



**HAL**  
open science

# In vivo , in vitro and in silico : an open space for the development of microbe-based applications of synthetic biology

Antoine Danchin

► **To cite this version:**

Antoine Danchin. In vivo , in vitro and in silico : an open space for the development of microbe-based applications of synthetic biology. *Microbial Biotechnology*, 2021, 15 (1), pp.42-64. 10.1111/1751-7915.13937 . hal-04542701

**HAL Id: hal-04542701**

**<https://hal.science/hal-04542701>**

Submitted on 11 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Special Issue Article

***In vivo*, *in vitro* and *in silico*: an open space for the development of microbe-based applications of synthetic biology**Antoine Danchin\* Kodikos Labs, Institut Cochin, 24 rue du Faubourg  
Saint-Jacques, Paris, 75014, France.**Summary**

Living systems are studied using three complementary approaches: living cells, cell-free systems and computer-mediated modelling. Progresses in understanding, allowing researchers to create novel chassis and industrial processes rest on a cycle that combines *in vivo*, *in vitro* and *in silico* studies. This design–build–test–learn iteration loop cycle between experiments and analyses combines together physiology, genetics, biochemistry and bioinformatics in a way that keeps going forward. Because computer-aided approaches are not directly constrained by the material nature of the entities of interest, we illustrate here how this virtuous cycle allows researchers to explore chemistry which is foreign to that present in extant life, from whole chassis to novel metabolic cycles. Particular emphasis is placed on the importance of evolution.

**Introduction: complementary avenues for the exploration of life**

It took a long time for science to understand that there was no difference between ‘organic’ and ‘inorganic’ matter. The nature of the chemical compounds of life had no specificity other than, perhaps, the fact that they were based on the chemistry of carbon. Still, the status of biology is somewhat different from that of standard chemistry, for living organisms are deeply connected with a surprising quality of ‘animation’, which at first sight seems alien to the laws of physics. As a consequence

of this observation, studies of living organisms were separated into *in vivo* (using living cells and organisms as the basic entities of interest) and *in vitro* (using parts of living entities combined together in test tubes) studies. More recently, with the enormous development of computing facilities, it has become possible to replace experimental studies using authentic material entities with studies of the living using computers. Although little developed before the advent of DNA sequencing techniques, this change in the study of life was greatly stimulated when it became apparent that knowledge of genomic sequences would rapidly become ubiquitous. The need to develop this endeavour was highlighted for the first time at a meeting of the European Biotechnology Action Programme in Tutzing, Germany, in 1989: ‘From the information technologies perspective, we are faced with the need to create databases where we will recognize relevant features in response to learning processes. The creation of descriptors, the genesis of exploratory hypotheses will add to the traditional methods of biology, *in vivo* and *in vitro*, experiments *in silico*’. Although this emphasis is already three decades old, it can be safely anticipated that *in silico* approaches will gain in importance in the coming decades, especially for developments in microbial biotechnology. In what follows, we briefly summarize the state of *in vivo* microbial studies, focusing on the forthcoming development of novel artificial chassis, then focus on current trends in cell-free syntheses and conclude with progress in *in silico* research. One of the conclusions of this study is that to understand the functions responsible for the apparent animation of living systems, the activities that dissipate energy in unanticipated ways should be given priority in the coming years.

**A future for *in vivo* microbial studies**

Microbial biotechnology exploits a substantial collection of microbes widely used to generate products/processes of interest to industry. Industrial applications go hand in hand with understanding the puzzles of their environmental niches and metabolism. We can anticipate that,

Received 11 September, 2021; accepted 14 September, 2021.

\*For correspondence. E-mail antoine.danchin@normalesup.org; Tel. +331 4441 2551; Fax +331 4441 2559.

*Microbial Biotechnology* (2022) 15(1), 42–64  
doi:10.1111/1751-7915.13937© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

using these same organisms, metabolic engineering will continue to evolve over the next few decades, now benefiting from rapid advances in synthetic biology (SynBio), with a likely emphasis on the use of large-scale DNA synthesis (Panke, 2020). This topic has been the subject of many studies [see, e.g. Calero and Nikel (2019)], and we do not explore it further here, except for a brief review of recent uses showing how novel living chassis will gradually gain impact in the near future. Following this pattern, we describe in this section how the emergence of novel microbial chassis is about to revolutionize microbial engineering.

#### *Brief landscape of the use of genetically modified natural microbial chassis*

Until recently, the majority of microbial biotechnology advances that did not rest on microbes with idiosyncratic properties – typically, specific synthesis of metabolites such as vitamin B12 (Balabanova *et al.*, 2021) – used a variety of *Escherichia coli* strains as the basis for elaborate metabolic engineering. This is not discontinued: sophisticated SynBio designs are already implemented in this organism, as seen, for example, in the use of optogenetics to control expression of the plant lycopene biosynthesis pathway in this organism (Raghavan *et al.*, 2020). Such innovative leads could soon become routine. The ubiquitous role of *E. coli* as a chassis was not unexpected due to the considerable accumulation of knowledge gained for this species. Relevant access to accurate sequence annotation is key for productive industrial developments. Alas, errors keep percolating in automatic genome annotation procedures that are in effect essentially based on the still imperfect knowledge of model organisms (Bell and Lord, 2017). Yet, complete, up-to-date annotation of the *E. coli* genome sequence is unfortunately not properly maintained on a perennial basis. The best public annotation of the sequence at this time is perhaps the one developed using orthologue-based annotation propagation (Paley *et al.*, 2021), but it is not adapted to the subtle requirements of industrial demands. This drawback implies that the use of chassis other than *E. coli* requires even more tedious manual annotation to account for the inevitable idiosyncrasies uncovered in the course of engineering [see the importance of ‘kludges’, by definition unexpected, in (Danchin, 2021)].

SynBio studies have long aimed at getting rid of seemingly superfluous gene-encoded functions, creating universal minimal chassis in which biotechnology-relevant pathways can be introduced. In line with this objective, the Mollicutes clade was perceived as an interesting paradigm because these organisms had already evolved streamlined genomes (Piñero-Lambea

*et al.*, 2020). Mollicutes have been and will still be used to understand the minimal constraints of cell survival and multiplication (Garcia-Morales *et al.*, 2020). However, because these organisms have a highly reduced genome and minimal macromolecule synthesizing machineries, they are intrinsically fragile and limited in their metabolic power, and thus will hardly be convenient for widespread industrial developments. In contrast, a large collection of organisms that thrive in more or less extreme environments are likely to take the limelight in the near future (Liu and Deutschbauer, 2018). Here is a brief list of some relevant candidates [see (Calero and Nikel, 2019) for a detailed account of an exhaustive list of non-traditional industrial microbes.]

The *Pseudomonas* genus is well suited to metabolize aromatic compounds and harbour novel metabolic pathways (Bitzenhofer *et al.*, 2021). The genome of *Pseudomonas putida* has already been streamlined and shown to allow improved heterologous gene expression (Lieder *et al.*, 2015). A variety of tools are available for this organism, which now becomes a widespread platform for SynBio constructs (Martin-Pascual *et al.*, 2021). For example, even though this organism has its metabolism poised to develop in the presence of dioxygen, strain KT2440 has been modified to multiply in anaerobic conditions (Kampers *et al.*, 2021). A companion, *Comamonas testosteroni*, a species belonging to a related clade, has become a promising chassis for bioremediation due to its natural pollutant-degrading capacity (Aksu *et al.*, 2021).

In the actinobacterial clade, *Corynebacterium glutamicum* has been used for decades as a major industrial source of vitamins, amino acids, carbohydrates and a variety of other metabolites. It is now the subject of a fair number of improvements to accommodate novel metabolic pathways and SynBio constructs (Chen *et al.*, 2021a; Wang *et al.*, 2021; Wolf *et al.*, 2021). With many industrial facilities already using this organism, it seems likely that further improvements will accumulate rapidly. In the same way, but in the Firmicutes clade, *Bacillus subtilis* has also been much discussed as an important industrial chassis (Liu *et al.*, 2020; Xiang *et al.*, 2020). Besides applications similar to those discussed previously with other organisms, as well as often ignored roles, such as that in heavy industry where it is used for improving concrete ageing *via* creating self-healing properties, for example (Nielsen *et al.*, 2020), the fact that Bacilli generate extremely resistant spores is likely to generate original methods for new types of applications. Using a 3D printer, Christopher Voigt and his colleagues printed *B. subtilis* spores within an agarose scaffold, showing that they germinated on its exterior surface, including spontaneously in cracks. The agarose-spores material sustained desiccation and returned to life after

rehydration. When containing spores of *B. subtilis* engineered to produce antibiotics, the material could be used to kill *Staphylococcus aureus*, a pathogen causing a variety of infections (González *et al.*, 2020). This family of advances are so interesting that they have even been used for 3D printing of a mixture of recombinant *E. coli*-expressing metallothionein/gold nanocomposites used as agitating paddles for the catalytic reduction of 4-nitrophenol (Long *et al.*, 2021).

Still in the domain of bacteria, Cyanobacteria, being able to fix carbon dioxide, have a distinctive status, especially under conditions when the steady increase in carbon dioxide in the atmosphere of our planet is a matter of concern. The species of this phylum are likely to become a preferred chassis for biotechnological developments in the near future (Wang *et al.*, 2020). However, they display specific structural features that make them difficult to manipulate. Their metabolism, resting on gasses and reactive metabolites, must be compartmentalized (de Lorenzo *et al.*, 2015; Flechsler *et al.*, 2021). This means that future progress with these organisms is likely to require considerable effort to account for the existence of compartments before they are suitable for versatile metabolic engineering. Nevertheless, the proliferation of research facilities based on this family of chassis is already showing us how these bacteria will soon be commonly used (Yang *et al.*, 2020; Cui *et al.*, 2021). Other photosynthetic bacteria have also been proposed as new chassis, especially for water treatment. Some non-sulphur bacteria, the oldest of the photomicroorganisms, not only treat different types of wastewater but are also non-toxic bioresources, containing many value-added products. These bioresources can be used as raw materials in the agricultural, food and medical industries (Lu *et al.*, 2021).

Finally, besides a few applications using Archaea [see, e.g. Aldridge *et al.* (2021)], the Eukarya domain of life is also a common source for microbial biotechnology. In this context, the model *Saccharomyces cerevisiae* plays a role similar to that of *E. coli*, but other yeasts are also widely used, such as *Yarrowia lipolytica*, for lipid-related applications (Park and Nicaud, 2020). The lines of research we have just outlined with Bacteria will also be developed with these organisms, including after their genomes have been reworked to facilitate genetic manipulation (Schindler, 2020). This is already seen in the high-level original developments such as those listed in Voigt (2020). These examples demonstrate that we are on the verge of an enormous number of applications based on SynBio *in vivo* metabolic engineering and material improvement of extant microbes, notwithstanding the variation associated with inevitable evolution of all living organisms (Wassenaar and Zimmermann, 2020).

#### *Alien chassis and xenobiotics*

Most contemporary advances in SynBio are variations on a theme based on reconstructed metabolic pathways derived from natural sources, and artificial constructs deployed in a variety of natural chassis. Conclusive innovation, however, will really come to life when heterologous implementation of known pathways will be replaced by counterparts that have not evolved in extant organisms, as well as chassis modified to be distinct from standard living cells. A considerable amount of work is being done in this direction. Here, we describe some of the corresponding endeavours, while recognizing that, for the time being, examples of modification of the core chemistry of extant chassis remain rare. Perhaps the best illustration of promising work is the creation of *Escherichia chlori*, an avatar of *E. coli* where thymine has been entirely replaced by 5-chlorouracil in DNA (Marlière *et al.*, 2011). However, there is only little data available about the stability of the construct. We should expect to understand before long the functions that have been altered in this organism, allowing it to grow on 5-chlorouracil while being inhibited by thymine. This entails understanding how these chassis will behave in the long run, but we can be confident that there is some hope for stability as demonstrated in stable evolution after 100 days of an *E. coli* strain after continuous co-culture with mammalian cells (Kunjapur *et al.*, 2021).

A large number of avenues are being explored to further modify the nature of the genetic material of cells based on xenonucleic acids (Fiers *et al.*, 2016). While there is continuous progress in the domain (Chaput and Herdewijn, 2019), the emergence of autonomous novel chassis amenable to industrial applications seems unlikely for the next decade or so. In the meantime, less dramatic changes continue to be designed. Since the beginnings of SynBio research, researchers attempted to reassign codons (Robertson *et al.*, 2021), often to non-natural proteinogenic amino acids, as for example illustrated early on (Döring and Marlière, 1998). The most efficient approaches attempt to device orthogonal constructs, for the time being devoted to a limited fraction of the translation machinery (Schmied *et al.*, 2018). Recent attempts have created an artificial genetic code of 68 codons that could allow the incorporation of four non-canonical amino acids into synthetic proteins (Dunkelmann *et al.*, 2021). In other advances, the synthesis of proteins containing d-amino acids,  $\beta$ -amino acids, phosphorylated amino acids as well as long-chain and cyclic amino acids in which the nucleophilic amino group is not in the  $\alpha$ -position has been achieved using modified peptidyl transferase regions in the 23 ribosomal RNA. Dipeptides and dipeptidomimetics of varying utility were also obtained (Hecht, 2021). Because this is

sophisticated work – relevant cells must maintain both the standard machinery and an orthogonal one – industry-relevant progress has been slow. With an increasing number of laboratories involved in these attempts, the number of successes is expected to increase rapidly. However, again, much of the corresponding works are more proofs of concept than ready to become, biotechnology-adapted synthetic microbes growing in industrial fermenters. More limited approaches have already been used to produce synthetic variants of proteins, for example, using analogues of amino acids such as selenomethionine for bioconjugations (Flood *et al.*, 2021) or variants of norleucine, known to replace methionine in proteins in *E. coli* for a very long time (Moroder and Budisa, 2010; van Eldijk and van Hest, 2018). Nevertheless, a creation somewhat similar in its concept to that of *E. chlori* was based on growth of cells on tryptophan analogues. In the earliest experiments, full replacement of tryptophan residues by an analogue could not yet be achieved (Bacher and Ellington, 2001). Later on, complete replacement was successful in a *B. subtilis* mutant that swapped all tryptophan residues for 4-fluorotryptophan after serial selection and mutagenesis (Yu *et al.*, 2014).

Yet another highly focused advance, isobiology, has been proposed. It is based on the observation that biological processes differentiate the isotopes of atoms used to form cells. This line of experiments involves a very small modification of the chassis. It is, therefore, likely to have minimal consequences. The most important effect is expected when the standard cell composition replaces hydrogen by deuterium, in particular when heavy water (D<sub>2</sub>O) replaces standard water (H<sub>2</sub>O). Notwithstanding its financial cost, this procedure limits the consequences of changes while allowing the fine properties of the cell chassis to be explored (Danchin, 2020). This type of effort is likely to be important for understanding functional phase transitions in cells (Choi *et al.*, 2020), a concept that emerged in parallel with the onset of SynBio, and which is likely to be crucial in the construction of novel chassis as it defines previously neglected membrane-less compartments (Gomes and Shorter, 2019). As an illustration of future progress, specific isotope-sensitive microbial models could be used for the development of biotechnological processes, based on the behaviour of thermophilic Archaea (Nishizawa *et al.*, 2016). This would be motivated, for example, by the fact that the use of carbon, nitrogen and oxygen isotopes is already well accepted – despite the cost of the experiments – as safe substitutes for tracer experiments (Davies, 2020). This opens up a niche for the industrial microbial production of amino acids and nucleotides (Nicolás Carcelén *et al.*, 2017).

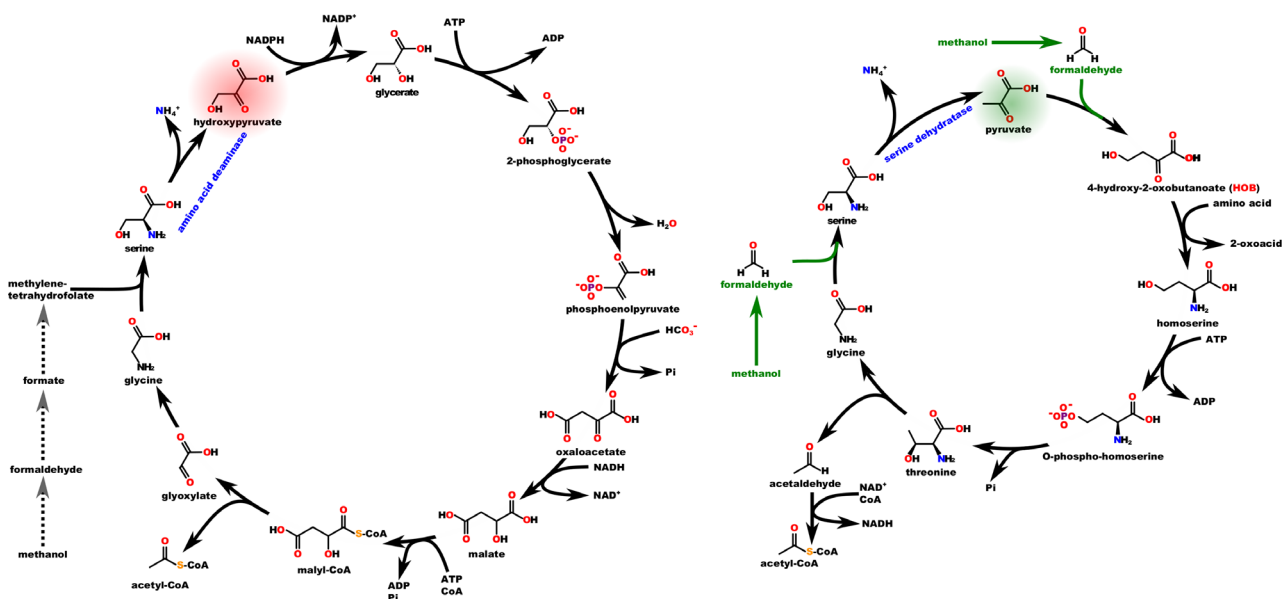
### Adaptive evolution

Life is a dynamic process, not only during the lifetime of an organism but also in the long term, as it keeps generating offspring. Evolution is an essential feature of life. Therefore, many SynBio projects have enabled constructs established in well-identified laboratory strains to evolve under various conditions [see, e.g. Li *et al.* (2021)]. Adaptive evolution has even been successfully developed for co-evolving communities of organisms (Konstantinidis *et al.*, 2021). In most cases, the operation has developed in large volume vessels, controlled as chemostats are. Among promising approaches, a variation on chemostats driven by growth medium dilution found that, in a continuous culture, growth chambers were invaded by mutants that escape dilution by attaching to the surface of the vessel. As expected, this trait is positively selected. It generates persistent subpopulations whose size and growth rate are uncontrollable, preventing the use of standard detection methods such as turbidity measurements to monitor growth. To overcome this limitation, a device with two growth chambers subjected to brief periodic alkaline sterilization phases has been designed (de Crécy-Lagard *et al.*, 2001; de Crécy *et al.*, 2007). Yet, because these vessels are at least a few millilitres in volume, this setting precludes carrying out large-scale experiments with many vessels working in parallel. This drawback triggered the development of continuous culture systems based on open reactors managing microfluidic droplets [see (Ito *et al.*, 2016) and the special issue on recent advances of molecular machines and molecular robots edited by Takinoue and Kawano, which describes advances in the domain, mainly in Japan (Takinoue and Kawano, 2020)]. The future of laboratory-set evolution is now likely to rest on nanolitre reactors such as those already used for production of important metabolites (Xu *et al.*, 2020), allowing *in vivo* directed evolution (Femmer *et al.*, 2020). This is at the cost of the evolving population's size, however, so that it is likely that a significant family of evolution devices will keep using sizeable vessels in the future.

### The design of alien metabolism

Intermediary metabolism is an intrinsic component of all live chassis. In most cases, researchers have attempted to streamline their pathways of interest in existing organisms with a small series of efforts to completely replace essential pathways with artificial ones. Still, extant pathways remain important: *E. coli* can be reprogrammed to fix carbon dioxide by simple modifications in its canonic carbon metabolism pathways (Satanowski *et al.*, 2020). However, an entirely novel pathway for carbon fixation could be implemented in this organism. The standard

core one-carbon metabolism has been reprogrammed with the design of an alternative cyclic pathway that substitutes 4-hydroxy-2-oxobutanoic acid (HOB), a compound absent from canonical metabolism – except perhaps as an accidental consequence of aminotransferases' promiscuity (Walther *et al.*, 2018) –, for the amino acids serine and glycine as omnipresent one-carbon donors. This pathway is based on two novel reactions, transamination of L-homoserine and transfer of a one-carbon unit from HOB to tetrahydrofolate, releasing pyruvate. Canonical reactions regenerate L-homoserine from pyruvate by carboxylation and subsequent reduction. In this way, any one-carbon moiety made available for anabolic reactions may originate from CO<sub>2</sub>. The HOB-dependent pathway was established in an *E. coli* auxotroph selected for prototrophy after enrichment in a long-term cultivation system (Marlière *et al.*, 2011). Monitoring its stability will be interesting, as L-homoserine, an analogue of authentic proteinogenic amino acids, interferes with translation (Jakubowski, 1997). Living organisms have found solutions to circumvent this obstacle, for example, via protection/deprotection systems (Danchin and Sekowska, 2015), and we should expect that, if the pathway's industrial role requires to raise its production level, such solutions will emerge before long in laboratory-set evolution experiments. Indeed, a cognate cycle that uses a homoserine phosphorylation step was successfully designed to use methanol to feed general SynBio-derived bioproduction processes, and it likely to be improved in the near future [(He *et al.*, 2020) and Fig. 1].



**Fig. 1.** Natural methanol assimilation compared with methanol assimilation in a HOB-dependent cycle.

The usual input of methanol/formate into carbon metabolism (left) is compared (right) with the novel HOB-dependent cycle described in He *et al.* (2020). Note that hydroxypyruvate is potentially toxic.

While it remains essential for SynBio developments to design integrated living units, the very fact that living cells are constantly evolving makes it crucial to bring out the symmetry of synthesis and analysis – breaking down cells into their components – both to understand the details of life processes and to implement industry-friendly activities that are more stable over time than inventions with life. This involves exploring the principles of synthetic life *in vitro*, bearing in mind that we must both master the choice and availability of relevant components and advance the conceptual understanding of all the relevant processes that take place in living cells in order to make them work *in vitro*.

### *In vitro* syntheses and evolution

Molecular biology is a combination of genetics and biochemistry (<https://plato.stanford.edu/entries/molecular-biology/>). Until these fields of biology were merged, the study of life was divided between the exploration of organisms *in vivo* – alive – and the chemical analysis of some of their components once synthesized in test tubes, *in vitro*. However, what was successful *in vivo* did not overlap much with what was discovered *in vitro*. The apparent 'animation' of cells and organisms remained a mystery. As in the early XIX<sup>th</sup> century, when chemicals were divided between organic and inorganic matter, some biologists believed – some still believe – that macromolecules, in particular, have mysterious properties, specific to life. The actual, purely chemical synthesis of macromolecules was, therefore, a prerequisite for

the exploration of life, accepted as belonging to the realm of physics and chemistry. This feat was achieved with small proteins such as insulin. Today, the chemical synthesis of proteins is a difficult challenge that remains relevant, not in response to philosophical debates, but as an industrial target (Hui *et al.*, 2021): high-level accuracy in the protein sequence of amino acids appears to be easier to control *in vitro* than *in vivo* (Karas *et al.*, 2021). The same situation prevailed with nucleic acids (Caruthers, 2013), with the same consequences in terms of industrial developments: For example, gene and vaccine syntheses require a very high precision, with accessible controls at each stage, a requirement which remains still difficult to achieve with *in vivo* constructs. While life seems to have evolved a way to keep error rate in many processes to some minimum (Bradley *et al.*, 2019; Mordret *et al.*, 2019), the trade-off between speed and accuracy in biosyntheses has been adopted by natural selection as a means to input resilience in the build-up of living entities (Danchin *et al.*, 2011), making a significant level of errors an inevitable consequence (and cause) of evolution (Médigue *et al.*, 1991; Piñeros and Tlustý, 2020; Jones and Uphoff, 2021).

#### *Cell-free protein synthesis, a half-century ongoing story*

Once it was recognized that there was no magic involved in enzymatic activity, it was widely – but not always – accepted that enzyme-based synthesis of molecules *in vitro* could be accepted as yet another artificial means of constructing macromolecules to complement purely chemical syntheses. In a first step – in an experiment with impact similar to that of the synthesis of urea by Woehler, as it established that there was no animation principle in RNA synthesis –, Sol Spiegelman used the replicase of RNA bacteriophage Q $\beta$  to synthesize phage RNA in the test tube. The ‘little monster’, thus, produced was physically and chemically indistinguishable from the original virus RNA. Indeed, to establish that the test-tube RNA was infective, Spiegelman injected it in a solution of all four nucleoside triphosphates and Q $\beta$  replicase, waiting for the RNA to replicate. He then incubated it with live bacteria, and demonstrated that the RNA led to generation of infective viral particles (Haruna and Spiegelman, 1965). In another family of *in vitro* experiments, RNA synthesis of more or less random sequence was successful with an enzyme that did not use nucleic acids as templates, polynucleotide phosphorylase (Grunberg-Manago, 1963). Much of the discovery of the genetic code table used synthetic RNA made with this enzyme, further blurring the line between chemistry and biology.

With synthesis of RNA and proteins available *in vitro*, the goal was to shift experiments to the upper level of

biosyntheses and set up the multi-level process of translation in the test tube. Mimicking macromolecule biosynthesis as it unfolds *in vivo* was a true challenge. To go further in the direction of authentic protein synthesis, it was important to demonstrate translation of a monocistronic message. This was established early on (Clark *et al.*, 1965). This difficult undertaking, which began in the early 1960s, is still under development as more and more components are integrated into the synthetic machinery of test tube. We will gradually gain a better understanding of the reasons for the difference between these processes when they take place *in vivo* and *in vitro*. With synthetic RNAs, the importance of the pH and the role of magnesium (Jiang *et al.*, 2021), polyamines (Fukuma and Cohen, 1975) and potassium (Fritz *et al.*, 2018) were identified as crucial for success (Hammerling *et al.*, 2020). Translation with ribosome preparations, endowed of progressively increasing specificity and accuracy, was much improved after the discovery of the need to start and stop translation, using initiation factors (Clark and Marcker, 1966) and subsequently other factors such as the ribosome release factor (Kung *et al.*, 1977). A further step was gained when DNA, much more stable than RNA, could be used as a template for *in vitro* gene expression, allowing protein synthesis *via* coupling transcription with translation in the test tube (Austin and McGeoch, 1973; Isturiz and Wolf, 1975).

Improved ‘cell-free’ RNA and protein synthesis systems based on bacterial and eukaryotic extracts then became common, with cloned or synthetic genes as DNA templates (Billerbeck *et al.*, 2013; Caschera and Noireaux, 2014; Li *et al.*, 2014). The knowledge developed in the early decades of transcription–translation coupling in the test tube has been used to design a prototype flow microreactor for synthetic biology *in vitro* providing solutions for industrial biosyntheses (Boehm *et al.*, 2013). Reconstituted cell-free protein synthesis systems such as the protein synthesis using recombinant elements (PURE) system allowed synthesis of a wealth of recombinant proteins. This family of cell-free systems keeps being improved (Grasemann *et al.*, 2021). Yet, despite continuous progresses (Zhang *et al.*, 2021b), sometimes for industrial purposes (at least for the synthesis of proteins used by research laboratories), the speed and accuracy of the translation process remained significantly below what happens *in vivo*.

Several biological functions were identified as critical for improving the outcome of cell-free syntheses. For example, proper protein folding was critical. In contrast to the widely accepted, but wrong, Anfinsen’s postulate that assumes that the whole folding potential is present in the protein primary sequence, it is now established that most unfolded proteins cannot spontaneously fold properly (To *et al.*, 2021). As a first step, the ribosome

itself is important as a folding device during translation, but the trigger factor (Piette *et al.*, 2011; Koubek *et al.*, 2021) and other molecular chaperones (Yang *et al.*, 2019) have often a decisive role. Many are dissipating energy (Boel *et al.*, 2019; Mallory *et al.*, 2019), which impacts the energy balance of the procedure, a feature rarely discussed. Furthermore, it appeared progressively that the role of water and molecular crowding were critical to ensure fidelity of translation [see, e.g. Garenne *et al.* (2020)]. The importance of these functions is not yet fully understood and it is likely that their role will impact future developments of cell-free syntheses. Their role is to discriminate between classes of entities (young and aged objects and also spatial positions or alternative structures, for example). This is illustrated, for example, by innate immune sensors such as RIG-I, which sensitively detect and respond to viral RNAs that enter the cytoplasm, while remaining unresponsive to the abundance of structurally similar RNAs that are the products of host metabolism (Ren *et al.*, 2019). Similarly, the SecA ATPase, which binds its substrates post-translationally, scans the ribosomal tunnel for potential substrates to secrete the correct ones (Knüpffer *et al.*, 2019). The information for discrimination is managed by the energy dissipation reflected by the ATP dependence of the enzyme (Boel *et al.*, 2019). In terms of SynBio developments, consideration of these agents, particularly ATP-dependent proteases such as protease Lon, will have a considerable impact on the production of canonical proteins (Tzeng *et al.*, 2021).

Yet another idiosyncratic feature of proteins has also been identified as having a likely impact on cell-free protein biosynthesis. While proteinogenic residues are universally considered to be 'amino acids', proline is a secondary cyclic amine (formerly, called 'imino acid'). This characteristic requires the translation machinery to take this chemical singularity into account. Fast and accurate translation of proline residues requires a specific variant of a post-translationally modified translation elongation factor, EF-P, or other counterparts depending on the organism (Hummels and Kearns, 2020). Yet, the explicit use of EF-P has seldom been explored as an explicit partner in cell-free protein synthesis (Li *et al.*, 2017; Hammerling *et al.*, 2020). Sequences of three consecutive prolines can fold into polyproline helices, structures that join alpha helices and beta folds as architectural motifs in protein configuration. Triproline helices are involved in protein–protein signalling interactions. Management of elongation factor EF-P in cell-free synthesis of proteins with series of proline residues should improve the outcome of a protein production line. This is not a minor point, as, for example, more than 4% of the human proteome contains series of three or more prolines (Morgan and Rubenstein, 2013). Furthermore,

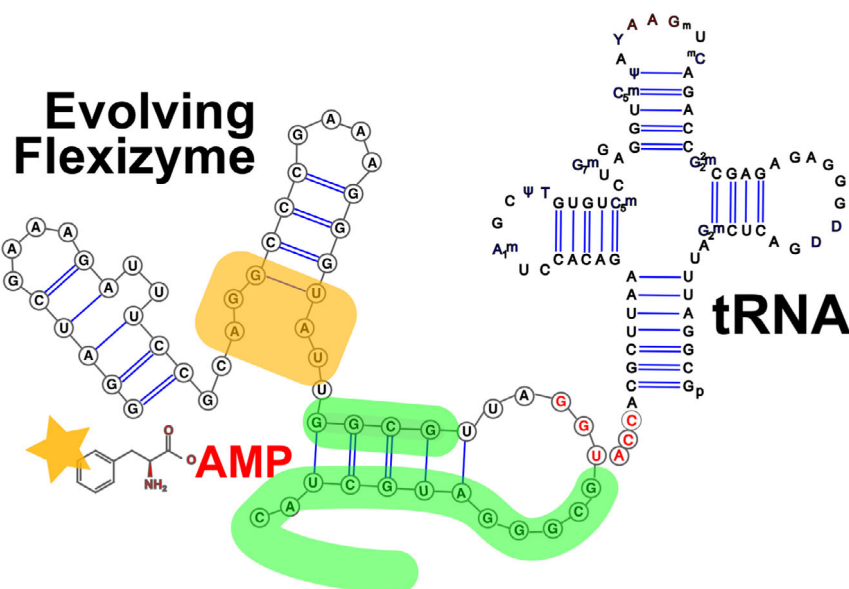
since EF-P binds to the 9 nt D-loop found in tRNA<sup>Pro</sup> isoacceptors, and not to proline itself (Kato *et al.*, 2016), the use of this factor can be expected to play a key role in the precision of cell-free translation engineering, a feature essential for SynBio developments. After half a century of hard work, the various obstacles to the development of *in vitro* synthesis continue to be improved, with the emphasis on speed (Burrington *et al.*, 2021). Much remains to be understood to allow the industrial process of translation to be stable in time, with proper maintenance activities in particular. We may expect that exploring these weak points will dominate research in the domain in the near future. This is more important because, during the spontaneous process of protein ageing – which is protein specific and spanning a considerable time frame, from minutes to centuries (Zhang *et al.*, 2020) –, proteins tend to aggregate after subtle conformation changes in the absence of the energy requirements that allow a subset of NTP-dependent activities to refold proteins in their proper shape, dissolving aggregates (Dewachter *et al.*, 2021).

In addition to using *in vitro* protein synthesis as a means of identifying all the key partners of the *in vivo* process, cell-free protein synthesis also allows proteins to be synthesized with a variety of non-proteinogenic amino acid analogues. Worth highlighting, the 'flexizyme' system based on ribozymes that are able to load a large number of chemical substrates onto tRNAs [(Passioura and Suga, 2014) and Fig. 2] eliminates the need to modify aminoacyl-tRNA synthetases prior to synthesis. We can expect that this family of cell-free approaches will make a considerable part of next-generation SynBio developments, in particular for the synthesis of polymers of industrial interest. This will enable the synthesis of polymers involving structurally disparate non- $\alpha$ -amino acids that bear little resemblance to canonical proteinogenic amino acids. Examples include long-chain carboxylic acids, substituted benzoic acids, 1,3-dicarbonyls, peptides and helical aromatic oligomer-peptide hybrids. For example, substituted benzoic acids are the precursors of aramids, a class of aromatic polyamides with a rich progeny. Similarly, long-chain amino acids are monomeric precursors of nylon (Tharp *et al.*, 2021).

These novel trends will benefit both from development of RNA SynBio and from streamlined cell-free systems amenable to scaling up.

As their name suggests, cells are compartments, specifically delimited by membranes built around lipid bilayers. Early on, as it became apparent that DNA-driven *in vitro* protein synthesis was often difficult to achieve, some researchers resorted to the spontaneous ejection of chromosome-less vesicles by some cells as an alternative to a purely *in vitro* system (Bose *et al.*, 2020). These vesicles, 'mini-cells', often kept plasmids in





**Fig. 2.** Flexizyme RNAs as substitutes for aminoacyl-tRNA ligases [from Ohuchi *et al.* (2007)]. Starting from an aptamer that binds phenylalanyl-AMP, synthetic variants (random sequences in the region highlighted in green) are constructed and submitted first for selection for binding to the 3'-CCA end of phenylalanine tRNA, then to a further selection after variants (region highlighted in orange) are proposed for the binding of analogues of phenylalanyl-AMP (orange star).

their cytoplasm, allowing protein expression from artificial constructs (Grindley *et al.*, 1977; Miller *et al.*, 1977; Bose *et al.*, 2020). Similarly, in animal cell-free systems, exosomes or cell-derived small extracellular vesicles have been used in therapeutic intervention (Pedrioli *et al.*, 2021). Lipid vesicles have also been used for some time and they are now used for glycosylation of proteins in cell-free systems (Hershewe *et al.*, 2021). Among the various hurdles still remaining to resolve are the synthesis of the lipid bilayer of the cell-free systems membranes (Bhattacharya *et al.*, 2021) as well as routine synthesis of active membrane proteins. Novel transcription–translation systems are now able to express integral membrane proteins. This allowed researchers to highlight the importance of lipid–protein interaction in maintaining structural–functional integrity of ion channels (Vaish *et al.*, 2018). Membrane-augmented cell-free systems have been used in a wide array of cellular processes from primary and secondary metabolite synthesis to electron transport and one-carbon metabolism (Kruyer *et al.*, 2021).

Following on from Spiegelman's early experiments, an important step in understanding the boundary between *in vivo* and *in vitro* syntheses, the synthesis of a fully active virus has been achieved (Rustad *et al.*, 2017). While viruses cannot be considered alive – they need a live cell to multiply – the very fact that they can be made *in vitro* blurs further the boundary between chemistry and biology. Still, the cell-free systems used to achieve this milestone are extracted from living cells, and are far

from sufficient to be able to duplicate spontaneously. Nevertheless, this achievement paves the road to identify the biological functions that are responsible for the signature 'animation' of live cells (Danchin, 2021). However, much work remains to be developed to implement them knowingly in cell-free systems. We can expect that this will be understood in a not too far away future, in particular if we understand better the role of apparently expletive energy-dissipating functions, such as those involved in the control of folding and aggregation (Boel *et al.*, 2019; Dewachter *et al.*, 2021).

#### Streamlined cell-free syntheses

In parallel with these advances, we are now in a new age of cell-free biology, allowing applications of three variations on the theme: Not only authentic cell-free transcription–translation systems, as just discussed but also protein-based and nucleic acids-based systems (Noireaux and Liu, 2020). Protein-based streamlined cell-free systems can be used to produce fine chemicals as well as precursors of heavy industry-relevant compounds. As these processes do not require the industrial setup to sustain life, they tolerate the presence of otherwise toxic molecules (Rasor *et al.*, 2021). Among these, a very innovative approach has been explored early on. Rather than reconstituting *in vitro* the conditions prevailing *in vivo*, the idea, intended to improve industrial metabolic engineering, was to find a way of isolating all the enzymes needed to put together a metabolic pathway in

a common container. Interestingly, the container could be of considerable size, an industrial reactor, for example. It is obviously time-consuming to synthesize and purify individually all the enzymes of a pathway of interest. Why not use the established outcome of natural selection, with all enzymes in correct proportions in relevant cells, while getting rid of all components not required in the pathway? The operation would develop into two steps: (i) growing cells at a high density under conditions when the pathway is fully functional and boosted for production of the metabolites of interest; and (ii) making a whole extract of the cells, finding some intelligent ways to separate the metabolic fraction from the bulk of the cell's machinery, and then finding a way to keep in a reactor only those enzymes that are required in the pathway.

This can be achieved after 'parasitic' reactions have been identified if one can discard the parasitic elements. Panke and co-workers proposed an astute approach to solve this quandary. In a first analysis, they boosted, *via* standard SynBio approaches, the pathway of interest. So doing, they also monitored all the changes that went up in parallel, drawing key resources away from the required pathway (Bujara and Panke, 2012). This interference was mediated by proteins, mainly enzymes and their regulators. How to get rid of them? Inactivating the corresponding genes directly would jeopardize production *via* destroying the network of interactions that make the cell functional. However, it is possible to tag the corresponding proteins by including in their genes, at positions that do not interfere with function (Oesterle *et al.*, 2017), sites that code for protease-sensitive peptide motifs. In this way, when the cell extract is obtained, it can be incubated with a protease acting specifically on the tagged regions and, thus, removing unwanted proteins. After a purification step, the extract essentially contains the enzymes of the pathway of interest and, with adequate stabilization, can be used in a chemical reactor.

In line with this approach, another family of advances is being developed. When extracted directly from living cells, cell-free systems always develop endogenous metabolic activity. The details of this activity are not yet understood but, as just developed, they interfere with the production of the compounds of interest. Using metabolomics to characterize time-dependent metabolic changes in cell-free systems and their components, including significant metabolic activity in central carbon and amino acid metabolism, Styczynski and colleagues developed several approaches to prepare extracts prior to production development. They observed that while changing the starting state of the reaction *via* pre-incubation of the lysate has an impact on protein production, its impact on the metabolic state was comparatively

small. Changes in lysate preparation had a greater effect on protein yield and time-dependent metabolic profiles, while general metabolic trends were maintained. Finally, while targeted supplementation of metabolic enzymes improved protein production, endogenous metabolic activity remained resistant to these enzymatic perturbations, calling for further development to improve the processes of interest (Miguez *et al.*, 2021). We can expect that in the next decade or so these families of approaches, combined in a variety of sequential reactors, including some involved in standard chemistry, will create novel designs for fine chemical syntheses.

### *Evolution in vitro*

Another crucial family of experiments in cell-free systems is the exploration of evolution, a feature recognized as critical to biology and commonly exploited *in vivo*. It is indeed possible to develop evolution experiments in cell-free conditions. As an illustration, Spiegelman again took an aliquot of the Q $\beta$  RNA synthesized in the test tube and added it to a second identical test tube, then repeated the procedure 15 times. At the end of this sequence, the original RNA was almost completely diluted out by the transfers, leaving in the final tube RNA that was all generated *in vitro*. Repeating these experiments allowed him to witness the evolution of RNAs. By varying the selective pressure on the system (such as temperature, nutrient mix or time allowed for reproduction), he could generate mutant RNA molecules with a wide range of properties. One of those, for instance, gradually got trimmed down from the original 3300 nucleotides to 470 – just sufficient to bind the replicase and replicate the RNA without further loss of its sequence. This small molecule had shed sections that coded for the virus' protein coat and other components it did not need in the artificial environment (Levisohn and Spiegelman, 1969). It should be noted that processes similar to this cell-free evolution also occur *in vivo* and are important for recombination, for example, with existing viruses such as SARS-CoV-2. This largely explains the ability of viruses to evade innate and acquired immunity, as observed several decades ago (Meyer and Southern, 1997). In parallel with functional copies of the authentic virus, cells keep shedding short counterparts that essentially kept the 3' and 5' end required to initiate and terminate replication (Gribble *et al.*, 2021). This indicates that the process of viral replication is prone to produce a large number of variants that could be exploited for use in cell-free expression systems. Based on this early work and recent observations, advances in the *in vitro* synthesis and evolution of DNA, RNA and polypeptides have already been designed to accelerate the construction of biopolymers, pathways and

organisms with new functions (Forster and Church, 2007). This trend is set to continue.

A most important sequel of Spiegelman's early experiment has been the systematic evolution of ligands by exponential enrichment (SELEX) procedure devised by Larry Gold and his colleagues (Tuerk *et al.*, 1992). After the discovery of ribozymes, it was clear that RNA molecules could fold into complex structures able to bind foreign molecules [aptamers, riboswitches and ribozymes (Ge and Marchisio, 2021)]. This prompted exploration of novel RNA-based activities, substrate-dependent ribozymes in particular. Using a variety of *ad hoc* selection procedures, SELEX allowed isolating functional RNAs from a pool of random sequences of RNA, in particular endowed of catalytic activities. Among the critical functions, as discussed above, the mimicry of aminoacyl-tRNA synthetases by ribozymes was an interesting target. Perhaps the first example of ribozymes capable of charging certain amino acids onto RNA was developed by Yarus and co-workers. They isolated synthetic ribozymes that catalyse aminoacylation of their own CCG-terminal 2'(3')-OH with phenylalanyl-AMP or tyrosyl-AMP as an aminoacylation donor (Illangasekare *et al.*, 1995). However, this early class of ribozymes was still unable to aminoacylate the CCA 3' end of a tRNA. Using a first family of ribozymes, Suga and co-workers designed variants of an artificial ribozyme able to bind CCA motifs and comprising random segments. These RNA molecules were submitted to a selection procedure meant to isolate those which got the ability to aminoacylate various tRNAs with phenylalanine or non-natural derivatives (Fig. 2). One of those was able to aminoacylate its cognate tRNA with a high degree of specificity, while failing to aminoacylate non-cognate tRNAs. Further evolution *in vitro* allowed the authors to load analogues of phenylalanine of a specific tRNA. Changing the tRNA load specificity could be easily programmed into such flexizymes, fashioning them into custom-made catalysts to generate non-natural aminoacyl-tRNAs (Ramaswamy *et al.*, 2004). This family of advances, based on *in vitro* RNA evolution, is likely to see considerable developments in the near future (Kofman *et al.*, 2021).

All these advances require complex laboratory facilities, with development time limited by the constraints of physical experiments, hours, days, months or even years. As with a considerable number of existing technological developments, it is now possible to replace approaches that involve wet laboratories with computer-based exploration. State-of-the-art computer modelling and design technologies allow us to achieve results that engineers could not even dream of a couple of decades ago. For example, the manufacture of environmentally interesting devices starts with the succession, concept, model, computer-aided testing of functions and

resilience, and finally the creation of a prototype, with feedback to material experiments. The conceptual development of an aircraft starts from a creative search stage: choice of style, composite approach, creation of sketches and drawing of the future model, and all this can be done with the help of computers (Abbasov, 2019). While laboratory experiments were the rule rather than the exception, it soon became apparent that, starting with data analysis and storage, true *in silico* experiments – using computers – could and should be developed.

### Synthetic biology *in silico*

*In silico* biology is now a major domain of SynBio developments, with one caveat. The exponential development of computers and the speed of related advances relative to the time required to obtain results from laboratory experiments have led to the creation of a large community of researchers and engineers who claim to be able to substitute *in silico* biology for experimental ('wet') biology. Alas, this is often at the cost of understanding what life is. The main consequence of the shift from *in vivo/in vitro* to *in silico* research is that a significant fraction of the community has turned inward, with its own journals and evaluation rules, reviewing and evaluating the work it produces without the critical steps that would require predictions or explanations – validation – based on authentic biological experiments. Furthermore, the very fact that *in silico* biology – modelling in particular – requires skills in both mathematics/computer sciences and biology means that a great many studies escape proper evaluation (when they are about biology, they are considered to be computer sciences, and vice versa). This results in a large amount of work which is not relevant for answering the questions posed by SynBio (Danchin *et al.*, 2018). This downside has a considerable impact when we have to follow the explosion of articles in the domain. It is virtually impossible to go beyond scraping the surface of the amount of literature claiming to be *in silico* biology. At the risk of missing important work, we have tried here to avoid this pitfall by giving an overview of a limited number of the work and trends that may provide constructive developments in the near future.

Mathematics/computer-based models of a variety of processes essential to life have long been developed, with two different aims. For diagnostic purposes, models were meant to produce a phenomenological account of the dynamics of processes of interest. The purpose of such *simulations* was not to understand their causes but to allow researchers to make predictions about the short-term and sometimes long-term outcome of the processes. This family of approaches is well established. It

is still witnessing substantial developments, in particular in the domain of health-related studies.

### *Simulation vs. explanation*

In the context of SynBio, simulations can be used as a first pass meant to measure whether a particular family of experiments behaves consistently as predicted, thus substantiating the relevance of the collected data and pointing out missing elements. The power of this family of models parallels the continuous advances of computing capacities associated with the continuous emergence of faster computers. In recent times, artificial intelligence (AI) approaches based on deep learning flourished as tools of choice in the domain of diagnosis (Akay and Hess, 2019). Many AI-based models are currently limited to diagnosis because successful predictions are not sufficient to tell correlation from causality, in particular when the underlying reasons for a prediction 'success' cannot be traced easily. To make the most of these models and use them as help to discovery requires that the outcome of their operation can be traced back to a causality chain. This restriction explains why legal regulatory instances have now asked creators of AI-based models to be able to highlight the internal chain of causality in their successful models. This is understood as a way to associate prediction with understanding [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 and for an example of the way understanding can be visualized in an AI model, see Prifti *et al.* (2020)]. This limitation should be extended to a vast number of models. Here is an illustration of the difference between phenomenology (simulations) and identification of causality (rational scientific modelling).

The behaviour of cells on plates is often used to illustrate SynBio construction [see a famous early illustration in (Elowitz and Leibler, 2000)]. However, a large number of phenomenological models can be used to account for this behaviour (McCallum and Potvin-Trottier, 2021). For example, it has long been observed that colonies on plates often generate patterns that have dendritic shapes (Fall *et al.*, 2004). This is generally 'explained' by models involving repulsion, multi-scale patterns, self-organization and all kinds of buzzwords [see, e.g. Deng *et al.* (2014)]. Yet, analysis of the causal chain of events that lead to these patterns demonstrates that the explanation of the apparent repulsion is not mediated by any 'self-organization' magic force. Notwithstanding a possible Marangoni effect, if movement involves surfactants and quorum sensing (Daniels *et al.*, 2006), 'repulsion' is driven by the – almost trivial – role of the nutrient source in the cell's behaviour: Bacteria need nutrients to multiply. They evolved a chemotactic response that makes them move towards nutrient sources (and, therefore, avoid

regions where they have been depleted). Understanding the authentic chain of causality is important: What appears to the naïve mind as a self-organized repulsion is in fact the distribution of moving cells in regions where nutrients are available (Sekowska *et al.*, 2009). This experiment illustrates the dichotomy of modelling, which can lead either to phenomenology or to understanding; the latter being the only way to generate real advances. To move beyond the phenomenology of simulations, one must interpret the data critically, looking for genuine causality and refraining from appealing to folk magic.

### *Standardization*

To make the most of computer-mediated exploration, it is critical that models are built on very precise definitions and cardinal foundations, not on fuzzy vocabulary. This topic, in particular, the need of accurate definitions, is discussed in details by Clark and Hicks in the field of chemistry, showing how the choice of models impacts the outcomes of research, including with modern AI approaches (Clark and Hicks, 2020). Because models rest on the collection of relevant data, the key question of standardization and nomenclature must be asked, with expedient approaches (Decoene *et al.*, 2018). A great many developments of SynBio are based on manipulation of nucleic acids. Because these polymers are made of the chaining of elementary building block, they are generally handled by computers as are abstract linear alphabetic texts. Intelligent storage and analysis of nucleic acid and also protein sequence data began four decades ago (Blumenthal *et al.*, 1982). Subsequently, an explosion of applications resulted in the creation of a new domain of biology, bioinformatics, with a slight change in the initial meaning of the word, now often restricted to genome-based studies (Hogeweg, 2011). Yet, while this restriction is commonplace, the advent of genomics, with its related 'omics' studies (Yadav, 2007), has considerably expanded the domain, with only fuzzy boundaries with recently fashionable domains such as AI and image analysis (Burian *et al.*, 2021).

Gene, genome and some transcriptomics data are collected, managed and released in a highly standardized way by the International Nucleotide Sequence Database Collaboration (DNA Database of Japan, Mishima, Japan; European Nucleotide Archive, Hinxton, UK; and GenBank, Bethesda, USA), which has developed standards for more than three decades (Arita *et al.*, 2021), and controlled vocabulary is implemented for proteins by UniProtKB (Feuermann *et al.*, 2021). The Gene Ontology Consortium maintains a knowledge base that '*develops a comprehensive, computational model of biological systems, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life*', also

based on controlled vocabulary (Gene Ontology Consortium, 2021). However, while the nomenclature of chemicals and chemical reactions tend to follow well-accepted rules [such as in CheBi (Hastings *et al.*, 2016) or Rhea (Lombardot *et al.*, 2019) for example], with efforts to allow interoperability (Navelkar *et al.*, 2021), the nomenclature of enzymes – besides the EC number standard – is still quite variable. Gene names are considerably dependent on the goodwill of authors and this has considerable negative impact on sequence annotation (and, therefore, discovery). Also, the data structure of specialized databases is still extremely disparate, which prevents interoperability (Danchin *et al.*, 2018). When dealing with SynBio, the situation is possibly less disappointing. The iGEM repository of Biobricks (<https://biobricks.org>) and, in Europe, the SEVA effort (Martínez-García *et al.*, 2020) are pushing hard for widespread standardization. A further incentive for an international coordinated effort in this direction would be more than welcome to support the development of SynBio in the next decades.

All the same, three major domains of *in silico* research for SynBio develop in parallel: 'omics-based' SynBio, whole-cell models and *in silico* integration of metabolic pathways.

#### 'Omics-based' synthetic biology

Besides the domain of genomics, born and diversifying into a number of subdomains, the fields of transcriptomics, proteomics and metabolomics that progressively sprouted out of this original domain (Tyo *et al.*, 2010; Pühler, 2012) are now key for SynBio developments. The high-dimensional datasets generated by these experimental approaches require considerable storage capacity and widespread sharing. A large number of analysis methods must also be made available. Sometimes it is possible to combine both storage and availability of analysis methods. In the case of bacterial genome data, this is the case, for example, with the MicroScope platform and its implementation of the MaGe methods (Vallenet *et al.*, 2020). Repositories such as GitHub with its community of open-source developers are progressively gaining more and more impact (see, e.g. its SynBio entry: <https://github.com/topics/synthetic-biology>). Cloud computing provides generally low-cost and highly flexible solutions in the domain of molecular modelling, omics data analytics (e.g. RNA sequencing, metabolomics or proteomics datasets) and for the integration, analysis and interpretation of phenomes (Koppad *et al.*, 2021). Merging highly heterogeneous datasets has been a prime goal early on (Arakawa and Tomita, 2013). This is critical because multi-omics may be used to improve systematic genome manipulation for

novel synthetic constructs (Fontana *et al.*, 2020) or improve industrial processes. Among the many illustrations of the approaches for SynBio applications, we may notice fine cellulose biosynthesis (Ryngajito *et al.*, 2020), treatment of natural biomass, in particular lignocellulose (Xie *et al.*, 2014; Chen and Dou, 2016), or in the present context of general concern triggered by climate change, genome manipulation carried out for the metabolic engineering of Cyanobacteria (Hagemann and Hess, 2018; Ng *et al.*, 2020).

The data introduced in modern omics analyses allow for efficient characterization of the genetic, regulatory and metabolic phenotypes of engineered microbes. Yet, designing genetic interventions to achieve the desired phenotype remains a challenge. With recent developments in genetic engineering techniques, the time frames associated with constructing and testing strain designs have been greatly reduced through *in silico* modelling. This has created an efficient design–build–test–learn loop iteration cycle between experiment and analysis. However, the scale and complexity associated with multi-omics datasets still require manual biological reasoning about the mechanisms driving phenotypic changes. As a consequence, the use of traditional statistical approaches can still reduce the dimensionality of these datasets and help compare strains of interest and improve metabolic engineering (Danchin *et al.*, 2018; St John and Bomble, 2019).

Besides modelling individual cells, modelling is also used to understand how microbial communities and their synthetic counterparts may form stable ensembles. Models allow researchers to better monitor the changes in microbiome composition and, for example, how this affects disease onset and development. Construction of models that integrate genomics, transcriptomics, proteomics and metabolomics allows researchers to visualize the meta-metabolism of the community. Constraint programming (Brown and Miguel, 2006) is used to couple meta-metabolism and multi-omics analyses (Ebrahim *et al.*, 2013). As an illustration, COBRA modelling combined with meta-omics analyses and multivariate statistical analyses are used as a tool to add value to clinical trials and ultimately propose therapeutic interventions (Heinken *et al.*, 2021). This type of work, when submitted to relevant constraints, allowed construction of a synthetic gut microbiome (Mabwi *et al.*, 2021).

Finally, we have seen that miniaturization was involved in developing novel approaches for parallel evolution of microbial cultures. This domain is also used to explore the consequences of the use of novel technologies based on new materials, for example, nanomaterials, which, because of their size, may be a matter of concern. Modelling using multi-omics approaches explored the impact of nanomaterials on cells or

pathways *via* interpretation of omics data (Mortimer *et al.*, 2021).

### Simulating whole-cell behaviour

In parallel with the sequencing of whole genomes, researchers proposed to shift efforts from *in vivo* biology to reconstitution of whole cell in computer-mediated models. Early attempts focused on *E. coli* at a time when DNA sequencing was not yet available (Weinberg and Berkus, 1971). The revolution brought about by sequencing resulted in a nucleic acid-, then genome-centric view of cells, but models were initially restricted to simulations based on the way macromolecule and intermediary metabolism could integrate functions previously identified from the study of individual genes.

As an early illustration of the integrated behaviour of cells, modelling was developed around a generic software environment that would explore, from the knowledge of their genome, how cells behave in all their biological properties. A first simulation software pipeline, E-Cell, was made publicly available in 1997, poised to exploit the then recently acquired genome data of *Mycoplasma genitalium* (Tomita *et al.*, 1997). This generic software environment for simulating a cell was subsequently improved to conduct *in silico* experiments based on numerical data obtained from *in vivo* experiments involving other organisms (Tomita *et al.*, 1997; Tomita, 2001; Ishii *et al.*, 2004). Among applications of the software was the exploration of an osmoregulatory switch in the growth properties of *E. coli* (Srividhya and Krishnaswamy, 2004) and the deciphering of red blood cells' metabolism (Yachie-Kinoshita *et al.*, 2010). The domain of mathematics involved in computation was based on differential equations. This preference is not without consequences, as it implies that biological objects are assumed to display a continuous behaviour, with their concentration represented as a real number not as a discontinuous integer. While this is good enough for objects having a concentration in the millimolar range or above, this is inappropriate when individual objects must be identified, such as regulatory molecules. For example, regulators or co-factors have often a concentration in the sub-micromolar range, or are present in cells such as an individual *E. coli* in less than a handful of individual molecules. Ten years after its start, the E-Cell project experienced a thorough reorganization, now focusing on simulations meant to help industrial applications. This focus on simulation rather than understanding may explain why, after a series of improvements, the project, while still existing (<https://www.e-cell.org/about.html>) progressively stalled. At present, the whole-cell modelling of *E. coli* follows another turn, still based on standard approaches but with some input of stochasticity to take

into account entities that are present at a low number in cells (Sun *et al.*, 2021).

More fitting to SynBio approaches, whole-cell models may be used as a basis for genome design based on previous knowledge. The bacterium *Mycoplasma genitalium* is represented by an updated whole-cell model (<https://github.com/CovertLab/WholeCell>). At this time, this framework combines 28 cellular submodels, integrating all the known functions of every gene and molecule in the cell, running individual models on multiple supercomputers (Rees-Garbutt *et al.*, 2021). This design allows researchers to propose knockouts of individual protein-coding gene, phasing out genes previously identified by their importance for growth and to test possible phenotypes to identify synthetic lethal associations. This semi-empiric work depends on known phenotypes, conducting whole-cell model simulations in three steps: design (the outputs of algorithms select possible gene deletions), simulation (behaviour of the genome minus those deletions) and testing (the behaviour of the resulting cell is analysed). Several simulations are conducted in parallel. Those that removed the most genes from the genome and produced cells that were still able to divide are kept for further study. Once this step is completed, the software proceeds to the next cycle, increasing the number of gene deletions and generating progressively smaller genomes which can be subsequently used as synthetic constructs. This bypasses the development of slower experimental approaches based on the construction of multiple knockout mutants, such as those which uncovered hidden reactions in *E. coli* (Nakahigashi *et al.*, 2009), showing how enzyme promiscuity could combine with paralogous metabolism (Chan *et al.*, 2014; Guzmán *et al.*, 2019) to generate unexpected pathways. Similar genome-scale modelling is also developed using Eukarya such as *Yarrowia* (Czajka *et al.*, 2021).

Based on genome data, these models are very general but not realistic. Their properties of symmetry do not take into account the asymmetry of the cell's geometry. Cells are generally not spheres, and taking into account the real time-dependent distribution of their metabolites is far from straightforward. Furthermore, compartmentalization is a critical feature of the development of metabolism (de Lorenzo *et al.*, 2015; Bohrer and Xiao, 2020) and this implies that realistic models must make the most of representations that consider the cells' shape and the way they divide. Exactly as illustrated previously with the case of populations of cells moving on plates, many phenomenological models try to fit with 3D observations to represent the cell's geometry. Bacteria are often rod-like structures. Analysis of the constraints that account for this family of shapes provides interesting views [see, e.g. Chang and Huang (2014); Banerjee *et al.* (2016)]. Yet, the ultimately disparate causes

resulting in bacteria being rods have a common source based on metabolism: Growth must put together compartments that have widely different spatial properties, while the core of metabolism that generates major building blocks is likely to grow in three dimensions. Indeed, growth of the cytoplasm is three dimensional, while membranes are two-dimensional, and the genome is one-dimensional. The major source of building blocks is expressed in the cytoplasm, and therefore, grows in three dimensions, unless specifically regulated or organized metabolic limitations prevent non-homothetic growth of the various cell's compartments. Remarkably, it is cytosine metabolism (and more generally *de novo* pyrimidine synthesis) that overcomes this obstacle, with all pyrimidine-related pathways converging on the activity of CTP synthetase (Ou *et al.*, 2020; Danchin, 2021). The consequences of this metabolic adaptation have not yet been input in models of bacteria *in silico*, but this will have considerable impact for the future of SynBio metabolic engineering developments. Many microbes have a structure that departs from that of spheres or bacillus shapes. They make filaments during their growth, either under natural circumstances (e.g. streptomycetes and filamentous fungi), but they also make filaments under stressful conditions or during ageing. This phenomenon is important for cell factories, and phenomenological models have been proposed to understand and control filamentation, with the aim of designing the morphogenesis of fungi (Meyer *et al.*, 2021) or Cyanobacteria (Yamamoto *et al.*, 2021). The process of filamentation has been studied for a long time, but the link to metabolism has generally not been explored, apart from the identification of targets such as LexA in bacteria, with limited *in silico* modelling (Bellio *et al.*, 2020).

Finally, in yet another register, modelling was used to improve cell-free protein synthesis, in a typical application of *in silico* approaches to complement *in vitro* experiments (Müller *et al.*, 2020). For example, the domain of heterologous protein expression benefited from approaches allowing codon optimization, a domain explored at the very beginning of genomics (Ang *et al.*, 2016). At a finer level of granularity, the evolution of proteins was modelled as a function of their abundance as well as environmental parameters such as temperature. Interestingly, in addition to their expression level, protein evolution is determined by the concentration of chaperones, highlighting the critical role of these neglected essential factors (Agozzino and Dill, 2018). More in line with larger scale physiological data, integration of a variety of approaches investigated the role of stress and helped in the improvement of recombinant protein synthesis (Chen *et al.*, 2021c). At this point, we may understand that we are still a long way from having whole-cell models that go far beyond simulation. Nevertheless,

modelling can be successfully used to guide SynBio designs. Moreover, based on semi-empirical approaches, intermediary metabolism is now more suitable for building predictive models that help streamline existing metabolism while exploring the novel avenues we mentioned earlier.

#### Whole-cell metabolism

It is common to assume that metabolism follows textbook pathways. Yet, unexpected metabolic pathways are the rule rather than the exception and the metabolic capacity of virtually all organisms is vastly underappreciated (Medema *et al.*, 2021). How can we connect our knowledge of genomes and whole-cell metabolism? Back in the XIX<sup>th</sup> century, August Weisman postulated that living organisms combined two major characters, with a *germen*, that was responsible for the hereditary properties of the organism and a *soma* that harboured the germen while displaying observable features. This distinction is still valid and recognized as the genotype/phenotype dichotomy. With the advent of DNA sequencing and identification of the genome sequence, this dichotomy is now often referred to as the genome/phenome question, and this is the subject of explicit *in silico* investigations (Norsigian *et al.*, 2020). Metabolic engineering, developed along an integrated view of the genome/phenome dichotomy, associates traditional metabolic engineering (Bailey, 1991) with systems biology, synthetic biology and evolutionary engineering (Jang *et al.*, 2019). It is enabling the development of microbial cell factories capable of producing efficiently a myriad of chemicals and materials, including biofuels, bulk and fine chemicals, polymers, amino acids, natural products and drugs. These developments are based on integrated circuits within metabolic networks, preferably using multi-omics high-throughput data collection (Chen and Li, 2016).

Combined with genome engineering and genome-wide metabolic simulations, as discussed above, many tools and strategies have been developed to generate microbial cell factories (Ko *et al.*, 2020). Here, we focus on a family of models that can be used as predictive tools to assess metabolic constructs. Based on flux balance analysis [FBA, (Edwards and Palsson, 1998)], genome-scale metabolic models [GSMM (Edwards and Palsson, 1999; Xu *et al.*, 2013)] have been developed to explore the metabolism of a large number of microorganisms of industrial interest, such as *P. putida* (Belda *et al.*, 2016) or *B. subtilis* (Belda *et al.*, 2013). Being based on the general constraints of stoichiometry (Moyer *et al.*, 2021), FBA is generic and may accommodate the network structure of any metabolic pathway (Antoniewicz, 2021). For example, the details of the complex metabolism of

actinomycetes are now clarified, allowing metabolic engineering in these organisms (Palazzotto *et al.*, 2019). Furthermore, these applications are not restricted to bacteria: Flux balance analysis guided the evolution of a yeast chassis (Pereira *et al.*, 2021).

Other genome-wide metabolism models have been implemented to explore *in silico* the metabolism of short-genome bacteria such as *Mesoplasma florum*. Using sequence and structure homology, the set of metabolic functions that its genome encodes was identified, allowing the reconstruction of a metabolic network representing one third of its protein-coding genes. Simplification of the growth medium allowed quantification of substrate uptake and product secretion rates, which, together with the experimental biomass composition, were integrated as species-specific constraints to produce a functional model of metabolism at the genome scale (Lachance *et al.*, 2021). At a larger scale, a whole-cell modelling with emphasis on metabolism is continuously improved for the model *E. coli* (Sun *et al.*, 2021). This effort is particularly important because *E. coli* remains the best known organism, with continued progress in understanding the functions encoded in its genome, based on laboratory experiments.

Modelling of metabolism is also useful to explore communities of organisms (Vázquez-Castellanos *et al.*, 2019). This family of approaches is particularly timely because we are beginning to understand, through the study of seawater, what metabolic 'currencies' are shared by the millions of microbes that coexist in this environment (Durham, 2021). Microbial co-cultures have been studied using FBA, with interesting observations involving cross-feeding (Konstantinidis *et al.*, 2021). Another type of interaction between microbes is witnessed by the role of viruses, which are able to reprogram the metabolism of their host (Jacobson *et al.*, 2021). Understanding the corresponding processes may allow construction of specific pathways for metabolic engineering.

### Evolution in silico

Many evolutionary experiments use model organisms, but they are slow (Sekowska *et al.*, 2016; Grant *et al.*, 2021) and the study of evolution *in silico* remains the approach of choice. Adaptive evolution in the laboratory has been developed for organisms of industrial interest. The process benefits greatly from *in silico* modelling (Lee and Kim, 2020). In this context, exploration of the potential deleterious effects of synthetic mobile elements in genomes has been studied *in silico* (Zamdborg *et al.*, 2015) and early work was designed to predict the evolution of the virulence of bacteria of interest to agriculture (Strauß *et al.*, 2016). In line with the objectives of

SynBio, based on families of enzyme variants, modelling is used to design synthetic counterparts with *ad hoc* catalytic properties (Broom *et al.*, 2020; Bunzel *et al.*, 2021). More generally, *in silico* modelling allows engineers to explore the evolutionary landscape along lines that would be impossible to follow in wet lab experiments, particularly when several alternative pathways are possible (Ambrus *et al.*, 2020). Furthermore, frequent evolution by gene duplication creates unexpected avenues that must be explored for the consequences of the regulatory properties they create (Marchant *et al.*, 2019).

Understanding gene networks is crucial for interpreting evolution and predicting evolutionary pathways (Gautam and Kumar Sinha, 2021). This task is difficult in the presence of noise, a pervasive feature, and requires special treatment (Vatsa and Agarwal, 2021). It is particularly important to understand the evolution of regulator binding sites. Comparative analysis of bacterial genomes has revealed that such sites are rapidly created at random during evolution, and then retained when they have a positive effect on the organism. Interestingly, the patterns of evolution depend on the nature of the regulator (Mrázek and Karls, 2019). A parallel software package, GeNESiS, was designed for the modelling and simulation of the evolution of gene regulatory networks (Kratz *et al.*, 2008). Various models of gene regulatory networks are available (Santibáñez *et al.*, 2020; Chen *et al.*, 2021b), in particular dynamic models will be of interest to take evolution into account (Handzlik *et al.*, 2021). Analyses involving co-evolution patterns *via* comparative genomics (Zhang *et al.*, 2021a; Guo and Amir, 2021) will considerably benefit SynBio studies.

### Conclusions

The development of SynBio is at a turning point. Over the past two decades or so, a large number of proofs of concept have been established, demonstrating the feasibility of a considerable number of SynBio developments. We are now at a point where even the most far-fetched ideas can be implemented in constructions of considerable industrial interest. The integration of approaches combining experimentation with living chassis, cell-free synthesis and modelling has created a virtuous cycle of design, construction, testing and learning loops between experiments and analysis, combining physiology, genetics, biochemistry and bioinformatics in a way that is constantly progressing. Perhaps the most promising approaches in terms of industrial production are experiments using building blocks that differ from those established during the evolution of life. It also seems important to open up strategies that combine biology and chemistry, for example, in factories where a succession of reactors combining *in vivo* fermenters and



biochemical and chemical reactors creates a production line starting with low-cost inputs and leading to high-value products. At the same time, understanding the processes will greatly improve the safety of the facilities, with prevention locks in place that will foil the unwanted spread of chemical or biological contamination. In particular, it is essential to remember that the closer we get to existing living organisms, the more difficult it is to control their spread. Contrary to popular belief, artifice is much less dangerous than nature (Danchin, 2016).

### Acknowledgements

This work benefited from discussions with members of the Stanislas Noria seminar.

### Funding information

No funding information provided.

### Conflict of interest

None declared.

### References

- Abbasov, I.B. (2019) Computer simulation in aircraft. In *Computer Modeling in the Aerospace Industry*. Abbasov, I.B. (ed). Hoboken, NJ: Wiley, pp. 1–21.
- Agozzino, L., and Dill, K.A. (2018) Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc Natl Acad Sci USA* **115**: 9092–9097.
- Akay, A., and Hess, H. (2019) Deep learning: current and emerging applications in medicine and technology. *IEEE J Biomed Health Inform* **23**: 906–920.
- Aksu, D., Diallo, M.M., Şahar, U., Uyaniker, T.A.L.P.A.Y., and Ozdemir, G. (2021) High expression of ring-hydroxylating dioxygenase genes ensure efficient degradation of p-toluate, phthalate, and terephthalate by *Comamonas testosteroni* strain 3a2. *Arch Microbiol* **203**: 4101–4112.
- Aldridge, J., Carr, S., Weber, K.A., and Buan, N.R. (2021) Anaerobic production of isoprene by engineered *Methanosarcina* species Archaea. *Appl Environ Microbiol* **87**: e02417–e2420.
- Ambrus, V., Hoffka, G., and Fuxreiter, M. (2020) Asymmetric dynamic coupling promotes alternative evolutionary pathways in an enzyme dimer. *Sci Rep* **10**: 18866.
- Ang, K.S., Kyriakopoulos, S., Li, W., and Lee, D.-Y. (2016) Multi-omics data driven analysis establishes reference codon biases for synthetic gene design in microbial and mammalian cells. *Methods* **102**: 26–35.
- Antoniewicz, M.R. (2021) A guide to metabolic flux analysis in metabolic engineering: methods, tools and applications. *Metab Eng* **63**: 2–12.
- Arakawa, K., and Tomita, M. (2013) Merging multiple omics datasets *in silico*: statistical analyses and data interpretation. *Methods Mol Biol* **985**: 459–470.
- Arita, M., Karsch-Mizrachi, I., and Cochrane, G. (2021) The international nucleotide sequence database collaboration. *Nucleic Acids Res* **49**: D121–D124.
- Austin, S., and McGeoch, D. (1973) The synthesis in vitro of RNA polymerase subunits of *Escherichia coli*. *Proc Natl Acad Sci USA* **70**: 2420–2423.
- Bacher, J.M., and Ellington, A.D. (2001) Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. *J Bacteriol* **183**: 5414–5425.
- Bailey, J.E. (1991) Toward a science of metabolic engineering. *Science* **252**: 1668–1675.
- Balabanova, L., Averianova, L., Marchenok, M., Son, O., and Tekutyeva, L. (2021) Microbial and genetic resources for cobalamin (Vitamin B12) biosynthesis: from ecosystems to industrial biotechnology. *Int J Mol Sci* **22**: 4522.
- Banerjee, S., Scherer, N.F., and Dinner, A.R. (2016) Shape dynamics of growing cell walls. *Soft Matter* **12**: 3442–3450.
- Belda, E., van Heck, R.G.A., José Lopez-Sanchez, M., Cruveiller, S., Barbe, V., Fraser, C., *et al.* (2016) The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. *Environ Microbiol* **18**: 3403–3424.
- Belda, E., Sekowska, A., Le Fèvre, F., Morgat, A., Mornico, D., Ouzounis, C., *et al.* (2013) An updated metabolic view of the *Bacillus subtilis* 168 genome. *Microbiology* **159**: 757–770.
- Bell, M.J., and Lord, P. (2017) On patterns and re-use in bioinformatics databases. *Bioinformatics* **33**: 2731–2736.
- Bellio, P., Mancini, A., Di Pietro, L., Cracchiolo, S., Franceschini, N., Reale, S., *et al.* (2020) Inhibition of the transcriptional repressor LexA: Withstanding drug resistance by inhibiting the bacterial mechanisms of adaptation to antimicrobials. *Life Sci* **241**: 117116.
- Bhattacharya, A., Cho, C.J., Brea, R.J., and Devaraj, N.K. (2021) Expression of fatty acyl-CoA ligase drives one-pot *de novo* synthesis of membrane-bound vesicles in a cell-free transcription-translation system. *J Am Chem Soc* **143**: 11235–11242.
- Billerbeck, S., Härle, J., and Panke, S. (2013) The good of two worlds: increasing complexity in cell-free systems. *Curr Opin Biotechnol* **24**: 1037–1043.
- Bitzenhofer, N.L., Kruse, L., Thies, S., Wynands, B., Lechtenberg, T., Rönitz, J., *et al.* (2021) Towards robust *Pseudomonas* cell factories to harbour novel biosynthetic pathways. *Essays Biochem* **65**: 319–336.
- Blumenthal, R.M., Rice, P.J., and Roberts, R.J. (1982) Computer programs for nucleic acid sequence manipulation. *Nucleic Acids Res* **10**: 91–101.
- Boehm, C.R., Freemont, P.S., and Ces, O. (2013) Design of a prototype flow microreactor for synthetic biology in vitro. *Lab Chip* **13**: 3426–3432.
- Boel, G., Danot, O., de Lorenzo, V., and Danchin, A. (2019) Omnipresent Maxwell's demons orchestrate information management in living cells. *Microb Biotechnol* **12**: 210–242.
- Bohrer, C.H., and Xiao, J. (2020) Complex diffusion in bacteria. *Adv Exp Med Biol* **1267**: 15–43.
- Bose, S., Aggarwal, S., Singh, D.V., and Acharya, N. (2020) Extracellular vesicles: an emerging platform in gram-positive bacteria. *Microb Cell* **7**: 312–322.

- Bradley, C.C., Gordon, A.J.E., Halliday, J.A., and Herman, C. (2019) Transcription fidelity: new paradigms in epigenetic inheritance, genome instability and disease. *DNA Repair (Amst)* **81**: 102652.
- Broom, A., Rakotoharisoa, R.V., Thompson, M.C., Zarifi, N., Nguyen, E., Mukhametzhano, N., *et al.* (2020) Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution *in silico*. *Nat Commun* **11**: 4808.
- Brown, K.N., and Miguel, I. (2006) Uncertainty and change. In *Handbook of Constraint Programming*. Foundations of Artificial Intelligence. Rossi, F., van Beek, P., and Walsh, T. (eds). Amsterdam: Elsevier Science, pp. 731–760.
- Bujara, M., and Panke, S. (2012) *In silico* assessment of cell-free systems. *Biotechnol Bioeng* **109**: 2620–2629.
- Bunzel, H.A., Anderson, J.L.R., and Mulholland, A.J. (2021) Designing better enzymes: insights from directed evolution. *Curr Opin Struct Biol* **67**: 212–218.
- Burian, A.N., Zhao, W., Lo, T.-W., and Thurtle-Schmidt, D.M. (2021) Genome sequencing guide: an introductory toolbox to whole-genome analysis methods. *Biochem Mol Biol Educ* **49**: 815–825.
- Burrington, L.R., Watts, K.R., and Oza, J.P. (2021) Characterizing and improving reaction times for *E. coli*-based cell-free protein synthesis. *ACS Synth Biol* **10**: 1821–1829.
- Calero, P., and Nikel, P.I. (2019) Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb Biotechnol* **12**: 98–124.
- Caruthers, M.H. (2013) The chemical synthesis of DNA/RNA: our gift to science. *J Biol Chem* **288**: 1420–1427.
- Caschera, F., and Noireaux, V. (2014) Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. *Biochimie* **99**: 162–168.
- Chan, C.M., Danchin, A., Marliere, P., and Sekowska, A. (2014) Paralogous metabolism: S-alkyl-cysteine degradation in *Bacillus subtilis*. *Environ Microbiol* **16**: 101–117.
- Chang, F., and Huang, K.C. (2014) How and why cells grow as rods. *BMC Biol* **12**: 54.
- Chaput, J.C., and Herdewijn, P. (2019) What is XNA? *Angew Chem Int Ed Engl* **58**: 11570–11572.
- Chen, B.-S., and Li, C.-W. (2016) Constructing an integrated genetic and epigenetic cellular network for whole cellular mechanism using high-throughput next-generation sequencing data. *BMC Syst Biol* **10**: 18.
- Chen, C., Cui, Z., Zhao, J., Li, S., Ren, X., Chen, T., and Wang, Z. (2021a) Improving diacetyl production in *Corynebacterium glutamicum* via modifying respiratory chain. *J Biotechnol* **332**: 20–28.
- Chen, C., Hou, J., Shi, X., Yang, H., Birchler, J.A., and Cheng, J. (2021b) GNET2: an R package for constructing gene regulatory networks from transcriptomic data. *Bioinformatics* **37**: 2068–2069.
- Chen, R., and Dou, J. (2016) Biofuels and bio-based chemicals from lignocellulose: metabolic engineering strategies in strain development. *Biotechnol Lett* **38**: 213–221.
- Chen, X., Li, C., and Liu, H. (2021c) Enhanced recombinant protein production under special environmental stress. *Front Microbiol* **12**: 630814.
- Choi, J.-M., Holehouse, A.S., and Pappu, R.V. (2020) Physical principles underlying the complex biology of intracellular phase transitions. *Annu Rev Biophys* **49**: 107–133.
- Clark, B.F., and Marcker, K.A. (1966) The role of N-formylmethionyl-sRNA in protein biosynthesis. *J Mol Biol* **17**: 394–406.
- Clark, J.M., Chang, A.Y., Spiegelman, S., and Reichmann, M.E. (1965) The *in vitro* translation of a monocistronic message. *Proc Natl Acad Sci USA* **54**: 1193–1197.
- Clark, T., and Hicks, M.G. (2020) Models of necessity. *Beilstein J Org Chem* **16**: 1649–1661.
- de Crécy, E., Metzgar, D., Allen, C., Pénicaud, M., Lyons, B., Hansen, C.J., and de Crécy-Lagard, V. (2007) Development of a novel continuous culture device for experimental evolution of bacterial populations. *Appl Microbiol Biotechnol* **77**: 489–496.
- de Crécy-Lagard, V.A., Bellalou, J., Mutzel, R., and Marlière, P. (2001) Long term adaptation of a microbial population to a permanent metabolic constraint: overcoming thymineless death by experimental evolution of *Escherichia coli*. *BMC Biotechnol* **1**: 10.
- Cui, J., Sun, T., Chen, L., and Zhang, W. (2021) Salt-tolerant *Synechococcus elongatus* UTEX 2973 obtained via engineering of heterologous synthesis of compatible solute glucosylglycerol. *Front Microbiol* **12**: 650217.
- Czajka, J.J., Oyetunde, T., and Tang, Y.J. (2021) Integrated knowledge mining, genome-scale modeling, and machine learning for predicting *Yarrowia lipolytica* bioproduction. *Metab Eng* **67**: 227–236.
- Danchin, A. (2016) Nature or manufacture: what should we fear most? *C R Biol* **339**: 329–335.
- Danchin, A. (2020) Isobiology: a variational principle for exploring synthetic life. *ChemBioChem* **21**: 1781–1792.
- Danchin, A. (2021) Three overlooked key functional classes for building up minimal synthetic cells. *Synth Biol* **6**: ys-ab010.
- Danchin, A., Binder, P.M., and Noria, S. (2011) Antifragility and tinkering in biology (and in business) flexibility provides an efficient epigenetic way to manage risk. *Genes (Basel)* **2**: 998–1016.
- Danchin, A., Ouzounis, C., Tokuyasu, T., and Zucker, J.-D. (2018) No wisdom in the crowd: genome annotation in the era of big data - current status and future prospects. *Microb Biotechnol* **11**: 588–605.
- Danchin, A., and Sekowska, A. (2015) The logic of metabolism. *Perspect Sci* **6**: 15–26.
- Daniels, R., Reynaert, S., Hoekstra, H., Verreth, C., Janssens, J., Braeken, K., *et al.* (2006) Quorum signal molecules as biosurfactants affecting swarming in *Rhizobium etli*. *Proc Natl Acad Sci USA* **103**: 14965–14970.
- Davies, P.S.W. (2020) Stable isotopes: their use and safety in human nutrition studies. *Eur J Clin Nutr* **74**: 362–365.
- Decoene, T., De Paepe, B., Maertens, J., Coussemont, P., Peters, G., De Maeseneire, S.L., and De Mey, M. (2018) Standardization in synthetic biology: an engineering discipline coming of age. *Crit Rev Biotechnol* **38**: 647–656.
- Deng, P., de Vargas Roditi, L., van Ditmarsch, D., and Xavier, J.B. (2014) The ecological basis of morphogenesis: branching patterns in swarming colonies of bacteria. *New J Phys* **16**: 015006.

- Dewachter, L., Bollen, C., Wilmaerts, D., Louwagie, E., Herpels, P., Matthay, P., et al. (2021) The dynamic transition of persistence toward the viable but nonculturable state during stationary phase is driven by protein aggregation. *mBio* **12**: e0070321.
- Döring, V., and Marlière, P. (1998) Reassigning cysteine in the genetic code of *Escherichia coli*. *Genetics* **150**: 543–551.
- Dunkelmann, D.L., Oehm, S.B., Beattie, A.T., and Chin, J.W. (2021) A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nat Chem*. in press.
- Durham, B.P. (2021) Deciphering metabolic currencies that support marine microbial networks. *mSystems* **6**: e0076321.
- Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013) COBRApy: COstraints-based reconstruction and analysis for python. *BMC Syst Biol* **7**: 74.
- Edwards, J.S., and Palsson, B.O. (1998) How will bioinformatics influence metabolic engineering? *Biotechnol Bioeng* **58**: 162–169.
- Edwards, J.S., and Palsson, B.O. (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem* **274**: 17410–17416.
- van Eldijk, M.B., and van Hest, J.C.M. (2018) Residue-specific incorporation of noncanonical amino acids for protein engineering. *Methods Mol Biol* **1728**: 137–145.
- Elowitz, M.B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335–338.
- Fall, R., Kinsinger, R.F., and Wheeler, K.A. (2004) A simple method to isolate biofilm-forming *Bacillus subtilis* and related species from plant roots. *Syst Appl Microbiol* **27**: 372–379.
- Femmer, C., Bechtold, M., Held, M., and Panke, S. (2020) In vivo directed enzyme evolution in nanoliter reactors with antimetabolite selection. *Metab Eng* **59**: 15–23.
- Feuermann, M., Boutet, E., Morgat, A., Axelsen, K.B., Bansal, P., Bolleman, J., et al. (2021) Diverse taxonomies for diverse chemistries: enhanced representation of natural product metabolism in UniProtKB. *Metabolites* **11**: 48.
- Fiers, G., Choukhi, D., Oswald, L., Al Ouahabi, A., Chan-Seng, D., Charles, L., and Lutz, J.-F. (2016) Orthogonal synthesis of xeno nucleic acids. *Chemistry* **22**: 17945–17948.
- Flechslers, J., Heimerl, T., Huber, H., Rachel, R., and Berg, I.A. (2021) Functional compartmentalization and metabolic separation in a prokaryotic cell. *Proc Natl Acad Sci USA* **118**: e2022114118.
- Flood, D.T., Hintzen, J.C.J., Knouse, K.W., Hill, D.E., Lu, C., Cistrone, P.A., et al. (2021) Selenomethionine as an expressible handle for bioconjugations. *Proc Natl Acad Sci USA* **118**: e2005164118.
- Fontana, J., Sparkman-Yager, D., Zalatan, J.G., and Carothers, J.M. (2020) Challenges and opportunities with CRISPR activation in bacteria for data-driven metabolic engineering. *Curr Opin Biotechnol* **64**: 190–198.
- Forster, A.C., and Church, G.M. (2007) Synthetic biology projects in vitro. *Genome Res* **17**: 1–6.
- Fritz, S.E., Haque, N., and Hogg, J.R. (2018) Highly efficient in vitro translation of authentic affinity-purified messenger ribonucleoprotein complexes. *RNA* **24**: 982–989.
- Fukuma, I., and Cohen, S.S. (1975) Polyamines in bacteriophage R17 and its RNA. *J Virol* **16**: 222–227.
- García-Morales, L., Ruiz, E., Gourgues, G., Rideau, F., Piñero-Lambea, C., Lluch-Senar, M., et al. (2020) A RAGE based strategy for the genome engineering of the human respiratory pathogen *Mycoplasma pneumoniae*. *ACS Synth Biol* **9**: 2737–2748.
- Garenne, D., Libchaber, A., and Noireaux, V. (2020) Membrane molecular crowding enhances MreB polymerization to shape synthetic cells from spheres to rods. *Proc Natl Acad Sci USA* **117**: 1902–1909.
- Gautam, P., and Kumar Sinha, S. (2021) Anticipating response function in gene regulatory networks. *J R Soc Interface* **18**: 20210206.
- Ge, H., and Marchisio, M.A. (2021) Aptamers, riboswitches, and ribozymes in *S. cerevisiae* synthetic biology. *Life (Basel)* **11**: 248.
- Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res* **49**: D325–D334.
- Gomes, E., and Shorter, J. (2019) The molecular language of membraneless organelles. *J Biol Chem* **294**: 7115–7127.
- González, L.M., Mukhitov, N., and Voigt, C.A. (2020) Resilient living materials built by printing bacterial spores. *Nat Chem Biol* **16**: 126–133.
- Grant, N.A., Abdel Magid, A., Franklin, J., Dufour, Y., and Lenski, R.E. (2021) Generations of experimental evolution with *Escherichia coli*. *J Bacteriol* **203**: e00469-20.
- Grasemann, L., Lavickova, B., Elizondo-Cantú, M.C., and Maerkl, S.J. (2021) OnePot PURE cell-free system. *J vis Exp*. in press.
- Gribble, J., Stevens, L.J., Agostini, M.L., Anderson-Daniels, J., Chappell, J.D., Lu, X., et al. (2021) The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog* **17**: e1009226.
- Grindley, J.N., Blumberg, R., and Nakada, D. (1977) Synthesis of R-plasmid-coded beta-lactamase in minicells and in an in vitro system. *J Bacteriol* **130**: 852–859.
- Grunberg-Manago, M. (1963) Enzymatic synthesis of nucleic acids. *Prog Biophys Mol Biol* **13**: 175–239.
- Guo, Y., and Amir, A. (2021) Exploring the effect of network topology, mRNA and protein dynamics on gene regulatory network stability. *Nat Commun* **12**: 130.
- Guzmán, G.I., Sandberg, T.E., LaCroix, R.A., Nyerges, Á., Papp, H., de Raad, M., et al. (2019) Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol Syst Biol* **15**: e8462.
- Hagemann, M., and Hess, W.R. (2018) Systems and synthetic biology for the biotechnological application of cyanobacteria. *Curr Opin Biotechnol* **49**: 94–99.
- Hammerling, M.J., Krüger, A., and Jewett, M.C. (2020) Strategies for in vitro engineering of the translation machinery. *Nucleic Acids Res* **48**: 1068–1083.
- Handzlik, J.E., Loh, Y.L., and Manu, N. (2021) Dynamic modeling of transcriptional gene regulatory networks. *Methods Mol Biol* **2328**: 67–97.

- Haruna, I., and Spiegelman, S. (1965) Autocatalytic synthesis of a viral RNA *in vitro*. *Science* **150**: 884–886.
- Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., *et al.* (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res* **44**: D1214–1219.
- He, H., Höper, R., Dodenhöft, M., Marlière, P., and Bar-Even, A. (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*. *Metab Eng* **60**: 1–13.
- Hecht, S.M. (2021) Expansion of the genetic code through the use of modified bacterial ribosomes. *J Mol Biol* **167211**.
- Heinken, A., Basile, A., Hertel, J., Thinnies, C., and Thiele, I. (2021) Genome-scale metabolic modeling of the human microbiome in the era of personalized medicine. *Annu Rev Microbiol.* in press.
- Hershewe, J.M., Warfel, K.F., Iyer, S.M., Peruzzi, J.A., Sullivan, C.J., Roth, E.W., *et al.* (2021) Improving cell-free glycoprotein synthesis by characterizing and enriching native membrane vesicles. *Nat Commun* **12**: 2363.
- Hogeweg, P. (2011) The roots of bioinformatics in theoretical biology. *PLoS Comput Biol* **7**: e1002021.
- Hui, J.O., Flick, T., Loo, J.A., and Campuzano, I.D.G. (2021) Unequivocal identification of aspartic acid and isoaspartic acid by MALDI-TOF/TOF: from peptide standards to a therapeutic antibody. *J Am Soc Mass Spectrom* **32**: 1901–1909.
- Hummels, K.R., and Kearns, D.B. (2020) Translation elongation factor P (EF-P). *FEMS Microbiol Rev* **44**: 208–218.
- Illangasekare, M., Sanchez, G., Nickles, T., and Yarus, M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**: 643–647.
- Ishii, N., Robert, M., Nakayama, Y., Kanai, A., and Tomita, M. (2004) Toward large-scale modeling of the microbial cell for computer simulation. *J Biotechnol* **113**: 281–294.
- Isturiz, T., and Wolf, R.E. (1975) *In vitro* synthesis of a constitutive enzyme of *Escherichia coli*, 6-phosphogluconate dehydrogenase. *Proc Natl Acad Sci USA* **72**: 4381–4384.
- Ito, M., Sugiura, H., Ayukawa, S., Kiga, D., and Takinoue, M. (2016) A bacterial continuous culture system based on a microfluidic droplet open reactor. *Anal Sci* **32**: 61–66.
- Jacobson, T.B., Callaghan, M.M., and Amador-Noguez, D. (2021) Hostile takeover: how viruses reprogram prokaryotic metabolism. *Annu Rev Microbiol* **75**. in press.
- Jakubowski, H. (1997) Aminoacyl thioester chemistry of class II aminoacyl-tRNA synthetases. *Biochemistry* **36**: 11077–11085.
- Jang, S., Kim, M., Hwang, J., and Jung, G.Y. (2019) Tools and systems for evolutionary engineering of biomolecules and microorganisms. *J Ind Microbiol Biotechnol* **46**: 1313–1326.
- Jiang, N., Ding, X., and Lu, Y. (2021) Development of a robust *Escherichia coli*-based cell-free protein synthesis application platform. *Biochem Eng J* **165**: 107830.
- Jones, E.C., and Uphoff, S. (2021) Single-molecule imaging of LexA degradation in *Escherichia coli* elucidates regulatory mechanisms and heterogeneity of the SOS response. *Nat Microbiol* **6**: 981–990.
- Kampers, L.F.C., Koehorst, J.J., van Heck, R.J.A., Suarez-Diez, M., Stams, A.J.M., and Schaap, P.J. (2021) A metabolic and physiological design study of *Pseudomonas putida* KT2440 capable of anaerobic respiration. *BMC Microbiol* **21**: 9.
- Karas, J.A., Wade, J.D., and Hossain, M.A. (2021) The chemical synthesis of insulin: an enduring challenge. *Chem Rev* **121**: 4531–4560.
- Katoh, T., Wohlgemuth, I., Nagano, M., Rodnina, M.V., and Suga, H. (2016) Essential structural elements in tRNA (Pro) for EF-P-mediated alleviation of translation stalling. *Nat Commun* **7**: 11657.
- Knüpfper, L., Fehrenbach, C., Denks, K., Erichsen, V., Petri-man, N.-A., and Koch, H.-G. (2019) Molecular mimicry of SecA and signal recognition particle binding to the bacterial ribosome. *mBio* **10**: e01317–e1319.
- Ko, Y.-S., Kim, J.W., Lee, J.A., Han, T., Kim, G.B., Park, J.E., and Lee, S.Y. (2020) Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem Soc Rev* **49**: 4615–4636.
- Kofman, C., Lee, J., and Jewett, M.C. (2021) Engineering molecular translation systems. *Cell Syst* **12**: 593–607.
- Konstantinidis, D., Pereira, F., Geissen, E.-M., Grkovska, K., Kafkia, E., Jouhten, P., *et al.* (2021) Adaptive laboratory evolution of microbial co-cultures for improved metabolite secretion. *Mol Syst Biol* **17**: e10189.
- Koppad, S., B, A., Gkoutos, G.V., and Acharjee, A. (2021) Cloud computing enabled big multi-omics data analytics. *Bioinform Biol Insights* **15**: 11779322211035920.
- Koubek, J., Schmitt, J., Galmozzi, C.V., and Kramer, G. (2021) Mechanisms of cotranslational protein maturation in bacteria. *Front Mol Biosci* **8**: 689755.
- Kratz, A., Tomita, M., and Krishnan, A. (2008) GeNESiS: gene network evolution simulation software. *BMC Bioinformatics* **9**: 541.
- Kruyer, N.S., Sugianto, W., Tickman, B.I., Alba Burbano, D., Noireaux, V., Carothers, J.M., and Peralta-Yahya, P. (2021) Membrane augmented cell-free systems: a new frontier in biotechnology. *ACS Synth Biol* **10**: 670–681.
- Kung, H.F., Treadwell, B.V., Spears, C., Tai, P.C., and Weissbach, H. (1977) DNA-directed synthesis *in vitro* of beta-galactosidase: requirement for a ribosome release factor. *Proc Natl Acad Sci USA* **74**: 3217–3221.
- Kunjapur, A.M., Napolitano, M.G., Hysolli, E., Noguera, K., Appleton, E.M., Schubert, M.G., *et al.* (2021) Synthetic auxotrophy remains stable after continuous evolution and in coculture with mammalian cells. *Sci Adv* **7**: eabf5851.
- Lachance, J.-C., Matteau, D., Brodeur, J., Lloyd, C.J., Mih, N., King, Z.A., *et al.* (2021) Genome-scale metabolic modeling reveals key features of a minimal gene set. *Mol Syst Biol* **17**: e10099.
- Lee, S., and Kim, P. (2020) Current status and applications of adaptive laboratory evolution in industrial microorganisms. *J Microbiol Biotechnol* **30**: 793–803.
- Levisohn, R., and Spiegelman, S. (1969) Further extracellular Darwinian experiments with replicating RNA molecules: diverse variants isolated under different selective conditions. *Proc Natl Acad Sci USA* **63**: 805–811.
- Li, J., Gu, L., Aach, J., and Church, G.M. (2014) Improved cell-free RNA and protein synthesis system. *PLoS One* **9**: e106232.

- Li, J., Zhang, C., Huang, P., Kuru, E., Forster-Benson, E.T.C., Li, T., and Church, G.M. (2017) Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *Translation (Austin)* **5**: e1327006.
- Li, Y., Cheng, Z., Zhao, C., Gao, C., Song, W., Liu, L., and Chen, X. (2021) Reprogramming *Escherichia coli* metabolism for bioplastics synthesis from waste cooking oil. *ACS Synth Biol* **8**: 1966–1979.
- Lieder, S., Nickel, P.I., de Lorenzo, V., and Takors, R. (2015) Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microb Cell Fact* **14**: 23.
- Liu, H., and Deutschbauer, A.M. (2018) Rapidly moving new bacteria to model-organism status. *Curr Opin Biotechnol* **51**: 116–122.
- Liu, Y., Su, A., Tian, R., Li, J., Liu, L., and Du, G. (2020) Developing rapid growing *Bacillus subtilis* for improved biochemical and recombinant protein production. *Metab Eng Commun* **11**: e00141.
- Lombardot, T., Morgat, A., Axelsen, K.B., Aimo, L., Hyka-Nouspikel, N., Niknejad, A., et al. (2019) Updates in Rhea: SPARQLing biochemical reaction data. *Nucleic Acids Res* **47**: D596–D600.
- Long, L., Pei, R., Liu, Y., Rao, X., Wang, Y., Zhou, S.-F., and Zhan, G. (2021) 3D printing of recombinant *Escherichia coli*/Au nanocomposites as agitating paddles towards robust catalytic reduction of 4-nitrophenol. *J Hazard Mater* **423**: 126983.
- de Lorenzo, V., Sekowska, A., and Danchin, A. (2015) Chemical reactivity drives spatiotemporal organisation of bacterial metabolism. *FEMS Microbiol Rev* **39**: 96–119.
- Lu, H., Zhang, G., He, S., Zhao, R., and Zhu, D. (2021) Purple non-sulfur bacteria technology: a promising and potential approach for wastewater treatment and biore-sources recovery. *World J Microbiol Biotechnol* **37**: 161.
- Mabwi, H.A., Kim, E., Song, D.-G., Yoon, H.S., Pan, C.-H., Komba, E.V.G., et al. (2021) Synthetic gut microbiome: advances and challenges. *Comput Struct Biotechnol J* **19**: 363–371.
- Mallory, J.D., Kolomeisky, A.B., and Igoshin, O.A. (2019) Trade-offs between error, speed, noise, and energy dissipation in biological processes with proofreading. *J Phys Chem B* **123**: 4718–4725.
- Marchant, A., Cisneros, A.F., Dubé, A.K., Gagnon-Arsenault, I., Ascencio, D., Jain, H., et al. (2019) The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs. *Elife* **8**: e46754.
- Marlière, P., Patrouix, J., Döring, V., Herdewijn, P., Tricot, S., Cruveiller, S., et al. (2011) Chemical evolution of a bacterium's genome. *Angew Chem Int Ed Engl* **50**: 7109–7114.
- Martínez-García, E., Goñi-Moreno, A., Bartley, B., McLaughlin, J., Sánchez-Sampedro, L., del Pozo, H.P., et al. (2020) SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic Acids Res* **48**: 3395.
- Martin-Pascual, M., Batiánis, C., Bruinsma, L., Asin-García, E., García-Morales, L., Weusthuis, R.A., et al. (2021) A navigation guide of synthetic biology tools for *Pseudomonas putida*. *Biotechnol Adv* **49**: 107732.
- McCallum, G., and Potvin-Trottier, L. (2021) Using models to (re-)design synthetic circuits. *Methods Mol Biol* **2229**: 91–118.
- Medema, M.H., de Rond, T., and Moore, B.S. (2021) Mining genomes to illuminate the specialized chemistry of life. *Nat Rev Genet* **22**: 553–571.
- Médigue, C., Rouxel, T., Vigier, P., Hénaut, A., and Danchin, A. (1991) Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol* **222**: 851–856.
- Meyer, B.J., and Southern, P.J. (1997) A novel type of defective viral genome suggests a unique strategy to establish and maintain persistent lymphocytic choriomeningitis virus infections. *J Virol* **71**: 6757–6764.
- Meyer, V., Cairns, T., Barthel, L., King, R., Kunz, P., Schmideder, S., et al. (2021) Understanding and controlling filamentous growth of fungal cell factories: novel tools and opportunities for targeted morphology engineering. *Fungal Biol Biotechnol* **8**: 8.
- Miguez, A.M., Zhang, Y., Piorino, F., and Styczynski, M.P. (2021) Metabolic dynamics in *Escherichia coli*-based cell-free systems. *ACS Synth Biol* **10**: 2252–2265.
- Miller, D.L., Andreone, T.L., and Donelson, J.E. (1977) Translation in *Escherichia coli* mini-cells containing hamster mitochondrial DNA-Co1E1 - Amp<sup>r</sup> recombinant plasmids. *Biochim Biophys Acta* **477**: 323–333.
- Mordret, E., Dahan, O., Asraf, O., Rak, R., Yehonadav, A., Barnabas, G.D., et al. (2019) Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. *Mol Cell* **75**: 427–441.e5.
- Morgan, A.A., and Rubenstein, E. (2013) Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PLoS One* **8**: e53785.
- Moroder, L., and Budisa, N. (2010) Synthetic biology of protein folding. *ChemPhysChem* **11**: 1181–1187.
- Mortimer, M., Wang, Y., and Holden, P.A. (2021) Molecular mechanisms of nanomaterial-bacterial interactions revealed by omics-The role of nanomaterial effect level. *Front Bioeng Biotechnol* **9**: 683520.
- Moyer, D., Pacheco, A.R., Bernstein, D.B., and Segrè, D. (2021) Stoichiometric modeling of artificial string chemistries reveals constraints on metabolic network structure. *J Mol Evol* **89**: 472–483.
- Mrázek, J., and Karls, A.C. (2019) In silico simulations of occurrence of transcription factor binding sites in bacterial genomes. *BMC Evol Biol* **19**: 67.
- Müller, J., Siemann-Herzberg, M., and Takors, R. (2020) Modeling cell-free protein synthesis systems-Approaches and applications. *Front Bioeng Biotechnol* **8**: 584178.
- Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., et al. (2009) Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol Syst Biol* **5**: 306.
- Navelkar, R., Owen, G., Muthukrishnan, V., Thiessen, P., Cheng, T., Bolerlton, E., et al. (2021) Enhancing the interoperability of glycan data flow between ChEBI, PubChem, and GlyGen. *Glycobiology* cwab078.
- Ng, I.-S., Keskin, B.B., and Tan, S.-I. (2020) A critical review of genome editing and synthetic biology

- applications in metabolic engineering of microalgae and cyanobacteria. *Biotechnol J* **15**: e1900228.
- Nicolás Carcelén, J., Marchante-Gayón, J.M., González, P.R., Valledor, L., Cañal, M.J., and Alonso, J.I.G. (2017) A cost-effective approach to produce 15N-labelled amino acids employing *Chlamydomonas reinhardtii* CC503. *Microb Cell Fact* **16**: 146.
- Nielsen, S.D., Koren, K., Löbmann, K., Hinge, M., Scoma, A., Kjeldsen, K.U., and Røy, H. (2020) Constraints on CaCO<sub>3</sub> precipitation in superabsorbent polymer by aerobic bacteria. *Appl Microbiol Biotechnol* **104**: 365–375.
- Nishizawa, M., Sakai, S., Konno, U., Nakahara, N., Takaki, Y., Saito, Y., *et al.* (2016) Nitrogen and oxygen isotope effects of ammonia oxidation by thermophilic Thaumarchaeota from a geothermal water stream. *Appl Environ Microbiol* **82**: 4492–4504.
- Noireaux, V., and Liu, A.P. (2020) The new age of cell-free biology. *Annu Rev Biomed Eng* **22**: 51–77.
- Norsigian, C.J., Fang, X., Palsson, B.O., and Monk, J.M. (2020) Pangenome flux balance analysis toward panphenomes. In *The Pangenome: Diversity, Dynamics and Evolution of Genomes*. Tettelin, H., and Medini, D. (eds). Cham: Springer.
- Oesterle, S., Roberts, T.M., Widmer, L.A., Mustafa, H., Panke, S., and Billerbeck, S. (2017) Sequence-based prediction of permissive stretches for internal protein tagging and knockdown. *BMC Biol* **15**: 100.
- Ohuchi, M., Murakami, H., and Suga, H. (2007) The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr Opin Chem Biol* **11**: 537–542.
- Ou, Z., Ouzounis, C., Wang, D., Sun, W., Li, J., Chen, W., *et al.* (2020) A path towards SARS-CoV-2 attenuation: metabolic pressure on CTP synthesis rules the virus evolution. *Genome Biol Evol* **12**: 2467–2485.
- Palazzotto, E., Tong, Y., Lee, S.Y., and Weber, T. (2019) Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. *Biotechnol Adv* **37**: 107366.
- Paley, S., Keseler, I.M., Krummenacker, M., and Karp, P.D. (2021) Leveraging curation among *Escherichia coli* pathway/genome databases using ortholog-based annotation propagation. *Front Microbiol* **12**: 614355.
- Panke, S. (2020) Taming the beast of biology: synthetic biology and biological systems engineering. *Chimia (Aarau)* **74**: 402–406.
- Park, Y.-K., and Nicaud, J.-M. (2020) Metabolic engineering for unusual lipid production in *Yarrowia lipolytica*. *Microorganisms* **8**: E1937.
- Passioura, T., and Suga, H. (2014) Flexizymes, their evolutionary history and diverse utilities. *Top Curr Chem* **344**: 331–345.
- Pedrioli, G., Piovesana, E., Vacchi, E., and Balbi, C. (2021) Extracellular vesicles as promising carriers in drug delivery: considerations from a cell biologist's perspective. *Biology (Basel)* **10**: 376.
- Pereira, F., Lopes, H., Maia, P., Meyer, B., Nocon, J., Jouhten, P., *et al.* (2021) Model-guided development of an evolutionarily stable yeast chassis. *Mol Syst Biol* **17**: e10253.
- Piette, F., D'Amico, S., Mazzucchelli, G., Danchin, A., Leprince, P., and Feller, G. (2011) Life in the cold: a proteomic study of cold-repressed proteins in the antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Appl Environ Microbiol* **77**: 3881–3883.
- Piñero-Lambea, C., Garcia-Ramallo, E., Martinez, S., Delgado, J., Serrano, L., and Lluch-Senar, M. (2020) *Mycoplasma pneumoniae* genome editing based on oligo recombineering and Cas9-mediated counterselection. *ACS Synth Biol* **9**: 1693–1704.
- Piñeros, W.D., and Tlustý, T. (2020) Kinetic proofreading and the limits of thermodynamic uncertainty. *Phys Rev E* **101**: 022415.
- Prifti, E., Chevaleyre, Y., Hanczar, B., Belda, E., Danchin, A., Clément, K., and Zucker, J.-D. (2020) Interpretable and accurate prediction models for metagenomics data. *Gigascience* **9**: g1aa010.
- Pühler, A. (2012) Genome-based microbiology—from -omics research to systems and synthetic biology. *J Biotechnol* **160**: 1–2.
- Raghavan, A.R., Salim, K., and Yadav, V.G. (2020) Optogenetic control of heterologous metabolism in *E. coli*. *ACS Synth Biol* **9**: 2291–2300.
- Ramaswamy, K., Saito, H., Murakami, H., Shiba, K., and Suga, H. (2004) Designer ribozymes: programming the tRNA specificity into flexizyme. *J Am Chem Soc* **126**: 11454–11455.
- Rasor, B.J., Vögeli, B., Landwehr, G.M., Bogart, J.W., Karim, A.S., and Jewett, M.C. (2021) Toward sustainable, cell-free biomanufacturing. *Curr Opin Biotechnol* **69**: 136–144.
- Rees-Garbutt, J., Chalkley, O., Grierson, C., and Marucci, L. (2021) Minimal genome design algorithms using whole-cell models. *Methods Mol Biol* **2189**: 183–198.
- Ren, X., Linehan, M.M., Iwasaki, A., and Pyle, A.M. (2019) RIG-I selectively discriminates against 5'-monophosphate RNA. *Cell Rep* **26**: 2019–2027.e4.
- Robertson, W.E., Funke, L.F.H., de la Torre, D., Fredens, J., Elliott, T.S., Spinck, M., *et al.* (2021) Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**: 1057–1062.
- Rustad, M., Eastlund, A., Marshall, R., Jardine, P., and Noireaux, V. (2017) Synthesis of infectious bacteriophages in an *E. coli*-based cell-free expression system. *J Vis Exp* **126**: 56144.
- Ryngajłto, M., Jędrzejczak-Krzepkowska, M., Kubiak, K., Ludwicka, K., and Bielecki, S. (2020) Towards control of cellulose biosynthesis by *Komagataeibacter* using systems-level and strain engineering strategies: current progress and perspectives. *Appl Microbiol Biotechnol* **104**: 6565–6585.
- Santibáñez, R., Garrido, D., and Martin, A.J.M. (2020) Atlas: automatic modeling of regulation of bacterial gene expression and metabolism using rule-based languages. *Bioinformatics* **36**: 5473–5480.
- Satanowski, A., Dronsella, B., Noor, E., Vögeli, B., He, H., Wichmann, P., *et al.* (2020) Awakening a latent carbon fixation cycle in *Escherichia coli*. *Nat Commun* **11**: 5812.
- Schindler, D. (2020) Genetic engineering and synthetic genomics in yeast to understand life and boost biotechnology. *Bioengineering (Basel)* **7**: E137.

- Schmied, W.H., Tnimov, Z., Uttamapinant, C., Rae, C.D., Fried, S.D., and Chin, J.W. (2018) Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**: 444–448.
- Sekowska, A., Masson, J.-B., Celani, A., Danchin, A., and Vergassola, M. (2009) Repulsion and metabolic switches in the collective behavior of bacterial colonies. *Biophys J* **97**: 688–698.
- Sekowska, A., Wendel, S., Fischer, E.C., Nørholm, M.H.H., and Danchin, A. (2016) Generation of mutation hotspots in ageing bacterial colonies. *Sci Rep* **6**: 2.
- Srividhya, K.V., and Krishnaswamy, S. (2004) A simulation model of *Escherichia coli* osmoregulatory switch using E-CELL system. *BMC Microbiol* **4**: 44.
- St John, P.C., and Bomble, Y.J. (2019) Approaches to computational strain design in the multiomics era. *Front Microbiol* **10**: 597.
- Strauß, J.F., Crain, P., Schulenburg, H., and Telschow, A. (2016) Experimental evolution in silico: a custom-designed mathematical model for virulence evolution of *Bacillus thuringiensis*. *Zoology (Jena)* **119**: 359–365.
- Sun, G., Ahn-Horst, T.A., and Covert, M.W. (2021) The *E. coli* whole-cell modeling project. *EcoSal plus* eESP00012020.
- Takinoue, M., and Kawano, R. (2020) Editorial on the special issue on recent advances of molecular machines and molecular robots. *Micromachines (Basel)* **11**: E1031.
- Tharp, J.M., Walker, J.A., Söll, D., and Schepartz, A. (2021) Initiating protein synthesis with noncanonical monomers in vitro and in vivo. In *Synthetic and Enzymatic Modifications of the Peptide Backbone*. Methods in Enzymology. Vol **656**. Amsterdam: Elsevier Science, pp. 495–519.
- To, P., Whitehead, B., Tarbox, H.E., and Fried, S.D. (2021) Nonrefoldability is pervasive across the *E. coli* proteome. *J Am Chem Soc* **143**: 11435–11448.
- Tomita, M. (2001) Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol* **19**: 205–210.
- Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., et al. (1997) E-CELL: software environment for whole cell simulation. *Genome Inform Ser Workshop Genome Inform* **8**: 147–155.
- Tuerk, C., MacDougall, S., and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci USA* **89**: 6988–6992.
- Tyo, K.E.J., Kocharin, K., and Nielsen, J. (2010) Toward design-based engineering of industrial microbes. *Curr Opin Microbiol* **13**: 255–262.
- Tzeng, S.-R., Tseng, Y.-C., Lin, C.-C., Hsu, C.-Y., Huang, S.-J., Kuo, Y.-T., and Chang, C.-I. (2021) Molecular insights into substrate recognition and discrimination by the N-terminal domain of Lon AAA+ protease. *Elife* **10**: e64056.
- Vaish, A., Guo, S., Murray, R.M., Grandsard, P.J., and Chen, Q. (2018) On-chip membrane protein cell-free expression enables development of a direct binding assay: a curious case of potassium channel KcsA-Kv1.3. *Anal Biochem* **556**: 70–77.
- Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., et al. (2020) MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res* **48**: D579–D589.
- Vatsa, D., and Agarwal, S. (2021) PEPN-GRN: a Petri net-based approach for the inference of gene regulatory networks from noisy gene expression data. *PLoS One* **16**: e0251666.
- Vázquez-Castellanos, J.F., Biclôt, A., Vrancken, G., Huys, G.R., and Raes, J. (2019) Design of synthetic microbial consortia for gut microbiota modulation. *Curr Opin Pharmacol* **49**: 52–59.
- Voigt, C.A. (2020) Synthetic biology 2020–2030: six commercially-available products that are changing our world. *Nat Commun* **11**: 6379.
- Walther, T., Calvayrac, F., Malbert, Y., Alkim, C., Dressaire, C., Cordier, H., and François, J.M. (2018) Construction of a synthetic metabolic pathway for the production of 2,4-dihydroxybutyric acid from homoserine. *Metab Eng* **45**: 237–245.
- Wang, F., Gao, Y., and Yang, G. (2020) Recent advances in synthetic biology of cyanobacteria for improved chemicals production. *Bioengineered* **11**: 1208–1220.
- Wang, Q., Zhang, J., Al Makishah, N.H., Sun, X., Wen, Z., Jiang, Y., and Yang, S. (2021) Advances and perspectives for genome editing tools of *Corynebacterium glutamicum*. *Front Microbiol* **12**: 654058.
- Wassenaar, T.M., and Zimmermann, K. (2020) How industrial bacterial cultures can be kept stable over time. *Letts Appl Microbiol* **71**: 220–228.
- Weinberg, R., and Berkus, M. (1971) Computer simulation of a living cell: Part I. *Int J Biomed Comput* **2**: 95–120.
- Wolf, S., Becker, J., Tsuge, Y., Kawaguchi, H., Kondo, A., Marienhagen, J., et al. (2021) Advances in metabolic engineering of *Corynebacterium glutamicum* to produce high-value active ingredients for food, feed, human health, and well-being. *Essays Biochem* **65**: 197–212.
- Xiang, M., Kang, Q., and Zhang, D. (2020) Advances on systems metabolic engineering of *Bacillus subtilis* as a chassis cell. *Synth Syst Biotechnol* **5**: 245–251.
- Xie, S., Syrenne, R., Sun, S., and Yuan, J.S. (2014) Exploration of Natural Biomass Utilization Systems (NBUS) for advanced biofuel—from systems biology to synthetic design. *Curr Opin Biotechnol* **27**: 195–203.
- Xu, C., Liu, L., Zhang, Z., Jin, D., Qiu, J., and Chen, M. (2013) Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. *Appl Microbiol Biotechnol* **97**: 519–539.
- Xu, Z., Wang, S., Zhao, C., Li, S., Liu, X., Wang, L., et al. (2020) Photosynthetic hydrogen production by droplet-based microbial micro-reactors under aerobic conditions. *Nat Commun* **11**: 5985.
- Yachie-Kinoshita, A., Nishino, T., Shimo, H., Suematsu, M., and Tomita, M. (2010) A metabolic model of human erythrocytes: practical application of the E-Cell simulation environment. *J Biomed Biotechnol* **2010**: 642420.
- Yadav, S.P. (2007) The wholeness in suffix -omics, -omes, and the word om. *J Biomol Tech* **18**: 277.
- Yamamoto, H., Fukasawa, Y., Shoji, Y., Hisamoto, S., Kikuchi, T., Takamatsu, A., and Iwasaki, H. (2021) Scattered migrating colony formation in the filamentous cyanobacterium, *Pseudanabaena* sp. NIES-4403. *BMC Microbiol* **21**: 227.

- Yang, H.-J., Lee, K.-H., Lim, H.J., and Kim, D.-M. (2019) Tandem cell-free protein synthesis as a tool for rapid screening of optimal molecular chaperones. *Biotechnol J* **14**: e1800523.
- Yang, J., Zheng, P., Li, Y., Wu, J., Tan, X., Zhou, J., et al. (2020) Landscapes of bacterial and metabolic signatures and their interaction in major depressive disorders. *Sci Adv* **6**: eaba8555.
- Yu, A.-S., Yim, A.-Y., Mat, W.-K., Tong, A.-Y., Lok, S.I., Xue, H., et al. (2014) Mutations enabling displacement of tryptophan by 4-fluorotryptophan as a canonical amino acid of the genetic code. *Genome Biol Evol* **6**: 629–641.
- Zamdborg, L., Holloway, D.M., Merelo, J.J., Levchenko, V.F., and Spirov, A.V. (2015) Forced evolution in silico by artificial transposons and their genetic operators: the ant navigation problem. *Inf Sci (N Y)* **306**: 88–110.
- Zhang, B., Jiao, J., Zhang, P., Cui, W.-J., Zhang, Z., and Tian, C.-F. (2021) Comparative analysis of core and accessory genes in coexpression network. *Methods Mol Biol* **2242**: 45–58.
- Zhang, L., Lin, X., Wang, T., Guo, W., and Lu, Y. (2021) Development and comparison of cell-free protein synthesis systems derived from typical bacterial chassis. *Bioprocess Bioprocess* **8**: 58.
- Zhang, T., Hansen, K., Politis, A., and Müller, M.M. (2020) An unusually rapid protein backbone modification stabilizes the essential bacterial enzyme MurA. *Biochemistry* **59**: 3683–3695.