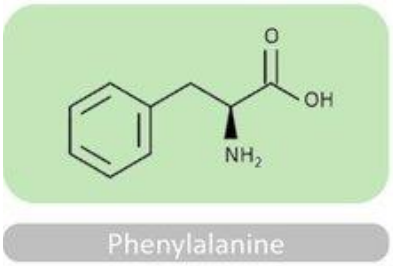
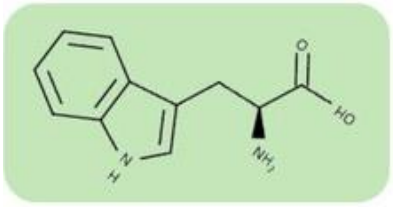
SYNTHETIC BIOLOGY FOR THE PRODUCTION OF BIOBASED MOLECULES

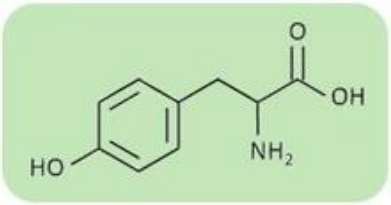
The ShikiFactory100 project aims to produce a collection of over 100 high-value compounds from the shikimate pathway, a metabolic pathway widely used by organisms in nature (e.g. bacteria, fungi and plants) to synthesise the aromatic amino acids (AAAs) phenylalanine (Phe), tryptophan (Trp) and tyrosine (Tyr) (Figure 1). These AAAs are important precursors for the production of cell metabolites[1] and other intermediates, which serve as branch points for the production of many interesting and valuable compounds.



Phenylalanine



Tryptophan



Tyrosine

Figure 1: The aromatic amino acids produced by the shikimate pathway.

To learn more about what is meant by **cell metabolism**, **amino acids** and the **shikimate pathway** click here.

In industry, many compounds derived from the shikimate pathway are used in food, pharmaceutical and cosmetic applications. They may be extracted directly from the organisms in which they are made (i.e. biomass extraction), but this tends to be costly due to low concentrations of target molecules typically present in natural feedstocks. Chemical synthesis on the other hand, is a more efficient and cost-effective method of manufacturing these compounds, however it is often associated with undesirable by-products and the use of non-renewable, oil-derived feedstocks.

The microbial production of complex chemicals offers an alternative to extraction and chemical synthesis. It involves the use of host organisms (also known as **cell factories**) to produce target molecules of bio-based origin. In nature, organisms may be able to produce a limited selection of target compounds from established biosynthetic pathways. However, through the intelligent redesign of **metabolic pathways**[2] and **cell chassis**,[3] microbes can become equipped to manufacture a wide range of possible chemicals for use in various applications. The ShikiFactory100 project aims to apply these same principles to the shikimate pathway, using *S. cerevisiae* and *E. coli* as host organisms to manufacture a collection of over 100 high-value compounds.

The redesign of organisms and their metabolic pathways can be explained by **synthetic biology**, a multidisciplinary area of research that applies principles of engineering and computing to biology, for the creation of novel biological systems that do not exist in the natural world. The ShikiFactory100 project will employ synthetic biology to generate novel pathways for the production of many interesting and valuable compounds. In order to understand how synthetic biology will be used in the ShikiFactory100 project, it is useful to gain an understanding into the background of this subject area.

A background to Synthetic Biology

All information contained cells is stored as **DNA**, but cellular functions are carried out by **proteins. Genes** are the fundamental units of hereditary information, described as short stretches of DNA that code for the production of specific proteins. By modifying genes and the types of proteins that are subsequently generated, microorganisms can therefore be engineered to express new, desirable

[1] Key molecules that are required to drive life-sustaining chemical reactions in organisms.
[2] Chemical reactions that occur within cells which lead to the conversion of substances into final products.
[3] Organisms that house and support genetic components by providing the resources that allow them to function.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 814408.

behaviours. However, the design of novel biological systems relies on an understanding of how information links to function, this is often referred to as the central dogma of molecular biology (Figure 2). In general, it can be described as DNA transcribing into an intermediate molecule of RNA, which is then translated into the final product, a protein.

Figure 2: The central dogma of molecular biology.

DNA is made up of nucleotides which consist of one of four bases (cytosine, guanine, thymine and adenine); connecting sugar moieties; and a phosphate backbone. Each of the four bases in DNA can pair with only one of the others (i.e. cytosine with guanine, and thymine with adenine), meaning that one strand defines the sequence of the other. The consequence of this is that a cell always has two copies of encoded information, which is important if one strand becomes damaged as the other will direct the repair and no cellular code is lost. The complementarity of both strands means that each can act as a template for the synthesis of new DNA (daughter helices).

The DNA and RNA elements highlighted in Figure 2 are very similar in terms of structure, the key difference being the presence of an -OH (alcohol) group and the use of uracil in place of thymine in RNA. As is the case for both strands of DNA, the conserved nature of RNA and DNA bases means that they can also base pair together. This complementarity ensures that molecular code is retained in in the transcription of DNA to mRNA (messenger RNA). Once synthesised, mRNA takes the genetic information to the cell's ribosome and it is here that the RNA sequence is translated into amino acids, the building blocks of proteins.

Three mRNA nucleotides encode for a specific amino acid. This group of three nucleotides is called a codon (Figure 3). RNA sequences dictate the amino acids produced, and amino acid sequences dictate the proteins generated. Two codons called start and stop codes signal the beginning and the end of translation. The final protein product is formed after the stop codon has been reached.

Figure 3: RNA codon table where U = uracil, C = cytosine, A = adenine, and G = guanine.

1st position	2nd position U	2nd position C	2nd position A	2nd position G	3rd position
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	stop	stop	A
	Leu	Ser	stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino Acids

Ala: Alanine
Arg: Arginine
Asn: Asparagine
Asp:Aspartic acid
Cys:Cysteine
Gln: Glutamine
Glu: Glutamic acid
Gly: Glycine
His: Histidine
Ile: Isoleucine
Leu: Leucine
Lys: Lysine
Met: Methionine
Phe: Phenylalanine
Pro: Proline
Ser: Serine
Thr: Threonine
Trp: Tryptophane
Tyr: Tyrosisne
Val: Valine

Enzymes are proteins, made as described above. They are biological catalysts that accelerate chemical reactions in living systems and are required by almost all metabolic processes to ensure that reaction rates are fast enough to sustain life. Metabolic pathways, such as the shikimate pathway, depend on enzymes to catalyse individual steps, they drive the conversion of substrates to products, also known as metabolites, which are subsequently used by cells in a variety of applications. When looking to

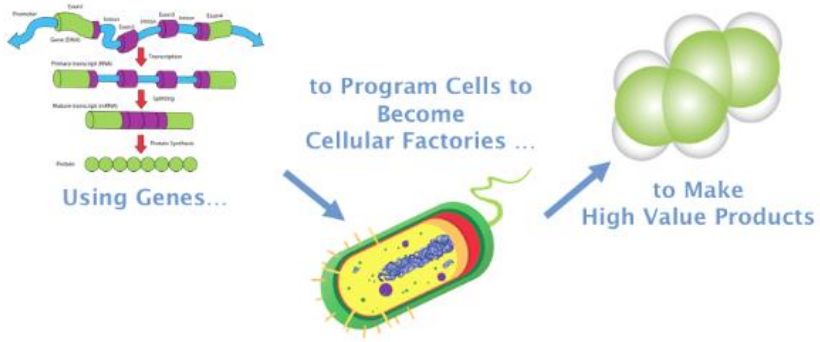
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 814408.

produce a range of other compounds from metabolic processes in nature, synthetic biology can be employed to create alternative pathways and generate suitable new enzymes to catalyse the production of valuable new products. This can be achieved by:

- Taking microbes (e.g. *S. cerevisiae* or *E. coli*) to produce enzymes, using DNA from other microbes that already do that job.
- Modifying metabolic pathways to get more or less of certain precursors/by-products.
- Designing new enzymes - based on software that fits molecules of interest into the active sites of enzymes and altering DNA code until the desired enzyme is created.

Figure 4 illustrates how genes can be used to programme cells, turning them into cell factories for the production of novel high value products.

Figure 4: The production of high value products using synthetic biology.



Design-Build-Test-Learn

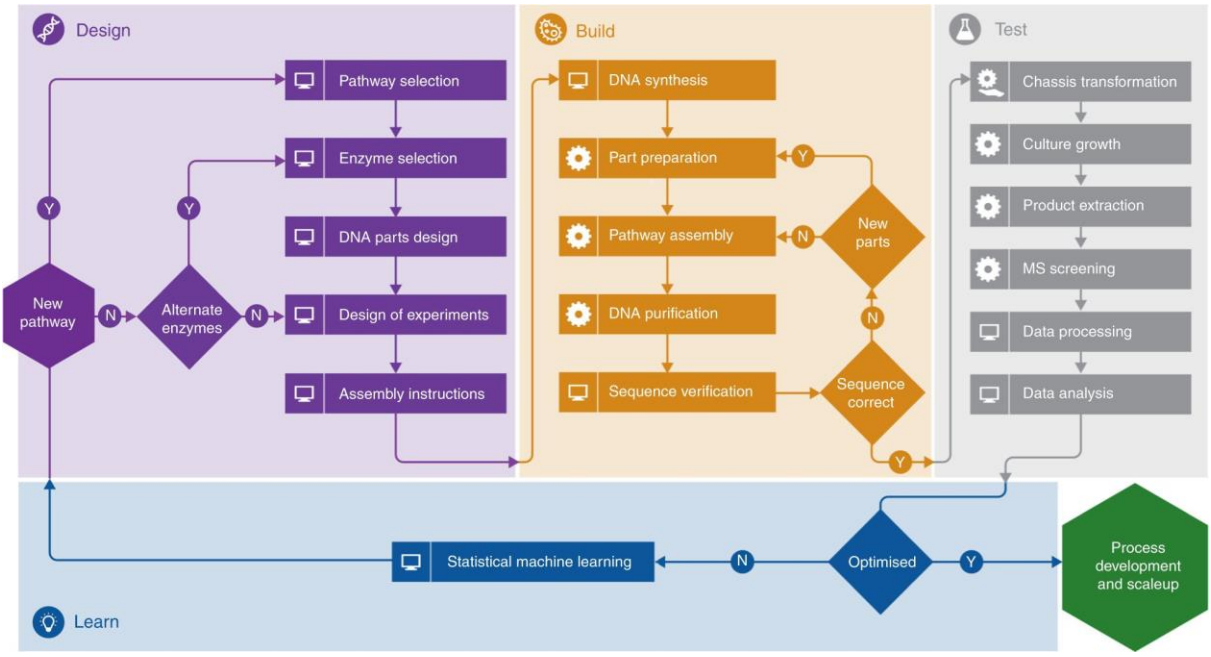
Systematic design is at the heart of synthetic biology. It relies on the understanding and creation of biological parts, devices and systems where '**parts**' represent pieces of DNA (genes) that encode single biological functions, they may then be combined into '**devices**' which can undertake more complex operations. '**Systems**' are collections of said devices, which perform as specified in the design phase. New engineered components must be inserted into a **chassis** (host organism), which will provide the necessary biology to transcribe and translate information. Typical chassis used in synthetic biology are *E. coli* bacteria and *S. cerevisiae* yeast, both of which are single-celled organisms.

Systematic design relies on the use of design cycles to carefully model new systems. The Design-Build-Test-Learn (DBTL) cycle provides a reliable framework for this to be achieved (

Figure 5). In the design phase, specifications must be outlined so that detailed plans can be drawn up, incorporating set criteria. This may include the design of new pathways or enzymes and typically involves software-enabled approaches i.e. computational modelling. Wet lab processes typically take large amounts of time to complete, whereas dry lab activities (i.e. modelling) are generally much faster, enabling researchers to quickly rule out inefficient designs and identify those with the greatest potential. Modelling may be used both before and after a system or device is assembled, to predict whether a design will work and to suggest other designs that may function more effectively. Modelling can simulate conditions that may be difficult to test otherwise, it can describe how proteins will bind to substrates, predict the how inputs will flow through metabolic pathways (carbon flux), and even outline how cells may interact with one another.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 814408.

Figure 5: The design build test learn cycle.[4]



Following the initial design of new systems or components, the build phase will take place. A mixture of wet and dry lab techniques will be employed to synthesise and purify DNA, and to prepare and assemble parts. Once built, new systems must be tested, both *in silico* and *in vivo*, to provide insight into how they are expected to behave and perform in reality. Finally, at the learn stage, data and experience from the design, build and test phases will be collected and examined. Mechanistic modelling, as well as artificial intelligence approaches e.g. machine learning, may be used to extract value from data sets and findings will be used to validate features and identify where improvements can be made in the next DBTL cycle.

ShikiFactory100 is using synthetic biology and the Design-Build-Test-Learn cycle to generate new pathways and chassis cells for the production of over 100 high-value compounds. To learn more about the types of molecules targeted by this project and the markets/applications in which they are associated, click here for more information.

[4] https://www.nature.com/articles/s42003-018-0076-9

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 814408.