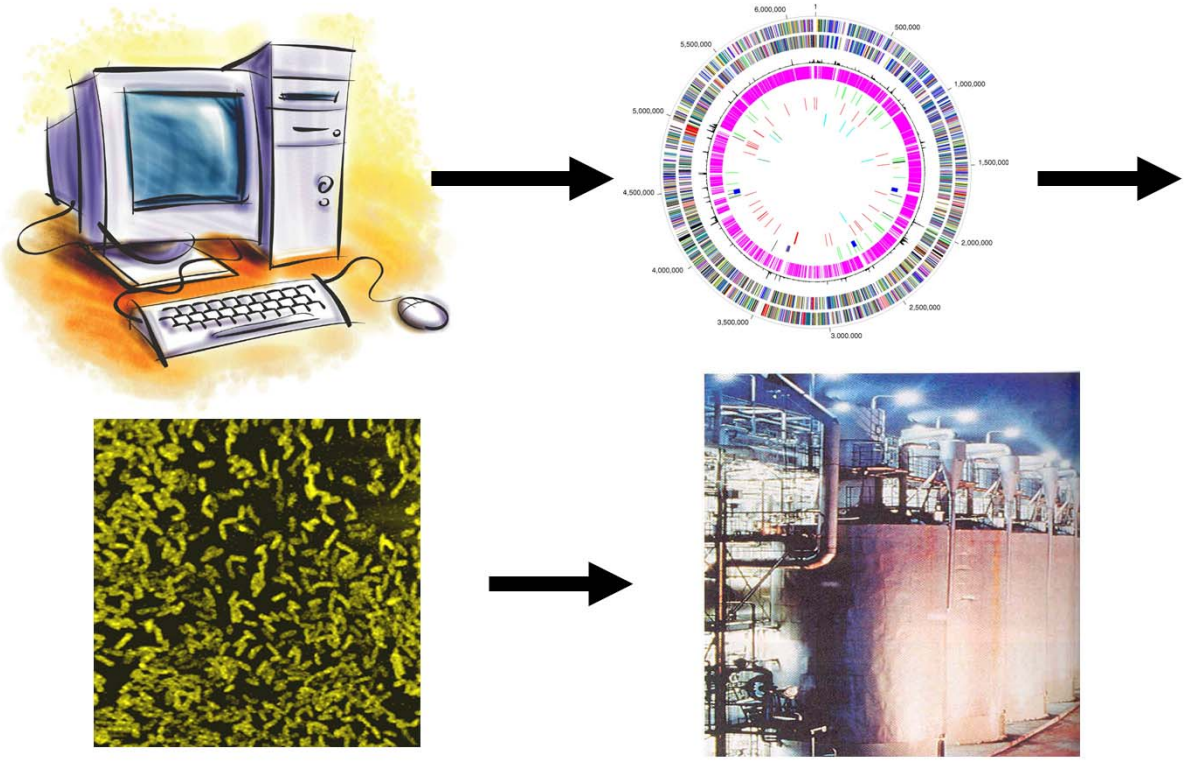

REPORT

Analysis of new developments in white (industrial) biotechnology

April 2016

Colophon

Title:	**Analysis of developments in white (industrial) biotechnology**
Client:	National Institute for Public health and the Environment Rijksinstituut voor Volksgezondheid en Milieu (RIVM) Antonie van Leeuwenhoeklaan 9 3721 MA Bilthoven Contact person: Petra Hogervorst 030 274 25 35 petra.hogervorst@rivm.nl
Contracted organisation:	Ameco Adviesgroep Milieubeleid (Ameco) Koningslaan 60 3583 GN Utrecht Contact person: Rik Kleinjans 030 254 58 40 r.kleinjans@ameco-ut.nl Authors: Hans Bergmans Wilma Vennekens
Picture cover page:	V. Martins dos Santos: "Ideal microbial factory from scratch"
Date report:	29 April 2016

Content

1 Executive summary

The underlying report presents the developments and innovations that are taking place within white biotechnology. The National Institute for Public Health and the Environment (RIVM) initiated this research project to investigate the risks, in terms of biosafety and biosecurity, which may result from applications of production organisms, and to evaluate whether the current risk assessment methodologies are adequate. Industrial or 'white' biotechnology is the application of biotechnology for the processing and production of chemicals, materials and energy. White biotechnology is based on microbial fermentation processes. This report focuses on techniques used for the development of enhanced production strains for white biotechnology under contained use. Similar reports were also prepared for red and green biotechnology.

Based on desk research of articles in Current Opinion in Biotechnology and the list of current grant awards of the British Biotechnology and Biological Sciences Research Council (BBSRC) a preliminary list of techniques was identified for which the most developments are expected in the coming years. Further information was obtained in interviews with 8 representatives of scientific and commercial organisations active in white biotechnology. Five techniques were singled out for further discussion, because of their innovative character in the field of white biotechnology or, in the case of the evolutionary techniques, because of their prominence in the interviews:

- Genome editing: CRISPR/Cas9
- 'Next-' (and 'third') generation sequencing techniques
- DNA building blocks: application of synthetic biology
- Techniques for DNA assembly
- Adaptive laboratory evolution and directed evolution

Genome editing: CRISPR/Cas9

CRISPR/Cas is a complex between the Cas moiety, which is a nuclease (DNA cutting enzyme) and a guide RNA molecule that guides the nuclease towards a specific position on a DNA molecule. The Cas nuclease cuts in both strands of a DNA molecule. The CRISPR/Cas nuclease is guided towards specific sequences in the DNA, and makes its cut at very precise places in the DNA. CRISPR/Cas is considered a major innovative technique. For optimal use, the full genome of the organism should be known in order to predict optimal places for CRISPR/Cas9 activity. The main advantage of the technique, next to its specificity, is in the potential of making several changes at different specific sites of the genome in one go. CRISPR/Cas9 is an available technique at the moment. It is mainly used for eukaryotic cells (fungi, yeasts), but prokaryotic organisms (bacteria, archaea) will probably follow. Still, the interviewed experts from industry were hesitant about the full-fledged use of the technique in industrial microbiology, also because other techniques are already available, at less cost. Also, the situation around Intellectual Property of CRISPR/Cas methodologies is not yet resolved.

'Next'- (and 'third') generation sequencing techniques

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. Sequencing techniques have developed over the years and include any method or technology that is used to determine the order of the four bases in a strand of

DNA. Sequencing techniques are very important in the context of strain improvement because of their potential to find new genes, to plan the genetic improvement of a strain, and to validate the strain after it has been constructed for the exact genetic composition. Major improvements in next and third generation sequencing that drive these techniques are speed, amount of data produced, and low cost. Third generation techniques offer the possibility to look at native DNA, but a barrier at the moment still is the error rate of third generation sequencing techniques. Next generation sequencing techniques are available at the moment, but may become superseded by third generation techniques, once the precision of that technique has improved. The techniques will yield enormous amounts of data, in the order of a terabase per day. This may in itself create problems because of the computing power that is necessary to analyse the data, and logistical problems around making the data available.

DNA building blocks: applications of synthetic biology

Synthetic biology is a vast field, where creative use is made of the knowledge of biological systems. DNA building blocks are designed with the aim to make 'standardised parts' that can be put together to biological systems, much like the motor of a car is produced from standardised parts. A standardised building block consists of a gene, together with regulatory elements that determine how the gene will be expressed in the cell. Fine tuning of regulatory networks in a cell is one of the main problems for efficient operation of genes within a metabolic pathway. The usefulness of DNA constructs depends on the possibilities to predict which (variants of) genes should be put together in order to introduce a desired process into a production strain. A barrier to the application of the techniques is the requirement for thorough knowledge of the functioning of the biological parts. The techniques are not specifically suited to trial-and-error approaches. The reverse consideration is the main driver for the use of the techniques: the possibility to rationally approach the design of more efficient production strains.

Techniques for DNA assembly

DNA assembly is necessary to fuse small DNA building blocks into larger arrays, and into entire chromosomes and even genomes. The possibilities of the assembly of DNA elements into larger arrays are limited by our understanding of the influence of DNA architecture on its functioning. Some techniques therefore aim at assembling DNA in various combinations to find out what works best. The technology of DNA assembly is already at the point that an entire bacterial genome can be put together. While assembling smaller DNA arrays is a prerequisite for white biotechnology already, the actual use of larger arrays like artificial chromosomes is only expected at medium term. DNA assembly will no doubt lead to the construction of 'minimal' organisms, with completely synthetic genomes. These developments are however only expected to become actual for production strains at a term of 10 years or more.

Adaptive laboratory evolution and directed evolution

Techniques that make use of genetic variability that is present in strains of production organisms have been used already for a long time (see for instance Novick and Szilard, 1950). These techniques are in principle unfocussed: the entire genome of an organism is mutagen-

ized, and mutants with the desired phenotype are obtained through selection. As this process may require hundreds or thousands of rounds of replication and growth, the micro-organisms have to be subcultured, or grown in continuous culture, for a long time.
There are new schemes to make the process more focussed by using schemes to mutate only the genes of interest. This can be done either *in vivo*, in adaptive laboratory evolution, or *in vitro*, for instance by error prone replication. The 'libraries' that are created in this way have to be put under selective pressure *in vivo*, and screened for favourable phenotypes.
These evolutionary methods may therefore be time consuming and costly. The high throughput methods that may be necessary for screening are too specialised for small companies. But, less time and cost intensive methods are available, and are used in practice already.

The analysis concludes that white biotechnology is evolving rapidly, because of the possibilities offered by new techniques and approaches. These can be used for enhancing the efficiency of metabolic processes that are already in use, or to devise novel metabolic processes for the production of a large variety of biological compounds. There is a clear distinction between the expectations of scientists who are developing new techniques and using them for specific biotechnological purposes, and companies active in white biotechnology that are inclined to only use new techniques if a business case can be construed that leads to a clear advantage. In industrial biotechnology techniques are chosen based on the requirements for the process and its cost effectiveness, not on the basis of the mere availability of the technique. The direction into which white biotechnology will move is therefore difficult to predict. The conclusions in this report on the time frame for application of the new techniques (see Appendix 1) are primarily based on the, rather cautious, expectations put forward by the interviewed experts from industry. Based on the efforts and enthusiasm encountered in projects in applied science, one could also be led to expect faster developments.

The techniques described in this report cause a paradigm change for risk assessment of the (GM) products of these techniques. Properties of newly produced organisms can be evaluated based on the properties of the parental organism and the properties of the genetic information that is introduced. These last properties can be assessed from the phenotypes of the strains from which the genetic information is derived. The new techniques (except the evolutionary techniques) create novel genes, and the resulting phenotypes have not been seen before. They can be predicted, for instance by the computational techniques of bioinformatics. The screening methods developed for detection of organisms with a useful phenotype, may also provide useful information for biosafety purposes, at an early stage of development of novel organisms.

2 Introduction

Industrial or 'white' biotechnology is the application of biotechnology for the processing and production of chemicals, materials and energy (EuropaBio, 2011). White biotechnology is based on microbial fermentation processes. Fermentation processes have already been known for a very long time, ever since they were used for food and drink preservation, e.g., by lactic acid or alcoholic fermentation. Industrial biotechnology uses micro-organisms (bacteria, archaea, yeasts, fungi, micro-algae) as production organisms. The aim of the technology is to optimise the processes in terms of variety and quality of products formed, requirements for raw materials and efficiency of the applied processes. This can be accomplished by improving the production process, e.g. by improving fermenter design, and by optimizing the suitability of the production organisms for the production process. Over the past decades our understanding of the functioning of the biological cell has increased dramatically. This has led to an enormous potential for using the cellular processes for the purpose of white biotechnology, by turning the biological cell into an ever more efficient production engine for white biotechnology. The expectations for the accomplishments of white biotechnology in the near future are high. EuropaBio, the European association for bio-industries, for instance, predicts (EuropaBio, 2011) the following vision for white biotechnology in 2025:

- *An increasing number of chemicals and materials will be produced using biotechnology in one of its processing steps. Biotechnological processes are used for producing chemicals and materials, otherwise not accessible by conventional means, or existing products in a more efficient and sustainable way.*
- *Biotechnology allows for an increasing eco-efficient use of renewable resources as raw materials for the industry*
- *Industrial biotechnology will enable a range of industries to manufacture products in an economically and environmentally sustainable way.*
- *Biomass derived energy, based on biotechnology, is expected to cover an increasing amount of our energy consumption.*
- *Rural bio refineries will replace port-based oil refineries wherever it is economically feasible.*
- *European industry will be innovative and competitive, with sustained cooperation and support between the research community, industry, agriculture and civil society.*
- *Green Biotechnology* [i.e., biotechnology focusing on sustainable processes] *could make a substantial contribution to the efficient production of biomass raw materials.*

This study was commissioned by RIVM, the National Institute for Public Health and the Environment. RIVM acknowledges that newly developed techniques for the improvement of production organisms may extend to the boundaries of genetic modification, e.g., the techniques used in synthetic biology. RIVM has initiated this research project to investigate the risks, in terms of biosafety and biosecurity, which may result from applications of these production organisms, and to evaluate whether the current risk assessment methodologies are adequate. As rapid developments are occurring in biotechnology, the Dutch Ministry of Infrastructure and the Environment (I&M) wants to develop policies proactively, in order to ensure adequate environmental and human safety without impeding developments in biotechnological processes. As a first step in this analysis this report presents an overview of current new developments regarding production organisms for white biotechnology that can have

an influence on the environmental risk assessment of these organisms. Similar reports were also prepared for red and green biotechnology.

The main research question underlying this report is:
'Which developments and innovations are taking place within white biotechnology on the short term (1-4 years) and the medium-long term (5-9 years)?'

3 Methodology

In order to investigate and analyse future developments in the white biotechnology sector the authors used the following tools to gather information.

- Desk research: Analysis of articles in the 2014 and 2015 volumes of Current Opinion in Biotechnology concerning topics broadly[1] related to the development of strains for white biotechnological processes. This approach was chosen because earlier search strategies in the entire literature, based on key words, e.g. 'innovative', '(white) biotechnology', did not yield consistent or meaningful results.
- Desk research: Screening the awards lists of the British Biotechnology and Biological Sciences Research Council (BBSRC), with search criteria 'current' and 'biotechnology'.
- Preparing a preliminary overview of (innovative) technologies in white biotechnology. Selections made for this list were based on expert judgement by the authors.
- Interviews with representatives of organisations (scientific and commercial) using the selected techniques.
- Selection of the most relevant techniques, based on information received and expert judgement.

Chapter 4 describes the results of the desk research and the interviews. Furthermore, based on the information received, chapter 5 presents a detailed description of the technologies selected.

The techniques are discussed on the basis of the following outline:

- General description of the technique
- Technical description: *i.e., a description of intended mode of action. Which alteration is made in the organism, or what is the way it affects gene expression (target and off-target effects)?*
- Impact of the technology, e.g., host effects: *what will be the impact of the technique, e.g., which new traits or effects can be realized in the organism?*
- Application areas: *e.g., which products can be made; what is the scope of application?*
- Barriers and drivers: *what factors can contribute to or counteract the success of the technology? Which 'supporting technologies' can contribute in what way to new developments?*
- At the horizon: Which new developments/innovations can be anticipated (short term and medium term)?

[1] Articles that focus on developments that are relevant for a specific goal, e.g., the optimization of a specific metabolic pathway, were not taken into consideration.

4 Desk research and interviews

4.1 Desk research

In order to get an overview of the currently relevant developments in white biotechnology, the 2014 and 2015 volumes of Current Opinion in Biotechnology, one of the more opinionating journals in the field, were analysed. Approximately 100 papers were taken into consideration. An overview of research & development issues that are currently being explored was obtained by screening the awards lists of the British Biotechnology and Biological Sciences Research Council (BBSRC), with the search criteria 'current' and 'biotechnology'. A total of 565 grant descriptions was found. 57 of these projects described or used techniques that were deemed relevant for this inventory[2]. Based on both these inputs, a preliminary list of techniques was identified and categorised into a number of themes related to 'production strain development' (see Table 1). Techniques were selected based on the prominence of their occurrence in the above mentioned articles and descriptions and articles. The division into themes was done based on expert opinion by the project team, in consultation with the advisory committee.

Table 1 Preliminary list of the techniques identified to be discussed in the interviews

Area of interest	Overarching technique	Examples of specific techniques
Genetic strain improvement	Genome editing	CRISPR/Cas9
	Artificial chromosomes	DNA assembly
		Minimal organisms (e.g. Sc2.0)
Finding new and improved genes	Metagenomics	Next generation sequencing (NGS)
	Bioinformatics	Prediction of gene function
	Artificial genes	Optimized genes
		Synthetic building blocks
	Directed evolution	Mutagenesis, selection and NGS
Metabolic pathway engineering	Analysis and design of pathways	Metabolic modelling
		Metabolic flux analysis
		Pathway transfer
Modifying gene expression	Engineering existing regulation	RNAi, CRISPRi
	New regulatory element	Synthetic genetic switches
Phenotypic testing	Testing under fermentation conditions	Biosensing
		Microfluidics

The table above was used as the main input for the interviews to verify with experts in the field whether the list was accurate, complete and presented all relevant techniques which are currently developing/innovative. The results from the interviews are presented in paragraph 4.2.

[2] The search can be repeated at http://www.bbsrc.ac.uk/research/grants-search/advancedsearch/, using the settings: Search Criteria: Award Type: Research Grants; Institute Projects; Fellowships; Studentships; Training Grants; Award Status: Current; Text Search: 'biotechnology'. The list used in the preparation of this report was derived from the database on 11 February 2016.

4.2 Interviews with scientists and companies active in white biotechnology

In order to get an understanding what techniques are of actual importance, a number of experts from scientific institutes, large private sector companies and small scale innovative start-ups involved in white biotechnology were interviewed. The overview of techniques as presented in Table 1 and a list of questions was presented in advance to the interviewed experts.

The following questions were leading during the interview:

- As a background to your information: which types of organisms are of interest for your organisation (bacteria/archaea/yeast/fungi/other eukaryotes)?
- Which techniques do you expect will be of importance for the development of production organisms for white biotechnology, in the near future (now – 5 yr.) and in the period of 5 -10 yr. from now?
- The list you received specifies the techniques that we expect will be of importance for the development of white biotechnology. Are important techniques missing on this list? Do you agree with the logic of this list?
- Most of these techniques are still under development. They have large potential value for white biotechnology: 'the sky is the limit'. But, what do you expect will be the actual importance of these techniques in future?
- What will be 'drivers' and 'barriers' for the application of these techniques in white biotechnology?
- Besides methodological drivers and barriers, does your organisation perceive other drivers and barriers (e.g., economic/social/regulatory/public perception)?

A total of seven industrial organisations (six in the Netherlands, one in the USA) and two scientific institutions were approached. All industrial organisations as well as one of the scientific institutions agreed to be interviewed. Industrial organisations ranged from a large company to small start-up companies, and the interest group of Dutch biotechnological industries. Results of the interviews are an integral part of the report. They are not presented as separate interviews but are grouped according to topics. All experts agreed to respond provided that input would be presented anonymously. Where relevant comments provided by interviewees are included in the detailed description of the techniques (chapter 5) as well as in the discussion (chapter 6).

<u>Areas of expertise of the interviewees:</u>
The interviewees covered various fields of industrial microbiology: the use of yeasts and fungi as versatile organisms for various fermentation purposes, and the use of yeasts or prokaryotic organisms for specialised processes, such as production of renewable energy sources and use as food fragrances. One company does not work with pure cultures, but with undefined mixed cultures; this company is at the moment not interested in the use of modified production strains, but indicates that they are following technical innovations in white biotechnology with interest for potential future use.

<u>General comments made by the interviewees:</u>
All commercial interviewees commented on the focus of the present project: techniques in white biotechnology and their expected role on the short and medium term. They indicated

that their focus is on the various processes that they are using. New techniques are screened for their potential use, but the main driver to adopt a technique will be the (potential) commercial benefit. The companies underline that they are working under contained use conditions. Their products, proteins or smaller molecular weight organic compounds will be marketed, but the products will not contain live organisms. The regulatory situation around contained use are not felt as too much of an impediment for the developments in white biotechnology. Hence, the main issue for biosafety in white biotechnology are the characteristics of the organism and its history of safe use. Still some experts underlined the importance of keeping the right perspective of technological innovations. They plead that innovative precise techniques are also developed with the aim of safety. They should also be seen in that context. This is particularly an issue for companies working in white biotechnology products for food and feed use.

Comments made by interviewees regarding the development and use of techniques in white biotechnology:

- The experts recognised the techniques presented to them in the preliminary list of techniques

- All experts agreed that CRISPR/Cas9 is a promising technique that is expected to cause changes in the field of white biotechnology. That being said, the experts from industry cautioned against too high expectations about the innovative breakthroughs of the use of CRISPR/Cas9. According to the scientific interviewees, CRISPR/Cas9 offers wide possibilities, especially for producing several genome edits in one go.

- Sequencing techniques are of the most importance, for instance for the validation of produced strains, and for learning which genetic changes are actually important for strain improvement. But, the sequencing power of in particular 3rd generation sequencing techniques will soon lead to an enormous increase in available data that have to be stored, handled and analysed. Also, the free availability of these data is an issue.

- Genome engineering is an important process. At this moment the issue is the reduction of complexity in a genome. Creating new artificial chromosomes is seen as a possibility for the longer term. Opinions differ about the expected implementation of minimal organisms in the industry. Minimal organisms could resolve efficiency issues in a production strain due to large redundancy. On the other hand it was commented that heterogeneity in an organism may be essential for robustness and genetic variation. Hence it is questionable whether the concept of minimal organisms will be a viable option for industry.

- Metabolic modelling was mentioned as a very important approach to strain development. But it was also commented that blockades for metabolic processes may not only be in the efficient operation of gene products in a pathway, but also in side processes such as transporters, or in cumulation of toxic compounds.

- Several experts from both scientific institutes as well as the industry pointed out the importance of classical techniques for strain improvement and screening. Classical microbiology remains a very important approach. The use of the classical production platforms,

bacteria, yeasts, fungi, will continue. Evolutionary approaches using classical techniques are essential. The obtained strains with the desired phenotype will have to be checked (by sequencing) for the mutations that have occurred and that are essential for the phenotype.

- High throughput screening of developed strains is important to find the optimal strains. Miniaturization and online non-invasive sensoring are important developments. However, very specialised techniques for high throughput screening, e.g., microfluidics, are not within the budget for small companies (yet).

- A number of subjects have been discussed that are expected to have an impact on white biotechnology at the long term (>10 years). Some examples that were mentioned were: automation and robotisation in strain development (but see the much more optimistic expectations of companies like Ginkgo Bioworks); the use of xenobiology; the development of in vitro, cell free, production systems.

Other comments:

- There is a tendency to make use of the large amount of already available studies and data in data bases, rather than to search for new solutions in nature.

- Off-target effects are a factor that should be taken into account, but they are not necessarily a biosafety issue. Some of the most advanced techniques, like xenobiology, 'alternative life', may even be intrinsically safe because of the use of not naturally occurring amino acids and nucleotides.

- Attention was drawn to interesting developments in 'biofabrication', the use of DNA in different kinds of technologies (i.e., not only in biotechnology, but also in, for instance, nanotechnological processes).

- The public perception of technological developments is an important factor. These perceptions differ in different societies.

4.3 Selection of techniques

Techniques to be discussed in detail were chosen because of their innovative character in the field of white biotechnology, or because of their effectiveness for the construction of novel production strains. These techniques are principally those that solve urgent problems in white biotechnology, or speed up current biotechnological processes in such a way that it is economically feasible and interesting to use them. The choice was made on the basis of expert opinions of the interviewees, combined with the expert opinion of the authors. During the interviews comments were made that all techniques mentioned table 1 are relevant, but not all of them are innovative, in the sense that they have been around for quite some years. Also, some techniques, e.g. bioinformatics, metabolic modelling, biosensing, microfluidics, are ancillary techniques for the development of new strains, but are not in themselves techniques for strain development. A number of ancillary techniques are discussed in paragraph 5.3.

Techniques presented in chapter 5:

- Genome editing: CRISPR/Cas9
- 'Next-' (and 'third') generation sequencing techniques
- DNA building blocks: application of synthetic biology
- Techniques for DNA assembly
- Adaptive laboratory evolution and directed evolution

4.4 Techniques and /or subjects not covered by the underlying report

Based on the outcome of the desk research, meetings with the advisory committee and the interviews, this report focuses on the contained use of micro-organisms for production purposes. Deliberate release into the environment of micro-organisms is not treated explicitly in the report. In deliberate release of micro-organisms, for instance for purposes of bioremediation or growth enhancement and pest control in agriculture, the focus is on activities with organisms under environmental conditions that cannot be controlled. This is not a fundamental difference from fermentation, but different approaches may be chosen, especially for the selection of the best performing organisms. Deliberate release of micro-organisms is subject to regulatory procedures different from contained use. However, if micro-organisms are handled under the lowest containment in contained use, they will also be scrutinised for their environmental impact, in a procedure much like the procedure for deliberate release.

The results of this report will be used in further investigations of the risks, in terms of biosafety and biosecurity, which may result from applications of novel production organisms for white biotechnology, and to evaluate whether the current risk assessment methodologies are adequate. The report therefore focuses on techniques to construct production organisms. Consequently, developments for improved reactor design are not taken into account. These developments can also have important aspects for biosafety and biosecurity. An example are reactors for light harvesting processes with photolithotrophic organisms. These reactors have to be transparent and are necessarily more fragile and sensitive to mechanical damage or vandalism.

Xenobiology, or 'alternative life', i.e., the use of not naturally occurring amino acids and nucleotides, is a field showing potential for the safe development and use of micro-organisms in white biotechnology (e.g.: Schmidt and De Lorenzo, 2016). This topic is still very much in its first stages of development. It was not recognized as a topic of current interest by the interviewed experts, and is not covered here. It is only mentioned 'at the horizon', in paragraph 5.3.

5 Description of relevant techniques

The table in Appendix 1 provides a summary overview of the techniques described in this chapter, their application areas, the perceived barriers and drivers for their use and the horizon of use.

5.1 Genome editing: CRISPR/Cas9

General description

Strain improvement in white biotechnology has always been an issue. Mutagenesis by chemical or physical means has always (knowingly or unknowingly) been the method of choice. After mutagenesis the desired phenotype can be selected by growth under selective conditions. As such, mutagenesis is a crude, i.e., not directed, type of genome editing. This type of random mutagenesis has a disadvantage that next to the desired mutations, many more mutations will be formed, that may impair the fitness of the strains. The techniques described in this paragraph offer various possibilities for rational approaches to genome edits, e.g., the introduction of point mutations that lead to amino acid changes in a protein. CRISPR/Cas technology offers the possibility to make such precise genome edits. CRISPR/Cas is a complex between the Cas moiety, which is a nuclease (DNA cutting enzyme) that introduces double stranded (DS) breaks and a guide RNA (gRNA, or single guide RNA (sgRNA), a combination of the two RNA species of the original CRISPR system as it operates in bacteria) molecule that guides the nuclease towards a specific position on a DNA molecule. The Cas nuclease cuts in both strands, of a DNA molecule. The acronym CRISPR ('Clustered regularly-interspaced short palindromic repeats') refers to the original role of the enzyme system as a defence mechanism against invading, e.g., viral, nucleic acid in bacteria, but has no meaning in the CRISPR/Cas process described here. The CRISPR/Cas nuclease is guided towards specific sequences in the DNA, and makes its cut at very precise places in the DNA. The specificity is achieved by means of a guide RNA molecule that recognizes a 20 nucleotide stretch of DNA, based on homology between the RNA and the targeted DNA. A 20 nucleotide stretch occurs, in principle, only once in every 10^{12} (i.e., once in 4^{20}) base pairs. For comparison: the length of the human genome is 3.10^9 base pairs. But, it has been observed that perfect homology is not a prerequisite for the recognition of a target site by the CRISPR/Cas nuclease. There are ways to improve the fidelity of the CRISPR/Cas system (see below). Cas, the 'CRISPR associated' nuclease, carries two domains responsible for making cuts in the DNA, one for each DNA strand. The nuclease can be converted by mutation of one or both of these domains, into a nuclease that only cuts one strand, or does not cut at all, but still associates with the target site in the genome. These properties can be used in modifications of the CRISPR/Cas process. A DS break in a genomic DNA molecule has lethal consequences for a cell. Therefore living organisms have various repair processes to join the ends of DS breaks. These repair processes are, mostly, error prone and lead to changes in the DNA sequence at the position where the break has occurred.

Technical description

The CRISPR/Cas9 system (e.g., Cong et al., 2013; Sander and Joung, 2014) makes use of the Cas9 nuclease that causes double stranded (DS) DNA breaks. The cleaving activity of the nuclease can be guided to specific locations in a genome by adding a guide RNA (gRNA). The gRNA typically contains 20 base pairs (bp) (the protospacer) that is homologous to the targeted site in the genome; the genome target should be adjacent to a PAM (Protospacer Adjacent Motif, i.e., NGG[3] for Cas9).

In yeast, the Cas9 nuclease can be expressed constitutively without damage to the cell, while the gRNA can best be expressed transitionally (DiCarlo et al., 2013). Repair by non-homologous end joining (NHEJ) of the DS break may lead to mistakes in the sequence of the targeted gene and inactivation of gene activity. The repair can be done by homologous DNA recombination (HDR), which is a rather efficient process if a donor DNA homologous to the DS ends at the break is present in the cell. HDR will supplant the resident DNA sequence at the recognition site by the added DNA sequence, which may carry small modifications, e.g. single nucleotide changes of small deletion, but also larger DNA insertions such as complete or truncated genes.

Impact of the technology; off-target effects

CRISPR/Cas9 as a genome editing tool is typically used in eukaryotic cells (e.g., DiCarlo et al., 2013), although prokaryotic cells may also be targeted (Jiang et al., 2013, Mougiakos et al., 2016). The application of CRISPR/Cas9 can have the following target effects (Ran et al.,2013):

- NHEJ is error prone and will lead to small deletions and insertions (indels) at the DS break, resulting in gene knockout when the break is targeted to an exon.
- HDR using an added donor DNA will lead to insertion of the sequence of the donor DNA at the DS break.
- Multiple CRISPR/Cas9 species with different gRNA can be used to achieve different edits at different, specific locations in the genome at the same time.

Off-target effects occur due to the fact that no perfect homology is necessary for the gRNA to interact with a DNA molecule. Hence a combination of CRISPR/Cas9 and a gRNA may recognize DNA sequences at other locations than the target site. Off target effects will comprise the same types of genome edits as the target effects. Off-target effects due to NHEJ are much like other spontaneous or chemically or radiation induced mutations, or spontaneous rearrangements of DNA. Off-target effects due to HDR lead to precise small edits or insertion of larger sequences of the added donor DNA, which would lead to increase of copy number of the inserted sequence. There are several ways to increase the fidelity of the site recognition of the CRISPR/Cas9 system, e.g., the use of two different gRNAs, or truncating of the 20 nucleotide rRNA (Ma et al., 2014; Fu et al., 2014).[4]

[3] Any nucleotide followed by two guanine ("G") nucleotides

[4] The use of CRISPR/Cas in 'gene drives' is an application that is considered for organisms that reproduce sexually, with the purpose of transmitting certain trait through an entire population. This particular application does not appear to have use in white biotechnology.

Application areas in white biotechnology

Current applications of CRISPR/Cas9 for strain improvement are mainly the production of mutations, either random indels or specific edits, at the specific location to which the Cas9 nuclease is directed. Other applications of mutant Cas9 protein in modification of gene regulation are discussed in paragraph 5.3. Application areas are the same as the application areas of chemically or radiation induced mutations, or the application areas of genetic modification by other GM techniques. The main difference, and advantage, compared to other techniques is the precision of the CRISPR/Cas9 method and the relative ease of use. The major advantage of the method is that more than one edit can be made in one go, using combinations of differently targeted Cas nucleases (Mans et al., 2015).

Barriers and drivers

CRISPR/Cas9 genome editing is a technique that allows for, in principle, very precise modifications of target genes. Most important, multiple changes can be induced in the one and the same round of CRISPR/Cas9 activity. The technique is rapidly becoming available, and appears to supersede the more traditional techniques for GM, and other editing techniques like ODM (Oligonucleotide Directed Mutagenesis), and the use of zinc finger nucleases or TALEN (Transcription activator-like effector nuclease). One driver may be the (expected) regulatory status as non-GM, of some types of products produced by the technique, e.g., the small indels.

At the horizon

The adoption of the technique for genome editing of eukaryotic organisms (yeasts, fungi) is ongoing and may be expected within the next 5 years. Although CRISPR/Cas9 can be applied to prokaryotes too, it is pointed out by some of the interviewees that the already available techniques for prokaryotes will probably be sufficient. The technique may therefore not become as prominent for prokaryotes. Further development of the CRISPR/Cas9 technology is mainly expected to occur in the field of therapeutic genome engineering (Hsu et al., 2014). These studies will yield more insight, for instance into the specificity and off-target effects of the technology (Hsu et al., 2014). These results will probably have an impact on the use of CRISPR/Cas9 in white biotechnology, too. A dead Cas (dCas) protein can be turned into a versatile tool for DNA interaction with DNA at a precise location. The Cas protein can be linked to effector domains, like a transcriptional activator domain, that can be used for regulatory purposes. For an overview of non-nuclease applications of CRISPR/Cas9 systems see Sander and Joung, 2014 and Gilbert et al. (2013).

5.2 'Next-' (and 'third') generation sequencing techniques

General description

The ease and scale of DNA sequencing has revolutionized our understanding of biology. From the knowledge of a DNA sequence predictions can be made of the gene products it (may) encode(s). But it can also be used for checking whether a change of a DNA sequence has been introduced into the genome in the way that was expected in genetic modification, and for elucidating mutations that have been introduced into a genome. The usefulness of a DNA sequencing technique depends on its speed and reliability, its ease of use and, of course, on the costs involved. 'Next generation sequencing techniques' (NGS) such as Illumina sequencing typically perform sequence analysis of very many (typically several millions) of DNA fragments per run, at a run time of hours to a few days (see, e.g., Reis-Filho, 2009). The length of the DNA molecules that can be read is however rather limited, typically 250. In 'third generation' sequencing the sequence of one single DNA molecule can be determined, at a speed in the order of 5 µsec per base, and read length of typically several 1,000 of bases. NGS and 3rd generation techniques can be used for direct sequencing of any DNA samples, without the need for generating a gene library by cloning of the DNA. 3rd generation sequencing (or single molecule sequencing) techniques may be refined to even see modifications, e.g., methylation, of bases in a genome. These novel sequencing techniques have become very affordable. The sequencing of the genome of a particular microbial strain can be done fast and relatively cheap (the '$ 1.000 genome'). This has made the checking of a genome of a particular strain for mutations and modification possible on a routine basis. The techniques have also been very successful in elucidating the complex genomic structure of unknown microbial communities, such as the 'metagenome' (Handelsman et al., 1998) of soil microbial communities. Metagenomes are seen as an important source of potentially interesting genes that function in diverse metabolic pathways. As the individual organisms in a metagenome cannot be cultured and therefore are basically unknown, DNA sequencing techniques are the only practical approach to get to know the metabolic functions in these organism, and the features of speed and cost effectiveness of NGS and 3rd generation techniques

Technical description

NGS methods have been designed to obtain vast amounts of DNA sequence data from any source of DNA (Mardis, 2013). Basically, the methods allow for separation of single DNA fragments, amplification of the fragments and determining the sequencing of each amplified fragment separately. NGS techniques can also be used for transcriptome (RNA) analysis (Mutz et al., 2013). Several NGS techniques, or 'massively parallel sequencing' methods are available. One process that is typically used for (meta)genome analysis is the Illumina sequencing technology[5]. The method relies on terminal addition of different oligonucleotides on each end of a ss DNA molecule, annealing of the single stranded DNA molecule to a complementary oligonucleotide that is sitting on a solid platform, amplification of each annealed DNA fragment to a group of amplification products closely clustered on the platform. The

[5] http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

groups will be separated from each other on the platform so that the subsequent reactions can be monitored photographically. The amplification products are turned into ss DNA molecules, that can be sequenced, first in one direction and subsequently in the reverse direction. Sequences are determined by stepwise addition of fluorescently labelled bases by DNA polymerase action of DNA polymerase. The typical length of one read is 250 nucleotides. The read sequences are assembled into larger arrays based on homology.
Third generation sequencing methods have been designed to determine DNA sequences based on direct observation of a DNA molecule. Of the various platforms, nanopore sequencing methods are successfully proving their value (Laszlo, 2014). The Oxford nanopore method, for instance, relies on a DNA molecule moving through a solid state pore or a bacterial membrane pore protein of nanometer pore size. During its passage through the pore there will be a change of the electric conductivity of the pore. Each base causes its own typical change pattern during its passage. At the moment, the nanopore devices produce sequences of thousands of bases in a single read. But the techniques are evolving rapidly, and production of data at more than a terabase per day are within reach (see an announcement on genomeweb[6]). An advantage of long reads typically produced by these methods is that they can potential resolve allelic variation in diploid, polyploid and aneuploid organisms.

Impact of the technology

If predictions turn out right, the technology will yield unprecedented amounts of data, that can be of immense use, but only if the computing power is available to do the required analyses. Bioinformatics techniques have been developed in the past, and can be adjusted to handling these big data (e.g., Miller et al., 2010, DePristo et al., 2011). But even on a more modest scale of, for instance, sequence data of a metagenome, the impact could be enormous. For a full benefit of the data, bioinformatic methods must be available, and their predictive power should be understood. The field of bioinformatics has developed into a science of its own, and is beyond the scope of this report. Suffice it to state here that ample techniques and experience are available. Potential host effects and off-target effects depend on the use of sequence information. Some of the application areas are summarized in the next paragraph. For these applications it is crucial that the sequence information is correct. Incorrect information would lead to wrong conclusions, and wrong predictions for the metabolic functioning of genes. The error rate of Illumina sequencing is in the order of 0.1% (Glenn, 2011; see also the '2014 field guide'[7]). The error rate of nanopore sequencing techniques is still high, and can be in the order of 10% but efforts are ongoing to reduce the error rate (Li et al., 2016).

[6] https://www.genomeweb.com/sequencing/oxford-nanopore-presents-details-new-high-throughput-sequencer-improvements-mini
[7] http://www.molecularecologist.com/next-gen-fieldguide-2014/

Application areas in white biotechnology

Some application areas in of sequence information in white biotechnology are the following fields:

- To plan the construction of genetically enhanced strains;
- Verification and checking for off-target effects, of the organisms resulting from strain improvement, for instance to check whether a genetic modification has led to the intended result;
- To establish the result of a genetic improvement step, e.g., to find which mutations have occurred in directed evolution;
- To establish the sequence of new genes or new metabolic pathways that could be used in a production organism.

Barriers and drivers

Drivers are low cost and the speed of sequencing, combined with the enormous amount of useful information that the techniques yield. On the short term a barrier is the error rate of 3rd generation sequencing (Li et al., 2016), but this barrier will probably be overcome soon. Also, the enormous load of data that is expected to be produced can be prohibitive for making optimal use of the data.

At the horizon

Analysis of the available sequence data will yield a continuously increasing understanding of how organisms work. The NGS techniques are available for use, and will continue to be used in the near future. They will however probably be superseded by the 3rd generation techniques, depending on how fast the reliability problems will be solved.

5.3 DNA building blocks: applications of synthetic biology

General description

The planning of the design and construction of enhanced production strains will start from a certain phenotype that is desired for the activity of the strain in a biotechnological process. The planning will be inspired by insights that have been acquired in various studies, e.g., in (meta)genomics, proteomics, metabolomics and regulomics studies, and from the results of metabolic modeling. All these studies will yield hypotheses for the design of modified resident genes or new artificial genes that accomplish (a step in) the realization of the required phenotype. Also the regulation of the newly designed (set of) gene(s) should be optimized for the phenotype. The original regulation of a resident gene may not be efficient for the biotechnological process, because of interference of other regulatory processes in the regulome of the cell. To overcome these problems, orthogonal genetic switches are designed, i.e. switches that do not 'cross-talk' with other regulatory signals in the cell. In synthetic biology, many logic switches have been designed already, that mimic switches as they are used in electronic systems (but see Kwok (2010) about the complexity and compatibility of the interplay between biological parts). The combination of a synthetic gene together with its regulatory signals can be considered as a DNA building block. The design of these DNA building blocks is a domain of synthetic biology. '*One of the main goals in Synthetic Biology is to assess the feasibility of building novel biological systems from interchangeable and standardized parts*' (Rouilly et al., 2007). The concept of BioBricks is one solution for this goal. BioBricks is a trade mark of the BioBricks Foundation, and BioBrick parts must conform to the established standard of the trade mark. Of course, similar parts can be made, and custom designed for integration into the genome of choice, without conforming to the BioBrick standard. Another important player in this field is iGEM, the International Genetically Engineered Machine, who run a registry of standard biological parts and a repository of these parts. The newly (re)modeled genes can be seen as building blocks, that can be arranged, in the way described in paragraph 5.4 on DNA assembly.

Technical description

Synthetic biology embodies the idea that the machinery that is at the basis of life processes can be designed in a way similar to mechanical engineering. It is defined as "the application of science, technology and engineering to facilitate and accelerate the design, manufacture and/or modification of genetic materials in living organisms" (SCHER, SCENIHR, SCCS, 2014). It is characterized as the "expanding toolbox for industrial biotechnology", and comprises protein engineering, metabolic engineering, '-omics' approaches and the *in silico* approaches of bioinformatics, and the toolbox of synthetic biology (Tang and Zhao, 2009). Santos et al. (2011) point out that the high throughput techniques used in this area yield a volume and complexity of data that need to be mined, interpreted, and need forms of modelling from the biologist's perspective. They provide a framework for modelling approaches.

An important aspect in metabolic modelling is the regulation of metabolic pathways. The regulation of resident pathways is fine tuned to the total regulatory ('regulomics') network

of the cell. This regulation may be engineered, for instance by specific interfering processes, e.g., RNA interference (RNAi, see for instance Tomer, M. et al., 2011, Qi et al., 2013), or CRISPR interference (CRISPRi, see for instance Qui et al., 2013). But the regulation of a (set of) gene(s) may be more fundamentally engineered by uncoupling it from the regulatory network of the cell. For this purpose, 'orthogonal' synthetic genetic switches are being designed (Brophy and Voigt, 2014). Orthogonality implies that the newly added parts and modules of a pathway should not cross-talk with each other in the engineered biological systems as well as the host genetic background.

"One of the main goals in Synthetic Biology is to assess the feasibility of building novel biological systems from interchangeable and standardized parts" (Rouilly et al., 2007). The development of such standardized parts is the subject of the remainder of this paragraph. Mechanical engineering makes use of standard mechanical parts, that can be put together to build a part of a machine. In a similar way, the individual genes, together with their regulatory elements, that have to be assembles into the machinery of a metabolic pathway, can be seen as standard parts. The idea of using standard parts that can be used in the construction of machinery has been worked out in the design of building blocks, like BioBricks™, BglBricks (Anderson et al., 2010), and the more recent Golden Gate cloning parts. A large number of building parts is described in registries (the iGEM registry[8], has some 20,000 parts). An example that provides a good impression of the potential force of the use of standard biological parts can be found at the page of the iGEM team of Aalto University[9]. The team's goal was to create an *E. coli* strain that produces propane from cellulose, as a renewable fuel. They used a synthetic pathway, described by Kallio et al. (2014), that brings together 10 enzymes from different organisms, for the enzymatic steps that convert glucose into propane. The feasibility of the pathway was checked by modelling the enzyme kinetics. The genes encoding the enzymes were assembled into two operons on two plasmids, under an inducible promoter. A third plasmid was put together that encodes three genes coding for secreted enzymes that degrade cellulose into glucose. These three genes were available as standard biological parts from the iGEM registry. To enhance the efficiency of the propane pathway, the enzymes responsible for the last two steps were fused to micelle forming proteins, so that they would be brought into close proximity, which enhances the speed of the reactions.

The salient features of this approach are:

- The use of enzymes from different organisms that can work together in a synthetic pathway to perform the requested process[10].
- A modelling step, to check and confirm the suitability of the chosen enzymes[11].
- Further fine tuning of two of the enzymes (i.e., into micelle forming proteins).
- The design of the separate genes into building blocks that can be assembled (see next paragraph) into synthetic autonomously replicating DNA.

[8] http://parts.igem.org/Main_Page
[9] http://2015.igem.org/Team:Aalto-Helsinki
[10] http://2015.igem.org/Team:Aalto-Helsinki/Project
[11] http://2015.igem.org/Team:Aalto-Helsinki/Kinetics

Impact of the technology

The technology enhances the ease of combining various 'parts', i.e., enzymes flanked by regulatory units, that have already been characterized in other studies, and that have proven their use. This will mainly facilitate the design of biological processes. The impact will therefore be not only on the fundamental possibilities of the design of production strains, but also on the speed with which genetic changes can be introduced into a strain.

Application areas in white biotechnology

The usefulness of the DNA constructs that are produced by these techniques depends on the possibilities to predict which (variants of) genes should be put together in order to introduce a desired process into a production strain. There are no restrictions to the use of the DNA constructs: any organism that is amenable to genetic modification can be used as a host of the DNA constructs that are produced by these techniques.

Barriers and drivers

A barrier to the application of the techniques is the requirement for thorough knowledge of the functioning of the biological parts. The techniques are not specifically suited to trial-and-error approaches, such as discussed in the next paragraph. The reverse consideration is the main driver for the use of the techniques: the possibility to rationally approach the design of more efficient production strains. How these approaches fit in the concepts of biotechnological companies is discussed further in chapter 6.

At the horizon

The use of DNA building blocks with synthetic genes and regulatory switches appears to be established already. What can be expected is the further design of new pathways for various biotechnological purposes, such as the example of the iGEM team of Aalto University, described above. The list of current projects of BBSRC (see Appendix 2) provides insight into the areas that are currently interesting for R&D. Which of these and similar projects will adopted by the white biotechnological industry remains to be seen (see discussion in chapter 6). It is hard to put a time frame on these developments, but the use of production strains produced by these techniques could certainly occur within 5 years, but it may take longer for industry to adopt these techniques in practice. The development of synthetic genetic elements is one possibility for 'alternative life': for instance the use of orthogonal regulatory switches that have no interaction with the resident regulation in the cell, or the introduction of amino acids that do not naturally occur in proteins.

5.4 Techniques for DNA assembly

General description
The small DNA building blocks described in the previous paragraph will have to be assembled into larger units, especially if a metabolic pathway, consisting of several gens, has to be introduced into a production strain. There is a trend build large replicable DNA molecules, such as artificial chromosomes, from smaller genetic building blocks, in order to achieve the most effective set of genes in the artificial chromosome of an organism, i.e., a minimal set of genes without superfluous gene that are not necessary for the basic functions of the organism under industrial conditions. The specific architecture of such chromosomes is a determining factor for their efficient functioning. DNA assembly, that is necessary to build this architecture, is perceived as a limiting technology in the advancement of synthetic biology. This has led to the development of different assembly methods that to obtain better functioning constructs. (Ellis et al., 2011).

Technical description

Available techniques for assembling DNA fragments into large, typically over 100 kb, constructs all rely on homology between the ends of the DNA fragments which will guide the DNA assembly. Examples of these techniques are Yeast co-transformation (e.g., Gibson et al., 2008), circular polymerase extension cloning (CPEC) (e.g., Quan and Tian, 2009), Gibson isothermal assembly (Gibson et al., 2009) and ligase cycling reaction (LCR) (e.g., Kok et al., 2014). In Yeast co-transformation, the DNA assembly will occur through homologous recombination *in vivo*. This technique leads to assembly and integration of DNA constructs at a selected site on the genome (Kuijpers et al., 2013). CPEC relies on annealing of homologous ss DNA ends and filling of gaps by DNA polymerase and ligase. In the Gibson isothermal assembly, the homologous DNA ends are turned into ss overhangs that can then be annealed and covalently joined. In LCR, ss DNA that is to be assembled is annealed to small ss DNA fragments homologous to the 3' end of one DNA fragment and the 5'end of the next fragment. After ligation of the two large DNA fragments, their homologous DNA strands are annealed and ligated. Golden Gate cloning allows assembling up to nine fragments at a time in a recipient plasmid. Cloning is performed *in vitro*, by mixing the various plasmid donors, the recipient vector, a type IIS restriction enzyme and DNA ligase. The cloning requires thermal cycling. The method is simple, but yields mostly the expected construct depending on the care that is taken in the design of the constructs. (Engler and Marillonnet, 2013).

Impact of the technology

The various DNA assembly techniques lead in principle to a DNA molecule (plasmid, artificial chromosome) that can be replicated in a host cell, depending on the type of origin of replication that is included. Depending on the experimental approach, the technique may aim to reassemble a previously known chromosome, but may also be used for automated high-throughput assembly of DNA parts into DNA constructs (e.g., Kok et al., 2014). The techniques are error prone, and may lead to other than the predicted assembly products, and

also to ‘mistakes’, like single-nucleotide polymorphisms and insertions or deletions. Optimization of the experimental conditions, e.g., choice homologous ends and hybridization conditions can optimize the fidelity of the assembly (Kok et al., 2014).

Application areas in white biotechnology

DNA assembly has been called a ‘key part of most metabolic engineering projects’ (Merryman and Gibson, 2012). Clearly, the value of these techniques lies in the possibility to merge different genetic building blocks into larger units (genomes, chromosomes). The techniques are a necessary step in the development of minimal genomes.

Barriers and drivers

A barrier used to be the cost of DNA synthesis, but currently prices for DNA synthesis are priced on the internet at $0,10 to $0,20 per bp. Drivers are the possibilities for aligning genes that would be scattered in different places on the genome and constructing optimally organised minimal genomes for 'minimal organisms'. The merits of minimal genomes, and specifically minimal genomes constructed by DNA assembly, are however disputed (see chapter 6; for a discussion on the merits of minimal genomes see Choe et al., 2016).

At the horizon

DNA assembly techniques are used to make constructs for GM, and this use will continue. The use of artificial chromosomes may be foreseen for the next five years. The techniques allow the development of entire genomes for organisms. The construction of minimal organisms can be done at the moment (see Hutchison III et al., 2016). This would allow for the construction of organisms whose genomes copy (part of) existing genomes, containing the necessary minimum of the ‘household genes’ for the organism, with other ‘custom designed’ genes for specific metabolic processes and pathways added. In this way a ‘synthetic’, ‘minimal’ organism may be designed, e.g., the ‘synthetic yeast’ Sc2.0 (Dymond et al., 2011; see also Synthetic yeast 2.0[12]). The practical use of these organisms is not foreseen for the next 10 years.

[12] http://syntheticyeast.org/

5.5 Adaptive laboratory evolution and directed evolution

General description

The previous paragraphs in this chapter focus on techniques that, in various ways, allow the rational design of production strains. These techniques require extensive knowledge of and experience with various facets of molecular biology, such as the '-omics' approaches. It was pointed out by several interviewees that traditional microbiological techniques for strain improvement are still very actual. Traditional techniques for evolutionary strain improvement are known as Adaptive Laboratory Evolution (ALE, for an overview see Dragosits and Mattanovich, 2013). ALE is based on spontaneous or induced mutations. Successful mutants will outcompete less successful variants. The approach by mutation and selection has been made more focused by the use of directed evolution. In directed evolution mutations are introduced into (a) selected gene(s), followed by expression of the mutant gene and selection of the mutant with the desired phenotype. ALE and directed evolution typically produce large number of mutants that have to be screened and analysed for favourable strains. This requires techniques for high throughput screening, biosensing and cell sorting to harvest the cells with the desired phenotype.

Technical description

There are various ways to perform ALE in the laboratory (Dragosits and Mattanovich, 2013). Generally, techniques involve either repeated subculturing in batch cultures, or continuous culture in chemostats. In batch cultures the conditions will change during growth, and selective pressure is not easily controlled and will fluctuate. Chemostat culturing offers more possibilities to manipulate growth conditions. Evolution experiments take many, typically hundreds or several thousands, generations (see tables 1 and 2 in Dragosits and Mattanovich, 2013), and consequently long time. Like ALE, directed evolution depends on selective pressure. In directed evolution, biological variety is created, *in vivo*, for instance by the use of mutator strains in combination with a bacteriophage carrying the gene(s) of interest, or *in vitro*, for instance by performing error prone PCR on the gene(s) of interest or even more advanced approaches like computational strategies (see table 1 in Packer and Liu, 2015). This library design, the size of the libraries that can be prepared and the rate at which screening or selection can be carried out are decisive steps in directed evolution (Cobb et al., 2010). Large numbers of cells carrying the mutagenized genes can be screened by high throughput methods, for instance by optic screening of compartmentalized single cells in combination with fluorescent biosensors (see table 2 in Packer and Liu, 2015). Cells with promising phenotype can be isolated by cell sorting. As it is intrinsic to the techniques that the genetic changes in the resulting organism are unknown, the final strain will need to be further characterized. Full genome sequencing of both the parental strain as the resulting strain will be an adequate and practically feasible method for the characterization.

Impact of the technology

Contrary to the techniques for strain development described in the previous paragraphs of this chapter, directed evolution is used to explore the possibilities of the genetic variation

that is offered by the organism itself. It is seen by some (see chapter 6) as an ultimate resort if other approaches fail. In the literature it is described as a powerful tool, and the approach is used in several BBSRC projects (see Appendix 2).

Application areas in white biotechnology

In principle the techniques can be used for any production strain, and for any process in white biotechnology. Portnoy et al. (2011) mention *inter alia*: activation of latent pathways to utilize non-native substrates or produce non-native products, improving the arsenal and rate of utilisable substrates, and adaptation to specific conditions in an industrial process.

Barriers and drivers

There may be technical and financial barriers, as the production of gene libraries and the efforts for expression of the new genetic products and the high throughput analysis require specialist technical equipment. The resulting strain may need to be further optimized for large scale production. As a consequence the practical realization of directed evolution is expected to be quite cost intensive.

At the horizon

Packer and Liu (2015) mention "*ambitious goals such as reprogrammed substrate selectivity and synthetically useful biocatalysts as benefits from innovative screens and selections that balance the need for throughput and accurate assessments of library members. New screens and selections that achieve higher throughput or carry out more continuous rounds of evolution can broaden the exploration of the fitness landscape, whereas novel mutagenesis strategies increase the search efficiency. Through computational techniques and creative molecular biology protocols, diversity is focused on residues and specific mutations that influence desired activities. New directed evolution methods will continue to generate proteins with useful new activities and specificities, as well as expand the scope of protein evolution to include even larger sets of chemical and biological functions.*"

6 Discussion and conclusion

"The fundamental force that drives the development and implementation of industrial biotechnology is the market economy, as biotechnology promises highly efficient processes at lower operating and capital expenditures. In addition, political and societal demands for sustainability and environment- friendly industrial production systems, coupled with the depletion of crude oil reserves, and a growing world demand for raw materials and energy, will continue to drive this trend forward" (Tang and Zhao, 2009).

White biotechnology is evolving rapidly, because of the possibilities offered by new techniques and approaches. These can be used for enhancing the efficiency of metabolic processes that are already in use, or to devise novel metabolic processes for the production of a large variety of biological compounds. There is a clear distinction between the expectations of scientists who are developing new techniques and using them for specific biotechnological purposes, and companies active in white biotechnology that are inclined to only use new techniques if a business case can be construed that leads to a clear advantage. In industrial biotechnology techniques are chosen based on the requirements for the process and its cost effectiveness, not on the basis of the mere availability of the technique.

The wide variety of projects that are interesting to the scientific community can be derived from the list of projects that are being sponsored by the BBSRC (see Appendix 2). It should be remembered that these projects are functioning in the realm of knowledge development, i.e., at an early stage on the roadmap towards commercialisation of biotechnological innovations. Companies in white biotechnology are much more focused on their specific production processes, with which they have ample experience, also in terms of the investments necessary to optimise process when technological innovations are brought in. They will balance these investments necessary for adoption of a new technique against the chances of success and the expected return on investment. Some of the interviewees were rather cautious in their answers with regards to the use of new techniques, while others expect implementation of techniques to occur faster. Hence, it is difficult to formulate clear predictions about the actual use of the techniques within the next five years, or within 5-9 years.

Against this background, specific comments can be made about the techniques dealt with in chapter 5:

The use of CRISPR/Cas9 for genome editing is widely seen as a very important development, although the interviewees expressed some reservations. Especially for prokaryotes there are other techniques available that have an excellent track record for making precise genome edits. However, the use of CRISPR/Cas9 to make several precise changes in one action is mentioned as a clear major advantage of the technique. Non-cutting mutants of Cas proteins (dead or dCas) can be used for RNA-guided regulation of transcription in yeast and other eukaryotes, or prokaryotes. This possibility was not specifically commented on by the interviewees. The optimal use of RNA guided nucleases requires careful design of the gRNA. This requires the availability of the complete genome sequence and extensive bioinformatics studies, and probably some trial-and-error experimenting. All prerequisites for the practical

use of CRISPR/Cas9 and related approaches are met, or under advanced study. The application of the technique can therefore occur already within the next five years. The actual application in industrial biotechnology depends on whether and what advantages are perceived by the biotechnological companies. The (potential) regulatory status of CRISPR/Cas9 products as non-GM is recognized by the interviewees, but they comment that regulatory status is not a very much an issue as they are working under contained use. One interviewee, operating in the food market, drew attention to negative connotations of some uses of CRISPR/Cas, e.g., for gene drive. Of course, this use is not actual in white biotechnology, but negative feelings appear to exist among part of the general public.

'Next' and in particular the 'third' generation sequencing techniques are likely to cause a revolution in the availability of sequence data. Most importantly, the third generation sequencing techniques can provide information of the native state of DNA sequences, e.g., the methylation patterns. For white biotechnology the most important role for DNA sequencing will be the possibility to characterise and validate modified strains by whole genome sequencing. At same time, the availability of whole genome sequences may raise unexpected questions, for instance if (putative) genes are found that are not reflected in the phenotype. An example are the occurrence of (silent) toxin genes in an organism that has a long history of safe use, or is 'generally regarded as safe' (GRAS). The application of NGS techniques is already current and can be expected to increase within the next five years, because of low cost and the speed at which data can be obtained. Third generation sequencing is still rather error prone. But these problems may be surmounted within the next years. The impact of the technique expected to be extensive, at least in the course of the next 5-9 years. The main problem that can be expected is the computing power that is needed for storage, and the availability of sequence data for bioinformatics analysis.

The use of DNA building blocks has been discussed above against the background of applications of synthetic biology. The power of the technique is in the standardisation of 'biological parts' that can be used for putting together metabolic pathways from the most eligible genes from different sources. The possibilities of synthetic biology have led to a burst of creativity, as is shown in the iGEM contest, for example the project of Aalto University, shown above. The practical application of DNA building blocks is already ongoing in scientific projects. The impression is that industry will make use of building blocks, but not necessarily in the standardised form. One interviewee pointed out that the use of genes from other sources may need more adjustments in the organisms, for instance in the availability of cofactors, e.g., NADH/NADPH. In general, there is a notion that the products of applications of DNA building blocks will need extensive fine tuning for optimal functioning under fermentation conditions. The use of genes from other sources and the use of synthetic parts, especially synthetic regulatory units, promoters and genetic switches, is actual in scientific projects. Their use in industry is already occurring, and there is a wish for better understanding of regulatory units. Industrial applications are expected at the longer term of 5-9 years.

Techniques for DNA assembly are used in scientific projects. A good understanding of the rules for optimal DNA architecture is needed for making optimal assembly products. This will lead to the production of artificial chromosomes which is a step towards minimal genomes

for 'minimal organisms'. From a scientific point of view the construction of a 'minimal' bacterium is an important feat (see the recently reported synthesis of the synthetic minimal genome of *Mycoplasma mycoides*, Hutchison III et al., 2016). From the point of view of applied white biotechnology the synthesis and assembly of a minimal genome would be very costly, at least in the foreseeable future (see for instance the number of people involved in the development of *M. mycoides*, Hutchison III et al., 2016, or the development of Sc2.0[13]). The development of reducing complexity in a genome is seen as important, for instance in the development of yeast strains, where there are examples of multiple genes for one function. On the other hand, it was also mentioned that redundancy can be of value for an organism, and even cannot be lost without negative consequences. Genetic variation is of importance for (directed) evolution. Techniques for DNA assembly are being used already in scientific projects. Their use for creating large DNA arrays like artificial chromosomes is expected on the longer term, 5-9 years. The use of minimal organisms is foreseen only at a term of > 10 years.

Adaptive laboratory evolution and directed evolution.
All interviewees drew attention to the importance of 'traditional' techniques in microbiology for strain improvement. ALE, mutation and selection (or election) of desired phenotypes is still a powerful technique. Directed evolution can be done in a more focused manner. Various approaches are available. Packer and Liu (2015) describe techniques that can be used *in vivo* or *in vitro* for creating diversity in gene libraries, which allow for a more or less focused approach. The more traditional evolutionary approaches are still in use and seen as very important by industry. Directed evolution is also mentioned as a very cost intensive approach.

Final remarks

This report presents an overview of techniques that are driving the developments in white biotechnology. It is a snapshot in time, largely based on a general literature survey, and on the basis of information obtained in interviews held with a selection of representatives from science and industry in white biotechnology. The results from the literature survey are based on a screening of papers published in the scientific journal *Current Opinion in Biotechnology*. This journal claims to help the reader by providing the views of experts on current advances in biotechnology, and evaluate the most interesting papers, annotated by experts, from the great wealth of original publications[14]. Attempts to review other literature sources for relevant papers on upcoming developments in white biotechnology did not produce the desired information, because none of the search keys tested yielded comprehensive and systematic results. Another useful resource was the list of the grants awarded by the BBSRC that were current. Appendix 2 presents a selection from this list of projects that focus on general developments in white biotechnology.

The direction into which white biotechnology will move is difficult to predict. One pragmatic example is the sustainable production of biofuel. This is a prominent topic in the BBSRC database, and a topic that speaks very much to the imagination given the discussion on bio-based economies. But the experts interviewed commented that for a realistic estimation of the feasibility of these projects one should take into account the oil price. At this moment the low

[13] http://syntheticyeast.org/
[14] http://www.journals.elsevier.com/current-opinion-in-biotechnology/

oil price is considered an economic barrier for the costly development of projects for sustainable fuel. This effect is not limited to biofuels; there is also a dwindling interest in the development of biotechnological processes for the production of other bulk or fine chemicals, which can also be produced from fossil fuel.

On the other hand it is clear that all techniques described in chapter 5 are actively being used at this moment, in public research, as is evident from Appendix 2, and developed for use in white biotechnology. It is hard to imagine that these developments will not have an impact in further stages of applied research and in commercial development. These applications will probably be in gene optimisation (genome editing, paragraph 5.1) and the use of building blocks for introducing new or enhanced metabolic pathways (paragraph 5.3), i.e., to produce relatively small additions, not entire synthetic chromosomes or whole synthetic organisms (paragraph 5.4). There is no doubt that the new sequencing techniques (paragraph 5.2) will be applied as soon as they are available, and that they will change our outlook on how to characterise (genetic modifications in) strains.

The techniques described in this report cause a paradigm change for risk assessment of the (GM) products of these techniques. Properties of newly produced organisms can be evaluated based on the properties of the parental organism and the properties of the genetic information that is introduced. These last properties can be assessed from the phenotypes of the strains from which the genetic information is derived. The new techniques (except the evolutionary techniques) create novel genes, and the resulting phenotypes have not been seen before. They can be predicted, for instance by the computational techniques of bioinformatics. The screening methods developed for detection of organisms with a useful phenotype, may also provide useful information for biosafety purposes, at an early stage of development of novel organisms.

7 References

Anderson, J.C. et al. (2010) "BglBricks: A flexible standard for biological part assembly" Journal of Biological Engineering 4:1 (doi: 10.1186/1754-1611-4-1)

Brophy, J.A.N. and Voigt, C.A. (2014) "Principles of genetic circuit design". Nature Methods 11, 508–520 (doi: 10.1038/nmeth.2926)

Choe, D. et al. (2016) "Minimal genome: Worthwhile or worthless efforts toward being smaller?". Biotechnol. J. 11, 199-211 (doi: 10.1002/biot.201400838)

Cobb, R.E. et al. (2012) "Directed evolution: an evolving and enabling synthetic biology tool". Current Opinion in Chemical Biology, 285–291 (doi: 10.1016/j.cbpa.2012.05.186)

Cong, L. (2013) "Multiplex genome engineering using CRISPR/Cas systems." Science 339, 819-23 (doi: 10.1126/science.1231143)

DePristo, M.A. et al. (2011) “A framework for variation discovery and genotyping using next-generation DNA sequencing data”. Nature Genetics 43, 491–498 (doi: 10.1038/ng.806)

DiCarlo, J.E. et al. (2013) “Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems”. Nucleic Acids Res.41, 4336-43. (doi: 10.1093/nar/gkt135)

Dragosits, M and Mattanovich, D. (2013) Microbial Cell Factories 12, 64-80 (http://www.microbialcellfactories.com/content/12/1/64)

Dymond, J.S., et al. (2011) “Synthetic chromosome arms function in yeast and generate phenotypic diversity by design” Nature 477, 471–476 (doi: 10.1038/nature10403)

Ellis, T. et al., (2011) “DNA assembly for synthetic biology: from parts to pathways and beyond”. Integr. Biol. 2109-18 (doi: 10.1039/c0ib00070a)

Engler, C. and Marillonnet, S. (2013) "Golden Gate Cloning". DNA Cloning and Assembly Methods 116, 119-131 (doi: 10.1007/978-1-62703-764-8)

EuropaBio (2011) Industrial or White biotechnology. A driver of sustainable growth in Europe. Working document from EuropaBio and ESAB to be used as input for the Industrial Biotechnology section of the European Technology Platform for Sustainable Chemistry. http://www.europabio.org/sites/default/files/report/industrial_or_white_biotechnology_-_a_driver_of_sustainable_growth_in_europe.pdf (accessed March 2016)

Fu, Y. et al. (2014) “Improving CRISPR-Cas nuclease specificity using truncated guide RNAs”. Nature Biotechnology 32, 279–284 (doi: 10.1038/nbt.2808)

Gibson, D.G. et al. (2008) “One step assembly of in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome”. Proc. Natl Acad. Sci. USA 105, 20404-20409 (doi: 10.1073/pnas.0811011106)

Gibson, D. et al. (2009) “Enzymatic assembly of DNA molecules up to several hundred kilobases”. Nature Methods 6, 343 – 345 (doi: 10.1038/nmeth.1318)

Gilbert, L.A. et al. (2013) "CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes". Cell 154, 442–451 (doi: 10.1016/j.cell.2013.06.044)

Glenn, T.C. (2011) “Field guide to next-generation DNA sequencers” Molecular Ecology Resources (2011) 11, 759–769 (doi: 10.1111/j.1755-0998.2011.03024.x)

Handelsman, J., et al. (1998) "Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products". Cell Chemical Biology **5**, R245–R249 (doi*:* 10.1016/S1074-5521(98)90108-9)

Hsu, P.D. et al. (2014) "Development and Applications of CRISPR-Cas9 for Genome Engineering". Cell 157, 1262-1278 (doi: 10.1016/j.cell.2014.05.010)

Hutchison III, C.A. et al. (2016) "Design and synthesis of a minimal bacterial genome". *Science* 351, Issue 6280 (doi: 10.1126/science.aad6253)

Jiang, W. et al. (2013) “RNA-guided editing of bacterial genomes using CRISPR-Cas systems”. Nature Biotechnology, 31, 233–239 (doi: 10.1038/nbt.2508)

Kallio, P. et al. (2014) "An engineered pathway for the biosynthesis of renewable propane". Nature Communications 5, article number 4731 (doi: 10.1038/ncomms5731)

Kok, de, S. et al. (2014) “Rapid and Reliable DNA Assembly via Ligase Cycling Reaction ACS” Synth. Biol. 3, 97−106 (doi: 10.1021/sb4001992)

Kuijpers, N.G. et al. (2013) "One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in Saccharomyces cerevisiae". FEMS Yeast Res. 13, 769-81 (doi: 10.1111/1567-1364.12087).

Kwok, R. (2010) “Five hard truths for synthetic biology” Nature 463, 288-290 (doi: 10.1038/463288a)

Laszlo, A.H. (2014) “Decoding long nanopore sequencing reads of natural DNA”. Nat Biotechnol. 2014 Aug; 32(8): 829–833 (doi: 10.1038/nbt.2950)

Li, C. et al. (2016) “INC-Seq: Accurate single molecule reads using nanopore sequencing“. BioRxiv preprint (Jan. 27, 2016, doi: http://dx.doi.org/10.1101/038042)

Ma, Y. et al. (2014) “Genome modification by CRISPR/Cas9”. FEBS Journal 281, 5186-5193 (doi: 10.1111/febs.13110)

Mans R. et al. (2015) "CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae". FEMS Yeast Res. 15 (doi: 10.1093/femsyr/fov004)

Mardis, E.R. (2013) “Next-Generation Sequencing Platforms”. Annual Review of Analytical Chemistry 6, 287-303 (doi: 10.1146/annurev-anchem-062012-092628)

Merryman, C. and Gibson, D. (2012) “Methods and applications for assembling large DNA constructs”. Metabolic Engineering 14, 196–204 (doi: 10.1016/j.ymben.2012.02.005)

Mougiakos, I. et al. (2016) "Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit". Trends Biotechnol. In press (doi: 10.1016/j.tibtech.2016.02.004)

Miller, J.R. et al. (2010) “Assembly algorithms for next-generation sequencing data”. Genomics 95, 315–327 (doi: 10.1016/j.ygeno.2010.03.001)

Mutz, K-O et al. (2013) “Transcriptome analysis using next-generation sequencing”. Current Opinion in Biotechnology, 24, 22-30 (doi: 10.1016/j.copbio.2012.09.004)

Novick, A. and Szilard, L. (1950) "Experiments with the chemostat on spontaneous mutations of bacteria". Proc. Natl. Acad. Sci. USA 36**,** 708-719

Packer, M.S. and Liu, D.R. (2015) "Methods for the directed evolution of proteins". Nature Reviews Genetics 16, 379-394 (doi: 10.1038/nrg3927)

Portnoy, V.A. et al. (2011) "Adaptive laboratory evolution - harnessing the power of biology for metabolic engineering". Nanobiotechnology and Systems Biology, 4, 590-594 (doi: 10.1016/j.copbio.2011.03.007)

Quan J and Tian J (2009) “Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways”. PLoS ONE 4, e6441. (doi: 10.1371/journal.pone.0006441)

Qi, L.S. et al. (2013) "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression". Cell 152, 1173–1183 (doi: 10.1016/j.cell.2013.02.022)

Ran, F.A. et al. (2013) “Genome engineering using the CRISPR/Cas9system”. Nature Protocols 8, 2281–2308 (doi: 10.1038/nprot.2013.143)

Reis-Filho, J.S. (2009) “Next-generation sequencing”. Breast Cancer Research 11 (Suppl 3): S12 (doi: 10.1186/bcr2431)

Rouilly, V. et al. (2007) “Registry of BioBricks models using CellML”. BMC Systems Biology 2007, 1 (Suppl 1), P79 (doi: 10.1186/1752-0509-1-S1-P79)

Sander, J.D. and Joung, J.K. (2014) “CRISPR-Cas systems for editing, regulating and targeting genomes”. Nature Biotechnology 32, 347–355 (doi: 10.1038/nbt.2842)

Santos, F. et al. (2011) “A Practical Guide to Genome-Scale Metabolic Models and Their Analysis”. Methods in Enzymology, Volume 500, 509-532 (doi: 10.1016/B978-0-12-385118-5.00024-4)

SCHER (Scientific Committee on Health and Environmental Risks), SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks), SCCS (Scientific Committee on Consumer Safety) (2014), "Opinion on Synthetic Biology I, Definition". http://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_044.pdf (accessed April 2016)

Schmidt, M. and De Lorenzo, V. (2016) “Synthetic bugs on the loose: containment options for deeply engineered (micro) organisms”. Current Opinion in Biotechnology 38, 90-96

Tang, W.L. and Zhao, H (2009) "Industrial biotechnology: Tools and applications". Biotechnol. J. 2009, 4, 1725–1739 (doi: 10.1002/biot.200900127)

Tomer, M. et al. (2011) "RNAi as a potential tool for biotechnological applications in fungi". Applied Microbiology and Biotechnology, 89, 501-512 (doi: 10.1007/10.1007/s00253-010-2928-1)

8 Appendices

Appendix 1: Overview techniques, application areas, barriers/drivers and horizon

Technique	Application areas	Barriers / drivers	Horizon
CRISPR/Cas9	Precise genome editing Knock-in of DNA fragments	Drivers: characteristics of the technique: - very precise modifications - multiple changes in one go - available, supersedes other techniques - regulatory status as non-GM Barrier: sufficient adequate traditional techniques available (prokaryotes)	1-5 y. (available now)
Next and 3rd generation sequencing	Fast generation of sequence information for - characterisation of genes - validation of genetic change	Drivers: - low cost, high speed, amount of information Barriers: - error rate of 3rd generation sequencing - management of the load of available data	NGS: 1-5 y. (available now) 3rd generation sequencing: 5-9 y., depending on finding solutions for the error rate
DNA building blocks	Fast and flexible application of results from synthetic biology	Driver: - making available the results of synthetic biology Barrier: - notion that the products of applications of DNA building blocks will need extensive fine tuning for optimal functioning under fermentation conditions	1-5 y. (available now)
DNA assembly	- assembly of building blocks - artificial chromosomes - minimal organisms	Drivers: - aligning genes on the genome - constructing artificial chromosomes - constructing 'minimal organisms'. Barrier: - products need extensive fine tuning	1-5 y. Artificial chromosomes: 5-10 y. Minimal organisms: > 10 y.
ALE and directed evolution	- making use of genetic variation - making use of traditional techniques	Drivers: - extension of available traditional techniques Barriers: - technical and financial barriers - products need extensive fine tuning	1-5 y. (already ongoing) But also mentioned as > 10 y.: cost intensive

Appendix 2: Selection of projects from the list of current projects of the BBSRC Awards list

Reference	Title	Investigator / Supervisor	Institution
BBS/E/F/00044440	The identification of yeast strains optimal for fermentation	Dr I Roberts	Institute of Food Research
BBS/E/W/10963A01D	Optimising energy output and biorefining	Professor Iain Donnison	Aberystwyth University
BB/J020427/1	Towards process development of bacterial strains able to convert renewables into biofuels and other useful chemical commodities	Professor Nigel Minton	University of Nottingham
BB/K00199X/1	Rapid Evolution of Enzymes and Synthetic Micro-organisms for the Development of Industrial Biocatalysts	Professor Nicholas Turner	The University of Manchester
BB/J00054X/1	Extending the Boundaries of Nucleic Acid Chemistry	Professor Andrew Turberfield	University of Oxford
BB/K002767/1	Synthetic approaches towards the production of biofuels from lignocellulosic feedstocks in yeast	Dr Mark Peter Ashe	The University of Manchester
BB/J01916X/1	A study of metagenomics informed biochemical functionality of microbial fuel cells using DDGS as a substrate	Dr Claudio Adolfo Avignone-Rossa	University of Surrey
BB/K014773/1	Development and Application of Next Generation Synthetic Biology Tools	Dr Neil Dixon	The University of Manchester
BBS/E/T/000GP016	Engineering Synthetic Microbial Communities for Biomethane Production	Dr David Swarbreck	The Genome Analysis Centre
BBS/E/W/10964A01D	Biomining the rumen for enzymes with industrial potential	Dr Alison Kingston-Smith	Aberystwyth University
BB/L003910/1	Cellulect: A Synthetic Biology Platform for the Optimization of Enzymatic Biomass Processing	Professor Alistair Elfick	University of Edinburgh
BB/K019791/1	The Sc2.0 UK Genome Engineering Resource (SUGER)	Dr Thomas Ellis	Imperial College London
BB/K006290/1	Genome Organisation for Optimising Synthetic Secondary Metabolism	Dr Thomas Ellis	Imperial College London
BB/L001284/1	SCILS - Systematic consideration of inhomogeneity at the large scale: towards a stringent development of industrial bioprocesses	Professor Chris Rielly	Loughborough University
BB/J001694/2	Extending the Boundaries of Nucleic Acid Chemistry	Professor Tom Brown	University of Oxford
BB/K00283X/1	GASCHEM: Optimising industrial gas fermentation for commercial low-carbon fuel & chemical production through systems and synthetic biology approaches	Professor Nigel Minton	University of Nottingham
BB/K003240/2	Engineering Synthetic Microbial Communities for Biomethane Production	Professor Orkun Soyer	University of Warwick
BB/K003356/1	A platform for rapid and precise DNA module rearrangements in Synthetic Biology	Professor Marshall Stark	University of Glasgow
BB/K011138/1	Development of an integrated continuous process for recombinant protein production using Pichia pastoris	Professor Stephen Oliver	University of Cambridge
BB/K017373/1	Towards predictive biology: using stress responses in a bacterial pathogen to link molecular state to phenotype	Professor Martin Woodward	University of Reading
BB/K019171/1	Towards predictive biology: using stress responses in a bacterial pathogen to link molecular state to phenotype.	Dr Peter Lund	University of Birmingham

Reference	Title	Investigator / Supervisor	Institution
BB/K019783/1	Continued development of ChEBI towards better usability for the systems biology and metabolic modelling community	Professor Pedro Mendes	The University of Manchester
BB/K020633/1	Sustainable bioenergy from microalgae: A systems perspective	Dr Seetharaman Vaidyanathan	University of Sheffield
BB/K016288/1	Design and construction of electrogenic cell-based biosensors for pathogens and toxins	Professor Martin Buck	Imperial College London
BB/L010798/1	Synthetic Biology for Biotechnology of Fine Chemicals – SynBioTech	Professor Nigel Scrutton	The University of Manchester
BB/L02683X/1	Synthetic Biology for Bioenergy and Biotechnology	Dr Claudio Adolfo Avignone-Rossa	University of Surrey
BB/L013789/1	PHYCONET: unlocking the IB potential of microalgae	Professor Saul Purton	University College London
BB/L013649/1	Network in Biocatalyst Discovery, Development and Scale-Up	Professor Nicholas Turner	The University of Manchester
BB/L007444/1	Metagenomics for new enzyme discovery and industrial biocatalysis	Professor John Ward	University College London
BB/L013754/1	Natural products discovery and bioengineering network (npronet)	Professor Jason Micklefield	The University of Manchester
BB/L020130/1	Advancing Microbial Electrochemistry: Biophysical Characterisation of the Electron-Transfer Interactome in S. oneidensis MR-1	Dr Lars Jeuken	University of Leeds
BB/M017982/1	Warwick Integrative Synthetic Biology Centre	Professor John McCarthy	University of Warwick
BB/L027801/1	The exploitation of metagenomics and meta-omics approaches in life science research; community network in metagenomics	Professor Elizabeth Wellington	University of Warwick
BB/L01386X/1	BrisSynBio: Bristol Centre for Synthetic Biology	Professor Dek Woolfson	University of Bristol
BB/L021056/1	Quantitative Genetics of Hybrid Yeasts: overcoming sterility and biotechnological exploitation of diversity	Dr Chris Powell	University of Nottingham
BB/L024152/1	Metabo	Dr Chris C Steinbeck	EMBL - European Bioinformatics Institute
BB/L013940/1	SBRC NOTTINGHAM: Sustainable Routes to Platform Chemicals	Professor Nigel Minton	University of Nottingham
BB/M026280/1	Multi-scale enzyme modelling for SynBio: optimizing biocatalysts for selective synthesis of bioactive compounds	Dr Marc van der Kamp	University of Bristol
BB/M027740/1	Accelerating Synthetic Biology Approaches to Renewable Chemicals and Fuels	Professor Nigel Minton	University of Nottingham
BB/M011259/1	PeriTune - a clonal optimisation platform	Dr Neil Dixon	The University of Manchester
BB/M006891/1	Enriching Metabolic PATHwaY models with evidence from the literature (EMPATHY)	Professor Pedro Mendes	The University of Manchester
BB/M029085/1	[14-ERA IB] MetaCat: A metagenomic collection of novel and highly efficient biocatalysts for industrial biotechnology	Professor Peter Golyshin	Bangor University
BB/M000265/1	Engineering new capacities for solar energy utilisation in bacteria	Krebs Professor of B Neil Hunter	University of Sheffield
BB/M005690/1	14-ERASynBio - IESY - Inducible Evolution of Synthetic Yeast genomes	Dr Yizhi Cai	University of Edinburgh

Reference	Title	Investigator / Supervisor	Institution
BB/M025640/1	Building national hardware and software infrastructure for UK DNA Foundries	Dr Yizhi Cai	University of Edinburgh
BB/M028917/1	A Systems Biology Approach to Optimisation of (Fed-)Batch and Continuous Fermentation Processes for Recombinant Protein Production	Professor Gary Black	Northumbria University
BB/M024202/1	Biosynthesis, Regulation and Engineering of Bacterial Carbon Fixation Machinery	Dr Luning Liu	University of Liverpool
BB/K020617/3	Using flow cytometry and genomics to characterise and optimise microalgal-bacterial consortia cultivated on Wastewater to produce biomass for Biofuel	Dr Carole Llewellyn	Plymouth Marine Laboratory
BB/L018616/2	Unifying metabolome and proteome informatics	Dr Andrew Dowsey	University of Liverpool
BB/M011712/1	Expansion and Further Development of the PSIPRED Protein Structure and Function Bioinformatics Workbench	Professor David Jones	University College London
BB/M012557/1	Advanced MS instrumentation for enhanced proteomics capabilities	Professor Claire Eyers	University of Liverpool
BB/M019985/1	Open source pipelines for integrated metabolomics analysis by NMR and mass spectrometry	Professor Mark Viant	University of Birmingham
BB/M019993/1	Software tools for structure elucidation of synthetic and natural product peptide mixtures by LC-IM-MS	Dr James Redman	Cardiff University
BB/M020118/1	PIT-DB: A Resource for Sharing, Annotating and Analysing Translated Genomic Elements	Professor Conrad Bessant	Queen Mary, University of London
BB/M020282/1	Open source pipelines for integrated metabolomics analysis by NMR and mass spectrometry	Professor Andrew Jones	University of Liverpool
BB/M027023/1	Design and Evolution of Artificial Enzymes with Non-Canonical Organocatalytic Residues	Dr Anthony Green	The University of Manchester
BB/N01037X/1	DeTOX - Productive whole cell biocatalysis by engineering resistance to toxic products and substrates	Professor J Green	University of Sheffield
BB/N007212/1	Enabling synthetic biology with an expanded library of engineered orthogonal genetic logic gates and switches	Dr Baojun Wang	University of Edinburgh
BB/N010256/1	Quiescent Microbial Cell Factories	Dr David Keith Summers	University of Cambridge
BB/M022374/1	Super-Beacons and Beacon-STORM: a new generation of small tunable photoswitching probes and Super-Resolution approaches.	Dr Ricardo Henriques	University College London