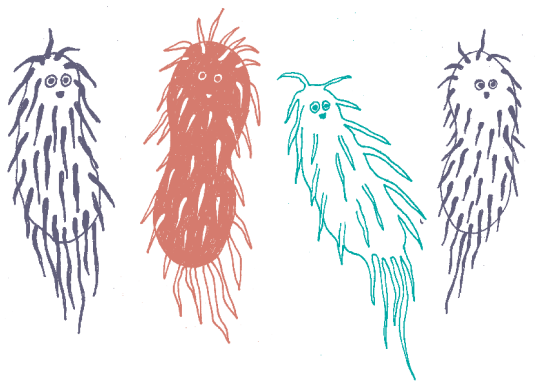
Genes and machines

Aleksandra Bartosik ★ Jakub Piątkowski ★ Marcin Ziemniak

Table of contents



Scientific method

☞ IN BRIEF

What criteria can we use to decide whether a statement is scientific or not? In this chapter we will discuss the criteria involved, and describe a proper course for conducting scientific investigations.

☞ KEYWORDS

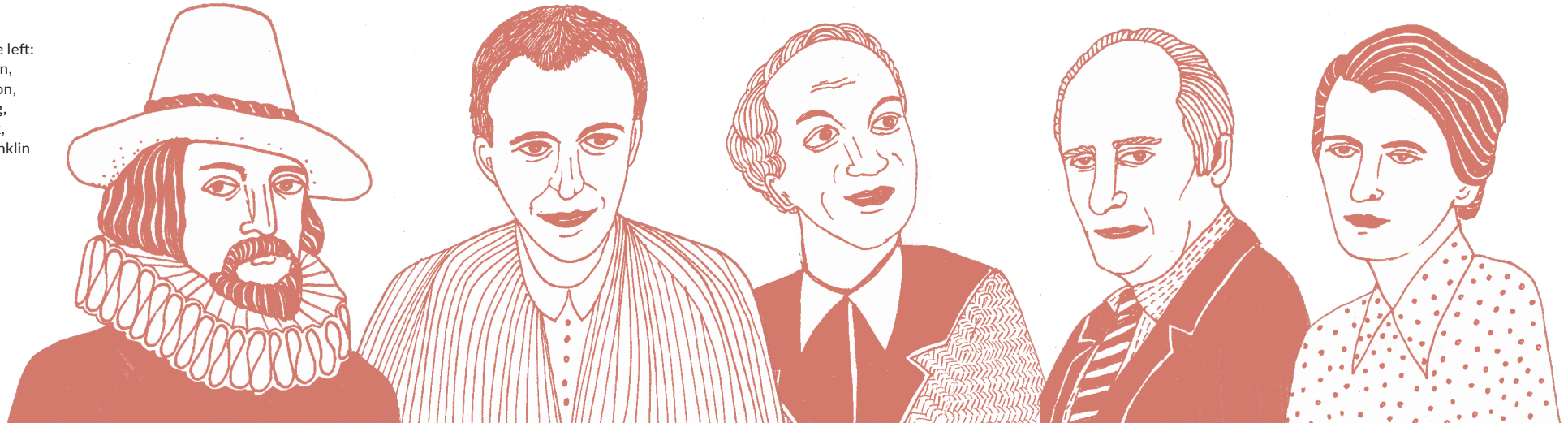
ROYAL SOCIETY
the second oldest scientific society in the world. It is still one of the most prestigious.

☞ IF YOU WANT TO KNOW MORE

1. Gauch, Hugh G., 2003, *Scientific Method in Practice*, Cambridge University Press
2. Watson J.D., Berry A., 2005, DNA. *The Secret of Life*, Wydawnictwo W.A.B.

The scientific method is a way of studying natural phenomena. Any study considered to be scientific should be based on empirical, measurable and replicable observations. The scientific method as we know it dates back to the seventeenth century, Francis Bacon is considered to be the father of modern scientific approach. His empirical and sceptical views inspired the foundation of one of the oldest scientific societies – the Royal Society. The application of the scientific method can vary depending on the scientific discipline in question. Generally, it can be said that the scientific method is based on the observation of natural phenomena in search of patterns, subsequently attempting to come up with possible explanations (hypotheses) and verification of these hypotheses by further observations, experiments and simulations. A scientific hypothesis must be falsifiable (potentially possible to overthrow by further observations and experiments). A hypothesis that repeatedly passed empirical verification and was widely accepted by the scientific community achieves the status of a scientific theory. It should be emphasized that the meaning of the word "theory" in the scientific context is completely different from its colloquial meaning. Thus, in the scientific debate practically no one uses the phrase 'it's just a theory' (one would rather say 'it's just a hypothesis'). A scientific theory is therefore a comprehensive explanation of natural phenomena, which has also passed numerous experimental verifications. This does not exclude the possibility of overthrowing a scientific theory on the base of future observations. A theory is considered instead to be the best available explanation for observable phenomena at a given time (although the probability of a total overthrowing long-standing and repeatedly verified theories – such as the theory of evolution – is extremely unlikely). From this perspective, the scientific method is a continuous process of observation, creating hypotheses, confirming or refuting them as well as improving and expanding upon the already existing theories (Fig. 2 and 3).

A good example of the use of the scientific method in molecular biology were the studies that led to the discovery of the structure of DNA. The chemical composition of the molecule as well as the fact that it is the carrier of genetic information were already known. But without the knowledge of its structure it was difficult to put forward a mechanism for the storage of genetic information, its replication and deciphering. Linus Pauling, Francis Crick and James Watson hypothesized that DNA has a helical structure. This hypothesis led them to predict that the X-ray diffraction pattern of the DNA strands will be X-shaped. That's the exact diffraction pattern that was obtained experimentally by Rosalind Franklin (Fig. 3). The final model of DNA was proposed by Crick and Watson after a detailed analysis of the results obtained by the Franklin and previously available data on the chemical composition of the molecule. This model has enabled the formulation of further hypotheses regarding the mechanism of DNA replication (which was then tested independently).



➡
Fig. From the left: Francis Bacon, James Watson, Linus Pauling, Francis Crick, Rosalind Franklin

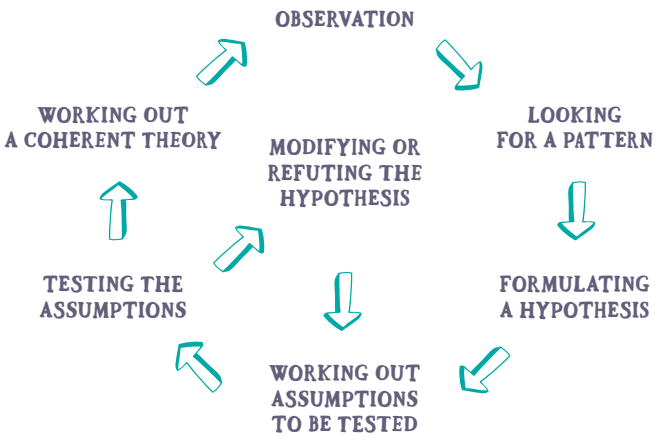


Fig. 1. Research conducted in accordance with the scientific method (simplified).

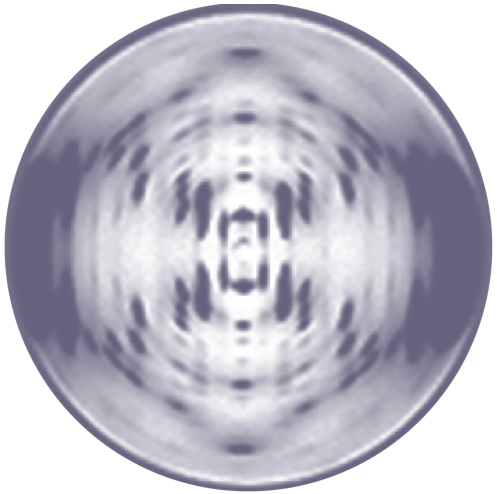


Fig. 2. X-ray diffraction pattern observed on DNA fibers obtained by Rosalind Franklin, also known as photo 51.

Fig. 3. The double helix DNA structure.

Introduction to molecular biology

Molecular biology is a branch of biology focused on the study of the activity, synthesis, structure and interactions between biomolecules such as DNA, RNA and proteins. It came into existence in the late 1930s, with the development of techniques of isolation and analysis of biomolecules (such as X-ray crystallography). The impulse that greatly accelerated its development was solving the structure of DNA in the 1950s and the development of basic tools of genetic engineering in the 1970s and 1980s. In this chapter we will briefly discuss the key findings in the field of molecular biology and its most important tools.

From 1944 until 1962, a number of breakthroughs took place in molecular biology. DNA has been identified as the molecule responsible for inheritance, its structure and mechanism of replication have been also discovered. Francis Crick, the co-discoverer of DNA structure, formulated the central dogma of molecular biology during this period. The central dogma states that the genetic information is transcribed from DNA to RNA (far less often in the opposite direction) and that proteins are synthesized on an RNA template. Solving the universal genetic code – that is elucidating how three nucleotides in the DNA/RNA correspond to specific amino acids in the protein sequence – can be considered the crowning achievement of this first phase of development. In the 1960s Francois Jacob and Jacques Monod also managed to unravel the first simple model of regulating the expression of genetic information in bacteria – the lactose operon.

The second phase of development of molecular biology symbolically opens in 1972, with the development of the basic tools for genetic engineering – restriction enzymes. These enzymes can be seen as a kind of molecular scissors, which allow us to precisely cut and merge different DNA segments. Fast and efficient reading of DNA sequences became possible in 1977, when Frederick Sanger developed his proprietary method of sequencing (now fully automated). In 1983, Kary Mullis developed PCR (Polymerase Chain Reaction), allowing for amplification of any DNA fragment in a vast number of copies. From this point onward, the set of basic tools for molecular biology and genetic engineering has been completed.

☞ IN BRIEF

Molecular biology is the study of the molecular basis of life. Over the past 50 years, life sciences has made great progress in this area explaining the mechanisms of fundamental processes, such as the expression of genetic information and its inheritance.

☞ KEYWORDS

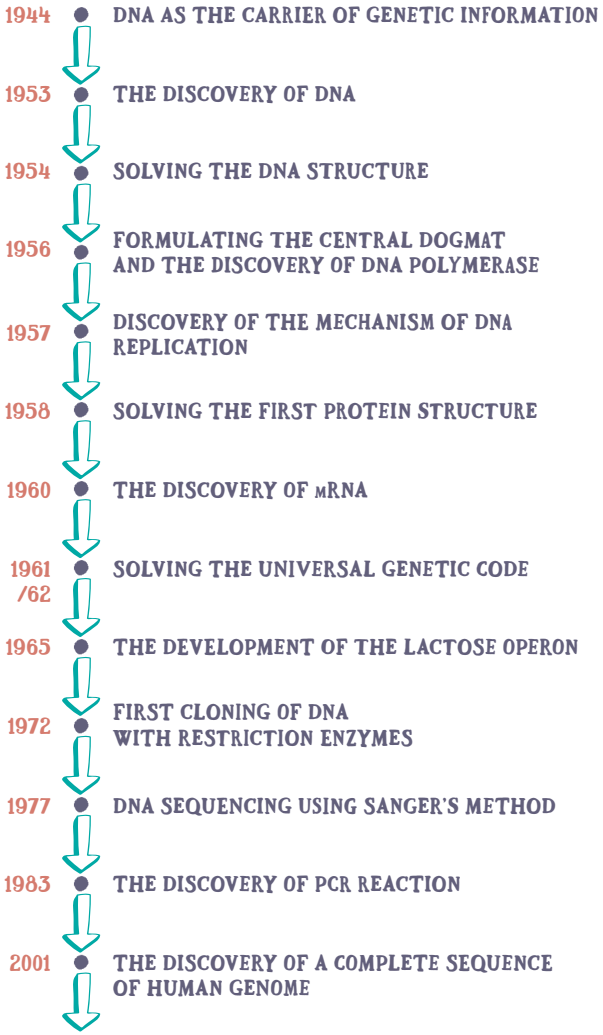
UNIVERSAL GENETIC CODE
the way of converting the sequence of nucleotide triplets into amino acids in the protein sequence. It is roughly the same in most living organisms (there are a few exceptions).

LARGE-SCALE METHOD
an automated method that generates an enormous amount of data on the genetic sequence, as well as RNA and proteins levels.

☞ IF YOU WANT TO KNOW MORE

1. http://www.dnai.org/timeline/index.html
2. Watson J.D., Baker T.A., Gann A., Levine M., *Molecular Biology of the Gene, Seventh Edition*; Cold Spring Harbor Laboratory Press
3. Watson J.D., Berry A., *DNA: The Secret of Life*; W.A.B.

In 2001, as a result of a joint effort by a government consortium and Craig Venters' company, the complete sequence of human genome has been published. Biology has entered the era of large-scale methods...



➡ **Fig. 1.** Major developments in molecular biology (on a timeline).

Introduction to macromolecules

From a general point of view, a biological molecule (biomolecule) is any molecule existing in a living body, including large macromolecules, such as proteins, nucleic acids, carbohydrates and lipids, as well as small molecules (metabolites, natural products). Biomolecules are typically endogenous (naturally existing in the body), but can also be exogenous (introduced into the body from the outside). Below, we are going to provide basic information about the most important (in the context of synthetic biology) of macromolecules, namely nucleic acids and proteins.

NUCLEIC ACIDS

Nucleic acid ssuch as DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are linear polymers composed of monomers called nucleotides. Each naturally existing nucleotide has three components: a sugar ring containing five atoms, a phosphate group and a nitrogenous base. If the sugar is deoxyribose, the polymer is DNA but, if the sugar is ribose – the polymer is RNA. Nucleic acids are pivotal to all known forms of life, because they mediate in coding, expressing and transfer of genetic information. the genetic information itself is encoded in the nucleic acid sequence (the linear sequence of nucleotides in the DNA or RNA molecule).

Deoxyribonucleic acid (DNA) is a nucleic acid containing the genetic instructions being necessary for the functioning of all living organisms. Parts of a DNA molecule containing the genetic information are called genes. Some other parts of DNA serve a structural function, or are involved in the regulation of gene expression. A DNA molecule contains two linear polymers composed of sugars and phosphate groups linked together by ester linkages. The two strands have opposite directions relative to each other (they are anti-parallel) and they form hydrogen bonds to form a helical structure. Each nucleotide contains one of four nitrogenous bases (usually abbreviated as bases). The sequence of these four bases directly encodes the genetic information, which itself determines the sequence of amino acids in all proteins. Prokaryotic organisms (bacteria) are relatively simple, so their DNA is stored in the cytoplasm. Contrastingly, eukaryotic organisms keep most of the DNA in the nucleus, and only a small portion of

☞ IN BRIEF

In this chapter we would like to briefly discuss three types of molecules being the main object of interest for molecular biologists, as well as their research techniques.

☞ KEYWORDS

ANTIPARALLEL ORIENTATION OF DNA STRANDS
DNA strands runs in opposite directions. Any DNA strand has two different ends called respectively 5 'and 3'. In the DNA double helix, the 3 'end of one strand is located near the 5' of the second strand.

METABOLITE
virtually every intermediate and final product of metabolism, however, the term is usually used only for small organic molecules

NATURAL PRODUCT
an organic compound isolated from natural sources, which is usually a product of metabolism

HYDROGEN BOND
a type of weak, non-covalent interaction between electronegative atom and a hydrogen atom bound to another electronegative atom

POLYMER
a large molecule composed of a number of smaller subunits.

their DNA exist in organelles like mitochondria and chloroplasts. In the nucleus, eukaryotic DNA is organised in long structures called chromosomes. During the cell division, chromosomes are duplicated (replication of DNA) to ensure that each daughter cell inherits a complete set of genetic information. Chromosomes also contain a set of protein which help to organise and compact DNA molecules.

Ribonucleic acid (RNA) plays an important role in the transmission of genetic information from genes to proteins. Three types of RNA are universal for all living organism, namely transport RNA (tRNA), messenger RNA (mRNA) and ribosomal RNA (rRNA). mRNA transfers genetic information, controlling the protein biosynthesis. rRNA is the active component of the ribosome, which catalyses peptide bond formation. Finally, tRNA delivers amino acids molecules required for protein biosynthesis. Moreover, there is a number of other RNA classes which are mostly involved in the regulation of gene expression.

TRANSCRIPTION
the first step of gene expression, wherein the segment of DNA is copied into mRNA by an enzyme called RNA polymerase

TRANSLATION
the process by which ribosomes form a protein using mRNA as a template

PROTEIN TURNOVER
a balance between protein synthesis and protein degradation in a cell

☞ IF YOU WANT TO KNOW MORE

1. J.M. Berg, J.L. Tymoczko, L. Stryer, 2005, *Biochemia* (Biochemistry), Wydawnictwo Naukowe PWN

PROTEINS

Proteins are large macromolecules comprised of at least one linear chains of amino acid residues. Such a polymer is called a polypeptide. Each protein contains at least one long polypeptide. Short polypeptides (having less than about 20–30 amino acid residues) are not considered proteins and are simply called peptides. The main difference between protein molecules lies in their amino acid sequences. Typically, the sequence is directly responsible for the protein to fold into a specific three-dimensional structure that determines its activity. Shortly after (or during) the protein synthesis some amino acid residues are chemically modified (post-translational modifications), which changes the structure and function of the protein. Proteins perform a wide range of functions, including metabolic reactions, DNA replication, response to external stimuli and transport of other molecules. Many proteins are enzymes that catalyse biochemical reactions, being essential for the metabolism. Proteins may also perform mechanical and structural functions. Typically, proteins work together existing in the form of larger, stable complexes consisting a number of protein molecules. Once formed, proteins only exist for a certain period of time and are then degraded and recycled by the cell through the process of protein turnover.

NOTEWORTHY METHODS OF STUDYING MACROMOLECULES

Electrophoresis – it allows to achieve size-dependent separation of macromolecules. This method can be applied to DNA, RNA and proteins. Studied molecules migrate through a polymeric gel (such as agarose or polyacrylamide) due to an input of electric field. Smaller molecules migrates faster and larger ones migrates slower (Fig. 2A).

Sanger DNA sequencing – it is based on the synthesis of the second DNA strand. In this method, one DNA is synthesised in the presence of dideoxynucleotides which quench the reaction. In the modern, automated version of the method dideoxynucleotide are fluorescently labeled, so it is possible to specify at which nucleus the strand synthesis was finished. In the previous, non-automated version four independent reactions were performed. In each of them, only one of the four nucleotides interrupted the reaction and was radiolabeled (then the reactions were electrophoretically separated).(Fig. 2B).

Mass Spectrometry (MS) – a technique allowing to identify various proteins. The proteins are enzymatically digested and their fragments are identified by their mass-to-charge ratio. This can be regarded as a type of "protein sequencing" (Fig. 2C).

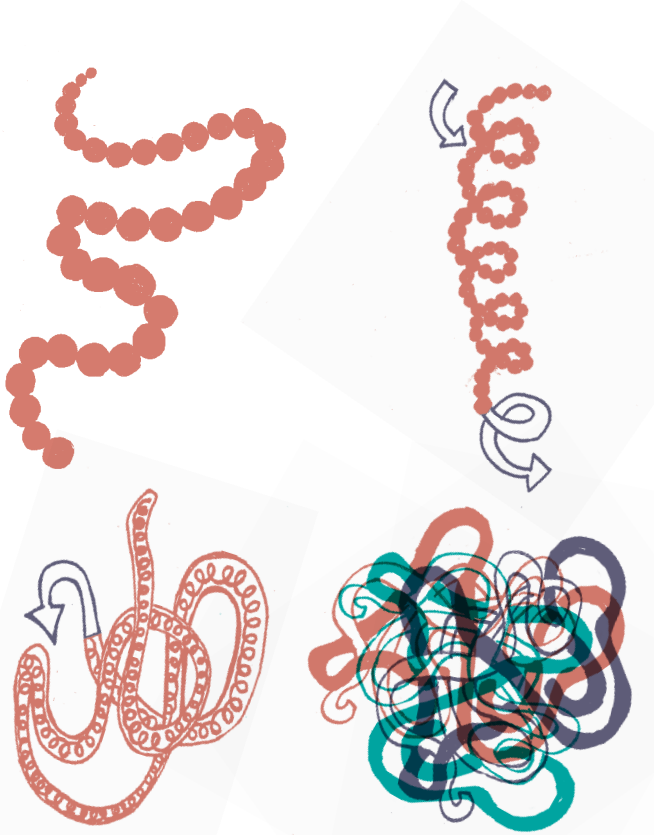


Fig. 1. The structure of a protein molecule.

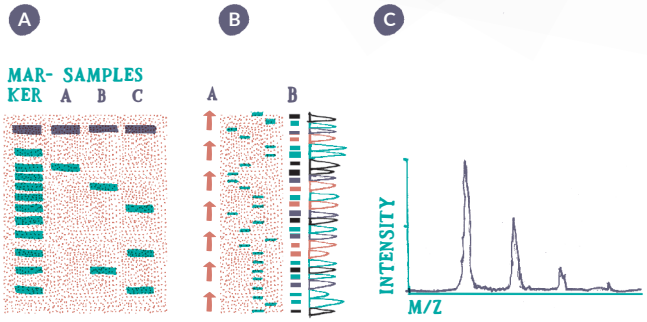


Fig. 2. Methods to study macromolecules. (A) Electrophoretic separation of DNA. (B) The Sanger sequencing; (C) An MS spectrum.

Biological databases

Biological databases are digital libraries that contain information in all fields of life sciences. They include data from individual experiments, larger research projects, literature searches, high-throughput experiments (those that result in large amount of data), and bioinformatics analyses. They may relate to genomic sequences, expressed levels of mRNA and proteins, interaction networks, regulatory networks and metabolic networks. Databases may also contain information about functions of genes, proteins, their location within the cell, relationship between genes from different organisms and clinical significance of specific mutations. Databases may concern many aspects of the above at once, or only their sections. They can also focus only on selected model organisms (e.g. budding yeast or fruit flies). According to the Nucleic Acids Research journal, in 2012 there were 1380 available biological databases. Three of the largest databases of DNA/RNA sequences are the American GenBank, European EMBL and Japanese DNA Data Bank. All three regularly exchange data to ensure their accuracy, timeliness and completeness.

OTHER IMPORTANT DATABASES INCLUDE:

- UniProt – a database of protein sequences
- PDB (Protein Data bank) – a database of protein structures
- Rfam – a database of RNA families
- Pathway Commons – a database of biological pathways
- KEGG – a database of metabolic pathways
- ArrayExpress – a database of RNA expression levels
- PhenCode – a database linking human mutations with phenotypes

The above list is not exhaustive by any measure. Every year new databases are made available. Some draw information directly from experimental results or literature, others integrate information from a pre-existing databases to help scientists analyse the data and formulate hypotheses.

IN BRIEF

Where do biologists get the necessary information at different stages of research projects? Below we discuss the basic types of biological databases and the types of information they contain.

KEYWORDS

NUCLEIC ACIDS RESEARCH a journal which publishes an annual free summary of available biological databases

IF YOU WANT TO KNOW MORE

1. Galperin, M.Y., & Fernández-Suárez, X.M. (2012). The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. Nucleic Acids Research, 40 (Database issue), D1–8. doi:10.1093/nar/gkr11965.

What is synthetic biology?

☞ IN BRIEF

Synthetic biology is a new and rapidly growing field of biology. At the same time, due to its engineering aspect, it goes far beyond what molecular biology traditionally meant.

☞ KEYWORDS

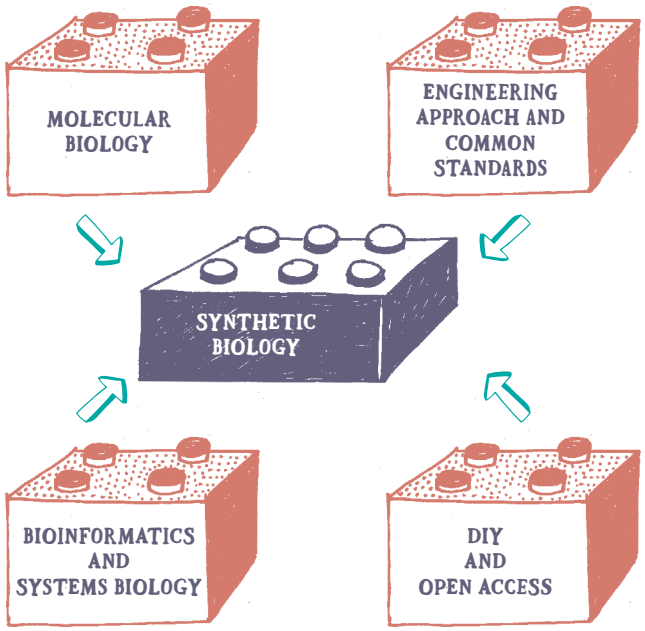
STANDARDIZATION
in the context of synthetic biology, it is an approach aimed at the applying a set of uniform, agreed upon and widely accepted biological standards.

SYSTEMS BIOLOGY
a multidisciplinary field that treats complex biological systems as a whole (and is thus the opposite of the reductionist approach).

☞ IF YOU WANT TO KNOW MORE

1. University of Washington, *Synthetic Biology*, http://synbio.washington.edu/

Synthetic biology is an attempt to approach molecular biology as an engineering science. But it is not just that. Scientists involved in synthetic biology are continually trying to design complete molecular mechanisms in a rational manner. They attempt to create complex molecular machineries – designed from the ground up. Until recently, geneticists have only introduced subtle, individual modifications into naturally occurring systems. Synthetic biology is an interdisciplinary project, combining molecular biology, genetic engineering, bioinformatics, and systems biology. Synthetic biology seeks to create uniform standards and data exchange systems that involve the largest possible number of researchers. Synthetic biologists also aim to ensure that the tools of molecular biology and genetic engineering are available to all, including bio-hobbyists. This is done to ensure that as many people as possible participate in the upcoming scientific breakthroughs.



➞ **Fig. 1.** Foundations of synthetic biology.

Mathematical models in life sciences

The process of building a biological model is a sort of multi-layered cycle, which requires a large number of variables, mathematical relationships and numerical parameters, thanks to which the calculations and other mathematical analyses allow for observing the properties of the system and predict its behavior in the future. In each cycle, the model is further developed and evaluated. It must be remembered that even the most sophisticated simulation methods are virtually useless if their initial structure is flawed.

Because of that, building a model is a rather time-consuming and tedious task. Importantly, some biological insight can be extracted at each stage of the model building. The first step is to analyse the existing experimental data (i.e. gene expression, interaction of biomolecules, etc.) which will be put into the model. Importantly, the model does not necessarily include everything we know about the system which we study. Sometimes a simpler model may be easier to understand and describe. Choosing the right type of models (and computational methods) is one of the most difficult steps in the process of model building. For example, including too many variables in the simulation may hinder the construction of the model as well as the quality of calculations. The next step in the model building is to find all possible interactions between the components.

Even a simple analysis of that network can provide important information on the systemic level. A deeper description also includes directional interactions between elements of the network and allows you to conceptualise the flow of information in the network. Finally, the network elements can be characterized quantitatively (properties such as the level of gene expression, stability of the protein, etc.). All stages of both the model building and computations must be documented carefully in order to further verify the model and allow fellow scientists to repeat the whole study. The biologists' community has already developed a set of guidelines for the correct description of the models, hence all data used to build a model should be also available. This data can be encoded in using broadly accepted standards, such as Systems Biology Markup Language (SBML) for models and Simulation Experiment Description Markup Language (SED-ML) for simulations and analyses. The introduction of these open

☞ IN BRIEF

Although traditionally mathematics and biology are seen as rather distant disciplines, due to the current progress in life sciences and medicine many biological experiments became essentially quantitative and may hence be described by mathematics.

☞ WHY THIS IS IMPORTANT

Engineering approach in science requires tools to build and analyse models. Depending on the discipline, different tools are required to create such models, but the goal is the same – to create a mathematical description of the studied system. A good model can predict the behaviour of the system in various circumstances. In the case of biological sciences, we need models emulating living organisms. In this chapter (and a few subsequent ones), we take a bit less ambitious, but equally important task – to answer a question how to model metabolic processes and regulate gene expression.

☞ IF YOU WANT TO KNOW MORE

1. Alon U., 2006, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman&Hall
2. Le Novera, Nicolas. *Quantitative and logic modeling of molecular and gene networks*, Nature Reviews Genetics 16.3 (2015): 146-158.

standards had a significant impact on the further development of systems biology, mainly by facilitating the automation of both computation and experiments.

COLLECTION OF EXPERIMENTAL DATA

Advances in molecular biology made at the beginning of this century led to a substantial increase in the amount of biological data collected by scientists. One of the most versatile methods used to visualise the various cellular processes is the use of reporter genes, such as green fluorescent protein (GFP). This protein does not disturb the cell functions allowing efficient to monitor the level of expression of specific genes or to determine the cellular localisation of studied proteins.

QUANTITATIVE DESCRIPTION OF THE CELL

From the engineering point of view, the cell can be treated as an integrated device, enclosing thousands of interactions between its building blocks (biomolecules). A significant part of these molecules are proteins. To simplify, we can say that each protein is a microscopic machine capable to perform specific tasks. For example, *Escherichia coli* cell is simple but still it comprises of several million protein molecules, which are divided into about 4000 types. During its existence, the cell encounters various environmental conditions wherein its mere survival will require expression of different sets of proteins. Therefore, cells must constantly monitor their environment – it is the only way to adapt to a dynamically changing environment. For example, if the genetic material of the cells is damaged the cell starts to produce specific proteins needed to repair the DNA. Dynamic regulation of protein biosynthesis occurs in virtually all cells and it is controlled by complex systems known as transcriptional networks.

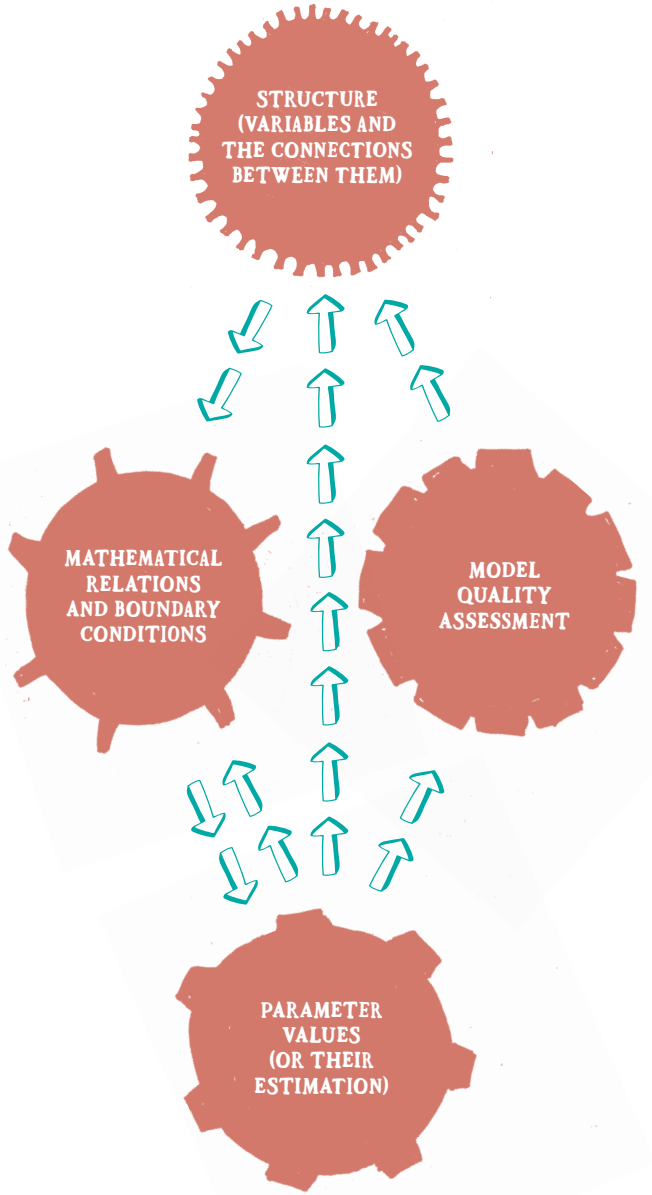


Fig. 1. Building a model of the biological process is a cyclic process. In each cycle, the model become more and more accurate.

Computational tools in synthetic biology

☞ IN BRIEF

Any engineering science, such as synthetic biology requires an appropriate mathematical framework. Modelling complex biological systems is difficult, but it has already demonstrated its first practical benefits.

☞ KEYWORDS

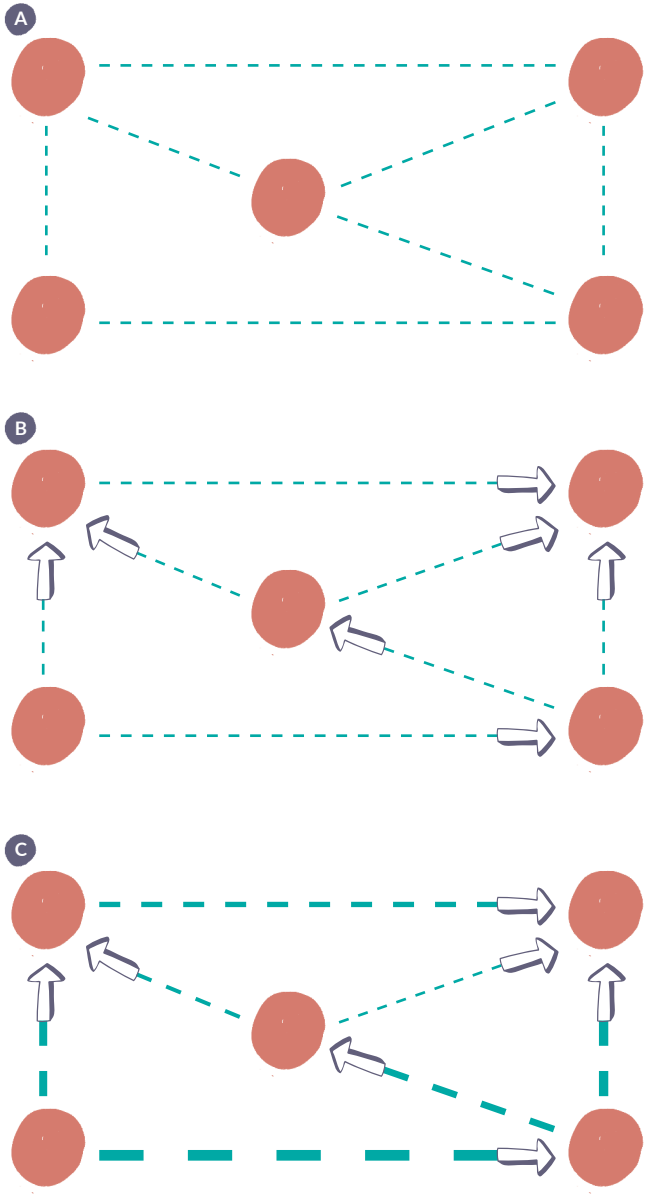
EMERGENT QUALITIES
qualities displayed by complex systems. They arise not due to qualities of individual components of the system, but from interactions between these components.

METABOLIC ENGINEERING
a field dealing with the optimization of the genetic and regulatory processes in the cell, in order to obtain the most efficient production of certain substances..

One of the basic goals of synthetic biology is the rational design of complete biological systems. This requires the use of appropriate computational tools. Even a relatively simple biological systems may in fact behave in a highly unpredictable and counter-intuitive way. The more complex the system, the greater is the probability that it will exhibit emergent properties (i.e. such that cannot be simply deduced from characteristics of its constituent elements, but rather interactions between them). This can be observed both on molecular and macroscopic scale (a good example would be social insects, where relatively simple rules that direct the behaviour of individual insects lead to the creation complex structures, such as beehives or termite mounds).

One of the key approaches is the creation of genetic and molecular interaction networks. Some networks may contain only information regarding the occurrence of interactions (they could be physical or genetic) between proteins or genes. Such networks can be constructed relatively easily, but their predictive capacity is limited – they allow us to make predictions regarding only the stability of the system or what may lead to its disruption. More sophisticated models can take into account the nature of interactions, their direction and quantifiable values (e.g. the strength with which the presence of one protein affects the level of another protein, Fig. 1). Such models allow us to make much more detailed predictions, but also require using of accurate experimental data and multiple stages of verification.

It should be noted that modelling of simple biological systems (such as molecular oscillators – Fig. 2) has been successful even at the early stages of synthetic biology development. However only recently, more advanced projects in the field of metabolic engineering became possible, creating hope for a more efficient and cheaper production of important substances, such as medical drugs. The latest spectacular success of this kind was the efficient production of anti-malarial drug, artemisinin in genetically modified budding yeast cells (*saccharomyces cerevisiae*).



☞ IF YOU WANT TO KNOW MORE

1. Novère Le N., 2015, *Quantitative and logic modelling of molecular and gene networks*, *Nature Reviews, Genetics*, 16(3), 146–58. doi:10.1038/nrg3885
2. Woods M.L., Leon M., Perez-Carrasco R., Barnes C.P., 2016, *A Statistical Approach Reveals Designs for the Most Robust Stochastic Gene Oscillators. ACS Synthetic Biology*, 5(6), 459–70. doi:10.1021/acssynbio.5b00179
3. Paddon, C.J., Keasling, J.D., 2014, *Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development*, *Nature Reviews, Microbiology*, 12(5), 355–67. doi:10.1038/nrmicro3240

←
Fig. 1. Biological networks: A – a simple interaction-only network; B – a network that takes directionality into account; C – a network that takes directionality and quantitative aspect (represented as the relative thickness of arrows) into account.

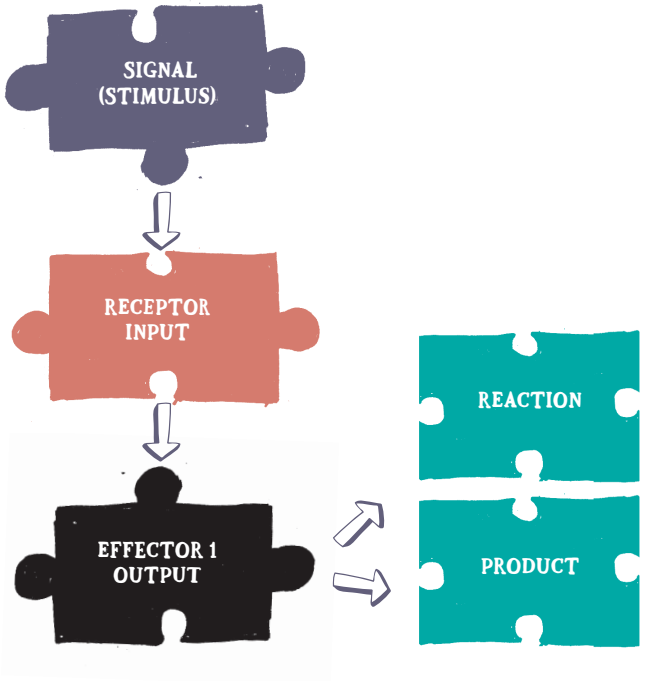
Regulation of biological systems

☞ IN BRIEF

Cells have to exchange matter, energy and information with surrounding environment. A signal from the outside is received by a cellular receptor located in the cell membrane. Afterwards, it is amplified and passed on by a chain of transmitters until it reaches the final effector. The most basic mechanism of network signal regulation is called feedback. It might be either positive or negative.

Although a cell is considered the basic unit of the living matter organisation, it is not a stand-alone. It has to exchange information with its environment: neighbouring cells, tissues and organs (if it is a part of the body of a multicellular organism) in order to survive. Unicellular organism signalling pathways are only a little bit easier to understand than multicellular ones. The cell needs to eat (scientists would say that it needs to provide substrates for chemical reactions) and excrete (remove unneeded products of metabolism). Also, it needs to reproduce, move and keep stable chemical environment in its interior. The cell is not aware why the stimuli appeared and what to do. Thus, a response must be fully automated.

Signalling systems are extremely diverse. Many of them are still unknown or their mechanism is not fully understood (so you still have a lot to discover!). Relationships between components of signalling systems and the information flow are studied by systems biology. In principle, most of the systems share a similar pattern as shown on Figure 1.



➡ **Fig. 1.** Signal transduction in cells is the response to a stimulus. There are many types of stimuli: vibration, temperature change, touch, pain, electric or magnetic field, change in atmospheric pressure, change in osmotic pressure or in the concentration of the detected molecule (called ligand). What type of ligand and receptors do we mean? They can be e.g. big molecules– protein, RNA, DNA or small molecules. They can be organic as well as inorganic (such as nitric oxide (II) causing widening of blood vessels). Among them, we can find neurotransmitters, hormones, transporters and also many different proteins.

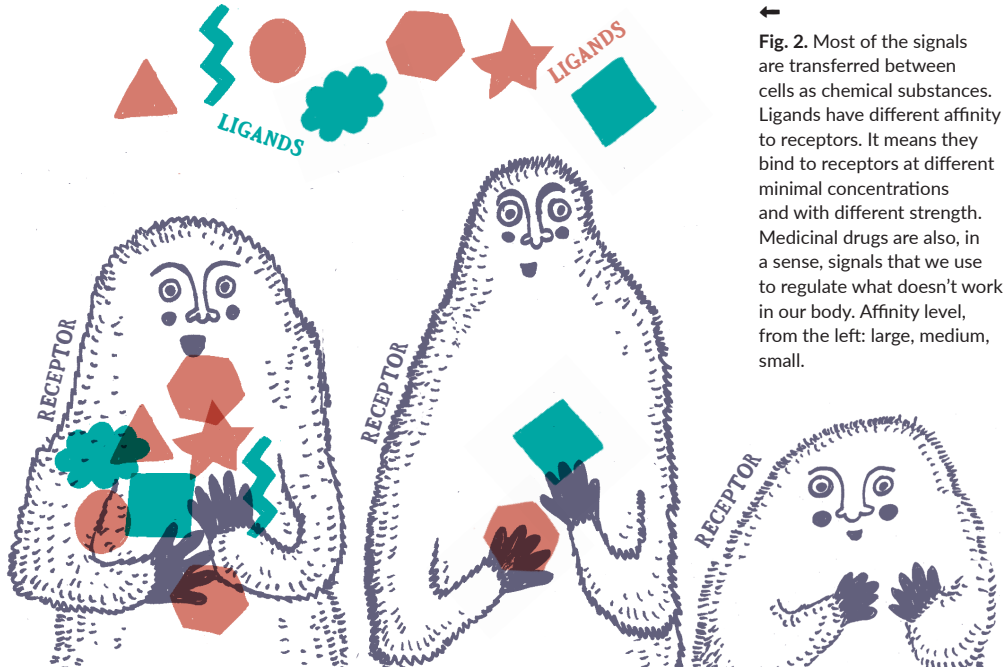
Highly potent drugs have a strong effect even in low concentrations. If the signaling molecule has low power, it needs a much higher concentration to produce the same effect upon binding to the receptor than in the case of a high-powered molecule. The effectiveness of the signaling molecule (drug) tells you how big cellular response it causes compared to the highest response ever observed. Signal molecules can bind to the receptor permanently (strong covalent bond) or impermanently (non-covalent bond). Internal signals are usually unstable. This is because ligand can detach and attach to the receptor repeatedly. A single receptor can bind several ligands at various locations or they may compete with each other for the same binding site. A ligand can activate or block the receptor. Usually, a receptor binds less signalling molecules in comparison to the effector molecules responsible for the final physiological effect. The reason is that the cell itself is an energy-saving system. Initial signal is amplified owing to a cascade of other signal molecules inside the cell. It looks a little bit like an avalanche and could be compared to people recommending a computer to friends. The more people hear about it, the more pass the news to others (Fig. 3).

☞ WHY THIS IS IMPORTANT

Once we know the mechanism of the signal transmission, we know how to modify it to get beneficial results.

☞ KEYWORDS

CELL SIGNALLING, RECEPTORS, A CASCADE OF TRANSMITTERS, DRUG POTENCY, EFFECTORS, FEEDBACK, SYSTEMS BIOLOGY, SIGNAL (STIMULUS), DRUG AFFINITY, DRUG EFFICACY



⬅ **Fig. 2.** Most of the signals are transferred between cells as chemical substances. Ligands have different affinity to receptors. It means they bind to receptors at different minimal concentrations and with different strength. Medicinal drugs are also, in a sense, signals that we use to regulate what doesn't work in our body. Affinity level, from the left: large, medium, small.

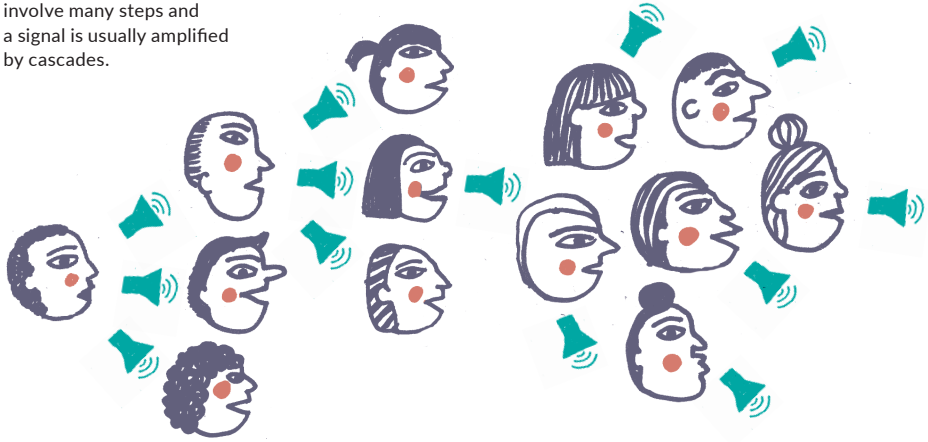
☞ IF YOU WANT TO KNOW MORE

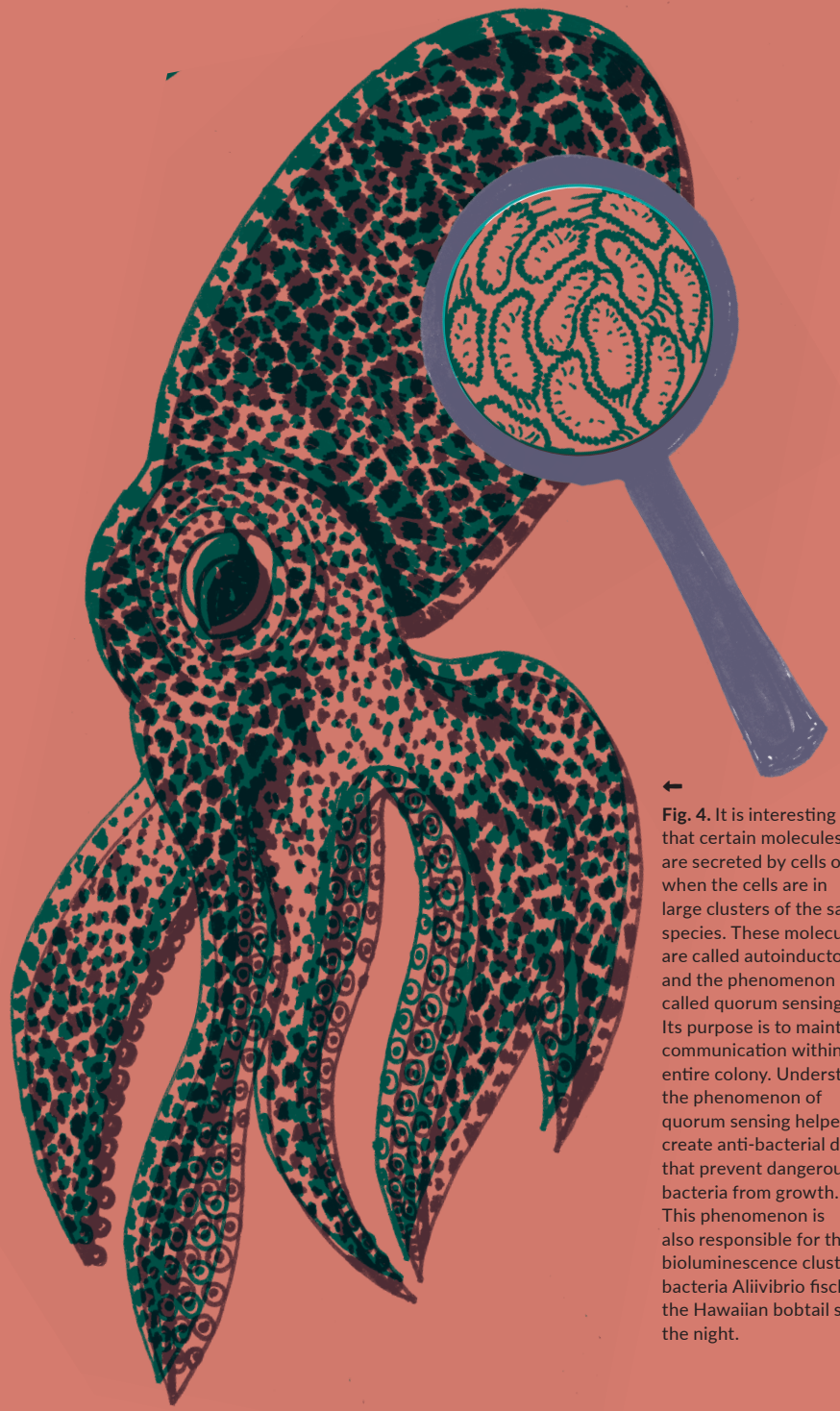
1. Cameron D.E., Bashor C.J., Collins J.J., *A brief history of synthetic biology*, *Nat Rev Microbiol*, 2014, 12(5):381-90
2. Konieczny L., Roterman I., Spólnik P., *Systems Biology. The strategy of a living organism*, 2010, PWN
3. *Molecular Genetics*, edited by P. Węgleńskiego, 2006, PWN
4. Khan Academy, *Overview of cell signalling*, http://www.youtube.com/watch?v=FQFBygnIONU

What happens to the signal along the way? Signal chains typically have multiple stages, and in several places there are signaling cascades that strengthen the signal. Some ligands are associated with the receptor longer, others detach (dissociate) quickly. A stimulus can induce signalling pathway response a couple of times.

However, a signal can be also weakened (attenuated) as a result of negative feedback, or the receptor can be totally turn off by the final effect of the cascade. Alternatively, it may be held still by the positive feedback if the product pathway has been activated. An example of positive feedback is the lactose operon in *Escherichia coli*. Living in an environment rich in lactose, *E. coli* produces an enzyme called β-galactosidase, which makes it possible to use lactose as an energy source.This is possible because lactose unlocks its production. If there is no lactose in the environment, β-galactosidase is not produced, because there is no need to do so. Tryptophan operon is another canonical example of a negative feedback. It can also be found in *E. coli*. Tryptophan is an essential amino acid. If it is present in the environment, its production is blocked in order to save energy of the cell. If it is not present, the lock is removed and the bacteria start to produce tryptophan.

➡
Fig. 3. Signalling pathways involve many steps and a signal is usually amplified by cascades.





⬅
Fig. 4. It is interesting that certain molecules are secreted by cells only when the cells are in large clusters of the same species. These molecules are called autoinductors and the phenomenon is called quorum sensing. Its purpose is to maintain communication within the entire colony. Understanding the phenomenon of quorum sensing helped to create anti-bacterial drugs that prevent dangerous bacteria from growth. This phenomenon is also responsible for the bioluminescence clusters of bacteria Aliivibrio fischeri in the Hawaiian bobtail squid in the night.

What are genetic networks and how can we use them?

☞ IN BRIEF

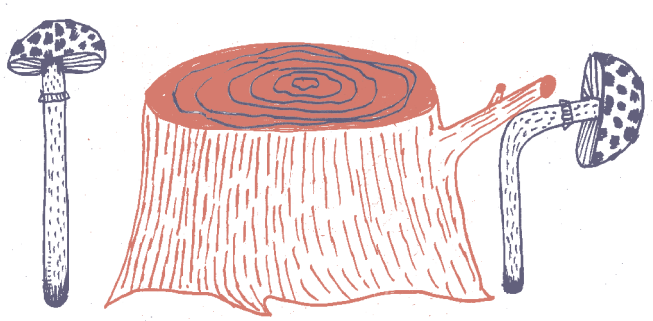
Development and functioning of an organism depend not only on its genotype. For this reason, designing a metabolic pathway requires considering a number of relationships between genes, proteins and environmental factors. Lactose operon is a simple example of how these relationships work.

☞ WHY THIS IS IMPORTANT

Drug action is influenced by many processes that are difficult to predict. Scientists hope that regulatory networks of genes will help to predict, prevent and diagnose diseases such as diabetes, cancer and cardiovascular diseases more efficiently.

☞ KEYWORDS

METABOLIC ENGINEERING, SYSTEMS BIOLOGY, BIG DATA, PERSONALIZED MEDICINE, GENOMICS, EPIGENETICS, LOGICAL GATES



In the previous chapters we have already seen how complicated gene regulation and signals transmission are. It is true for both intra – and intercellular signalling. Genetic interaction networks shape all the events in an organism lifetime. They also have their regulators, which we will discuss in more detail in the chapter on the design of genetic circuits in synthetic biology..

In this chapter, we hope to scare you a little with the overwhelming complexity of biological systems. However, shortly after that, we aim to calm your fear of genetically modified monsters running around the streets or the grim reality of Gattaca movie. First, some basics – what is a gene and how does it work? Let's look at operon lactose – a team of 3 genes regulated together. In this case, they are also related to each other according to what they do. (Fig. 1). It looks quite complicated, doesn't it? Making it even worse, in plants and animals, transcription and translation work differently and they are more complicated. Let's go to the "gene" definition then and hope it is clearer.

A gene is the most basic unit of heredity that determines the formation of protein or RNA recorded in the nucleotide sequence of the DNA. However, this definition been discussed since 1909 when it was introduced by a botanist Wilhelm Johannsen. In 2016, it turned out that genetic information is stored not only in the DNA sequence. It is also kept in DNA strand structure inside nucleus.

GENE REGULATION ON A LARGER SCALE

Genetic networks are thousands times more complex than a single operon. Fortunately, thanks to high throughput sequencing methods and development of big data analysis we are witnessing a huge progress in understanding of genetic interaction over the last 15 years.

Many researchers expect that the gene regulatory networks could help us find biomarkers (i.e. molecular markers, especially diseases). That knowledge could improve diagnosis of cancer, diabetes and other complex metabolic diseases. In the future, models of epigenetic network probably will take into account the impact of the environment on genotype as well. However, it is quite distant future. These research projects will surely be expensive and require global cooperation to create new drugs.

The biggest hope is the use of regulatory networks in medicinal drug design. On a larger scale, logic gates, similar to those used in electronic systems, will also be used. Regulatory networks may additionally vary in activity depending on different environmental conditions, organism status or age. Today, of course, there are huge networks being created, but the possibilities are much bigger. This approach allows biologists to block some biochemical pathways, increase the productivity of others, and obtain the desired substances in other organisms than those from which they are derived, e.g. human insulin in bacteria.

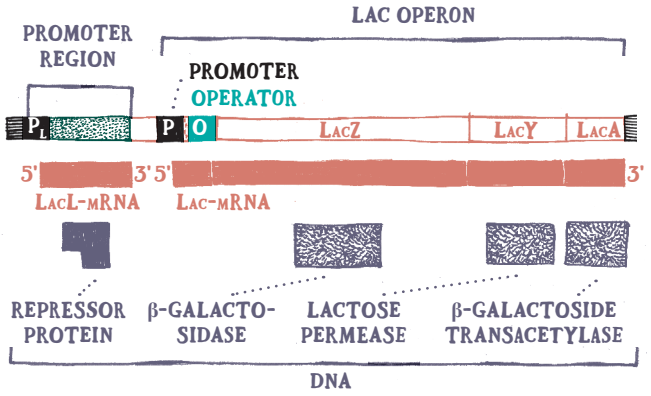


Fig. 1. Lactose operon of Escherichia coli. Let's look first at the promoter region (P). It regulates the start of transcription of all the other genes. The operator gene of the lac (O) is a place where lactose repressor protein binds in the absence of lactose (R). Then transcription of the genes is blocked. Gene lacZ encodes the enzyme β-galactosidase decomposing lactose to galactose and glucose. Both these sugars used as an easily available energy source for the bacterium. The gene product is secreted inside the cell, and therefore lactose transport to the cell is required. Gene lac Y encodes lactose permease, which is a transporter protein that transfers lactose into the cell. Gene lac A encodes galactoside o-acetyltransferase responsible for the transport of lactose molecules into the cell where it is hydrolysed. Neighbouring genes are not relevant at this point and are marked in grey). Operator is marked in green. It overlaps promoter, so repressor protein could physically block transcription. Now let's look at the coding region of the repressor protein on the left. Repressor promoter (P) enables gene transcription of repressor protein (lacI). DNA (blue line) is transcribed into mRNA (messenger RNA– red line). After translation from mRNA, lactose repressor protein binds to operator, so transcription enzyme is unlocked.

←

☞ IF YOU WANT TO KNOW MORE

1. http://sbrg.ucsd.edu/InSilicoOrganisms/OtherOrganisms
2. Silva-Rocha R., Lorenzo de D., *Mining logic gates in prokaryotic transcriptional regulation networks*, FEBS Letters, 2008, 582(8): 1237–44
3. Emmert-Streib F., Dehmer M., Haibe-Kains B., *Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks*, *Frontiers in ell and Developmental Biology*, 2014, 2:38, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4207011/

➡ **Fig. 2.** Interactom (means all the genetic interactions of a cell) of Treponema pallidum bacterium. Dots represent genes and lines represent interactions. It is estimated that the humans have roughly 20 000 genes and yeasts approximately 6 000 genes. Analysis of vast interaction networks allows to extract only the most important connections.

Why are cells so complex?

In March 2016, employees of J. Craig Venter Institute (JCVI) and Synthetic Genomics Inc. (SGI) announced that they were successful in creation of a bacterium JCVI-syn3.0. Its genome contained 531560 base pairs and only 473 genes, out of which 149 have an unknown function. A part of Minimal Cell Project was the creation of synthetic *Mycoplasma mycoides* bacterium genome. However, the main aim of the project was to test how many genes are really necessary for a bacterium to survive. By now, it is the smallest synthetic bacterium grown in the laboratory. Natural smallest bacterial genomes have around 0.13 million base pairs and the largest genome discovered was 14 million base pairs. However, most bacteria have genomes of more than 2 million base pairs and less than 5 million pairs.

Reading, modifying, and duplicating the genetic material of the cell is subjected to specific rules. There are three main levels, at which molecular interactions can occur (Fig 2.). When it comes to a design of an organism all of them have to be taken into consideration.

☞ IN BRIEF

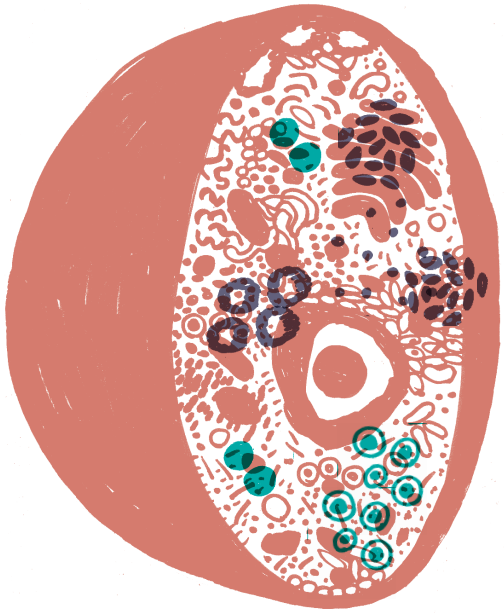
Inner environment of each cell is a very crowded and dynamic place. Despite this complexity, scientists have managed to synthesise DNA of bacterial and viral genomes. In the cell, DNA is transcribed (rewritten) into mRNA and finally the mRNA is translated into a protein.

☞ WHY THIS IS IMPORTANT

Many cellular processes vary between different cell types. Thus, it is difficult to prepare a good and versatile models of their functioning.

☞ KEYWORDS

MINIMAL CELL PROJECT, CENTRAL DOGMA OF BIOLOGY, TRANSCRIPTION, TRANSLATION



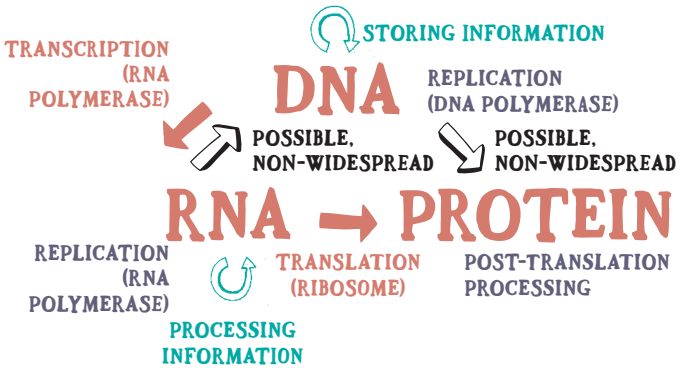
⬅ **Fig. 1.** Schematic structure of an eukaryotic cell

☞ IF YOU WANT TO KNOW MORE

1. *Fundamentals of Cell Biology*, pod red. B. Albertsa, 2009, PWN
2. *Kinesin Walking on a Microtubule*, www.youtube.com/watch?v=y-uuk4Pr2i8
3. *Biology Cell Structure*, https://www.youtube.com/watch?v=URUJD5NEXC8
4. *Fundamentals of Biology*, MIT, *OpenCourseWare*, https://www.youtube.com/watch?v=PzY0MWEEE6U
5. *Minimal Cell Project*, press note, http://www.jcvi.org/cms/fileadmin/site/research/projects/minimal-cell/minimal-cell-release.pdf

Designing organisms is a big challenge for synthetic biologists. What is more, eukaryotic cells are far more complex than their prokaryotic (bacterial) counterparts. Although we know how the basic building blocks work, the effects of these interactions combined often appear to be different from those expected. Recently many DNA fragments called "junk DNA" turned out to be of great importance for the cell, but still their actual function remains unknown (or too subtle to be assigned a specific function in the colloquial understanding of this word). Among these sequences, there are repeated sequences and inactive transposable elements (when they are active, they can "jump" through the genome – cut and paste in random locations). Also, even if synthetic genes can take over control and do the "magic" inside a bacterium, they must be somehow kept in its genome. This means they have to be not only useful for the bioengineer, but also for the bacterium, so that they are passed to future generations. A cell is like a huge city that works reasonably well, but you can never avoid traffic jams from time to time. In the cell, there is practically no free space (it is quite literally very crowded). For this reason, it is impossible to avoid interactions between nearby molecules.

Intracellular signal transmission occurs very rapidly as changes in response to stimuli from the environment happen constantly. Cell behaviour is even more complicated in the case of tissues and organisms – groups of cells are very different from each other. All these factors make the development of biotechnology into an expensive and time-consuming process requiring the cooperation of specialists from many different fields.



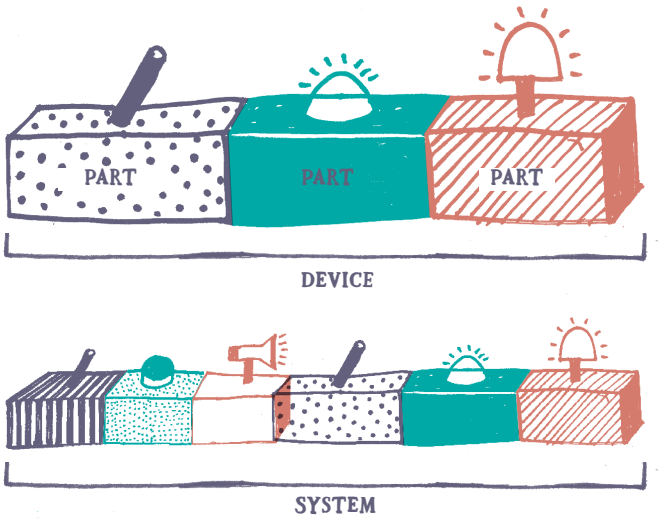
➞ **Fig. 2.** The central dogma of biology. In cells of a vast majority of organisms, DNA is transcribed into mRNA and then translated into protein. In the case of eukaryotic cells transcription occurs in the nucleus while translation takes place in the cytoplasm.

How to build a genetic circuit?

Synthetic biology is sometimes called 'turbo genetic engineering' by journalists. The main goal of this discipline is to standardise the design of biological systems using a set of rules that have been used in design of electronic circuits for a long time. In cooperation with programmers synthetic biologists have even created a brand new programming language called SBOL (Synthetic Biology Open Language). Its main purpose is to define genetic circuits, including their mathematical description.

Repressilator is an example of a small artificially-designed device (see Fig. 1). Optimal parameters of the repressilator system are selected by mathematical modelling, as the system itself is very sensitive to changes. Matching all parameters experimentally would be way too time-consuming and impractical.

The genetic puzzles (or bricks) – BioBricks™ are fragments of DNA having a standardised sequences at their ends. This is an easy way to allow using them in a variety of systems. BioBricks™ are placed on the so-called plasmid backbones. They can becompared to a memory card, which can store any files you want (Fig. 2).



← **Fig. 1.** Biological system consists of devices, which are composed of parts called BioBricks™. The figure shows modularity of such a system. This way of thinking about biology is a basic method of standardisation in chaotic biological world.

☞ IN BRIEF

This chapter contains basic information on the design of genetic circuits – biological systems that modify the functioning of the host organism, depending on the function assigned to them.

☞ WHY THIS IS IMPORTANT

Synthetic biology is all about designing genetic circuits, so the knowledge of their operation and design principles is crucial..

☞ IF YOU WANT TO KNOW MORE

1. Alon U., 2006, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman&Hall/CRC
2. Myers J.C., 2009, *Engineering Genetic Circuits*, Chapman&Hall/CRC

The basic techniques that are extremely useful in synthetic biology are: PCR (polymerase chain reaction) which is used to create libraries of DNA sequences, RNA interference, restriction endonuclease cleavage, site-directed mutagenesis and genome editing. In addition, synthetic biology make use some methods used in organic chemistry and material such as: oligonucleotide synthesis, DNA nanotechnology, atomic force microscopy (AFM), electron microscopy (EM), introduction of alternative base pairs, DNA modifications (including modifications of its phosphate backbone).

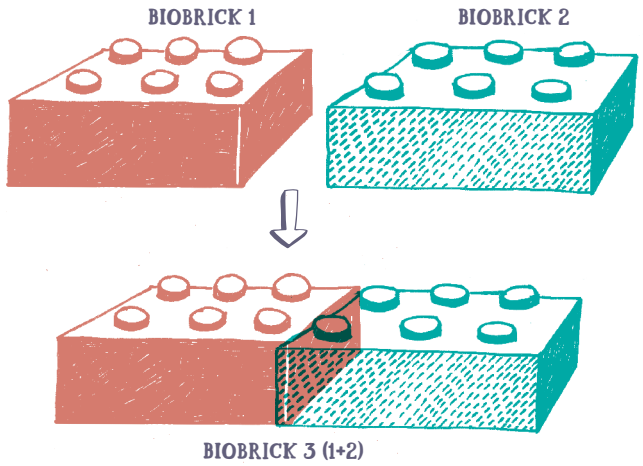


Fig. 2. Repressilator is a system based on three genes linked together in a feedback loop. Each gene represses the following gene, and itself is repressed by the gene proceeding it. Thus, visible oscillations can be followed by green fluorescent protein emitting light that increases and decreases in time. It looks

Fig. 3. BioBricksTM allow for rearranging complicated genetic systems, even those come from different species, or they are artificially created. Many of them were created and described during the international competition in synthetic biology called iGEM (International Genetically Engineered Machine), organized by the Massachusetts Institute of Technology in Boston.

Application of synthetic biology – outline

What I cannot build, I cannot understand – these are words of Richard Feynman, a brilliant American physicist. Another brilliant scientist, Craig Venter encoded these words in the genome of the first artificial bacteria. The above citation epitomises the major challenge of synthetic biology, which is to build a whole organism from the scratch. Going even further, to build an organism using alternative letters of genetic code, not present in the nature.

Also, many scientists hope to create safer vaccines and new antibiotics, which could protect us against mutating pathogens, e.g. a HINIFLU virus vaccine from Novartis. In 2015 the Nobel Prize in medicine went to Tu Youyou, a Chinese pharmaceutical chemist for discovering antimalarial drug called artemisinin. Naturally, it can be obtained from plant sweet wormwood (*Artemisia annua*). Because artemisin is not feasible to extract from natural sources (or to synthesise) an operon allowing production of this drug has been synthetically designed and introduced to *E. coli*.Giving another example, the mechanism of inhibition of B-cell maturation has been discovered using synthetic biology tools. Patients suffer from severe immunodeficiency because B-cell receptor is not properly located on the surface of B lymphocytes cells. During the study, B-cell maturation process was completely rebuilt from smaller fragments. Such approach helped to precisely identify the cause of the maturation block. The use of artificial genetic circuits has also made it possible to improve the treatment of certain bacterial infections with antibiotics such as quinolones or etionamide. There are also modified phage-destroying bacteria that form biofilms. Another interesting idea was to create adenoviruses carrying a synthetic operon, which function is to attack cancer cells. These cancer cells are deprived of the protein p53 being a repressor of that synthetic operon. Normal cells produce this protein, so the deadly operon remains inactive when adenovirus infect the cells.

In addition, designing synthetic pathways can be extremely promising when it comes to production of chemicals and drugs e.g. proteins having numerous posttranslational modifications. In that case efficacy can be levelled up, while maintaining or even lowering the total cost. However, the question is what organism should become the producer. Popular hosts are yeasts and *E. coli*. Nonetheless, proteins that require human post-translational

☞ IN BRIEF

Synthetic biology is believed to provide a real breakthrough in medicine, biofuels industry, chemical industry and agriculture. This chapter is aimed at reviewing of some applications of synthetic biology, and furthermore we would like to consider its possible effects on biological producers (called host organisms).

☞ WHY THIS IS IMPORTANT

Twenty-first century is the time of extremely rapid development of biomedical sciences. The pace of change is huge and it is definitely worth developing awareness about it. Biologists could not imagine more interesting times. The twenty-first century will undoubtedly be the age of development for synthetic biology. Drew Endy, a professor of biology at Stanford University, said that it synthetic biology is a science that can "change our relationship with nature". That is why it is extremely important to assure that the next brilliant and cost-effective technical solutions will include social and ethical issues. Only the cooperation of biologists, MDs, computer scientists, journalists, lawyers, entrepreneurs and authorities will allow safe, efficient and harmonious

development of this field. Science never exists in isolation from the socio-economical relations and cultural change.

☞ IF YOU WANT TO KNOW MORE

1. Khalil A., Collins J., *Synthetic Biology: Applications Come of Age*, *Nat Rev Genet*, 2010, 11(5): 367–379.
2. König H., Frank D., Heil R., Coenen C., *Synthetic Genomics and Synthetic Biology Applications Between Hopes and Concerns*, *Current Genomics*, 2013, 14, 11-24

processing are not suitable for the production in bacteria. The reason is they could (due to difference in posttranslational modification scheme) induce very strong immune response upon introduction to human system. The safest way is to use an organism having very similar biochemistry. Then the only challenge is to optimise the process. Unfortunately, optimisation is often a choice between diminished stability of the process in the new host and lower productivity in the natural organism. A common method to increase productivity is to change gene promoter to a stronger one – e.g. using the lactose operon. Alternatively, total production of the entire colony can be manipulated using more advanced genetic modifications.

As it was mentioned before, gene expression in bacteria, and in plants may differ. This is an important issue in case of synthetic rubber production. Synthetic rubber is a good alternative both to the petroleum-based rubber and the natural one produced from isoprene rubber tree. Unfortunately, necessary genes cannot be simply transferred from a rubber tree to E. coli, but with some pathway engineering even that formidable task is impossible.

There is a world of products, which price depend on the price of crude oil, so there is a huge interest in the industry of biofuels. One of recent achievements achieved by researchers from DSM company was to provide enzymatic cocktail, which breaks down mixture of lignin and cellulose (found in all plants). Then yeasts (created by tools of synthetic biology) ferment the mixture of those simple products into biofuel.
Another interesting idea came from researchers working at Purdue University. The goal was to produce some cheap biofuels by modifying yeasts to produce 40% more ethanol than usually. On the other hand, Lawrence Berkeley Lab focused on copying termites' metabolic pathways, as they could digest cellulose into simple sugars, and put that pathway into different host organisms.

DIY biology

The idea of DIY (Do It Yourself) covers a wide range of activities, from independent art and craft, and to electronic projects. Representatives of the DIY movement are largely motivated by values derived from the 'hacker ethic'. At this point, we must stop in order to clarify terminology. The terms hacking/hackers will be used here in the sense which largely differs from the one we usually encounter in mainstream media. We will not apply it to people who break into security systems in order to show off their skills nor to dangerous criminals, or even to the recent high-profile actions of idealistic "hacktivists". To understand this different, older (and also recognized as more correct) meaning of the term hacking, we need to go back to the 60' at MIT. It is there that we can trace back the roots of the movement that focused on creating, modifying, and sharing of new solutions in the field of programming. The theme of work wasn't as crucial as the manner in which it was conducted – through constantly testing the boundaries of what can be achieved, or coming up with a completely new, surprising applications for existing technologies.

Biohacking draws inspiration from that movement and seeks to provide basic tools of genetic engineering to hobbyists. Those hobbyists may be biologists involved in projects unrelated to their main research topic, or scientists from other fields, who were attracted the interdisciplinary character of synthetic biology. However, biohacking also attracts an ever wider circle of people unconnected with science – such as readers of popular science press or even artists seeking new forms of expression.

Many projects are conducted in cooperation with traditional research institutions, while one of the key objectives of DIY activity is to be independent. The first obstacle for an aspiring independent biohacker is the cost of laboratory equipment. While it is possible to obtain several key instruments (pipettes, thermocycler) for less than $1000 by buying second hand equipment on sites such as eBay (Fig. 1), this type of offer is still mostly inaccessible to people outside of the United States. Reagents are also a problem – their costs are often prohibitively high and the companies that produce them will sometimes refuse to deliver them to private homes. To overcome these obstacles, a number of initiatives were created. Some aim to provide low-cost substitutes for professional equipment,

☞ IN BRIEF

Can molecular biology be practiced as a hobby? What are the difficulties facing amateur biologists and what are their prospects? What is the exact meaning of DIY (Do-it-yourself) biology?

☞ KEYWORDS

BIOART
area bordering on biology and art. It applies the tools of modern biology as a form of artistic expression.

REGISTRY OF STANDARD BIOLOGICAL PARTS
an open repository of genetic elements that can be used for projects in the field of synthetic biology.

☞ IF YOU WANT TO KNOW MORE

1. Ledford H., 2010, *Life Hackers*, *Nature*, 467
2. France G., 2011, *A Lab of Their Own*, *Science*, 333
3. Stallman R., *Hackers – Wizards of the Electronic Age*, an interview
4. Berg, P., Singer, M., 1995, *The recombinant DNA controversy: twenty years later*, PNAS, 13(10), 1132–4.

with the use of open standards (often sold as kits to be assembled by the buyer), others take advantage of the increased availability of 3D printing.

Recently OpenBiotechnology appeared on the market. It is a company that provides relatively cheap reagents, enzymes, plasmids and cell lines for amateur biologists. It is also worth to mention groups such as GeneSpace and BioCurious, which organise workshops for future biohackers and provide laboratory space for independent projects. The community of DIY biologists also benefits from the Registry of Standard Biological Parts operated by MIT.

Matters of biosafety, legal restrictions and social prejudice are a whole different issue. Experiments on microorganisms carried out independently of major research institutions are a relatively new phenomenon and, as such, can cause violent reactions on the part of government agencies. Most biohackers have not met with any drastic repressions, although even as recently as 2010, Jason Bobe (organizer of the first DIYbio rally) joked that of the 2000 subscribers on the DIYbio mailing list "30% are spammers, and 70% are government agents." Over time, FBIS interest in biohacking movement decreased, partly due to active measures biohackers themselves had undertaken to control each other (an approach called "neighborhood watch"). However, the legislation of many countries still does not allow performing the simplest of genetic manipulations (such as introducing a plasmid into a bacterium) outside of research units specifically designed for that purpose.

It seems that at least in the near future, fear of genetic engineering may be one of the major obstacles to the development of DIY biology. On the other hand, it is must be remembered that as recently as in the 80', there was considerable fear of computer technology. While today Hollywood studios like to scare the audience with the monstrous effects of genetic manipulation, in that time they relied on apocalyptic consequences of computer hacking (in the colloquial sense). In those times few people suspected that a personal computer can be useful for someone aside from specialists and a handful of enthusiasts. Does this mean that the mini-biolabs will one day be as common as personal computers are now? It is difficult to predict. In the near future a much more likely scenario is that

synthetic biology will join the ranks of fields such as astronomy and ornithology, where no one questions the importance of the contribution by passionate hobbyists.amatorów.

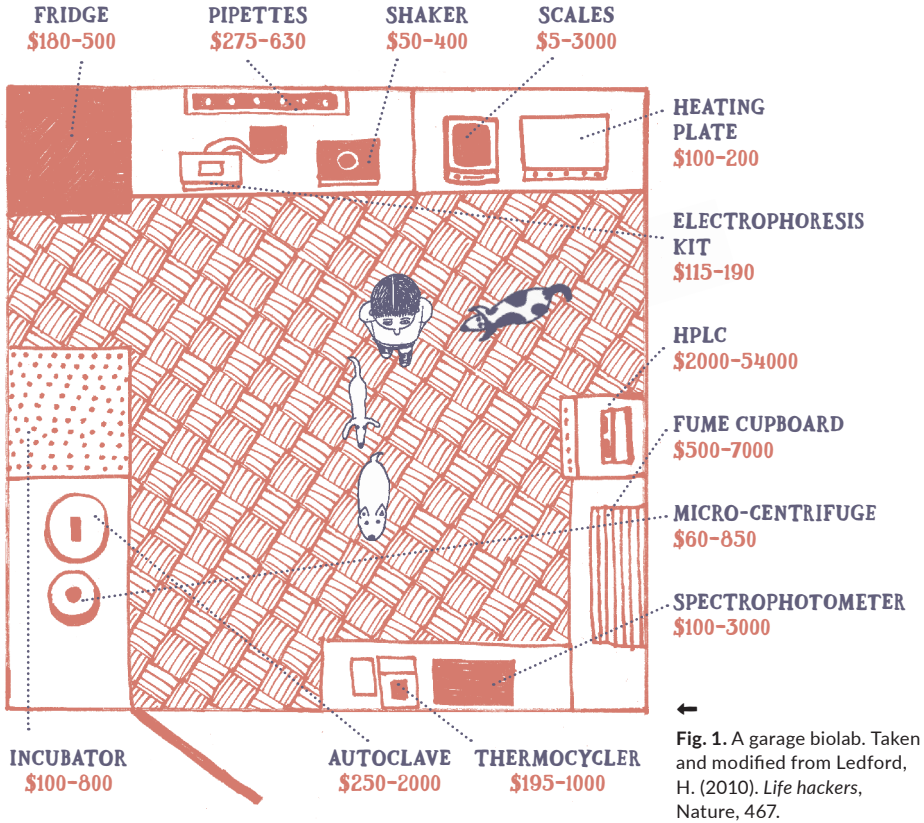


Fig. 1. A garage biolab. Taken and modified from Ledford, H. (2010). *Life hackers*, Nature, 467.

Ethical issues of synthetic biology

☞ IN BRIEF

The rapid advances in genetic engineering paved the way for its commercial application. Genetically modified organisms (GMOs) found applications in the preparation of medications such as insulin. Moreover, in many countries they are used in the form of GM crops. In spite of that, many people mistrust genetic engineering, often due to lack of access to reliable scientific information about GMO. Another issue that aroused a lot of controversy is the potential use of genetic engineering to modify the human genome. This issue has yet to be resolved, however, in the imminent future it may have the utmost importance for us.

Some aspects of modern molecular biology are controversial for certain groups of people. Most of the controversy is caused by lack of understanding of scientific concepts, such as molecular cloning, or even the very definition of a genetically modified organism. However, some technologies (and their far-reaching consequences) may be subjected to serious ethical discussion. One of the most significant sources of controversy, concerning not only synthetic biology but entire genetic engineering is genetically modified food (GM food). The debate focuses on the use of food and other products derived from genetically modified crops as well as general use of genetic engineering in food production. The debate involves consumers, farmers, biotech companies, government officials, non-governmental organizations and the scientists themselves. Key questions need to be addressed are: 1) whether such products should be labelled 2) what is the role of government regulators 3) how to conduct objective scientific research and publish the results 4) what is the impact of GM crops on human health and the environment (including the impact on resistance to pesticides) and 5) what is the socio-economic impact of such crops. It also needs to be remembered that GM organisms already play an important role in the production of biofuels and pharmaceuticals such as insulin. Some people think that eating GMOs can be harmful, but there is no scientific evidence that such food can have any detrimental effect on human health. All research concerning safety of genetically modified foods is carried out by relevant regulatory authorities. It starts with an assessment of whether the tested food is synonymous with its non-GMO counterparts, which have already been found suitable for human consumption. So far, there are no documented reports of any illnesses caused by the consumption of genetically modified food. Hence, there is a scientific consensus, saying that currently available GM food products do not pose more threat to human health than conventional food. Despite this, every product derived from GMO available on the market must be thoroughly tested for possible harmful effects on human health. Other potential threats caused by the presence of GMOs in the biosphere are: potential crossbreeding of genetically modified and non-genetically modified organisms, contamination of natural food by GMOs and the impact of GMOs on the natural environment.

Some radical groups and social movements such as Greenpeace constantly try to manipulate public opinion in order to spread false information claiming that the aforementioned risks were not properly identified and studied. In addition, these groups question the objectivity of the regulatory authorities and scientific institutions, which deal with genetic engineering. Because of these harmful activities, more and more people start to see GM food as unsafe. Legal status and regulations of GM food may be different in different countries, some of them prohibit or restrict the cultivation of GMOs and studies on them.

Genetic engineering can be theoretically used to modify the physical appearance or metabolism of human being and even improve their physical or mental capabilities. The debate on the engineering of human embryos has led to proposing several possible approaches to tackle this problem: 1) each foetus has a right to remain intact – no genetic modification 2) parents have the right to genetically modify their offspring 3) every child has the right to be born free of genetic diseases that can be prevented by the use of genetic engineering. The last point seems to be very relevant. Hence, many scientists believe that some restrictions should be introduced, but engineering germ cells and embryos must not be prohibited. Genetic engineering of adult people can be seen as an health-improving technology similar to diet, physical exercises, education, use of cosmetics or plastic surgery. Possible regulatory systems include a total ban, accessibility for all interested people or partially restricted access to that technology. For example, the American Medical Association's Council on Ethical and Judicial Affairs stated that "genetic interventions to enhance traits should be considered permissible only in severely restricted situations: 1) clear and meaningful benefits to the foetus or child; 2) no trade-off with other characteristics or traits; and 3) equal access to the genetic technology, irrespective of income or other socioeconomic characteristics".

From the very onset of biotechnology, there were group of scientists protesting against any attempt to modify the human germ cells. This approach is still common in most countries. With the advent of new techniques like CRISPR, in March 2015 a group of scientists urged a worldwide moratorium on clinical use of gene

☞ KEYWORDS

GENETICALLY MODIFIED ORGANISM (GMO)
any organism whose genetic material has been altered using genetic engineering techniques.

GENETICALLY MODIFIED FOOD
any kind of food made from genetically modified organisms.

GERM LINE
a group of cells being sufficiently differentiated that they can pass the genetic material to the progeny.

GENE THERAPY
herapeutic delivery of nucleic acid molecules (DNA or RNA) into cells of the patient to treat a given condition.

GENE EDITING
a type of genetic engineering, wherein the genomic DNA is modified, deleted or replaced using molecular biology tools.

GENE DOPING
a hypothetical application of gene therapy to improve athletic performance in competitive sports events.

CRISPR
a system used by prokaryotic organisms to defend from bacteriophages and other forms of foreign germs. Currently used as an effective method of editing genomes.

editing technologies to modify the human genome in a way that could be inherited. Importantly, in April 2015 a huge controversy sparked after the publication of research the DNA of non-viable human embryos using CRISPR. These studies did not lead to the creation of live embryos having modified genomes.
Currently, there are concerns that some athletes may benefit from gene therapy to improve performance. Up to now there was no documented case of such gene doping. Some ethicists even believe that gene doping could potentially increase the overall performance of athletes and should be legal (if everyone will have equal access to this technology). On the other hand, critics argue that any therapeutic intervention to enhance the athletic performance compromises the foundations of ethical sport and therefore should be banned.

Biological chemistry

Biological chemistry (often also called chemical biology) is a scientific discipline on the intersection of chemistry, biology and physics. As in the case of synthetic biology, biological chemistry is one of several interdisciplinary disciplines aimed at providing a more holistic description of living organisms. Biological chemistry involves the use of chemical techniques and compounds obtained by organic chemists, in order to explore and modify biological systems. This approach stemmed from medical chemistry and bioorganic chemistry but it also took some elements of pharmacology, genetics, biochemistry and metabolic engineering. Research carried out by chemical biologists is often more related to cell biology than typical biochemistry. Biochemists are usually interested in studying chemistry of biomolecules and regulation of biochemical pathways, while scientists working in biological chemistry put more emphasis on using new chemical reactions and compounds in biological research. Biological chemistry also seeks the answers to purely biological questions related to functioning of living systems at the molecular level (e.g. regulation or monitoring of metabolic processes in a cell using artificially introduced compounds). In contrast to the biochemistry or synthetic biology, where the main methods of probing and modulating living systems are genetic techniques, biological chemistry research is based on the use of specific organic compounds and reactions. In spite of the fact that most research in the field of biological chemistry concerns biomolecules, this discipline often refers to synthetic biology. This is due to the fact that all biologically relevant molecules can be chemically modified to create a "synthetic" living system that has some properties designed in advance by researchers. For example, molecular recognition of DNA is mainly based on the phosphate backbone which means that nucleotides may be chemically modified without altering the physiological role of DNA. Few years ago, one research group created the expanded genetic code with eight new base pairs, which can be incorporated into DNA by PCR reaction. Innate amino acids can be regarded as somewhat poor building blocks for protein molecules. They do not operate independently of each other and, the relationship between linear amino acid sequence and spatial structure of the protein is quite complex and still poorly understood. Chemical biology aims at modeling,

☞ IN BRIEF AND WHY THIS IS IMPORTANT

Biological chemistry is a relatively new interdisciplinary field that has an ambitious goal: to use of existing compounds and chemical reactions in order to study the living organisms at the molecular and cellular level. Chemical biology techniques allow us not only to better understood the cell biology but also to perform subtle manipulations of cellular physiology. Similarly to synthetic biology, biological chemistry also relies on engineer-like approach. The combination of these two disciplines can open new possibilities in the creation of artificial biological systems as well as modifying existing organisms.

☞ KEYWORDS

MOLECULAR RECOGNITION
specific interaction between two or more molecules occurring through non-covalent interactions such as hydrogen bonding.

BIOORGANIC CHEMISTRY
a branch of chemistry that merge elements of organic chemistry and biochemistry.

MEDICAL CHEMISTRY
a discipline between synthetic organic chemistry and pharmacology aimed at the design, preparation, and commercialization of medicinal drugs.

☞ IF YOU WANT TO KNOW MORE

1. Miller A.D., Tanner J., *Essentials of Chemical Biology: Structure and Dynamics of Biological Macromolecules*, 2008, Wiley

design and synthesis of chemically modified peptides to evaluate their biological functions. What is more, fully formed proteins can be combined in a number of new ways which may allow the formation of protein complexes not seen in the nature. These complexes could also find potential applications. One can easily imagine the economic benefits from the use of technologies which use the chemically modified proteins to prepare drugs or new materials in a more efficient way. Finally, the signaling pathways can be turned on or off by adding a specific organic compound. It creates a whole new approach to studies on genetic circuits in synthetic biology, and significantly expands the range of tools available for biologists interested in synthetic biology.

Nanobiotechnology: design of biological macromolecules

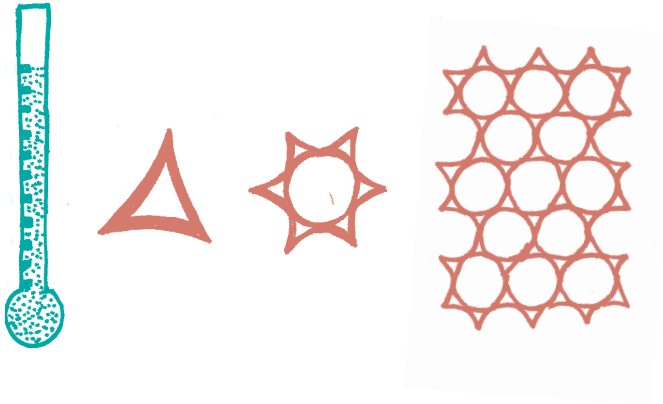
If DNA and RNA sequence can be modified with new, non-naturally occurring nucleotides, why not trying that with a protein? There is a plethora of amino acids that are not found in living organisms. How can these changes affect protein structure and functions? What part of it can be freely "spoiled" without any consequences? Can two or more proteins be joined together? Can we just add one synthetic part to a natural protein backbone? Two main approaches can be used in protein engineering: rational design or directed evolution. The first one is based on previously mentioned site-directed mutagenesis. The idea is to swap the amino acids at selected locations based on the knowledge of the sequence of the protein. Unfortunately, it is not the ideal solution, because not all proteins have known structures. Thus, not always can viable computational models be created. But there is another way, called targeted evolution. In this approach amino acids are mutated randomly and only proteins displaying the desired properties are for selected. The process is repeated until optimal results are obtained. An example of a protein designed from scratch is *Top7*. It is stable protein, and it is composed of 93 amino acids. However, its structure does not resemble any other existing proteins and it has no biological function. Artificial proteins that are really useful, belongs to restriction enzymes family. These enzymes act just like molecular scissors for nucleic acids and they were obtained using directed evolution method.

☞ IN BRIEF

Synthetic biologists do not only design nucleic acids and protein sequences to store information. These macromolecules can be also used as a building material. The structure and function of the protein can be altered by substituting amino acids in a processes called site-directed mutagenesis and through targeted evolution. It is also possible to use some of their biologically active components, as mimicked by peptidomimetics. Interestingly, DNA and stable forms of RNA can form various 2D and 3D structures, which are used in medicine and electronics.

☞ WHY THIS IS IMPORTANT

Currently, most patented drugs are biological macromolecules. Often synthetic elements of proteins show some interesting physicochemical properties and biological activity not known before.



←

Fig. 1. Polymeric RNA nanostars are resistant to high temperature generated by prof. Peixuan Guo and William Farish of the UK College of Pharmacy and the Markey Cancer Center.

☞ KEYWORDS

SITE-DIRECTED MUTAGENESIS, EVOLUTION-ORIENTED, PEPTIDOMIMETICS, HYBRIDIZATION OF DNA, DNA NANOTECHNOLOGY, RNA NANOTECHNOLOGY

☞ IF YOU WANT TO KNOW MORE

1. Luisi P.L., Chiarabelli C. eds., *Chemical Synthetic Biology*, Wiley 2011

2. H. Li et al, *RNA as a Stable Polymer to Build Controllable and Defined Nanostructures for Material and Biomedical Applications*, *Nano Today*, 2015 10, 631–655

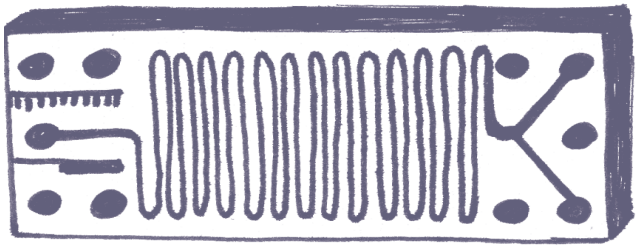
3. http://seemanlab4.chem.nyu.edu/, a website of Ned Seeman, who was the first in 1982 who suggested the use of DNA as a materia

4. phys.org/news/2014-04-team-potential-rna-heat-resistant-polymer.html

Peptidomimetics are also increasingly popular in drug design. They "pretend" to be a part of a protein to join the natural receptor and induce a cellular response. Their advantage is that they are easier to obtain (faster chemical synthesis) and stronger than natural proteins.

Engineering of the DNA structure is less expensive than protein engineering and DNA modelling is also easier thanks to the reduced number of components: there are 20 natural amino acids building up proteins and only four nucleotides building DNA. Nanobiotechnology is a branch of science aimed at using nucleic acids as a building material to design 2D and 3D structures. The fact that nucleic acids carry genetic information can be neglected here. DNA has been used as a building material for the first time in 1982. Currently, there are two approaches to building such nanostructures.. In the first one, a given DNA structure is used as a template for nanoparticles, e.g. gold. What's important, this method is highly specific, and its DNA-nanoparticle connections are reversible. The second approach, relies on assembling various spatial structures using hybridisation between at least two different DNA strands. So far a number of two-dimensional and three-dimensional polyhedral structures resembling origami, nanostars, bricks and tiles was obtained. These structures have a lot of potential applications starting from drug delivery to electronic microchips.

Design of spatial RNA structures may seem like a huge challenge because of its high susceptibility to degradation, in comparison to the DNA. Fortunately, stable forms (even at high temperatures) of RNA have already been obtained. Microsponges, beams, polygons, and labels can be constructed using these RNA forms. Such polymeric (DNA is also a polymer) structures can be an interesting subject of study from a material engineering point of view.



➡

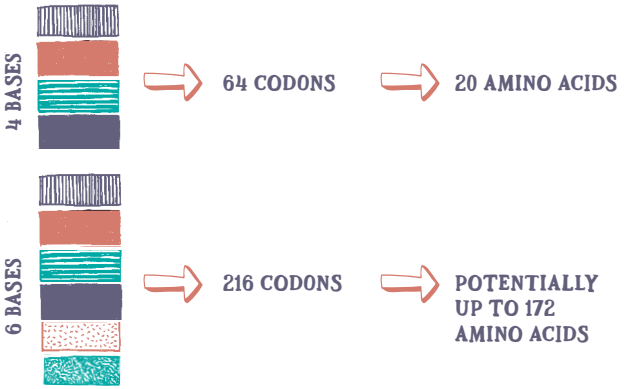
Fig. 2. A two-dimensional DNA mesh that may be used to develop a lab-on-a-chip assay to select macromolecules according to their size.

Modifications of nucleic acids and extended genetic code

The DNA sequences of all living organisms contain only four bases that can form two base pairs A-T and C-G. Triplets of bases within a single DNA strand encode proteins. Protein sequence, in turn, consists of 20 biogenic amino acids. Expanding the number of available base pair and amino acid would permit us to create completely new protein molecules with unprecedented properties.

Numerous research groups have been working on the discovery of unnatural base pairs. The first study of its kind was conducted in the late 1980s, but it was only in 2014 that researchers managed to introduce an unnatural base pair into a DNA sequence (along with standard base pairs) in such a manner, that it underwent stable replication in a bacterial cell. An extra base pair increases the theoretical number of different amino acids that can be encoded in a DNA sequence to 172.

Unnatural base pairs are therefore a milestone in the development of expanded genetic code, but they are not indispensable whatsoever. Since the genetic code is degenerate (i.e. more than one codon corresponds to a single amino acid) and there are three STOP codons, it is possible to assign a one codon to a non-standard amino acid (Fig. 1). Using this method, researchers have been able to add up to 71 non-standard amino acids into proteins synthesized by bacterial, yeast and mammalian cells. However, this is not an easy process. It requires modifying basic elements of cellular machinery – tRNA molecules and their corresponding aminoacyl tRNA synthetases (enzymes that add amino acids residues to their corresponding tRNAs).



☞ IN BRIEF

Synthetic biology has reached the stage where it is possible to modify the basic cellular machinery in order to create biomolecules never before present in nature.

☞ KEYWORDS

UNNATURAL BASE PAIRS
base pairs other than those normally found in DNA.

BIOGENIC AMINO ACIDS
20 amino acids that make up proteins in all living organisms.

☞ IF YOU WANT TO KNOW MORE

1. Malyshev D.A., Chen T., Dhami K., Lavergne T., Foster J.M., Corrêa I.R., Dai N., Romesberg F.E., May 7 2014, *A Semi-Synthetic Organism with an Expanded Genetic Alphabet*, *Nature*, 509: 385–8. Doi:10.1038/Nature13314

2. Wang L., Brock A., Herberich B., Schultz P.G., April 2001, *Expanding the Genetic Code of Escherichia Coli*, *Science*, 292 (5516): 498–500

3. Xie J., Schultz P.G., 2005, *Adding Amino Acids to the Genetic Repertoire*, *Current Opinion in Chemical Biology*, 9 (6): 548–54

⬅

Fig. 1. Possible further expansion of genetic code by introduction of an unnatural base pair.

Big dreams – synthetic genomes and artificial life

☞ IN BRIEF

Large-scale DNA synthesis and methods for assembling long stretches of DNA have recently enabled researchers to create entire synthetic genomes. How far are we from creation of synthetic life?

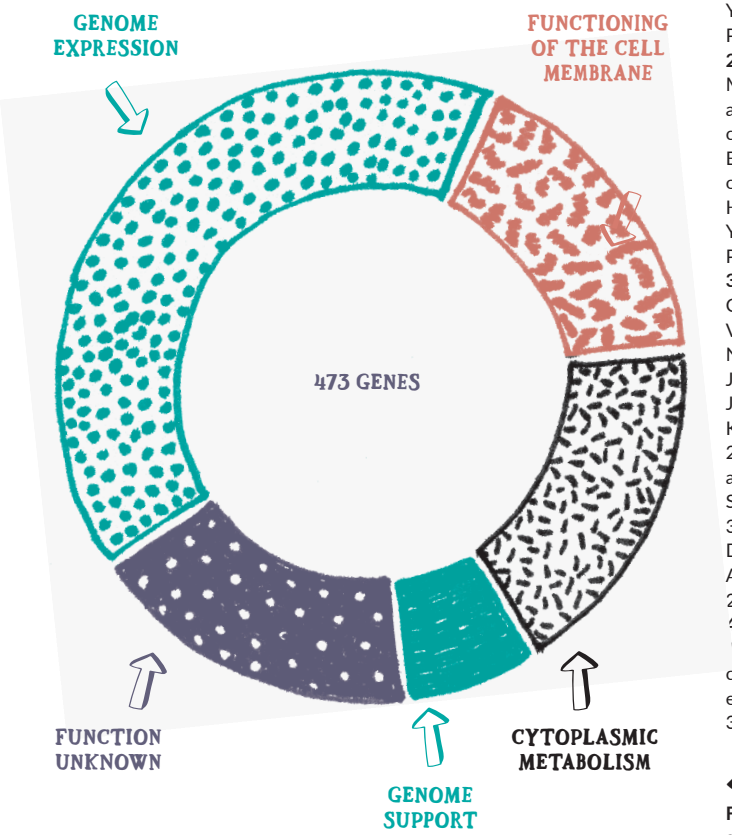
☞ KEYWORDS

SYNTHETIC GENOMICS
field of synthetic biology dedicated to creating new genomes, based on those already existing in nature.

MINIMAL GENOME
a genome comprising the smallest number of genes enabling the survival of the bacterial cell line.

Creating a synthetic genome, even for the simplest single-celled organism was until recently an unimaginably difficult goal to achieve. Inserting that genome into a living cell, so that it could begin function and replicate, was also considered a great challenge. Both of these tasks were undertaken by the Craig Venter Institute. The bacterium *Mycoplasma genitalium* was chosen as the object of study as at the time it had the smallest number of genes among known organisms. The goal of the project was to create a synthetic minimal genome. A minimal genome would be a milestone in the development of synthetic biology. While a bacterium with a minimal suit of genes would not necessarily be the most efficient tool for biotechnology, extreme simplification of genetic interaction network could facilitate efforts aimed at the rational design and creation predictable biological machines from scratch. In 2016, Craig Venter and his team managed to create a synthetic genome which contained 473 genes (*Mycoplasma genitalium* initially had 482 genes), based on the genome of *Mycoplasma mycoides*, and insert it into a cell of *Mycoplasma capricolum*, creating a new synthetic organism called Syn 3.0 (Fig. 1). That bacterium, along with its minimal genome has been able to survive and divide. It should be emphasized that while the genome Syn 3.0 was entirely synthetic, the cellular machinery that enabled its replication and expression was not. Syn 3.0 cannot therefore be considered an example of fully artificial life. In the meantime, scientists strive to create an even further reduced genome based on the genome of *Mycoplasma genitalium*, potentially narrowed down to 382 genes. The synthetic organism thus obtained is to be named *Mycoplasma laboratorium*.

While on the subject of additional base pairs and non-standard amino acids we should also mention XNA – nucleic acids where non-canonical sugars form the scaffolding of the molecule. These XNA molecules can (in theory) incorporate both natural and unnatural base pairs. They differ from DNA and RNA in stability and cannot be replicated by standard RNA or DNA polymerases, thus remaining "invisible" for the unmodified cellular machinery. "X" in the XNA stands for "xeno" or foreign. Researchers managed to create a synthetic replication machinery, capable of replicating certain types of XNA, making them an alternative template for the transfer of genetic information. XNA is therefore extremely interesting subject of research – both from the perspective of synthetic biology and medicine (some XNA can interact with the natural nucleic acids), and evolutionary biology (as a model of alternative paths that life might have taken at the earliest stages of evolution).



☞ IF YOU WANT TO KNOW MORE

1. Hutchison CA, Montague MG, 2002, *Mycoplasmas and the minimal genome concept*, Molecular Biology and Pathogenicity of Mycoplasmas (Razin S, Herrmann R, eds.). New York: Kluwer Academic/ Plenum. pp. 221–54.
2. Hutchison CA, Montague MG (2002). "Mycoplasmas and the minimal genome concept". Molecular Biology and Pathogenicity of Mycoplasmas (Razin S, Herrmann R, eds.). New York: Kluwer Academic/ Plenum. pp. 221–54.
3. Hutchison, Clyde A.; Chuang, Ray-Yuan; Noskov, Vladimir N.; Assad-Garcia, Nacyra; Deerinck, Thomas J.; Ellisman, Mark H.; Gill, John; Kannan, Krishna; Karas, Bogumil J. (2016-03-25). "Design and synthesis of a minimal bacterial genome". Science. 351 (6280)
3. Longest Piece of Synthetic DNA Yet, Scientific American News, 24 January 2008
4. Pinheiro, Vitor B. et al, Synthetic genetic polymers capable of heredity and evolution", Science, 2012, 341–344.

Fig. 1. Syn 3.0 genome along with gene functions (According to Thomas Shaffe [2] Creative Commons 4 International)

Future perspectives

☞ WHY THIS IS IMPORTANT

In this final chapter we would like to present some interesting perspectives for the future development of synthetic biology and related disciplines. Mayba it could also be interesting to ask a few questions about the synthetic biology itself as well as its possible impact on our lives in the future: What makes synthetic biology very distinctive among other disciplines of biology and biotechnology? Which future applications could be realised through synthetic biology? What can you expect from synthetic biology taking into account both benefits that it provides and the threats that it creates? Now, we can indicate several possible directions for synthetic biology, which we found either interesting or promising.

IMPROVED BIOSYNTHESIS OF NATURAL COMPOUNDS

Natural products still play a main role both in the search for new drugs and basic medical research. Nonetheless, recent advances in chemistry and metabolic engineering allow us to introduce advanced modifications to naturally existing pathways. It may offer new opportunities for obtaining analogues of natural products. Examples include 1) extended substrate specificity of enzymes, 2) transposition of metabolic pathways of natural products from one organism to another. In the near future we may expect that metabolic engineering of natural products will be further improved, mainly owing to advances in biological chemistry and protein engineering.

NEW STRAINS OF MICROORGANISMS AND THEIR APPLICATIONS

Microbial production has a significant advantage over chemical catalysis: it uses thousands of enzymes already adapted by nature to efficient production of a wide range of chemical compounds. Therefore, in recent years, we put more emphasis on the production of biofuels and other materials using bacteria and other microorganisms.

Advances in metabolic engineering have increased the range of available chemical products obtained from biological processes, including drugs, precursors of plastics and biofuel component. However, there are still some problems to solve. Naturally existing enzymes are often sensitive to temperature and pH, and what is more a number of them do not function well in organisms commonly used in the industry. Introduction of synthetic metabolic pathways to model organisms is a possible solution to this problem. Artificially created metabolic pathways must fulfill a number of criteria, such as the maximum use of already existing metabolic reactions, minimal number of synthetic steps and maximization of the yield. In an ideal world, synthetically designed strains should behave in accordance with the results of previously conducted simulation of the metabolism. This requires, most of all, accurate and reliable control of gene expression. Fortunately, many genetic elements have already been characterised well. The pitfall is, however, that some properties of these elements are often different in distinct contexts (other host organism, environmental conditions, etc.). Therefore, refining microbial production usually requires a lot of experimental work.

A significant body of work has been focused on the introduction of foreign enzymatic pathways to model organisms. Another important goal is to introduce new organisms, which have some desirable features (ease of cultivation and efficient metabolism) to the industry. The real challenge, however, is the introduction and expression of foreign metabolic pathways in a chosen organism, and simultaneous elimination of unwanted metabolic pathways existing in that organism. Fortunately, a large number of genes encoding desired features (i.e. enhanced biofuel production or increased tolerance to toxic by-products) has been identified. Furthermore, now we can more and more precisely regulate the expression of metabolism-controlling genes more and more precisely.

In the future, we may expect creation (and detailed characterization) of large libraries containing modified metabolic networks and even whole genomes. Therefore, further development of methods being able to analyse large data sets will be necessary in order to extract valuable information. Another essential goal is to systematically eliminate unwanted enzymatic reactions lowering the yield of the synthesis of target compounds.

XENOBIOLOGY

Xenobiology (XB) is part of synthetic biology (this term is derived from the Greek word *xenos*, which means "alien"). XB describes hypothetical forms of metabolism which do not exist in nature. From practical point of view xenobiology focuses on research at the molecular level. Until recently, this discipline was considered to be rather theoretical, but in the recent years we have created the first biological systems thatcan be described as "xenobiological". One such example (which was described in a previous chapter) is the use of XNAs as carriers of genetic information. The far-reaching perspectives of the discipline (much more than in the case of other branches of synthetic biology) could undermine the existing legal regulations for genetic engineering. The current legal system is adapted only to genetically modified organisms and does not take into account organisms which have artificially created systems of gene expression or an alternative metabolism. Fortunately, organisms that are truly biochemically “alien” most likely will not be created within next few years, so scientist and policymakers should have enough time to prepare themselves for the upcoming challenges.

SOME FURTHER CHALLENGES AND PERSPECTIVES

Synthetic biology started to develop rapidly at the beginning of this century. The development of modern genetic engineering techniques have enabled the design and creation of the first genetic circuits. Further progress in the field of synthetic biology resulted in the creation of more advanced biological systems such as synthetic genomes or XNA-based metabolism. Some scientists and engineers believe that synthetic biology will help in solving many global challenges, such as access to cheap energy, the treatment of tropical diseases or remediation of contaminated areas. At the moment, we are unable to give an answer whether solutions for these problems will be found soon. But we can be sure that socio-economic consequences of further development of synthetic biology will be the subject of a fierce scientific debate.

The brochure was co-produced by iGEM Team Warsaw (Aleksandra Bartosik, Marcin Ziemniak, Jakub Piatkowski) and the biological laboratory of the Copernicus Science Center (Elżbieta Turek, Stanisław Łoboziak), within the SYNENERGENE program – Responsible SPACE Research and Innovation (RRI) in Synthetic Biology.

This publication aims to popularize the knowledge about synthetic biology. This is a new fast-growing scientific discipline that was inspired by the idea of designing and creating artificial biological systems. They can be modeled on systems that occur in nature or built from the scratch using genetic engineering.

Authors
☞ Aleksandra Bartosik
☞ Marcin Ziemniak
☞ Jakub Piątkowski

English version edit
☞ Grzegorz Piątkowski

Design and layout
☞ Katarzyna Walentynowicz

COPERNICUS SCIENCE CENTRE
synenergene
Responsible Research and Innovation in Synthetic Biology