divergence of a nominal value) and allowance (deviation introduced on purpose for the sake of flexibility and hence modularity, i.e. fitting together with a variety of other components) are key aspects of standardization that need to be brought to biological design. These should endow functional building blocks with a pre-specified level of confidence for bespoke biosystems engineering. However, in the absence of more fundamental knowledge, fine-tuning necessarily relies on evolutionary/combinatorial gravitation toward a fixed objective.

DOI 10.1002/bies.201400091

[1)] Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain
[2)] Fundació General, University of Valencia, Valencia, Spain
[3)] AMAbiotics SAS, ICM, Hôpital de la Pitié-Salpêtrière, Paris, France
[4)] National Center of Biotechnology, CSIC, Campus Cantoblanco, Madrid, Spain

***Corresponding author:**
Víctor de Lorenzo
E-mail: vdlorenzo@cnb.csic.es

Abbreviation:
IGR, intergenic region.

Keywords:
allowance; standardization; synthetic biology; tolerance

> "... A good door needs no lock – yet no one can open it. Good binding requires no knots – yet no one can loosen it ..."
>
> Lao Zi (IV century BC)

Introduction

Although Molecular Biology is considered to have been founded by physicists [1–3], this circumstance did not result in the quantitative culture and an accurate and standardized descriptive language that is characteristic of the hard sciences. On the contrary, with very few exceptions (see, e.g. the attempts of Jacques Monod to quantify bacterial physiology [4], and early work published in the Cold Spring Harbor Symposia on Quantitative Biology) the molecular biology and genetics that developed since that time seldom took advantage of the opportunity to formalize the mechanisms and functions present in living systems with accurate languages and codes – let alone a nearly complete disregard for metrology: the theory and practice of quantitative measurement. The result has been decades of complete mess in the nomenclature of genes and ways to gauge and place numbers on biological activities, not to mention the explosion of all types of DNA vectors for genetic manipulation of the experimental systems under study. The organized and systematic ethos of physics is often at odds with the free-minded and typically informal culture of molecular biology and its spinoffs. Biology has in fact, thus far, produced very few quantitative codes: those that exist are mostly limited to EC numbers for enzymes and rules for measurement/annotation of enzymatic activities. The fixed bases present in the sticky ends of the DNA sequences after digestion with Type II-restriction enzymes also impose an involuntary format on habitual gene cloning procedures.

This state of affairs has recently been shaken by the onset of synthetic biology (SynBio) and its view of extant biological systems as similar to engineered artifacts, and thus potentially (re)engineer-able objects [5]. The fresh angle of SynBio is that engineering transcends the status of a *metaphor* to become a veritable *methodology* (both conceptual and material) for both understanding and designing biological systems – in the latter case with enhanced or altogether new-to-nature properties [6]. If the central dogma of molecular biology highlights the unidirectional information transfer from DNA to RNA and from RNA to proteins, the tenet of SynBio is that *parts make devices and devices make systems* (Fig. 1). The two interpretative frames are utterly compatible, but also patently different.

Standards are the basis of engineering

Modern engineering relies to a great extent on the adoption of standards of various types that allow uncoupling in space and time of the various steps involved in the construction of an object (Box 1). Furthermore, standards let different people work together even if they do not personally meet or know each other. Interestingly, the word *standard* comes from the military notion of many following a banner: Anglo-French: *estandard*, itself from old English and old German *standanhart*. The meaning of "all standing for the same" but also the concept of leadership are somewhat embedded in the genealogy of the word. When combined with assembly lines, standardization underpins the overwhelming success of mass production, one of the best examples being the Ford Model T one century ago, the first affordable automobile [7]. From the British Empire telegraphic network to today's computer operating systems, standards have reached every aspect of our society and are the basis of global industrial production. But how much of this can be imported into the biological realm? It is crucial to examine which functions of live systems are amenable to a concerted standardization effort.

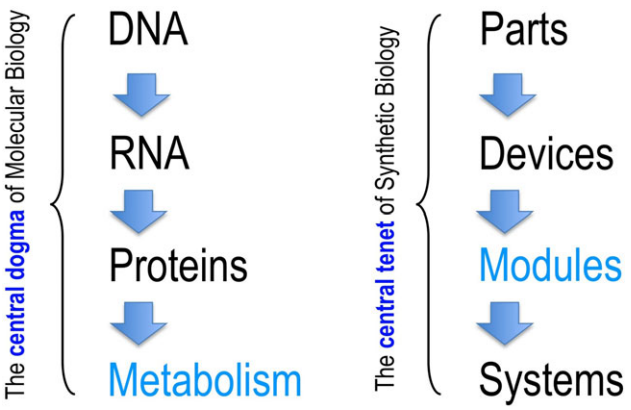


Figure 1. Molecular Biology and Synthetic Biology address live systems from different interpretative frameworks. The central dogma of molecular biology (left) focuses on the transfer of residue-to-residue information from DNA to proteins and (as recently proposed: [61]) deploying such an information flow in examining metabolic networks. In contrast, SynBio (right) tackles the compositional raison d'être and the assembly hierarchy of biological objects for both understanding their functioning and re-shaping them to create new properties. The emphasis of SynBio is thus on the relational and structural logic of existing and to-be-made biological systems, rather than on their evolutionary origin.

Box 1

What is a standard?

In the world of engineering, the terms *standard* and *standardization* refer mostly to: (a) the adoption of a shared semantic and graphic language for annotating the nature and the properties of the components of a system, (b) the definition of units of measurement of relevant properties and parameters as well as the conditions and procedures to calculate them (e.g. Amperes for current, Ohms for resistance, etc), (c) the specification of geometric shapes and size formats for the physical assembly of the components of a man-made system (e.g. the sizes, pitches, and shapes of helical threads), and (c) implementation of unambiguous protocols for the manufacture of the engineered objects. These standards allow the abstraction of the properties of the components of a system, their precise description with a suitable – also standardized – quantitative language, and the construction of the blueprint of the designed object with identical representation methods. A big bonus in this respect is the possibility of disengaging the detailed design of a product from the fabrication of its components and the final assembly of the artifact.

Can biological objects be built with construction standards?

The somewhat facile starting point of standards for bioengineering is the *establishment of rules for the physical composition of genetic devices*, in particular expression devices. The main instructions encoded in the genomes of prokaryotes such as *Escherichia coli* for expression of a given gene or operon rely on four types of adjacent DNA sequences: a promoter, an untranslated 5′ region (5′ UTR) that determines inter alia the binding of ribosomal machinery to the coding sequence (CDS) and a 3′ UTR that determines transcription termination. As sketched in Fig. 2, a minimalist prokaryotic expression unit typically includes a regulatory gene *R* along with its own expression signals (promoter P_R and ribosome binding site or RBS), a target promoter (P_{out}) and an activity module (encoded by default by a reporter product), which is endowed with upstream and downstream UTR sequences and contains genes (A, B, and C) separated, where necessary, by intergenic regions (IGRs). For the sake of composition, each of these functional sequences can be punctuated by sites that facilitate the assembly of many device variants and their cloning in vectors tailored to the same end for determining the input/output functions in a high-throughput fashion [8, 9]. These devices can then be moved from the assembly/measurement plasmids into deployment for either stable chromosomal implantation [10,

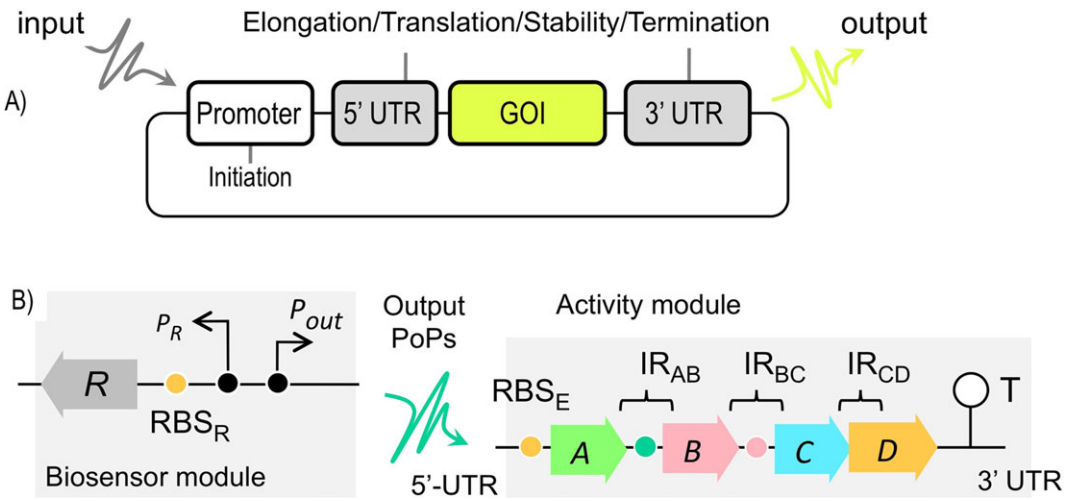


Figure 2. Organization of archetypal parts-based SynBio devices. **A:** Minimal prokaryotic expression devices are composed of the functional parts shown: a promoter, an untranslated 5′ region (UTR) bearing the ribosomal binding sequence (RBS), the gene of interest (GOI), and a 3′ UTR. The last determines transcription termination and often stability. The mRNA sequence can itself be punctuated by 3D motifs and secondary structures that determine its stability or its availability for translation. As shown in (**B**) such devices often involve not just one gene, but a poly-cistronic operon that is conditionally expressed upon exposure to an external signal that triggers the transcriptional activity. The output of the biosensor module is a given rate of productive transcription initiation. This activity – which is reminiscent of *current* in electric circuits [5] could be quantified as polymerase per second (PoPs), which is set by the amount of RNA polymerase molecules that pass a specific non-return position of promoter DNA each second. In this way, both the input and output transfer function of a regulatory node or module can be accurately described [62].

11] or for maintenance in multi-copy vectors [12]. DNA composition standards thus facilitate the assembly of individual parts and devices into systems. Ideally, such devices are inserted into an equally standardized genomic chassis [13] that provides the basic scaffold for the biological engineering exercise.

The emphasis on physical composition standards is one of the trademarks of the parts-based Genetically Engineered Machine competition (iGEM: http://igem.org). This branch of SynBio relies altogether on the so-called Registry of Standard Biological parts (http://parts.igem.org). Under this scheme, any particular DNA encoding a specific function (i.e. a BioBrick™) is formatted in such a way that given parts can be recursively composed (often in an automated fashion) with other parts in similarly standardized cognate vectors [8, 14, 15]. The resulting composition itself becomes an interchangeable element that can further be put together with other BioBrick™ parts for creating complex genetic devices [16]. Not unlike assembly lines, such standard parts can be used as building blocks for the making of increasingly intricate function-bearing DNA sequences [17]. The registry, founded in 2003 at MIT, accumulates a very large number of parts and devices that have been employed by hundreds of teams of undergraduate students for developing educational SynBio projects. Alas, assembling DNA parts following composition standards does not mean that the encoded functions are preserved in the new construct. Many such expression-related sequences adjacent to genes proper have been characterized in detail in their specific context [9, 18, 19], but once excised from their native cellular milieu and combined with other functional parts, they more often than not behave differently [20]. This poses in all its magnitude the problem of context-dependency of engineered biological functions [21], a feature long appreciated in the biotech industry but rarely in the limelight. As shown in Fig. 3, the performance of every expression device in a cell is subject to at least seven contextual layers ranging from the immediate mutual influence of adjacent DNA sequences all the way to environmental physicochemical conditions. Does this mean that any standardization effort is doomed to fail?

Although the collection is largely used and produced by undergraduate students and quality control is not perfect [22], the registry of parts is one of the most helpful resources for SynBio. But the practice over one decade has revealed

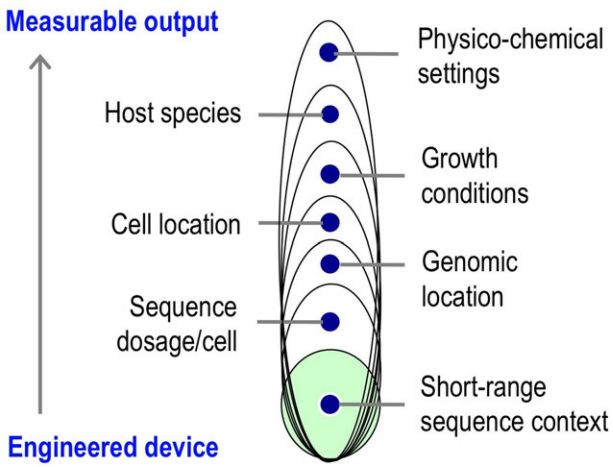


Figure 3. Biological context of designed genetic devices. The figure sketches how functionality of a given construct has to go through various layers of biological and physicochemical contexts before the output can be measured by the observer.

that the mere physical composition of parts is no guarantee of merging the encoded traits, let alone their kinetic parameters. Furthermore, quite frequently the constructs display outlying – if not altogether unexpected – behaviors. While this does not come as a surprise to molecular geneticists, it has created many misgivings about standardization efforts in biology [22]. Even well accredited parts, such as constitutive promoters, are liable to yield rather different outputs depending on circumstances and methods of measurement [23]. Recent efforts have been directed at exploring the combinatorial space of key parts involved in gene expression, in particular promoters and RBSs. The purpose of such an endeavor is to provide users with a repertoire of useful activity windows for a given project. But available data [18–20] indicate that the functional composability of parts seems to be no more than approximate, even in the best case. These developments raise again the question of the viability of directly translating engineering concepts to the biological world. As suggested in [24], we may need to develop an ad hoc engineering paradigm for biological constructs that looks fundamentally different from those of other disciplines that have benefited from engineering approaches. Before we get to such a grand endeavor, we may wish to address some angles of standards that have been largely ignored, and which could help to re-orient the attempts made thus far and those planned for the future.

Engineering standards incorporate tolerance and allowance

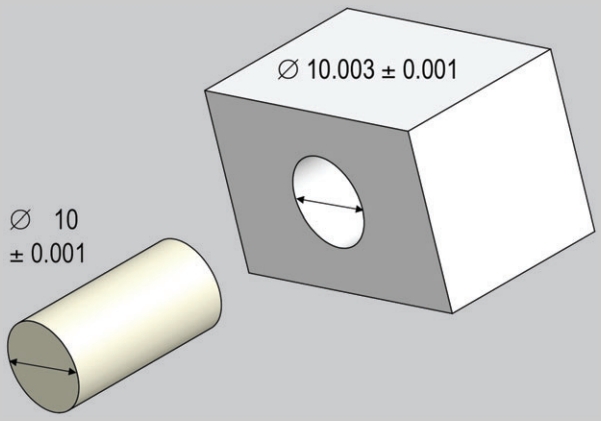
That a large number of parts and devices have been characterized does not seem to help the translation of physical assembly into functional composition very much – and it is unlikely that more efforts of the kind will rid the endeavor of its uncertainties [20]. Will the intrinsic variation of biological circuits and their feared context-dependence make the bioengineering endeavor impossible? We argue that there is still room for improvement by re-interpreting in biological terms some notions of construction standards that are commonplace in mechanical engineering. The bottom line is that physical composability might be ultimately impossible to translate into functional assembly if the specifications of the components at stake conform to an absolute value.

In engineering, the production processes are never perfect, thereby causing (tolerable) variations in the actual values of a building block versus the nominal values. *Tolerance* thus refers to an inevitable discrepancy of a standardized component of an object with respect to a token value: the range of variation permitted for maintaining the overall structure for machining a piece. Usually, the pair of upper and lower tolerance is used to define a non-perfect range within which an actual dimension may fall while still being acceptable for the sake of the assembled object. But, in order to be feasible, any modular construction must consider not only the tolerance (random deviation from a nominal value) but also the *allowance* (deliberate deviation from a nominal value; Box 2) of its building blocks, because only a degree of flexibility at the boundary between the components will allow a realistic connection.

Box 2

Tolerance and allowance

If a shaft has to be inserted in a Ø10 mm hole of a machine then 10 mm is the *nominal* dimension. But the fabrication process will unavoidably introduce an error that will make the shaft not exactly 10 mm, but, e.g. 10 ± 0.001 mm, the tolerance being plus or minus 0.001 mm. As long as the shaft diameter falls between 9.999 and 10.001 mm, the nominal Ø10 mm is acceptable for the purpose. But this is not the whole picture: for connecting the shaft to the receiving component, we also need to consider the dimensions of the hole proper and thus *plan* a deviation from its nominal or theoretical dimension. In order to accommodate the $Ø10 \pm 0.001$ mm shaft above, the tolerance of a Ø10 mm hole cannot be the same, because an, e.g. 10.001 mm shaft would not fit in a 9.999 mm hole. This calls for deliberately introducing a degree of *allowance* in the hole, for example 10.003 mm, with a manufacturing tolerance of ± 0.001 mm. This means that holes will actually be Ø10.002–10.004 mm, as they will have to accommodate shafts of 9.999–10.001 in size. By doing this, even the least favorable combination (hole of Ø10.002 mm, shaft of Ø10.001 mm) would allow the shaft to fit in the hole.



Biological design needs flexibility

How does all this translate into parallels in the biological realm? The straight consequence is that using fixed restriction sites for connecting DNA parts (as in the iGEM and related SynBio branches) is likely to fail in terms of functionality because this sets an absolute fit with no room for biological tolerance or allowance. As discussed above, tolerances are needed in engineering to define acceptable boundaries in the midst of unavoidable variation. And these boundaries need to be small enough to allow suitable interactions between parts. But in Biology, small variations in the physical composition of the parts at stake (whether a DNA sequence or a protein structure) may result in large variations in their activity. The standardization challenge in this case is therefore that of functional tolerances, i.e. establishing boundaries outside which variations would not be compatible with the other

parts. In other words, the functional (bio)shaft must enter the functional (bio)hole. This is a good metaphor, but what are the specific issues at stake?

The first challenge is that unlike man-engineered objects, the *functional* modularity of the components of live systems does not necessarily correspond to any *physical* modularity. In fact, the issue of modules in Biology is a controversial one [25, 26]. Typical biological functions (e.g. those performed by given metabolic blocks) are quite shapeless from a material point of view and often lack well-defined physical or chemical boundaries. Every component of metabolic networks (whether enzymes or small molecules) is prone to develop multiple interactions with others, which makes definition of limits very problematic. Inspection of archetypal pathways (e.g. amino acid biosynthesis) reveals that evolution of metabolic networks has been framed by the need of avoiding toxicity derived from the reactivity of the chemicals at issue with its neighbors [27]. A related phenomenon is the so-called paralogous metabolism [28], an often inconspicuous activity of many enzymes that can, at low rate, catalyze non-canonical reactions on unexpected substrates. By the same token that cells fix errors in DNA replication, there also seem to be active mechanisms by which cells correct metabolic mistakes [29]. These features of metabolic blocks are still poorly understood, but they pinpoint the phenomenal challenge of fixing physical boundaries and thus defining functional tolerances.

A second issue is that SynBio assumes, without stating so explicitly, that the underlying matter of live systems is similar to that found in industrial contraptions, including electronic devices, i.e. that we deal with *hard* stuff. However, biology is intrinsically made of soft materials. Functional equivalents to nuts and bolts do exist, but they are far from rigid: biology works in a world of glues, gels, and elastic entities. The challenge is thus standardizing soft matter, known to be prone to manifest emergent properties [30]. This requires classifying the world of biological objects into a hierarchy of *hardness* levels, taking into consideration composite materials as well. The basic building blocks, amino acids, nucleotides, coenzymes, etc., are fairly rigid in their definition, but as soon as these building blocks are combined together they tend to make material entities that are prone to be deformable. In addition, there are (tolerable) errors in transcription and translation that result in what one could call *informational softness* that cells need to handle.

The last question is that engineered devices have to be placed somewhere in the 3D architecture of the biological system. Bespoke genetic and metabolic networks are real material entities and the functions they deliver must be associated with an address in the space of the cell, where crowding imposes drastic constraints [31]. It is everyone's experience that the same promoter sequence placed at diverse locations along the chromosome will behave differently. This is not random, however, because such positional diversity allows integration of individual expression modules into the whole though genomic location [32–34]. But once produced, proteins as well as the biological equivalents to nuts and bolts have to be directed to specific places. Functions, structures, and processes are associated with addresses. It will be important to establish a classification of proteins into those that can distribute evenly in the cell, versus those that have a precise address, and see what makes the difference. This should help understand the way cell functions are organized. In parallel, it will be important to understand the principles (typically, post-translational) that rule such positioning.

What's in a DNA sequence?

All these considerations take us back to the general organization of SynBio expression devices (Fig. 2), in particular the DNA sequences of the parts and the nature of the IGRs. Recent studies [35, 36] have shed light on the amazing density of regulatory and structural instructions that are encrypted in the DNA (and thus mRNA) sequences. Precise measurements of genome-wide absolute, protein production rates in *E. coli* reveal that such sequences not only encode protein structure (and thus activity and lifetime) but also determine specific translation speed for each polypeptide that enable a precise stoichiometry of multi-protein complexes. At the same time, the levels of transcription factors ensure a balance between production cost and activity demands [37]. On this background, the approach of decreasing complexity by stitching standardized CDSs to similarly formatted promoters and IGRs seems not to be the way to go. When the regulatory regions of the T7 phage genome were decompressed to make it amenable to forward engineering [38], the resulting virus was certainly infective. However, its subsequent evolution in vivo toward recovering the fitness level of the wild-type phage erased ~40% of man-made modifications [39]. In contrast, naturally occurring systems are robust, and maintain their performance across time and space.

As a matter of fact, the informational density imprinted in DNA spans not only the boundaries between genes, but also within the sequences of the genes proper. One original feature of the genetic program is that it allows for overlapping codes. What appears as a *redundancy* of the genetic code permits variations in the third base of triplets, which may be used for encrypting specific signals overlapping with CDSs. Along with availability of amino acid-loaded tRNAs, codon distribution through different portions of structural genes determines the folding kinetics of cognate proteins [40]. On this basis, it is no surprise that existing programs for "improving" codon usage during heterologous expression of recombinant proteins frequently result in unpredictable outcomes, because such underlying codes are often unknown. In the same way, proteins interact with regulatory regions on one side of the double helix, leaving room for interaction of other proteins on the other side, which implies the construction of discontinuous "words" that may be intertwined in the genome text [41]. Other sequence codes in the transcripts are believed to earmark specific mRNAs for translation close to the subcellular location where the activity of their encoded products is needed [42]. Issues that remain to be clarified also include the uneven genomic distribution of palindromic sequences (as present in most restriction sites used for cloning). This indicates that there are unrecognized selection pressures associated with their presence/absence. The case of GATC in *E. coli* is enlightening in this respect, because this sequence is clearly counterselected in most phages and other natural

mobile elements [43]. Since phages are the archetypal success of heterologous gene expression in nature, one could think of avoiding, e.g. palindromic sequences in synthetic constructs. Such deeply biological considerations are, alas, generally absent in SynBio projects: it is urgently necessary to convert fundamental knowledge on such phenomena [44] into adequate cloning methods for multi-gene assemblies that result in optimal activities.

Allowance in bioengineering: The devil is in the boundaries

While engineering biological tolerance might be difficult with the degree of knowledge that we have now, translating the concept of allowance into SynBio might be more readily attainable. As a matter of fact, rather than representing an obstacle for standardization, development of allowances in genetic constructs could be an opportunity for a significant improvement of assembly rules in bioengineering. This calls for the development of allowance-compatible cloning methods that permit the molecular *nesting* in time and space of the products of the engineered device into the pre-existing biochemical and structural frame of the receiving cell. One possibility to this end could be extending the current strategies of designing [45, 46] and diversifying IGRs [47] toward also varying the distances from the promoter to the genes to be transcribed, as well as the size of the IGRs between different genes arrayed in an operon. This could allow immediate selection of the optimal combination among those displaying the boundaries specified in the cloning strategy. A second possibility would be to have connecting IGRs that bring about a fixed stoichiometry between the products encoded in a poly-cistronic mRNA, taking into account the spatial requirements of the gene products. Typically, a cytoplasmic protein must be more abundant than a membrane protein, as reflected in the lactose operon, for example [48, 49]. In most naturally occurring multi-gene transcriptional units, such IGRs are not as suitably arrayed as shown in the ideal case of Fig. 2. Transcriptional pausing [35] and translational coupling [50] result in specific expression kinetics of the products at stake and in their relative stoichiometry. Finally, it could also be possible to graft interface-interaction patches in the sequences of the proteins to be expressed so that they assemble with a pre-fixed stoichiometry and/or 3D architecture [51–53]. Taken together, these and other flexibility-enabling approaches would deliver a repertoire of functionalities that would be selected by the adaptive worth of a particular allowance within a pool of different values. But this may look like a return to evolutionary, random exploration of a solution space that is often perceived to be at odds with the forward-engineering agenda of SynBio. Is there a solution to the conundrum?

In case of emergency: The Gaudí principle

Both tolerance and allowance relate to some inherent uncertainties in the physical versus functional connectivity of parts for forming a whole. As just discussed, this concept is present in mechanical engineering, but it becomes one of the key challenges of SynBio, because both factors involve not only the most proximal connecting partners but also many different contextual layers (Fig. 3). With the current knowledge it might not yet be feasible to calculate or define such complex parameters for ensuring the expected functional coupling of the parts of a SynBio device. Fortunately, traditional technology has often dealt with the need for assembling intricate constructs long before suitable formalisms were available to this end. One extreme example of such calculations is provided by the techniques developed by the Spanish architect Antoni Gaudí (1852–1926) for constructing some of the most complex buildings of his time, long before the era of computer simulations and sophisticated modeling. As sketched in Fig. 4, he addressed the problem of nearby and distal connectivity between the components of a quite convoluted edifice by making string models in which weights were positioned at given places for exposing the impact of local connections on the shape of the whole construct and vice versa. By creating an upside-down image of such a string-weight model, the arches and angles for maintaining a sturdy structure could be rigorously determined. Such a non-mathematical method thus allows one to find optimal parameters for distributing components in a difficult assembly, so that nature itself (i.e. gravity in Gaudi's case) provides a way of solving a multi-scale problem that could not be addressed with available formalisms (it can now [54]). Gravitation toward functional optimality in Biology can be brought about by either directed evolution [55, 56] or by selection from a pool of combinatorial values [20, 57]. Within this framework, it seems unavoidable that the connections between components of a pre-assembled biological device must be set to allow exploration of a variety of input–output transfer functions (e.g. by introducing functional flexibility in

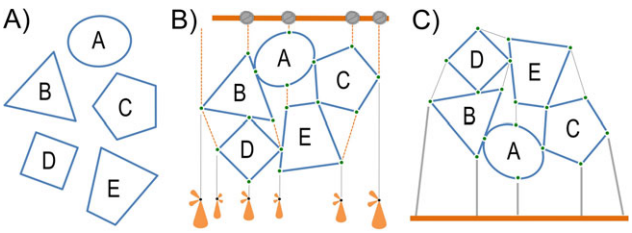


Figure 4. The Gaudí principle. *String-weight engineering* is an attempt to determine complex parameters for the construction of a modular object in which proximal interactions have an effect on the system as a whole and vice versa. Under this approach, the collection of components A–D (**A**) is first linked physically to form an object that is then hung with weights at the sites that will later be the pinnacles of the whole piece (**B**). Gravity then deforms the overall shape of the object to reach an optimal distribution of angles and masses. Turning the model upside down (**C**) provides the parameters that endow stability to the construct. This method, which was exploited by architect A. Gaudí in many of his buildings, allowed him to solve multi-scale modeling problems that were not tractable with the formalisms of the time. We argue that such an approach – which takes the principles of tolerance and allowance to their limit should be incorporated as a formal tool in SynBio constructs, which are afflicted by the spontaneous and often non-predictable connectivity and stickiness of biological building blocks.

boundary sequences as discussed above) until given optima are reached. To this end it is crucial not to fix the connecting DNA sequences between parts, but rather to let them fluctuate in a fashion that is reminiscent of the allowance discussed above. The approach that we could call the *Gaudí principle* could thus be formulated as: *calculate the architecture of multi-component (bio)systems to the degree that you can and enable nature do the fine-tuning*.

Merging rigorous calculations with a degree of flexibility for the sake of building reliable biosystems may allow us to revisit the breach between forward engineering and random tinkering that Jacob formulated time ago [58]: engineers often adopt tinkering as a solution-seeking strategy. And frequently either way reaches the same outcome. It is possible that the viability space for a physico-chemical-architectural challenge, whether biological or not, has a limited number of attractors/solutions – which can be reached through various itineraries. In this respect, the way NASA deals with the uncertainties on the site of landing of its missions to Mars (http://mars.nasa.gov/mer/technology/is_entry_descent_landing.html) is quite inspiring. To this end, the entirety of the rover module is cushioned in a flexible airbag that once parachuted to the planet's surface, bounces from one place to the other until setting in a stable spot. The key in this case is the design of a device for resolving the uncertain steps of the process. There is much to learn in SynBio from such a way of entering flexibility and dealing with unknowns that cannot be calculated upfront.

Conclusions and outlook

The matter of standards in SynBio (and in biology in general) includes – but is by no means limited to – physical arrangement of DNA parts. But being able to assemble functional building blocks into predictable devices should be one of the first steps toward converting SynBio into a rigorous engineering discipline. Although biological systems may be inherently prone to a degree of messiness (see, e.g. the case of enzymes [59]), the main conclusion of our analysis above is that any compositional standard based on fixing boundary sequences by means, e.g. of restriction sites is likely to fail, because such types of rigid boundaries do not incorporate the tolerance and allowance that are mandatory for realistic construction of multi-modular objects. Although computer-assisted design (CAD) of IGRs between the genes of an expression unit can help improve the state of affairs, we still argue that sequence elements allowing a degree of evolutionary gravitation towards optimality should be incorporated in SynBio devices [60]. By doing this, we would do nothing that engineers and architects have not done before in situations where the knowledge of the time did not allow them to calculate the parameters from first principles.

Acknowledgements

This study was supported by the BIO Program of the Spanish Ministry of Science and Innovation, the ST-FLOW and ARYSIS Contracts of the EU, the ERANET-IB Program, and the PROMT Project of the CAM. Sven Panke is thankfully acknowledged for useful comments on the manuscript.

The authors have declared no conflict of interest.

References

1. **Fox Keller E**. 1990. Physics and the emergence of molecular biology: a history of cognitive and political synergy. *J Hist Biol* **23**: 389–409.
2. **Holliday R**. 2006. Physics and the origins of molecular biology. *J Genet* **85**: 93–7.
3. **Stent GS, Calendar R**. 1978. *Molecular Genetics: An Introductory Narrative*. San Francisco, CA, USA: WH Freeman and Company.
4. **Ullmann A**. 1978. *Selected Papers in Molecular Biology by Jacques Monod*. Oxford: Elsevier Science.
5. **Endy D**. 2005. Foundations for engineering biology. *Nature* **438**: 449–53.
6. **de Lorenzo V, Danchin A**. 2008. Synthetic biology: discovering new worlds and new words. *EMBO Rep* **9**: 822–7.
7. **Argyres N, Bigelow L, Nickerson JA**. 2014. Dominant designs, innovation shocks, and the follower's dilemma. *Strat Manag J*, in press. DOI: 10.1002/smj.2207
8. **Shetty RP, Endy D, Knight TF, Jr.** 2008. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* **2**: 5.
9. **Canton B, Labno A, Endy D**. 2008. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* **26**: 787–93.
10. **Martinez-Garcia E, Calles B, Arevalo-Rodriguez M, de Lorenzo V**. 2011. pBAM1: an allsynthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol* **11**: 38.
11. **St-Pierre F, Cui L, Priest DG, Endy D**, et al. 2013. One-step cloning and chromosomal integration of DNA. *ACS Synth Biol* **2**: 537–41.
12. **Silva-Rocha R, Martinez-Garcia E, Calles B, Chavarria M**, et al. 2013. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* **41**: D666–75.
13. **de Lorenzo V**. 2011. Beware of metaphors: chasses and orthogonality in synthetic biology. *Bioeng Bugs* **2**: 3–7.
14. **Vick JE, Johnson ET, Choudhary S, Bloch SE**, et al. 2011. Optimized compatible set of BioBrick vectors for metabolic pathway engineering. *Appl Microbiol Biotechnol* **92**: 1275–86.
15. **Zucca S, Pasotti L, Politi N, Cusella De Angelis MG**, et al. 2013. A standard vector for the chromosomal integration and characterization of BioBrick parts in *Escherichia coli*. *J Biol Eng* **7**: 12.
16. **Rokke G, Korvald E, Pahr J, Oyas O**, et al. 2014. BioBrick assembly standards and techniques and associated software tools. *Methods Mol Biol* **1116**: 1–24.
17. **Appleton E, Tao J, Haddock T, Densmore D**. 2014. Interactive assembly algorithms for molecular cloning. *Nat Methods* **11**: 657–62.
18. **Mutalik VK, Guimaraes JC, Cambray G, Mai QA**, et al. 2013. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat Methods* **10**: 347–53.
19. **Mutalik VK, Guimaraes JC, Cambray G, Lam C**, et al. 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* **10**: 354–60.
20. **Kosuri S, Goodman DB, Cambray G, Mutalik VK**, et al. 2013. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci USA* **110**: 14024–9.
21. **Rocha EP, Guerdoux-Jamet P, Moszer I, Viari A**, et al. 2000. Implication of gene distribution in the bacterial chromosome for the bacterial cell factory. *J Biotechnol* **78**: 209–19.
22. **Vilanova C, Porcar P**. 2014. iGEM 2.0 – refoundations for engineering biology. *Nat Biotechnol* **32**: 420–4.
23. **Singh P, Batth TS, Juminaga D, Dahl RH**, et al. 2012. Application of targeted proteomics to metabolically engineered *Escherichia coli*. *Proteomics* **12**: 1289–99.
24. **Gardner TS, Hawkins K**. 2013. Synthetic biology: evolution or revolution? A co-founder's perspective. *Curr Opin Chem Biol* **17**: 871–7.
25. **Silver PA, Way JC, Arnold FH, Meyerowitz JT**. 2014. Synthetic biology: engineering explored. *Nature* **509**: 166–7.
26. **Porcar M, Latorre A, Moya A**. 2013. What symbionts teach us about modularity. *Front Bioeng Biotechnol* **1**: 14.
27. **de Lorenzo V, Sekowska A, Danchin A**. 2014. Chemical reactivity drives spatiotemporal organization of bacterial metabolism. *FEMS Microbiol Rev*, in press, DOI: 10.1111/1574-6976.12089

28. **Danchin A, Sekowska A**. 2014. The logic of metabolism and its fuzzy consequences. *Environ Microbiol* **16**: 19–28.
29. **Linster CL, Van Schaftingen E, Hanson AD**. 2013. Metabolite damage and its repair or pre-emption. *Nat Chem Biol* **9**: 72–80.
30. **Doi M**. 2013. *Soft Matter Physics*. Oxford: Oxford University Press.
31. **Klumpp S, Scott M, Pedersen S, Hwa T**. 2013. Molecular crowding limits translation and cell growth. *Proc Natl Acad Sci USA* **110**: 16754–9.
32. **Bryant JA, Sellars LE, Busby SJ, Lee DJ**. 2015. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res*, **42**: 11383–92.
33. **Elati M, Nicolle R, Junier I, Fernandez D**, et al. 2013. PreCisIon: PREdiction of CIS-regulatory elements improved by gene's positION. *Nucleic Acids Res* **41**: 1406–15.
34. **Junier I, Martin O, Kepes F**. 2010. Spatial and topological organization of DNA chains induced by gene co-localization. *PLoS Comput Biol* **6**: E1000678.
35. **Larson MH, Mooney RA, Peters JM, Windgassen T**, et al. 2014. A pause sequence enriched at translation start sites drives transcription dynamics *in vivo*. *Science* **344**: 1042–7.
36. **Li GW, Burkhardt D, Gross C, Weissman JS**. 2014. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**: 624–35.
37. **You C, Okano H, Hui S, Zhang Z**, et al. 2013. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**: 301–6.
38. **Chan LY, Kosuri S, Endy D**. 2005. Refactoring bacteriophage T7. *Mol Syst Biol* **1**: 2005.0018.
39. **Springman R, Molineux IJ, Duong C, Bull RJ**, et al. 2012. Evolutionary stability of a refactored phage genome. *ACS Synth Biol* **1**: 425–30.
40. **Saunders R, Deane CM**. 2010. Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res* **38**: 6719–28.
41. **Larsabal E, Danchin A**. 2005. Genomes are covered with ubiquitous 11 bp periodic patterns, the "class A flexible patterns". *BMC Bioinformatics* **6**: 206.
42. **Bibi E**. 2012. Is there a twist in the *Escherichia coli* signal recognition particle pathway? *Trends Biol Sci* **37**: 1–6.
43. **Henaut A, Rouxel T, Gleizes A, Moszer I**, et al. 1996. Uneven distribution of GATC motifs in the *Escherichia coli* chromosome, its plasmids and its phages. *J Mol Biol* **257**: 574–85.
44. **Marsh JA, Hernandez H, Hall Z, Ahnert SE**, et al. 2013. Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**: 461–70.
45. **Salis HM**. 2011. The ribosome binding site calculator. *Methods Enzymol* **498**: 19–42.
46. **Farasat I, Kushwaha M, Collens J, Easterbrook M**, et al. 2014. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol Syst Biol* **10**: 731.
47. **Casini A, MacDonald JT, De Jonghe J, Christodoulou G**, et al. 2014. One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Res* **42**: E7.
48. **Choudhary K, Oehler S, Narang A**. 2014. Protein distributions from a stochastic model of the *lac* operon of *E. coli* with DNA looping: analytical solution and comparison with experiments. *PLoS One* **9**: E102580.
49. **Eames M, Kortemme T**. 2012. Cost-benefit tradeoffs in engineered *lac* operons. *Science* **336**: 911–5.
50. **Levin-Karp A, Barenholz U, Bareia T, Dayagi M**, et al. 2013. Quantifying translational coupling in *E. coli* synthetic operons using RBS modulation and fluorescent reporters. *ACS Synth Biol* **2**: 327–36.
51. **Fletcher JM, Harniman RL, Barnes FRH, Boyle AL**, et al. 2013. Self-assembling cages from coiled-coil peptide modules. *Science* **340**: 595–9.
52. **Gradišar H, Božič S, Doles T, Vengust D**, et al. 2013. Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat Chem Biol* **9**: 362–6.
53. **Bosshart A, Panke S, Bechtold M**. 2013. Systematic optimization of interface interactions increases the thermostability of a multimeric enzyme. *Angew Chem Int Ed Engl* **52**: 9673–6.
54. **Huerta S**. 2006. Structural design in the work of Gaudi. *Archit Sci Rev* **49**: 324–39.
55. **Marliere P, Patrouix J, Doring V, Herdewijn P**, et al. 2011. Chemical evolution of a bacterium's genome. *Angew Chem Int Ed Engl* **50**: 7109–14.
56. **Chou HH, Keasling JD**. 2013. Programming adaptive control to evolve increased metabolite production. *Nat Commun* **4**: 2595.
57. **Pfleger BF, Pitera DJ, Smolke CD, Keasling JD**. 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**: 1027–32.
58. **Jacob F**. 1977. Evolution and tinkering. *Science* **196**: 1161–6.
59. **Tawfik DS**. 2010. Messy biology and the origins of evolutionary innovations. *Nat Chem Biol* **6**: 692–6.
60. **Nikel PI, Martinez-Garcia E, de Lorenzo V**. 2014. Biotechnological domestication of pseudomonads using synthetic biology. *Nat Rev Microbiol* **12**: 368–79.
61. **de Lorenzo V**. 2014. From the selfish gene to selfish metabolism: revisiting the central dogma. *BioEssays* **36**: 226–35.
62. **de Las Heras A, Carreno CA, Martinez-Garcia E, de Lorenzo V**. 2010. Engineering input/output nodes in prokaryotic regulatory circuits. *FEMS Microbiol Rev* **34**: 842–65.