Chapter 3

Research infrastructure challenges for synthetic biology

While many of the fundamental laboratory techniques of biology and biotechnology are also applicable to synthetic biology, the major departure from the biological sciences tradition is in the development of technologies for the synthesis of large DNA sequences (of the gene and operon scale and above). Currently the cost of DNA synthesis lags a considerable way behind the spectacular advances in lowering the cost of DNA sequencing, although progress is being steadily made. In line with the aspirations to bring engineering standardisation to synthetic biology, there is a pressing need for new software developments, especially in design and manufacture. Chassis organisms, usually microorganisms engineered to be "minimal" life forms, are being developed as hosts for synthetic biology applications to reduce the noise and interference that is typical in biology. The bottleneck in synthetic biology is now shifting from DNA synthesis to dealing with the massive amounts of genetic and digital data being produced. If there is any role for co-ordinated international research infrastructure, it is to deal with this issue.

Introduction

The synthetic biology laboratory contains many of the same materials and equipment as a general molecular biology laboratory. However, technical barriers currently inhibit the widespread implementation of synthetic biology. To understand these barriers it is useful to recall the three core technology areas of synthetic biology: DNA synthesis and assembly, sequencing, and modelling.

A major goal in synthetic biology is to design and construct new metabolic pathways within a producer cell. This requires addressing three important obstacles (Notka et al., 2011):

1. For a stable and efficient series of reactions, the enzymes involved must be expressed in a highly concerted manner. As for other engineering technologies, this requires the availability of standardised regulatory parts and elements, e.g. promoters, ribosome binding sites, terminators, DNA-binding proteins (see Annex3.A1).
2. Fast and efficient formation of new gene clusters or operons requires the simultaneous assembly of such parts in a robust, yet flexible way.
3. Establishing an extrinsic biochemical pathway within a living cell must always be perceived in the context of its entire metabolism. For an industrial production organism, its metabolism should be limited to prevent interference from other pathways.

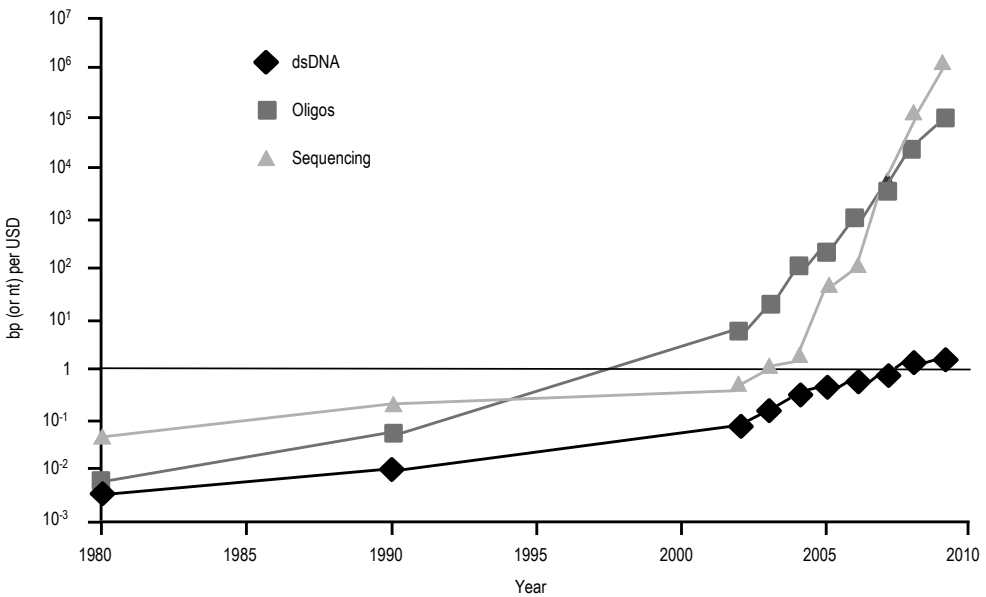
The most important technical barrier to progress in synthetic biology today is the cost and speed of fabrication of synthetic sequences. The need for routine large-scale synthesis of DNA hinders the ability to construct ever larger genetic devices and systems. By contrast, modelling does not require the development of entirely new technologies. This is the province of software design and construction and the field is progressing rapidly. DNA sequencing technologies have also moved rapidly in the last ten years, and are technically less demanding than large-scale synthesis.

Gene synthesis, the financial bottleneck

Maximising the potential of synthetic biology will require the development of cost-effective, high-throughput, high-fidelity methods of synthesising *de novo* DNA sequences of ever-increasing length and complexity. While the cost of sequencing has tumbled, the cost of gene synthesis had levelled out around USD 0.50 per base pair (bp) in 2010 (Jewett and Forster, 2010), of which USD 0.10 per bp for oligonucleotide synthesis (Cheong et al., 2010). This is prohibitive for most researchers at the genome level.

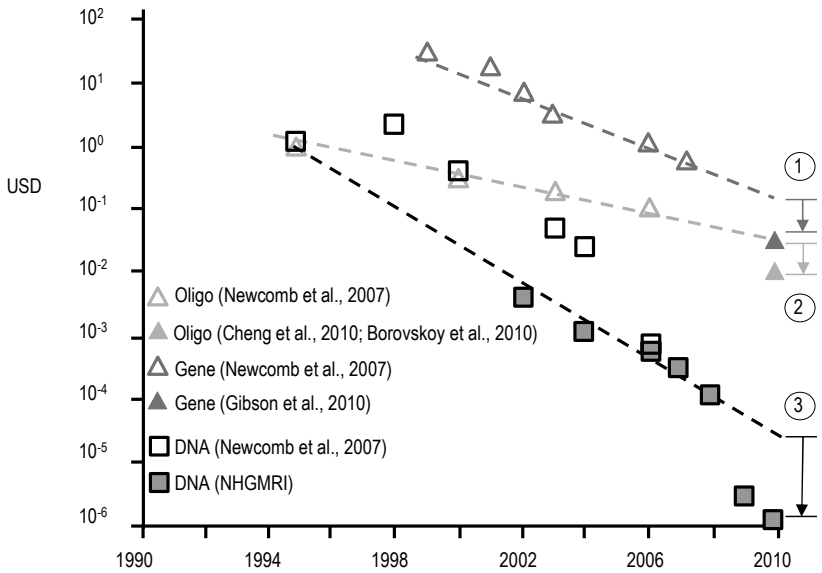
Figure 3.1 shows that gene synthesis (dsDNA), while improving rapidly, lags both oligonucleotide synthesis and sequencing. Figure 3.2 shows the decline in prices. Carlson gives data showing costs and productivity up to October 2012,[1] at which point the costs of synthesis were some four times the costs of sequencing. By February 2014, there was an apparent slowdown in the tumbling of prices, indicating that there may now be a phase in which prices will plateau.

Figure 3.1. Efficiency trends in synthesis and sequencing over the past 30 years (base pairs per dollar)



Source: Carr, P.A. and G.M. Church (2009), "Genome engineering", *Nature Biotechnology*, Vol. 27, pp. 1151-1162.

Figure 3.2. Declining cost trends for oligonucleotide synthesis, gene synthesis and DNA sequencing



Notes: Data sources shown on the graph. The recent declines are marked by downward pointing arrows drawn from the projected trend lines (based on about a decade of data) to the 2010 data points.

Source: Mitchell, W. (2011), "Natural products from synthetic biology", *Current Opinion in Chemical Biology*, Vol. 15, pp. 1-11.

Oligonucleotide synthesis

All gene synthesis technologies rely on the chemical synthesis of oligonucleotides to supply the building blocks for enzymatic assembly (Hughes et al., 2011). The most commonly used method is the cyclical, four-step phosphoramidite synthesis method developed in the 1980s. During the synthesis process, side reactions limit the quality and yields of oligonucleotides above 100 nucleotides (Hall et al., 2009). But the most crucial factor in DNA synthesis protocols today is the high error rate of oligonucleotide synthesis (Czar et al., 2009). Even an error rate as low as 1 in 10 000 bp can be a major concern if the product of interest is of that scale (10^4) or larger.

The robustness of solid phase phosphoramidite synthesis makes it easily amenable to automation, and this method is now used in almost all commercially available DNA synthesisers. However, for the assembly of long genes or even whole genomic sequences, the cost of the starting oligonucleotides alone can be prohibitive (for large-scale synthesis projects, of the order of hundreds of thousands of US dollars). Smaller-scale synthesis strategies are

needed to bring down the cost of starting oligonucleotides before *de novo* gene synthesis will be widely adopted.

DNA microarrays went part of the way to solving the problem. Depending on the chip platform used, several thousand to several hundred thousand distinct oligonucleotides can be synthesised on a single chip. In principle, these massively parallel microarrays can reduce the cost of oligonucleotides by orders of magnitude. However, microarrays produce very small amounts of oligonucleotides, and there are problems with purity and quality. Lee et al. (2010) reported the development of a microfluidic synthesis platform capable of generating a number of oligonucleotides in parallel for gene assembly. This system addresses many of the limitations associated with microarray technology. It is claimed that it can greatly reduce the cost of gene synthesis by reducing reagent consumption (by 100-fold) on a scale that removes the need for amplification before assembly.

Gene assembly

The yield of chemically synthesised oligonucleotides becomes exceedingly poor and the synthesis error rate increases with oligonucleotide length. To circumvent these limitations, methods have been developed to assemble relatively short synthetic oligonucleotides into longer gene sequences. They can be roughly grouped into ligation-mediated assembly and PCR-mediated assembly methods.

Ligation-mediated assembly has an inherently low mutagenesis rate (no errors due to DNA polymerase) and is relatively easy to use. For example, Blue Heron[2] uses a solid-support-based, ligation-mediated oligonucleotide assembly process to synthesise customer-supplied DNA sequences. The technology assembles a DNA duplex sequence on a solid support by iterative annealing and ligation of oligonucleotide pairs. This process is repeated until the entire gene sequence is sequentially assembled. As the technology has been fully automated, it allows for efficient, high-throughput synthesis of DNA sequences at commercial scale.

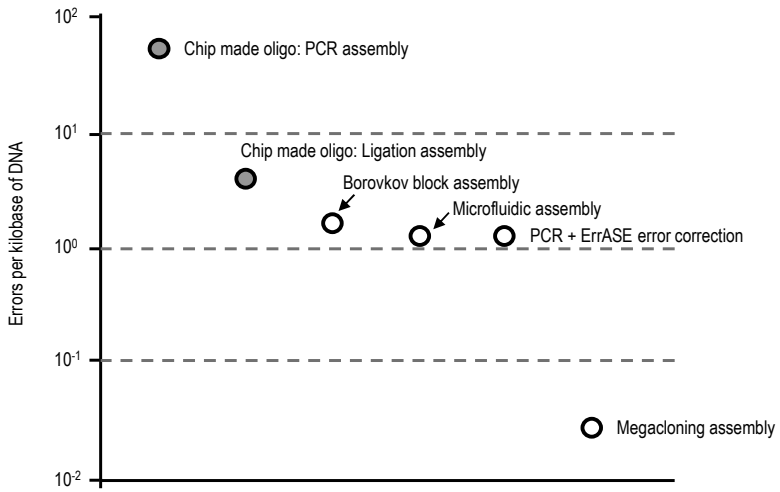
The most commonly used gene synthesis techniques currently rely on the polymerase chain reaction (PCR) to mediate assembly of a desired DNA sequence from short oligonucleotides. The desired gene product is assembled (often as a mixture of PCR products of varying lengths) in a single enzymatic reaction or as multiple-step assemblies that first divide the gene into separate sub-assembly reactions. In these methods, the various sub-assemblies are then mixed and joined in a series of thermal cycling reactions to yield the fully assembled gene products.

High error rates are a problem for many PCR-based gene synthesis techniques (Xiong et al., 2008). Errors are the result of mutations introduced during DNA polymerase-mediated synthesis and oligonucleotide synthesis procedures. At present, several methods of error correction of synthesised DNA exist, but they add complexity and cost to the process. A detailed discussion of these methods is beyond the scope of this paper, but some examples of research in the area are presented.

Improvements to PCR-based gene synthesis are published regularly. Mao et al. (2011) described a process they call Quikgene, a method that can assemble several hundreds of base pairs of genes. It can create complete *de novo* genes or extend or modify existing genes. The final genes are directly synthesised on desired vectors without any ligation or sub-cloning steps. Cheong et al. (2010) presented a simple, highly efficient, universal automatic kinetics switch gene synthesis method that enables synthesis of DNA up to 1.6 kbp (thousand base pairs) from 1 nano Mole oligonucleotide with just one PCR process.

A proof-of-concept experiment has demonstrated that the so-called "megacloning" method can reduce error rates by a factor of 500 (Figure 3.3) compared to the starting oligonucleotide pool generated by microarray (Matzas et al., 2010). In principle, with future development of platform automation, millions of oligos can be sequenced and sorted in a single megacloner run. This is paving the way to gene construction up to megabases in length (Ma et al., 2012).

Figure 3.3. Technology changes and reduction in error rates during DNA assembly



Source: Mitchell, W. (2011), "Natural products from synthetic biology", *Current Opinion in Chemical Biology*, Vol. 15, pp. 1-11.

A gene synthesis tipping point

Carr and Church (2009) describe a gene synthesis tipping point: the point at which commercial gene synthesis would be on par with synthesis of synthetic oligos, with similar costs and turn-around time, typically same-day shipping. When this occurs, there will be a landmark shift in the way many laboratories work. Much of the laborious work currently done to manipulate DNA will be phased out. Instead of cloning into vectors stored in laboratories, custom or standard vectors would simply be re-synthesised on demand. This will free up space and money: a lot of resources dedicated to deep-freezing of these materials will no longer be needed. Moreover, a much larger number of laboratories could undertake relatively large synthesis projects. In particular, small and spin-out companies would be relieved of this infrastructural expenditure and a new wave of synthetic biology companies may appear. The shift will also open the field to designers who need not be experts in traditional DNA manipulation techniques. It has been postulated that DNA fabrication will even lead to abiotic applications used in computing, detection or smart materials that will have very little in common with traditional biotechnology products (Czar et al., 2009).

Examination of the literature and research trends suggest that the tipping point will be reached in the near future. It seems that key developments and advances will come from exploitation of the advantages of miniaturisation that microfluidics offers, increased levels of automation, and high-quality error-correction methods.

The savings in expensive reagents is the most obvious advantage of microfluidics. Microfluidic systems that cover all the necessary unit operations for cellular assembly and analysis from oligonucleotide synthesis through to -omics analyses have now been designed (Szita et al., 2010). Other advantages include shorter analysis times and higher sensitivities. Moving forward, there should be a drive to integrate more steps into a single system and towards automation and parallelisation to increase experimental throughput.

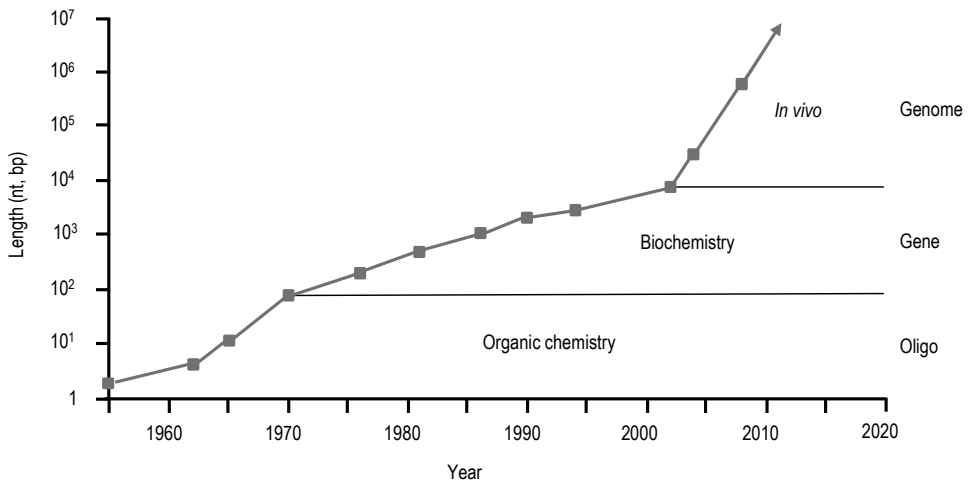
At the industrial scale, an increasing degree of automation of gene synthesis is mandatory to cut labour costs. Some steps are also simply no longer manageable by humans, such as the move from 96- to 384-well plates or the decrease of reaction volumes below 1µl. There is flawless interaction between automated pipetting and laboratory information management systems (LIMS). Notka et al. (2011) argue that the interplay between automation, LIMS and miniaturisation is the way to proceed to gene synthesis at industrial scale. Automation at least is not very technically demanding; current technology for robotics and automation should suffice.

Several companies (e.g. Gen9[3] and DNA2.0[4]) appear poised to make significant breakthroughs in high-throughput, automated production of DNA sequences at lower cost and higher accuracy than currently available. DNA2.0 now offers a rush service for DNA synthesis of <1 kilobase in five days, advertised as the fastest turnaround time in the industry, but still a way from overnight shipping for oligonucleotide synthesis.

Debugging of constructs

While testing, debugging and maintenance may appear of lesser importance than the actual synthesis operations described, they reportedly account for 80% of all software development costs. The Amyris five-year, USD 20 million artemisinin experiment reportedly spent 95% of its time trying to find and fix unintended interactions between parts (Henckel and Maurer, 2007).

Figure 3.4. Milestones in the sizes of *de novo* synthesized DNA



Note: Length is nucleotides (nt) for oligos before 1970, base pairs (bp) for double-stranded DNA from 1970 on. *In vitro* biochemical processing steps enabled the leap from oligos to genes, and *in vivo* processing steps (multiple cycles of cloning, sequencing and assembly) made possible the leap from genes to genomes.

Source: Carr, P.A. and G.M. Church (2009), "Genome engineering", *Nature Biotechnology*, Vol. 27, pp. 1151-1162.

Unfortunately, debugging a biological machine does not so far follow an apparent logical formula. Vast amounts of a genome can be completely deleted without apparent harm to the organism and even yield improved performance (this is, of course, an objective of minimal genome and cell research). At the same time, very modest changes can reduce performance,

and single-point mutations can easily be fatal. Biological complexity is the issue. Carr and Church (2009) describe two hierarchies of debugging:

1. All the separate genetic parts of a designed system should be tested singly in parallel, or in as simple a representation as possible. Where possible, combinations of simple parts into larger units should be performed along lines of linked function, so that these combinations can also be tested en route to final assembly. This is not unlike testing and debugging in other forms of manufacturing.
2. The ultimate testing environment is necessarily *in vivo* owing to biological complexity (Figure 3.4). Problems encountered at the drawing board or *in vitro* stages are likely to indicate real concerns for the *in vivo* context.

The chassis, or the minimal genome and cell concept

Although metabolic engineering has traditionally involved the manipulation of pre-existing cellular genomes, there is another way to think about the construction of industrial microbes. It involves the concept of a minimal genome: the minimum number of genes required to support basic life (Mushegian, 1999). The only minimal genome used as a starting point to date is the organism with the smallest known genome that can be cultivated under laboratory conditions, the bacterium *Mycoplasma genitalium* (Gibson et al., 2008; Glass et al., 2006). Precisely 100 of the 482 *M. genitalium* genes were deemed non-essential by genome-wide transposon mutagenesis. Deletion of these genes resulted in a strain with improved growth rates, as less energy is expended on non-vital cellular processes. The objective is to minimise the metabolic burden on the cell, so the remaining cellular energy can be directed towards the production of a desired industrial product, such as an industrial chemical or pharmaceutical drug (Pyne et al., 2011). Minimising the number of components required to support biological synthesis from synthetic DNA circuits or genomes enables adequate control of its function. The approach may also yield insights into the function of early cells, which were conceivably much simpler than modern cells (Stano et al., 2011). Insofar as the cell is the minimal form of the bioreactor, simpler, even artificial, cells make for more reliable bioreactors (Pohorille and Deamer, 2002). A further spin-off technology could be *in vitro* genome replication to replicate very large segments of DNA with high fidelity (Forster and Church, 2006).

For biotechnology applications, reducing the genomes of *E. coli* and other biotechnology workhorses is more useful than reduced-genome *M. genitalium* owing to the fragility and much slower growth rate of the latter (Jewett and Forster, 2010). Future work on *E. coli* and others will replace most current commercial bacterial strains, because in an industrial ferment-

er, an aerobic environment is usually desired and maintained, the nutrient concentrations are maintained within narrow ranges, and attachment to the vessel is not desirable. *E coli* is normally found in the (anaerobic) gut of mammals, although the genes required for survival in the gut may not be the same as those required for optimum industrial application (Sharma et al., 2007). Therefore many projects have aimed to reduce the size of the *E. coli* genome. For example, in 2006 targeted deletions removing 15% of the *E. coli* genome were not only viable, but also improved its properties for applications in molecular biology (Posfai et al., 2006). Synthetic genomics will be particularly helpful for redesigning microbes that possess potential biotechnology applications but have poor native genetic tools available.

Other chassis organisms

E. coli is the most commonly described chassis organism. However, despite its flexibility and its very low risk level, it is not always possible to ensure efficient transcription/translation of a heterologous gene in *E. coli*, and post-translational protein modification does not occur in prokaryotic[5] production systems, hence the development of specialised eukaryotic production hosts such as yeast. Such organisms have characteristics that lend themselves to use as a chassis organism.

The genus *Bacillus* has a long history in the biotechnology sector and various species have been used over the years to produce industrial enzymes such as amylases and proteases. About 60% of commercially available enzymes are produced by *Bacillus* species (Westers et al., 2004). Several species are non-pathogenic and have long been approved as safe to use as production hosts. A particular advantage is that they naturally secrete significant quantities of protein from the cell into the environment (Schallmay et al., 2004). They are easy to grow, and the genetics are well researched; the prototype species *B. subtilis* is second only to *E. coli* in terms of understanding both of its genetics and physiology. There are problems, however, of plasmid stability and there is no post-translational modification.

The use of yeast expression systems combines many advantages of complex mammalian hosts and prokaryotic hosts. Expression hosts such as *Saccharomyces cerevisiae* (the wine, beer and baker's yeast) are efficient at post-translational modification of other eukaryotic proteins, while, like *Bacillus*, they are non-pathogenic and can be grown in large volumes on simple growth media and can secrete proteins. In fact, yeasts are the most exploited group of industrial microorganisms (Fell and Phaff, 2003). In-depth knowledge of *Saccharomyces cerevisiae* genetics, genetic engineering, physiology and biochemistry has been accumulated, and industrial-scale fermentation technologies are readily available (for a review, see Nevoigt, 2008).

DNA sequencing: A challenge overcome?

Fortunately, sequencing technology is no longer the barrier to the development of synthetic biology that it once was, especially when compared to the difficulties of gene synthesis and assembly.

In 2004, the National Human Genome Research Institute of the National Institutes of Health (NIH – NHGRI) announced a total of USD 70 million in grant awards for the development of DNA sequencing technologies that would reduce the cost of sequencing the human genome from USD 3 billion, the amount spent on the public Human Genome Project, to USD 1 000 by 2014.[6] Already by 2008 massively parallel DNA sequencing platforms had become widely available, reducing the cost of DNA sequencing by over two orders of magnitude and bringing it within the grasp of individual investigators, not just genomics centres (Shendure and Ji, 2008).

As of December 2013 the routine cost of a human genome sequence had dropped to around USD 5 000, with the possibility of reaching USD 1 000 sometime in 2014. At least one company claims to have a technology that should soon become available that would bring the cost of the sequence of a human genome to USD 100.

Next-generation sequencing has shifted the bottleneck from sequencing to the best way to extract biologically meaningful or clinically useful insights from very large amounts of data (Shendure and Ji, 2008). The Short Read Archive at the US National Centre for Biological Information is soon expected to exceed a petabyte (National Academy of Sciences, 2013). As more and more high-throughput sequencers are deployed, not just in research but also in hospitals and biotechnology facilities and companies, growth of data on genomic information will be even faster.

Software infrastructure

Software infrastructure, the "unseen" infrastructure in the synthetic biology laboratory, deserves special attention for an essential reason. The "wet" technologies of synthetic biology described above will, at least in the medium term, be limited to research institutions and companies where oversight and regulation will be possible. However, as synthetic biology gains momentum specific types of software are likely to be increasingly accessible to non-experts working from a home computer who may start to use software to design parts. While in itself this does not represent a danger, subsequent construction of the designed part may. Therefore software use by non-experts (and experts, for that matter) represents a regulatory concern and may be far more difficult to monitor than the wet technologies of synthetic biology.

Technological advances have shown the utility and importance of using software tools that facilitate various engineering processes, such as computer-aided design (CAD). The application of computational tools in synthetic biology has not reached the stage at which the design and construction of biological parts has become routine, and some argue that there is a need for an integrated design environment for the synthetic biologist that is similar to CAD systems (Marchisio and Stelling, 2009). In recent years, many computational standards and tools have been developed, especially in the field of systems biology (Wierling et al., 2007), and most of these tools could be used in synthetic biology applications. While laboratory procedures are now borrowed from genetic engineering, concepts such as abstraction and interchangeable parts come from computer science and electrical engineering (Endy, 2005).

Since synthetic biology is its initial stages of development, best practices for the design, use and reuse of existing parts have not been widely established. The software infrastructure for synthetic biology at present raises several challenges, which have to be addressed in an efficient way to attain rapid growth and promote knowledge among young professionals wanting to enter this field.

Overview of existing tools and challenges

Computational tools that allow design and construction of model organisms *in silico* using scripts or visual interfaces have been developed in recent years (Table 3.1). A comprehensive list can be found at: www.sbml.org.

Biological computation and integration

Challenges such as programming life, with applications in DNA computing and synthetic biology, are already being addressed and represent a frontier for the convergence of computing with biology. For example, the Biological Computation Group at Microsoft Research[7] is working on projects that include designing molecular circuits made of DNA and programming synthetic biological devices to perform complex functions over time and space. The tools being developed are being integrated into a common software environment, which supports simulation and analysis on multiple scales and across many domains. This environment may in time serve as the foundation for a common language runtime for biological computation.

Table 3.1. Computational design tools for synthetic biology

Circuit design and implementation	
Biojade	http://web.mit.edu/jagoler/www/biojade/
Tinkercell	www.tinkercell.com/Home
Asmparts	http://soft.synth-bio.org/asmparts.html
ProMoT	www.mpimagdeburg.mpg.de/projects/promot
GenoCAD	www.genocad.org/genocad/
GEC	http://research.microsoft.com/gec
TABASCO	http://openwetware.org/wiki/TABASCO#TabascoSimulator
Hy3S	http://hysss.sourceforge.net/index.shtml
Circuit optimisation	
Genetdes	http://soft.synth-bio.org/genetdes.html
RoVerGeNe	http://iasi.bu.edu/_batt/rovergene/rovergene.htm
DNA and RNA design	
Gene Designer	www.dna20.com/index.php?pageID=220
GeneDesign	www.genedesign.org
UNAFold	www.bioinfo.rpi.edu/applications/hybrid/download.php
Mfold	http://mfold.bioinfo.rpi.edu/download/
DINAMelt	http://dinamelt.bioinfo.rpi.edu/
Vienna RNA package	www.tbi.univie.ac.at/~ivo/RNA/
Vienna RNA web servers	http://rna.tbi.univie.ac.at/
Zinc Finger Tools	www.scripps.edu/mb/barbas/zfdesign/zfdesignhome.php
Protein design	
Rosetta	www.rosettacommons.org/main.html
RAPTOR	www.bioinformaticssolutions.com/products/raptor/index.php
Hhpred	http://toolkit.lmb.uni-muenchen.de/hhpred
Modeler	http://salilab.org/modeller/
PFP	http://dragon.bio.purdue.edu/pfp/
Autodock 4.2	http://autodock.scripps.edu/
HEX 5.1	http://webloria.loria.fr/~ritchied/hex/
Integrated workflows	
SynBioSS	http://synbioss.sourceforge.net/
Clotho	http://biocad-server.eecs.berkeley.edu/wiki/index.php/Tools
Biskit	http://biskit.sf.net

Source: Marchisio, M.M. and J. Stelling (2009), "Computational design tools for synthetic biology", *Current Opinion in Biotechnology,* Vol. 20, pp. 479-485.

Standardisation and interoperability

Active efforts to develop synthetic biology tools have raised questions regarding standardisation and interoperability. In the field of synthetic biology, standardisation aims to allow researchers to exchange designs electronically, to send designs to fabrication centres for assembly, and to allow storage of designs in repositories and for publication purposes.

Building synthetic circuits involves design and simulation tools that combine standard parts to introduce or modify biological functions, akin to the way in which engineers design new machines. One of the defining missions of the MIT Registry of Standard Biological Parts is to store and share the list of standard parts and devices to make this process easier.

Standardisation in engineering disciplines allows components to be combined easily to form larger systems, an approach that relies on the modularity of those components. A prevailing assumption in synthetic biology is that biological components should be modular as well. However, characterisation, standardisation and modularity are affected by cellular context (Purnick and Weiss, 2009), and it cannot be assumed that a functional module in one cell type will work the same way even in a closely related cell type (Bagh et al., 2008). Therefore, quantitative characterisations of component functions are necessary for efficient network design (Canton et al., 2008). However, biological knowledge and design capabilities are not yet at the level of sophistication needed for *a priori* design and production of a prototype with a reasonable chance of success (Alterovitz et al., 2010).

In synthetic biology, network standardisation should be given equal weighting with component-centred standards. Systems Biology Markup Language (SBML) is a machine-readable format for representing computational tools in systems biology. It was developed to exchange biological process information in the systems biology community (Hucka et al., 2003). Many other standards such as Cellular Markup Language (CellML) (Lloyd et al., 2004), MIRIAM (Novere et al., 2005) and Systems Biology Graphical Notation (SBGN) (Novere et al., 2009) represent a set of conventions to depict biological processes in graphical notation to facilitate efficient and clear communication among biologists. The heterogeneity of approaches to addressing network standards by using specialised formats for data management within synthetic biology sub-groups must be addressed.

Interoperability is a somewhat vague term, but it is generally regarded as necessary for the diffusion of innovation. As used in information and communications technology (ICT), it may be described as the ability to transfer data and other information across systems (which may include organisations), applications or components (Gasser and Palfrey, 2007). The benefits for synthetic biology are clear: standard parts that, when put together in a

working system, will function across different systems and organisations. It is analogous to digital music that can be played by different music players.

The public sector is likely to play a limited role in bringing interoperability to synthetic biology. In ICT cases, the private sector largely can and does achieve a high level of interoperability on its own. The public sector may help by playing a convening role, or even in mandating a standard on which there is widespread agreement within industry after a collaborative process. In a very few cases, the public sector may need to ensure that market actors do not abuse their positions.

Biological noise control

Biological noise is a problem both at component and network level. Genetic circuits tend to mutate rapidly and become non-functional (Tucker and Zilinskas, 2006). In general, combining disparate components requires the tuning of biochemical parameters such as affinities or rate constants, which is often difficult to do in biological circuits. The development of synthetic gene networks is still difficult and most newly created genes are non-functioning owing to intrinsic parameter fluctuation, external disturbances and functional variations of intra- and extra- cellular environments. The design methodology for a robust synthetic gene network that works properly in a host cell under these conditions of noise and fluctuation is therefore a high priority (Lee and Chen, 2010).

International distributed research infrastructures

These infrastructures tend to be large, expensive international facilities such as CERN, the European Organisation for Nuclear Research.[8] While such an experimental facility for synthetic biology would be hard to envisage, the case for a central facility to house publicly available databases from which data can be distributed is easier to justify. For example, the National Center for Biotechnology Information[9] facility, and in particular its BLAST facility,[10] provides a harmonised method for searching a large range of genomes. Box 3.1 describes the fundamental requirements that governments need to be aware of (also see OECD Global Science Forum, 2010).

Box 3.1. Requirements for an international distributed research infrastructure (IDRIS)

An IDRIS should have:

- An identity and a name.
- A set of international partners that are, typically, research institutes, academic institutions, foundations or other research-oriented organisations from the public or private sectors. Often, only parts of these entities make up the infrastructure.
- A formal agreement by the partners to contribute resources, expertise, equipment, services or personnel to achieving a common scientific purpose. The agreement does not necessarily need to define a new legal entity or be legally binding.
- A strategic plan, or work programme, that conveys the rationale for establishing the IDRIS and its added value over and above the separate activities of the partners.
- A governance scheme (for decision making, at a minimum) and a set of officers (not necessarily salaried staff) with well-defined responsibilities.
- A focus on the provision of services to members and users.

In addition, an IDRIS may have:

- An independent legal status (or an equivalent legal identity under the terms of an existing intergovernmental agreement).
- A common fund and rules for acquisition/spending of funds.
- A secretariat.
- A host institution.
- A central entry point for users.
- Explicit policies for access by users to research resources and to data and for managing any generated intellectual property.

Source: Michalowski, S. (2013), presentation at the OECD Marine Biotechnology workshop, Paris, 7 November.

The reason why a computational and not an experimental IDRIS would be important in synthetic biology is the drive towards parity of price and delivery time between DNA sequencing and DNA synthesis. Achieving the tipping point would bring the costs of synthetic biology research to a spectrum of researchers well beyond the initial centres of excellence. However, even now, the bigger infrastructure challenge is electronic storage and distribution of the huge quantities of data being produced (National Academy of Sciences, 2013).

Conclusion

The biggest challenge – that of long, accurate DNA synthesis – is gradually being addressed, although the latest figures on cost show that the price may be starting to plateau. In fact, both synthesis and sequencing costs have stopped falling precipitously. Owing to the technical similarities between wet biotechnology and wet synthetic biology, many other infrastructural challenges can be said to have been overcome (at the laboratory level). The vast amount of sequence data is now shifting the bottleneck towards data storage and management. This mirrors what practitioners also regard as training and educational bottlenecks; the future synthetic biologist will be more skilled in mathematics and data handling, and more familiar with engineering concepts than focused on biotechnology laboratory skills.

Notes

1. www.synthesis.cc/.
2. www.blueheronbio.com/Services/Gene-Synthesis.aspx.
3. http://gen9bio.com/.
4. www.dna20.com/.
5. A prokaryote (bacteria) has no nucleus to contain its genetic material. A eukaryote, a higher form of life, has a nucleus.
6. www.genome.gov/12513210.
7. http://research.microsoft.com/en-us/groups/biology/.
8. http://home.web.cern.ch/.
9. www.ncbi.nlm.nih.gov/.
10. http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome.

References

Alterovitz, G., T. Muso and M.F. Ramoni (2010), "The challenges of informatics in synthetic biology: from bimolecular networks to artificial organisms", *Briefings in Bioinformatics*, Vol. 11, pp. 80-95.

Bagh, S., M. Mazumder, T. Velauthapillai, V. Sardana, G.Q. Dong, A.B. Movva, L.H. Lim and D.R. McMillen (2008), "Plasmid-borne prokaryotic gene expression: sources of variability and quantitative system characterization", *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, Vol. 77, 021919.

Canton, B., A. Labno and D. Endy (2008), "Refinement and standardization of synthetic biological parts and devices", *Nature Biotechnology*, Vol. 26, pp. 787–793.

Carr, P.A. and G.M. Church (2009), "Genome engineering", *Nature Biotechnology*, Vol. 27, pp. 1151-1162.

Cheong, W.C., L.S. Lim, M.C. Huang, M. Bode and M-H Li (2010), "New insights into the de novo gene synthesis using the automatic kinetics switch approach", *Analytical Biochemistry*, Vol. 406, pp. 51-60.

Czar, M.J., J.C. Anderson, J.S. Bader and J. Peccoud (2009), "Gene synthesis demystified", *Trends in Biotechnology,* Vol. 27, pp. 63–72.

Endy, D. (2005), "Foundations for engineering biology", *Nature*, Vol. 438, pp. 449-453.

Fell, J.W. and H.J. Phaff (2003), "Yeast", *Encyclopedia of Science & Technology*, 9th edition, McGraw-Hill.

Forster, A.C. and G.M. Church (2006), " Towards synthesis of a minimal cell", *Molecular Systems Biology*, doi: 10.1038/msb4100090.

Gasser, U. and J. Palfrey (2007), "When and how ICT interoperability drives innovation", *Berkman Publication Series*, November, The Berkman Centre for Internet & Society, Harvard University Research Centre for Information Law and University of St. Gallen.

Gibson, D.G., G.A. Benders, C. Andrews-Pfannkoch, E.A. Denisova et al. (2008), "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome", *Science*, Vol. 319, pp. 1215-1220.

Glass, J.I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M.R. Lewis, M. Maruf, C.A. Hutchison, H.O. Smith and J.C. Venter (2006), "Essential genes of a minimal bacterium". *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 103, pp. 425–430.

Hall, B., J.M. Micheletti, P. Satya, K. Ogle, J. Pollard and A.D. Ellington (2009), "Design, synthesis, and amplification of DNA pools for in vitro selection", *Current Protocols in Nucleic Acid Chemistry*, Chapter 9, Unit 9.2, Wiley Online Library.

Henkel, J. and S.M. Maurer (2007), "The economics of synthetic biology", *Molecular Systems Biology,* Vol. 3, pp. 117-121.

Hucka, M., A. Finney, H.M. Sauro, H. Bolouri et al. (2003), "The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models", *Bioinformatics*, Vol. 19, pp. 524–531.

Hughes, R.A., A.E. Miklos and A.D. Ellington (2011), "Gene synthesis: methods and applications", *Methods in Enzymology*, Vol. 498, pp. 277-309.

Jewett, M.C. and A.C. Forster (2010), "Update on designing and building minimal cells", *Current Opinion in Biotechnology*, Vol. 21, pp. 697–703.

Lee, C-C., T.M. Snyder and S.R. Quake (2010), "A microfluidic oligonucleotide synthesizer", *Nucleic Acids Research*, Vol. 38, pp. 2514–2521.

Lee, H-C. and B-S. Chen (2010), "Synthetic biology: construction of genetic circuits for engineering applications", *International Journal of Systems and Synthetic Biology*, Vol. 1, pp. 71-86.

Lloyd, C.M., M.D. Halstead and P.F. Nielsen (2004), "CellML: its future, present and past", *Progress in Biophysics and Molecular Biology*, Vol. 85, pp. 433–450.

Ma, S., I. Saaem and J. Tian (2012), "Error correction in gene synthesis technology", *Trends in Biotechnology*, Vol. 30, pp. 147-154.

Mao, Y., J. Lin, A. Zhou, K. Ji, J.S. Downey, R. Chen and A. Han (2011), "Quikgene: A gene synthesis method integrated with ligation-free cloning", *Analytical Biochemistry*, Vol. 415, pp. 21–26.

Marchisio, M.M. and J. Stelling (2009), "Computational design tools for synthetic biology", *Current Opinion in Biotechnology,* Vol. 20, pp. 479-485.

Matzas, M., P.F. Stähler, N. Kefer, N. Siebelt, V. Boisguérin et al. (2010), "High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing", *Nature Biotechnology*, Vol. 28, pp. 1291–1294.

Michalowski, S. (2013). Presentation at the OECD Marine Biotechnology workshop, Paris, 7 November.

Mitchell, W. (2011), "Natural products from synthetic biology", *Current Opinion in Chemical Biology*, Vol. 15, pp. 1-11.

Mushegian, A. (1999), "The minimal genome concept", *Current Opinion in Genetics & Development* Vol. 9, pp. 709–714.

National Academy of Sciences (2013), *Frontiers in Massive Data Analysis*, National Academies Press, Washington, DC.

Nevoigt, E. (2008), "Progress in metabolic engineering of *Saccharomyces cerevisiae*", *Microbiology and Molecular Biology Reviews,* Vol. 72, pp. 379–412.

Notka, F., M. Liss and R. Wagner (2011), "Chapter 11: Industrial Scale Gene Synthesis", *Methods in Enzymology*, Vol. 498, pp. 247-275.

Novere, N.L., A. Finney, M. Hucka, U.S. Bhalla, F. Campagne et al. (2005), "Minimum information requested in the annotation of biochemical models (MIRIAM)", *Nature Biotechnology*, Vol. 23, pp. 1509-1515.

Novere, N.L., M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin et al. (2009), "The systems biology graphical notation", *Nature Biotechnology*, Vol. 27, pp. 735-741.

OECD Global Science Forum (2010), "Establishing large international research infrastructures: issues and options", www.oecd.org/sti/scitech/47027330.pdf.

Pohorille, A. and D. Deamer (2002), "Artificial cells: prospects for biotechnology", *Trends in Biotechnology*, Vol. 20, pp. 123-128.

Posfai, G., G. Plunkett, T. Feher, D. Frisch, G.M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S.S. Sharma et M. de Arruda (2006), "Emergent properties of reduced-genome *Escherichia coli*", *Science*, Vol. 312, pp. 1044-1046.

Purnick, P.E.M. and R. Weiss (2009), "The second wave of synthetic biology: from modules to systems", *Nature Reviews Molecular Cell Biology*, Vol. 10, pp. 410-422.

Pyne, M., K. Sukhija and C.P. Chou (2011), "2.08 - Genetic Engineering", in *Comprehensive Biotechnology* (2nd edition), Vol. 2, pp. 81-91.

Schallmey, M., A. Singh and O.P. Ward (2004), "Developments in the use of Bacillus species for industrial production", *Canadian Journal of Microbiology*, Vol. 50, pp. 1-17.

Sharma, S.S., F.R. Blattner and S.W. Harcum (2007), "Recombinant protein production in an *Escherichia coli* reduced genome strain", *Metabolic Engineering*, Vol. 9, pp. 133–141.

Shendure, J. and H. Ji (2008), "Next-generation DNA sequencing", *Nature Biotechnology*, Vol. 26, pp. 1135-1145.

Stano, P., T.P. Souza, M. Allegretti, Y. Kuruma and P.L. Luisi (2011), "New and unexpected insights on the formation of protocells from a synthetic biology approach: The case of entrapment of biomacromolecules and protein synthesis inside vesicles", in P. Stano and P.L. Luisi (eds.), *Minimal Cells*, Springer, Dordrecht. pp. 195-216.

Szita, N., K. Polizzi, N. Jaccard and F. Baganz (2010), "Microfluidic approaches for systems and synthetic biology", *Current Opinion in Biotechnology*, Vol. 21, pp. 517–523.

Tucker, J.B. and R.A. Zilinskas (2006), "The promise and perils of synthetic biology", *The New Atlantis*, Spring, pp. 25-45.

Westers, L., H. Westers and W.J. Quax (2004), "*Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism", *Biochimica et Biophysica Acta*, Vol. 1694, pp. 299-310.

Wierling, C., R. Herwig and H. Lehrach (2007), "Resources, standards and tools for systems biology", *Briefings in Functional Genomics and Proteomics*, Vol. 6, pp. 240-251.

Xiong, A.S., R.H. Peng, J. Zhuang, J.G. Liu, F. Gao, J.M. Chen, Z.M. Cheng and Q.H. Yao (2008), "Non-polymerase-cycling-assembly-based chemical l gene synthesis: Strategies, methods, and progress", *Biotechnology Advances*, Vol. 26, pp. 121–134.

Annex 3.A1

Synthetic biology part types

The following selection of typical parts was taken from the iGEM Registry of Standard Biological Parts (http://parts.igem.org/Catalog?title=Catalog).

Composite parts: Composite parts are combinations of two or more BioBrick parts.

DNA: DNA parts provide functionality to the DNA itself. DNA parts include cloning sites, scars, primer binding sites, spacers, recombination sites, conjugative transfer elements and transposons.

Plasmid backbones: A plasmid backbone is defined as the plasmid sequence beginning with the BioBrick suffix, including the replication origin and antibiotic resistance marker, and ending with the BioBrick prefix.

Plasmids: A plasmid is a circular, double-stranded DNA (dsDNA) molecule typically containing a few thousand base pairs that replicates within the cell independently of the chromosomal DNA.

Primers: A primer is a short single-stranded DNA sequence used as a starting point for PCR amplification or sequencing.

Promoters: A promoter is a DNA sequence that tends to recruit transcriptional machinery and lead to transcription of the downstream DNA sequence.

Protein coding sequences: Protein coding sequences encode the amino acid sequence of a particular protein. Some protein coding sequences only encode a protein domain or half a protein. Others encode a full-length protein from start codon to stop codon.

Protein domains: Protein domains are portions of proteins cloned in-frame with other proteins domains to make up a protein coding sequence. Some protein domains might change the location of the protein, alter its degradation rate, target the protein for cleavage, or enable it to be readily purified.

Ribosome binding sites: A ribosome binding site (RBS) is an RNA sequence found in mRNA to which ribosomes can bind and initiate translation.

Terminators: A terminator is an RNA sequence that usually occurs at the end of a gene or operon mRNA and causes transcription to stop.

Translational units: Translational units are composed of a ribosome binding site and a protein coding sequence. They begin at the site of translational initiation, the RBS, and end at the site of translational termination, the stop codon.