Aalto University
School of Chemical Engineering

Laboratory course in Biosystems and Biomaterials

Theory: Part II and III

September 26th – December 1st, 2023

Organizers and instructors:

Karoliina Elving (Biomolecular materials)

Mengjie Shen (Biomolecular materials)

Salla Koskela (Biomolecular materials)

Laura Niemelä (Molecular Biotechnology)

Heli Viskari (Molecular Biotechnology)

Rahul Mangayil (Biomolecular materials)

Alexander Frey (Molecular Biotechnology)

Department of Bioproducts and Biosystems | Aalto University

Table of Contents

1. Overall aims of the laboratory course (Parts II and III)

During the course, you will obtain an overview and hands-on experience on many commonly used laboratory techniques.

In part I you will learn about BioBricks and Synthetic Biology. We will explore the potential of BioBricks and how the use of standardized building blocks might revolutionize genetic engineering. In these experiments you will be familiarized with some of the most common molecular biology techniques.

The part of the course will introduce the yeast *Saccharomyces cerevisiae* as a model for studying conserved cellular processes in eukaryotes and as production organism. In the experiment we will show how genetic engineering can be used to generate improved cellular factories.

The experiments cover a full cycle starting from strain engineering to the analysis of the engineered strains. You will learn how specific aspects of cellular production systems can be tested and targeted for improvements.

2. Research topics addressed in the experiments.

Part II: BioBricks for assembling genetic constructs.

We will use BioBricks to learn what standardization in Molecular biology means and how standardized genetic parts can be utilized to assemble new functional units.

Part II: Yeast as an expression system

We will analyze if and how Saccharomyces can be genetically modified to enable production of recombinant proteins and how productivity might be improved.

3. Organization of the course

The experimental work is carried in groups of two students. Parallel to the experiments, we will have lectures/seminars. During the lectures we discuss the experimental techniques and approaches, look at your results and possible problems.

4. Timetable

The course timtable is indicated in MyCourses. Note we will not use all the reserved time slots!

- Several days will not be used at all.
- On several days we will be able to finish the work before the official end of the lab sessions. However, please note that the indicated times are only estimates!

5. Health & Safety

During the course, students will work with, and create genetically modified organisms. All the included experiments can be carried out in a biosafety level 1 laboratory. During practical work in the laboratory, it is mandatory for students to wear a lab coat, safety goggles, protective gloves, and facemasks. Also, good care needs to be taken to ensure a clean workplace, and problems or a spilling should be reported immediately. Some of the experiments may include strong acids as reagents and chemical mutagens.

SYBR Safe® DNA gel stain has emerged as a safer substitute for EtBr. In a series of mutagenicity tests, including the Ames test, involving various strains of *Salmonella typhimurium* SYBR Safe® caused substantially fewer mutations than EtBr. However, all DNA intercalating substances are carcinogenic, therefore special care is required for working with SYBR Safe. Also, during handling of the scalpels used for cutting out gel pieces, students need supervision and scalpels should be disposed immediately after usage.

4. Introduction

4.1. *S. cerevisiae* as a model eukaryotic cell

Yeasts are unicellular microorganisms classified in the kingdom Fungi. The term yeast is commonly used to refer to the species *Saccharomyces cerevisiae* (Fig. 1). However, more than 700 other yeast species have been described. Most of them belong to the division *Ascomycota* except for some classified as *Basidiomycota*. The ascomycete *S. cerevisiae* has been used in wine, bread, and beer production for thousands of years due to its ability to efficiently ferment sugar to carbon dioxide and ethanol. A few yeasts, such as *Candida albicans* (an ascomycete) or *Cryptococcus neoformans* (a basidiomycete), are opportunistic human pathogens and can infect people with a compromised immune system.

Figure 1. Electron micrograph of S. cerevisiae

Fungi are eukaryotic organisms and comparisons between S. cerevisiae and higher eukaryotes have revealed that many cellular processes and mechanisms are conserved. Thus, yeast cells are often ideal model organisms to investigate basic processes of a eukaryotic cell, both genetically and biochemically. During the last two decades, a large number of molecular biologists have adopted yeast as their primary research system, resulting in continuing investigations of all aspects of its molecular and cell biology. Furthermore, completion of the complete S. cerevisiae genome sequence as the first eukaryotic genome in 1996 (Goffeau et al., 1996) has altered the general way molecular and cell biologists approach and carry out their studies and has paved the way to systems biology. Despite the progress made in other eukaryotic model systems, S. cerevisiae is still at the front of cell biological research mainly due to its cost-effective and easy cultivation and 'the awesome power of yeast genetics' (see below).

4.2. Life cycle and genome

The genome of *S. cerevisiae* consists of 16 chromosomes, has a size of 1.2 x 107 base pairs and was the first eukaryotic genome sequenced. Duplicated parts of the genome suggest that *S. cerevisiae* did undergo a genome duplication with subsequent rearrangements. The genome sequence predicts around 5600 protein-encoding genes, mostly without introns which facilitates gene annotation and cloning. Although yeasts have greater genetic complexity than bacteria, containing 3.5 times more DNA than *Escherichia coli* cells, they share many of the technical advantages that permitted rapid progress in the molecular genetics of prokaryotes and their viruses. Some of the properties that make yeast particularly suitable for biological studies include rapid growth, dispersed cells, the ease of replica plating and mutant isolation, a well-defined genetic system, and most important, a highly efficient and versatile DNA transformation system. Being nonpathogenic, yeast can be handled with little precautions (Biological safety level BL-1). Due to its rapid growth in liquid culture, yeast can be easily grown in sufficient quantities for biochemical studies. Unlike most other yeast or fungi, the life cycle of *S. cerevisiae* comprises both a stable haploid and stable diploid state (Fig. 2). Thus, recessive mutations can be conveniently isolated and manifested in haploid strains, and complementation tests can be carried out in diploid strains. Haploid yeast cells have two different "genders" or mating types (MAT a and MAT alpha). Wild type *S. cerevisiae* strains are able to switch their mating type by a sophisticated genomic transposition mechanism involving the so-called HO-endonuclease. Laboratory strains of *S. cerevisiae* carry a mutation in the gene coding for this enzyme and are thus unable to switch. When two haploid yeast cells of opposite mating type fuse, plasmogamy is immediately followed by karyogamy resulting in a diploid cell or zygote. The diploid cells propagate by budding similar to the haploid progenitors. Other yeasts like *Schizosaccharoymces* pombe do not have a stable diploid state and multiply, like bacteria, by fission rather than budding. Haploid yeast cells propagate by forming buds (“budding"). Yeast cells of opposite mating types can fuse to form a diploid cell (“mating"). Diploid cells multiply also by budding. When diploid cells are starved, the cells are driven into meiosis.

During meiosis, an ascus containing four haploid ascospores is formed. The spores can be separated by micromanipulation and analyzed for their genotype (tetrad or meiotic analysis).

Tetrad analysis is an important method for following the meiotic segregation and recombination of genes. In addition, 'wild' S. cerevisiae strains that have not been selected for dispersed growth in the laboratory are able, both in the haploid and the diploid state, to undergo morphological transitions to filamentous growth forms (invasive growth and pseudohyphal growth, respectively).

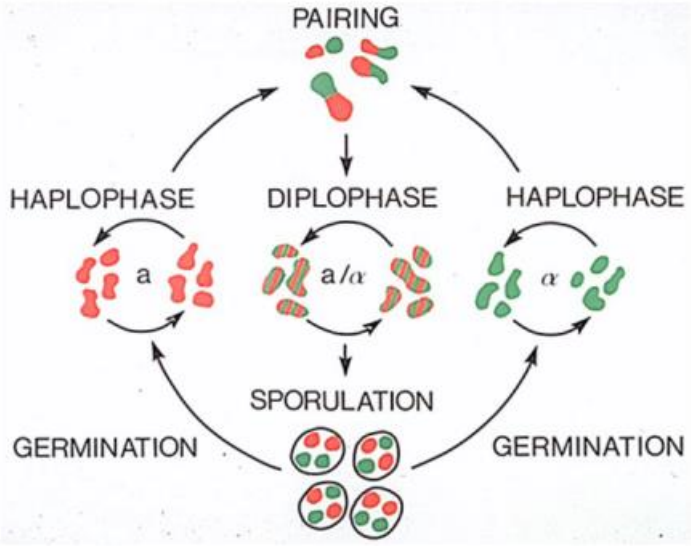

Figure 2. Lifecycle of yeast

Before we proceed it is necessary to introduce the *S. cerevisiae*-specific genetic nomenclature. *S. cerevisiae* genes are named with a three-letter code and a number. The status or the product of a particular gene is given by a specific style given below. Unfortunately, many genes in *S. cerevisiae* have multiple names due to their identification in independent genetic screens.

Example using *ARG2* (acetyl-CoA:L-glutamate N-acetyltransferase, involved in arginine biosynthetic pathway)

Gene symbol	Definition
ARG+	All wild type alleles controlling arginine requirement
ARG2	One of the genes involved in arginine biosynthesis. Gene symbol indicates that it is a dominant allele.
arg2	Recessive allele conferring arginine requirement
Arg+	A strain not requiring arginine
Arg-	Mutant strain requiring arginine
Arg2p or Arg2protein	A protein coded by *ARG2*
arg2Δ::LEU2 or *Δarg2::LEU2*	Replacing the complete *arg2* with functional *LEU2*
arg2::LEU2	Disrupting the loci with functional *LEU2*
arg2Δ	Complete deletion
arg2-(number)	Indicating mutated gene

A good source for information about yeast genetics and gene is the *Saccharomyces* genome database (https://www.yeastgenome.org/).

4.3. Episomal genetic elements

As indicated above, yeast cells have become popular due to their ease of genetic manipulation. A very convenient feature of *S. cerevisiae* is its ability to maintain genetic information on self-replicating episomal DNA elements or plasmids, like bacteria. A naturally occurring plasmid in *S. cerevisiae*, the so-called 2-micron plasmid, was used to construct the first *E. coli*-yeast shuttle vectors. These vectors contain sequences (origin of replication, selection marker) for propagation both in *E. coli* and *S. cerevisiae* and transform at high frequency (Fig. 3). Plasmids derived thereof are called YEp (Yeast episomal plasmids). These are maintained in a high copy number (20-50 copies) per cell and quite stably transmitted from mother to daughter cell but only in [cir+] strains carrying the natural endogenous 2-micron plasmid. Low copy number shuttle vectors, which are faithfully transmitted to the progeny, carry, instead of the 2-micron sequences, yeast chromosomal origins of replication, so-called autonomous replication sequences (ARS), and centromer sequences (CEN). Latter sequences are, in contrast to other fungi and higher eukaryotes, very compact (app. 150 bp) in *S. cerevisiae*. Such ARS-CEN-vectors or YCp (Yeast centromeric plasmids) are maintained in 1-5 copies per cell. Both YEp and YCp have a size of app. 5 kb and can accommodate up to 10-15 kb of additional DNA. For larger DNA fragments e.g., for genomic DNA libraries, yeast artificial chromosomes (YACs), which contain in addition to ARS and CEN telomeric sequences (TEL) and are linearized before transformation into *S. cerevisiae*, are available.

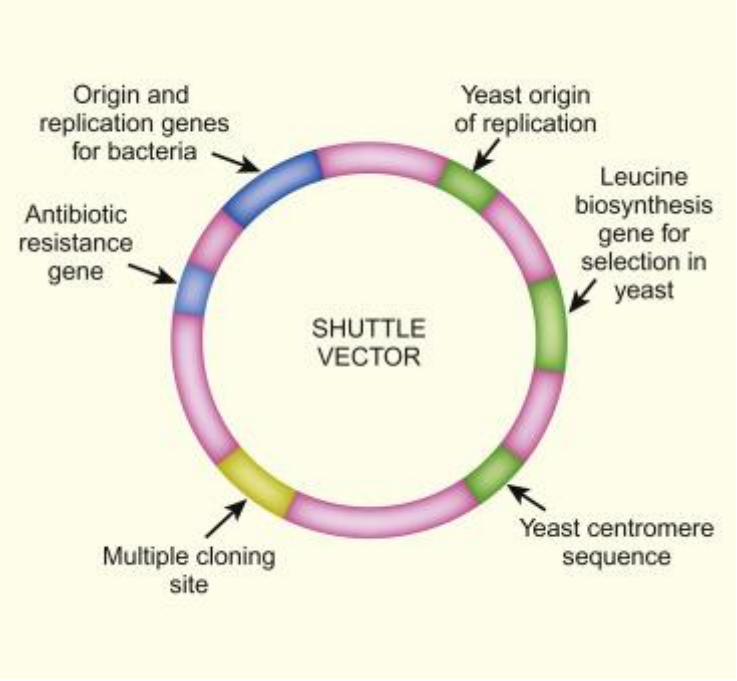

Figure 3. Scheme of E. coli-yeast shuttle vector.

4.4. Homologous recombination

In contrast to most other organisms including other fungi, recombination of transformed DNA in yeast proceeds exclusively via homologous recombination. This special feature of *S. cerevisiae* can be exploited for various purposes. The observation that homologous recombination in yeast can be efficiently achieved using linear DNA has led to the rapid development of powerful methods for DNA manipulation. One method is the cloning of DNA fragments by in vivo recombination. In this case, a DNA fragment whose ends are homologous to the ends of a linearized shuttle vector can be directly recombined in yeast, alleviating the need for an in vitro ligation reaction (Fig. 4). Corresponding DNA fragment and linearized shuttle vector are cotransformed into *S. cerevisiae* and transformants are selected based on the yeast selection marker on the vector. The cells will form colonies only upon recombining the DNA fragments to a functional self-replicating stable plasmid. For amplification and further analysis, the plasmid is isolated from yeast cells and transformed into *E. coli*. The minimal length of the recombinogenic end is only 25 bp. This is short enough to be built into a PCR primer. This cloning by in vivo recombination in yeast is very precise and independent of restriction sites (except for the linearization site of the vector).

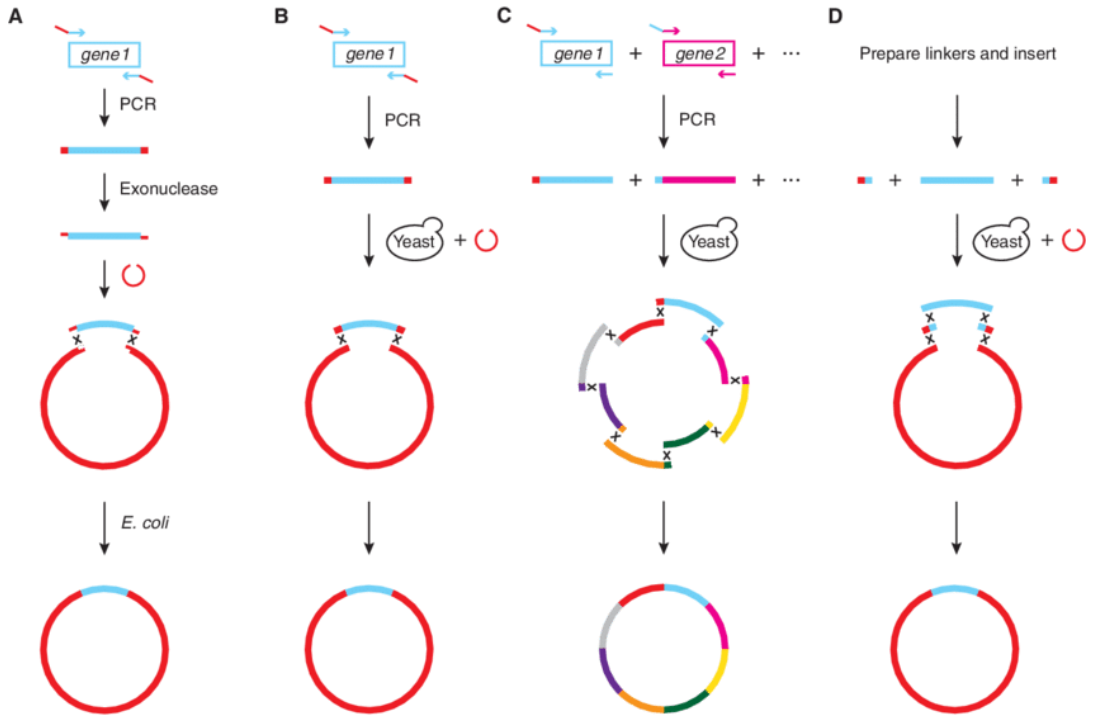

Figure 4. Cloning by homologous recombination. (A) In vitro homologous recombination of a linearized yeast plasmid and an insert through the creation of complementary single-stranded DNA overhangs by means of an exonuclease. (B) In vivo recombination of a linearized yeast plasmid and an insert in budding yeast using homologous double-stranded DNA sequences introduced by PCR. (C) In vivo recombination of multiple fragments in budding yeast using homologous double-stranded DNA sequences introduced by PCR. Instead of using a linearized

yeast plasmid as one fragment, the selection marker(s) and replication origin(s) can be introduced on different DNA fragments. (D) In vivo recombination in budding yeast using homologous sequences present on separate DNA-linker fragments. In this case, the DNA linkers have homology regions to both the acceptor vector and the insert and are useful for cloning inserts from different sources when PCR amplification is difficult or undesirable.

A second type of application is the targeted integration of exogenous DNA into the yeast chromosome. Depending on the design of the recombinogenic DNA, the exogenous DNA is either inserted at a specific location of the genome (so-called 'knock-in') or the recombination leads to replacement of a specific genome sequence by the exogenous DNA (so-called 'knock-out'). These applications allow genetic alterations such as gene replacements, promoter exchanges or insertion of epitope tags on the chromosome in single copy. Such chromosomal alterations have contributed significantly towards understanding the function of genes and proteins in vivo. A number of commercially available tools, such as a collection of knock out strains in all nonessential genes, facilitate such studies.

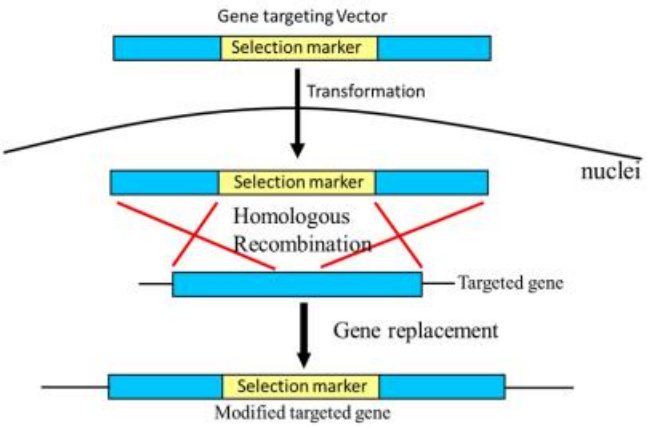

Figure 5. Gene replacement by homologous recombination A specific yeast open reading frame (ORF) is replaced by a linear DNA fragment carrying homologous flanking regions (red) and a marker gene coding for G418 resistance (kanMX4). The recombination leads to deletion of the yeast ORF, a process called gene knock out.

4.5. Selection and counterselection

In bacteria, resistances against antibiotics are usually used as selection markers. In *S. cerevisiae* selection is usually based on amino acid or nucleotide biosynthetic genes in combination with respective *S. cerevisiae* mutants which are no longer capable of synthesizing the respective compound. The auxotrophic strains to be transformed are usually grown in complete medium and transformants are selected on minimal media lacking the respective compound. Usual selection markers are URA3 (Uracil biosynthesis), LEU2 (Leucine biosynthesis), HIS3 (Histidine biosynthesis) and TRP 1 (Tryptophan biosynthesis), and common laboratory *S. cerevisiae* strains such as SS328 (MATα ade2-101 his3Δ200 lys2-801 ura3-52) are mutant for several of these genes. Alternatively, dominant selection markers against a variety of antifungal drugs are available. Usually, these markers consist of the coding

sequence of bacterial resistance genes under control of a strong, constitutive yeast promoter and terminator. Common dominant markers are resistances against Hygromycin (HYG), Neomycin (= G418, an antifungal Kanamycin derivative) and Nourseothricin (NAT).

In addition to these so-called positive selection procedures (growth requires uptake of DNA), there are also several negative or counterselection procedures available for *S. cerevisiae* (growth requires loss of DNA). The most often used counterselection is based on the fact that URA3 cells are sensitive to the drug 5-fluoroorotic acid (5-FOA) whereas ura3 cells are not. The non-toxic compound 5-FOA is thereby converted into the toxic compound 5-fluorouracil by the URA3 encoded enzyme. Such a counterselection can be exploited for various purposes. One application is the so-called "plasmid shuffling" procedure. In this procedure, an ura3 cell is first transformed with a URA3-containing plasmid by positive selection on minimal medium without uracil. Upon further manipulation of the transformant, e.g., mutagenesis, the URA3-plasmid can be "shuffled out" by streaking the transformants on media containing 5-FOA and low amounts of uracil. Application of this plasmid shuffling procedure to functionally test alleles of a gene that is essential for vegetative growth is subject to Part 1 of this course. Another counterselection using L-canavanine, a non-protein amino acid of certain leguminous plants, can be exploited to counter select against diploids and the original haploids. The structure of this compound is related to the proteinaceous amino acid, L-arginine. Canavanine is accumulated primarily in the seeds where it serves both as a defensive compound against herbivores and a vital source of nitrogen for the growing embryo. Organisms that consume it can mistakenly incorporate it into their own proteins in the place of arginine, thereby producing structurally aberrant proteins that may not function properly or not at all. Some specialized herbivores tolerate L-canavanine either because they metabolize it efficiently or avoid its incorporation into their own nascent proteins. Wild-type yeast cells are susceptible to canavanine whereas mutants in an arginine-transporter, encoded by the CAN1-gene, are partially resistant.

5. Synthetic Biology, what is it?

At the start of genetic engineering in 1973, when Herbert Boyer and Stanley Cohen used restriction enzymes and DNA ligase to insert the gene encoding frog ribosomal RNA into the vector pSC101, the possibilities of genetic engineering were thought to be almost endless and the optimism regarding the potential to quickly produce new medicines, achieve higher crop yields and produce high yields of biofuels from biomass, flourished. Although genetic engineering has allowed for big discoveries, like the production of synthetic insulin, conferred insect resistance of Bt corn plants (Saxena et al., 1999) and subunit vaccines from the FMD virus (McKenna et al., 1995), four decades later, the initial optimism in regard to the full potential of genetic engineering has tempered. This can partly be explained by the complexity of biological systems. For instance, genes are often part of complex networks, and alterations of a single gene could lead to a plethora of downstream effects. Also, a strong regulation

policy due to limited knowledge of long-term effects of genetically modified organisms (especially in Europe), and a lack of standardization have played had influenced discoveries.

Now, as synthetic biology makes its way as a relative new field in science, it is anticipated with the same optimism. So, to what degree is the field of synthetic biology capable of living up to the expectations?

The words “synthetic biology” was used for the first time by Stéphane Leduc, a French biologist, in 1910. But until this very moment, there is no set definition of synthetic biology. For instance, chemist Eric Kool described synthetic biology in 2000 as “The synthesis of unnatural organic molecules that function in the living system”. Whereas the Royal Academy of Engineering defines as “Synthetic biology aims to design and engineer biologically based parts, novel devices and systems as well as redesigning existing, natural biological systems”.

This diverging takes on definitions could be explained by the interdisciplinary nature of the field. The field of synthetic biology combines the efforts of biologists, chemists, and engineers (all with a different background and view) who work together to engineer complex artificial biological systems. Although the definition of synthetic biology might vary depending on one’s scientific or educational background, the goal does not. Synthetic biology aims to extend or modify the behavior of organisms and engineer them to perform new tasks.

To modify organisms and engineer them to perform specific tasks, the complexity of the cellular systems has to be reduced. Figure 6 displays a hierarchy of complexity present in cellular systems, ranging from tissues or cultures at the top to genes and proteins at the bottom, by comparing it to computer networks. Thus, to design new characteristics that occur in the top of the hierarchy, designs have to be implemented at the bottom, i.e., at a genetic or proteomic level (whereas features in a computer subsequently change by alterations to the number of transistors for example).

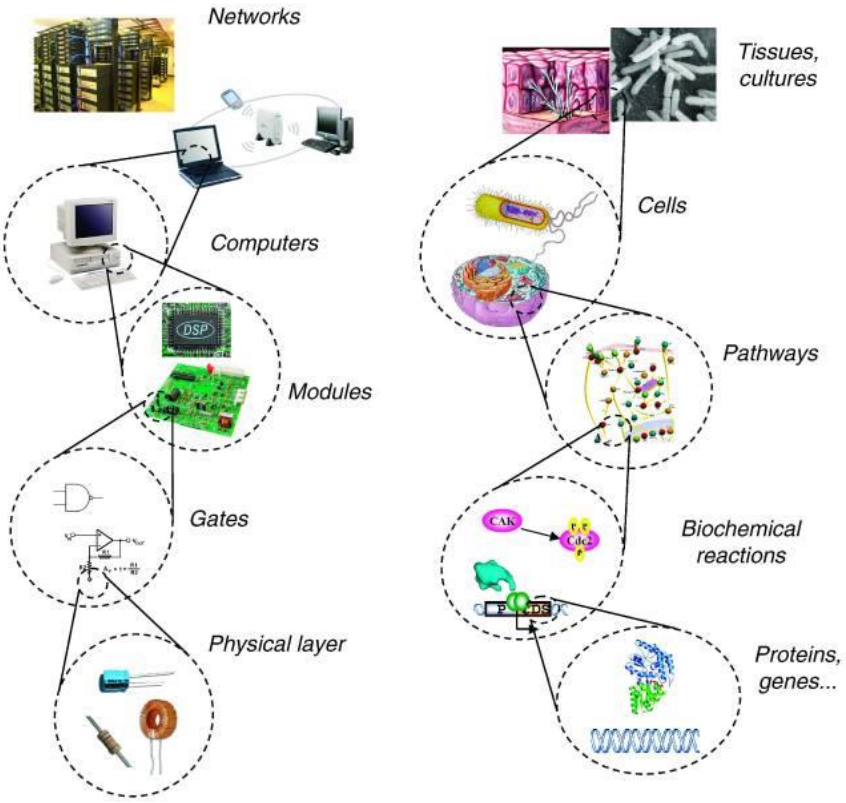

Figure 6. A comparison of system complexity between computer networks and cellular tissues. Andrianantoandro et al., (2006).

5.1. Research Areas

5.1.1. Metabolic Engineering

Metabolic engineering is the process of altering product-specific enzymes, or even whole pathways with the goal to increase the microbial production of natural products. It stimulates the production of many chemicals, bulk chemicals, and fuels. Although microorganisms have a long history of producing alcohols like ethanol and butanol, metabolic engineering has produced many larger, branched chain alcohols, which provide a better source of fuel (Holtz and Keasling, 2010). The most famous example of the use of metabolic engineering is the production of artemisinic acid by *S. cerevisiae*. The antimalarial drug artemisinin is an active pharmaceutical ingredient (API) that is produced from the plant *Artemisia annua*. In this instance, *S. cerevisiae* was used to produce the precursor artemisinic acid. This reduced the costs of artemisinin production significantly, making the drug more widespread available (Ro et al., 2006).

5.1.2. Minimal Genomes

An important (top down) attempt to reduce complexity in cellular responses is the creation of minimal genomes. By eliminating everything that is not essential, the simplest and basic system is created, which in turn can be used as optimized cell factories. Another view on this is the construction of special cellular compartments, or even whole cells and genomes

synthetically (bottom up). The most significant project currently attempted is the construction of a synthetic yeast genome http://syntheticyeast.org/.

5.1.3. Regulatory circuits

The natural activity of cells is controlled by circuits of genes analogous to electronic circuits. Thus, another approach in imparting novel cellular functionalities relies on creating novel internal circuitry to alter their pattern of activity. Using well-understood genetic components that act as molecular switches it should be possible to devise artificial gene networks. Linked together and implanted into natural systems such networks could be used to control what those systems do, when, and how frequently. Integrated into suitable cells an artificial network might be used to sense and correct metabolic disturbances of the kind found in diabetes.

5.1.4. De Novo protein design

The de novo design of proteins has become a crucial approach to elucidate the relationship between an amino acid sequence and the three-dimensional protein structure. In 1973, Anfinsen proposed that proteins fold according to the principle of minimum free energy (which describes that the structure shall deform to a position that minimizes the total potential energy) and that if a model of the free energy is available, the structure of proteins can be predicted, with as an ultimate goal the identification of amino acid sequences that fold into proteins with a desired function (Dahiyat and Mayo, 1997).

5.2. Biobricks

One of the largest attempts to introduce standardization into molecular biological research are the Biobricks that are used and produced in the iGEM competition. Every year, hundreds of teams from all over the world combine DNA parts to create increasingly complex genetic circuits from new or from the catalog of parts made in previous years (http://parts.igem.org/Catalog). This catalog now contains thousands of circuits, exposing new teams to a vast variety of options for future design plans.

Teams are given total freedom on the ideas that they want to implement but have to follow strict rules in the designs of their plasmids. This is done to ensure that all the parts that are admitted are compatible and can be implemented and combined with all the other parts in the database. There is a big similarity between the Biobricks and lego bricks (Figure 7). Lego exists in all sorts of varying shapes, sizes, and colors, but all the lego parts are compatible with each other. Together, these parts are combined and recombined to create increasingly complex structures.

Figure 9. Lego blocks provide a good example of how simple structures can lead to a wide variety of applications and structures.

5.3. Escherichia coli – bacterial workhorse for biotechnology

5.3.1 Characteristics

A bacterial species called *Escherichia coli* was first discovered in 1885 by Theodor Escherich, a German bacteriologist. It has since been commonly used for biological laboratory experiments, research, and many applications in biotechnology. These rod-shaped bacteria (Figure 10) measure approximately 0.5 µm in width by 2 µm in length and can be commonly found in animal feces, lower intestines of mammals, and even on the edge of hot springs. It has facultative characteristics, i.e., it can grow in both aerobic and anaerobic environments. *E. coli* makes ATP by aerobic respiration if oxygen is present, but it is capable of switching to fermentation or anaerobic respiration if oxygen is absent. They grow best at 37°C. *E. coli* is a gram-negative (does not retain crystal violet dye) organism that possesses a cell wall that consists of an outer membrane containing lipopolysaccharides, a periplasmic space with a peptidoglycan layer, and an inner cytoplasmic membrane. As *E. coli* cannot sporulate, it is easy to eradicate by simple boiling or basic sterilization. *E. coli* can be classified into hundreds of strains on the basis of different serotypes.

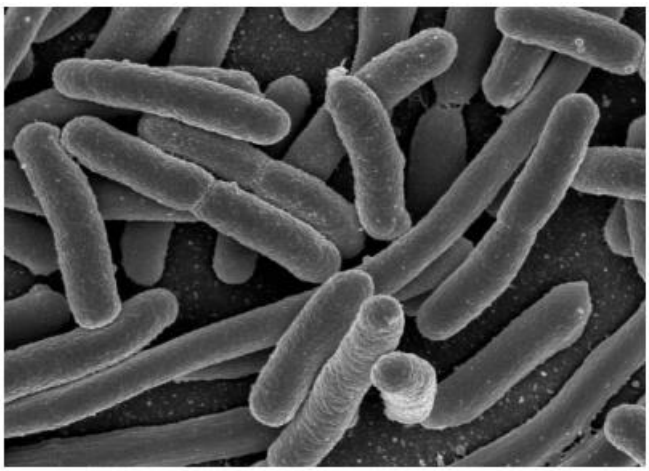

Figure 10. A microscopic view of Escherichia coli.

Some of the enteric *E. coli* can cause several intestinal and extra-intestinal infections such as urinary tract infection and mastitis. However, E. coli are not always harmful to human bodies or other animals. Most *E. coli* live in our intestines, where they help our body breakdown the

food we eat as well as assist with waste processing, vitamin K production, and food absorption.

5.3.2 Genome Structure

E. coli has only one circular chromosome, some along with one or several circular plasmids (Figure 11). The first complete DNA sequence of an *E. coli* genome (laboratory strain K-12 derivative MG1655) was published in 1997. E. coli was one of the first organisms to have its genome sequenced. The genetic material of *E. coli* was found to be a circular DNA molecule 4.6 million base pairs in length, containing almost 4300 annotated protein-coding genes that are organized into 2584 operons. In addition, seven ribosomal RNA (rRNA) operons and 86 transfer RNA (tRNA) genes were identified. Despite having been the subject of intensive genetic analysis for about 40 years, a large number of these genes were previously unknown. The coding density was found to be very high, with a mean distance between genes of only 118 base pairs. The genome was observed to contain a significant number of repeat elements, cryptic prophages, transposable genetic elements, and bacteriophage remnants. Because so many *E. coli* strains have had their genomes sequenced, we can also compare the DNA sequence of genes in different *E. coli* strains. Comparing gene sequences gives clues to the function of genes, their relative importance and the changes they have undergone over time.

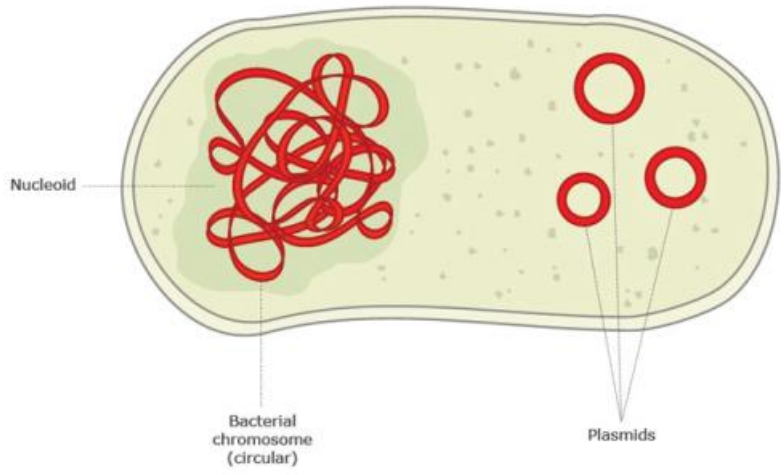

Figure 11. Bacterial DNA. http://www.biotechlearn.org.nz

5.3.3. Application to Biotechnology

Since the birth of molecular cloning, *E. coli* has been used as a host for introduced DNA sequences. In 1973, Herbert Boyer and Stanley Cohen showed for the first time that two short pieces of bacterial DNA could be 'cut and pasted' together and returned to *E. coli*. They went on to show that DNA from other species, such as frogs, could also be introduced to *E. coli*. Over time, it became the bacterium of choice for virtually all molecular cloning. Under ideal conditions, individual *E. coli* cells can double every 20 minutes. At that rate, it would be possible to produce a million *E. coli* cells from one parent cell within about seven hours. Fast growth means that experiments involving *E. coli* can be done quickly, conveniently, and cheaply. Today, *E. coli* is used in laboratories worldwide as a host for foreign DNA sequences

and their protein products. There is a wealth of knowledge and comprehensive tools for *E. coli* systems, such as expression vectors (Figure 12), production strains, protein folding and fermentation technologies that are well tailored for industrial applications.

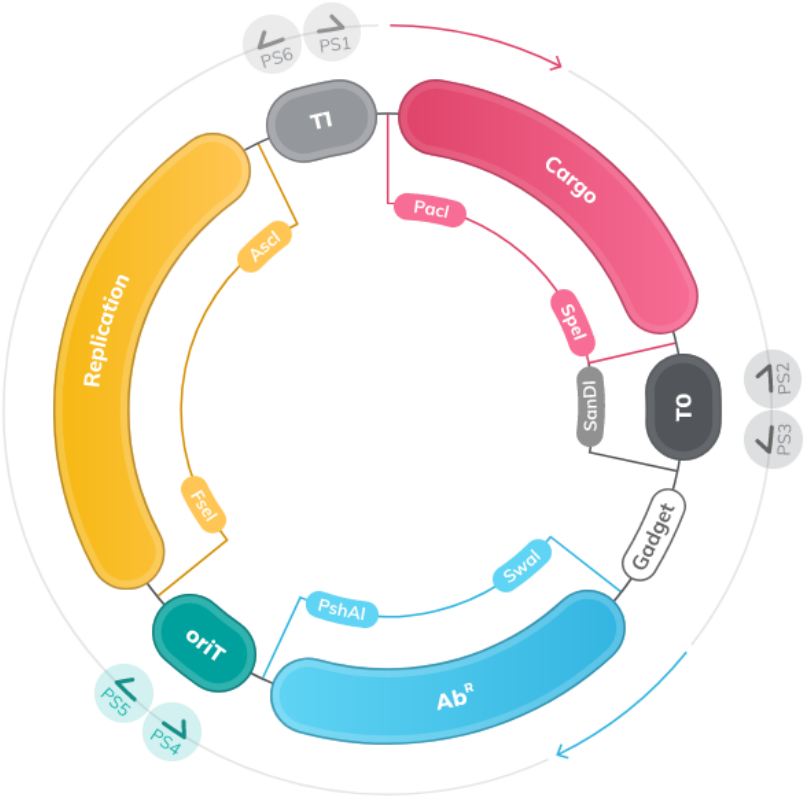

Figure 12. An E. coli replicative plasmid map. The vector map, based on Standard European Vector Architecture (https://seva-plasmids.com), presents origin of replication (yellow), origin of transference (oriT, green; for conjugation between gram-negative donor and recipient bacteria), selection marker (blue), transcriptional terminators (black and grey) and DNA portion which confers functionality to the vector (red).

Due to its well-characterized genetics, rapid growth and high-yield production, *E. coli* has been a preferred choice and a workhorse for expression of numerous commercially valuable compounds in the biotech industry. With metabolic engineering it has been possible to create strains that make efficient recombinant biocatalysts for the production of high-value organic acids like succinic acid, lactic acid, and glucaric acid as well as alcohols like xylitol, mannitol, and glycerol as well as several platform chemicals, such as aspartic acid, glutamic acid, and sorbitol, just to mention few. Another large recombinant production entity are non-glycosylated proteins. For example, nearly 30% of currently approved recombinant therapeutic proteins are produced in *E. coli*.

References

Andrianantoandro, E., Basu, D., Karig, D.K. and Weiss R. 2006. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology 2:2006.0028.

Blattner FR, et al (1997). "The complete genome sequence of Escherichia coli K-12". Science (New York, N.Y.) 277 (5331): 1453–62.

Dahiyat, B. I. and Mayo, L.M. 1997. De Novo Protein Design: Fully Automated Sequence Selection. Science. Vol. 278 no. 5335 pp. 82-87.
Holtz, W. J. & Keasling, J. D. Engineering static and dynamic control of synthetic pathways. Cell 140, 19–23 (2010).
Huang CJ, et al (2012). Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. J Ind Microbiol Biotechnol. 39(3):383-99.
McKenna TS, Lubroth J, Rieder E, Baxt B, Mason PW. 1995. Receptor binding site-deleted foot-and-mouth disease (FMD) virus protects cattle from FMD. J. Virol. 69:5787–5790.
Ro, D. K., Eric M. Paradise Mario Ouellet, Karl J. Fisher, Karyn L. Newman, John M. Ndungu, Kimberly A. Ho, Rachel A. Eachus, Timothy S. Ham, James Kirby, Michelle C. Y. Chang, Sydnor T. Withers, Yoichiro Shiba, Richmond Sarpong and Jay D. Keasling. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940-943.
Saxena, D., Flores, S. and G. Stotzky. 1999. Insecticidal toxin in root exudates from Bt corn. Nature. Vol. 402.
Yu C, et al (2011). Metabolic engineering of Escherichia coli for biotechnological production of high-value organic acids and alcohols. Appl Microbiol Biotechnol. 89(3):573-83.
https://microbewiki.kenyon.edu/index.php/Escherichia_coli.
http://biotechlearn.org.nz/themes/bacteria_in_biotech/e_coli_the_biotech_bacterium.