Literature Review
Synthetic Circuit Design
CPSC 791

Galen Michael Seilis
[redacted]

November 6, 2021

Contents

1 Introduction **3**

2 Fundamentals of Molecular Biology **3**

2.1 Information-Containing Macromolecules 4

2.1.1 DNA 4

2.1.2 RNA 7

2.1.3 Protein 7

2.2 The Central Dogma 11

2.2.1 Transcription 13

2.2.2 Translation 14

2.3 Summary 16

3 The *lac* Operon **16**

4 Genetic Circuit Design **20**

5 Advancements **22**

6 Challenges **24**

7 Conclusions **25**

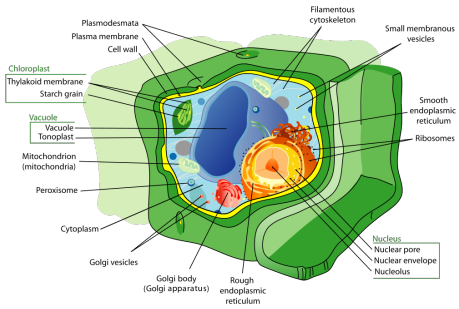
1 Introduction

Synthetic biological circuits, frequently called "*genetic circuits*" in the literature, are network models of molecular components that are inspired from logical circuits found in silica-based electronic circuits [1]. In this work the term "*genetic circuit*" if preferred over the term "*synthetic biological circuit*". We will return to genetic circuits in a later section, however some general background in biology is helpful in understanding and motivating the subject. An effective way to obtain an understanding of designed genetic circuits in their ontology and functioning is to consider their analogs *in vivo*.

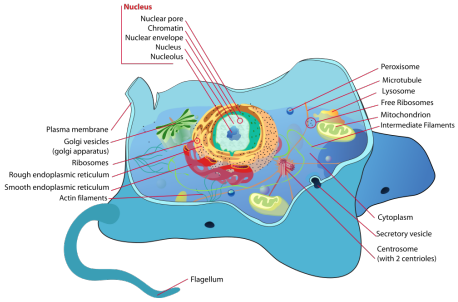
To begin, Section 2 covers fundamental concepts in Molecular Biology for unfamiliar readers. In Section 3 we will go into detail on a classic example of a gene regulatory network involving only a handful of genes. Section 4 returns to the subject of genetic circuit design, and will focus on the motivation and goals of the subject. With some background on gene regulation and genetic circuits in place, Section 5 will summarize the accomplishments of the discipline and Section 6 will discuss a couple of the most persistent challenges to the field. Section 7 gives some final remarks on the major takeways of the field, and how the literature should be further reviewed.

2 Fundamentals of Molecular Biology

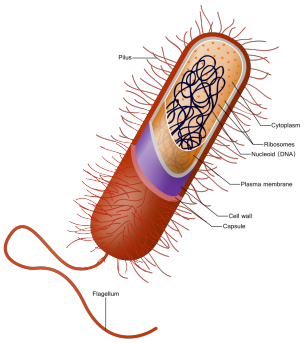
For readers that are unfamiliar with Molecular Biology, this section will give a short overview of fundamental concepts that are otherwise considered implicit background in the sections that follow. This section lightly touches on only a handful of molecules and processes for the sake of brevity, and we will defer detailed expositions to [2]. We will not cover the distinctions between Eukaryotes (such as animals and plants) from Prokaryotes (such as Bacteria and Archaea) because the core notions are the similar while the molecular details would vastly expand this exposition. An important reminder is that everything we are about to discuss is happening within biological cells (See Figure 1), but we also eschew many details pertaining to the structural compartments of a cell. For our purposes, if two or more molecules are discussed as interacting, it is safe to assume that there exists a cellular compartment in which they do so.



(Public Domain)



(Public Domain)



(CC-BY-SA 4.0 Ali Zifan)

Figure 1: Cartoon diagrams of (a) a plant cell (eukaryote), (b) an animal cell (eukaryote), and a bacterial cell (prokaryote). These diagrams are not proportional to each other. A prokaryotic cell is likely to be 5-100 times smaller than a typical eukaryotic cell.

Broadly speaking, there are four categories of macromolecules (i.e. large molecules) that Molecular Biology primarily concerns itself with studying: nucleic acids, carbohydrates, lipids, and proteins. We will exposit introductory concepts on two types of nucleic acid (DNA and RNA) and proteins.

Understanding carbohydrates and lipids are essential parts to understanding living systems in general, but we will treat them as background topics in favour of briefly communicating how information can 'flow' from DNA to RNA to protein via transcription and translation. Here we mean "information" imprecisely as an pattern in the structure of DNA that molecular machinery in a cell interacts with. This is related, but not equivalent, to formal notions of information such as *information entropy* developed by Claude Shannon [3]. This notion of information 'flow' will be made clear through exemplification rather than through formalism.

Numerous processes including RNA splicing, post-translation modification, and vessicle transport can also be found in [2], but are likewise ignored for brevity.

2.1 Information-Containing Macromolecules

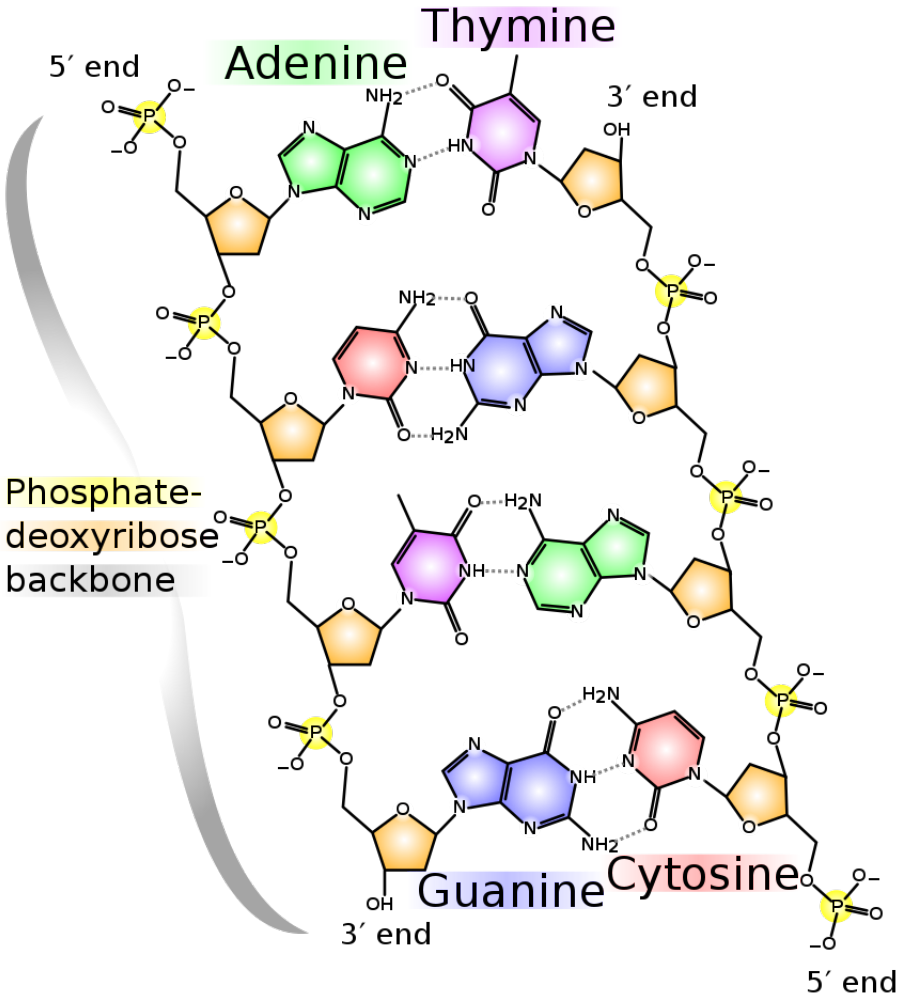
2.1.1 DNA

"*DNA*" is an acronym that has become so ubiquitous in the public vernacular that it is often used without demarcation: "**D**eoxyribo**N**ucleic **A**cid". While there are numerous exceptions, many molecules found in Biochemistry and Molecular Biology have names that in some sense correspond to their structure. As the nomenclature suggests, DNA is an acid. But DNA is not the sort of acid that will dissolve your body in any substantial sense, and is considered an acid in the technical sense of Brønsted and Lowery [4]. DNA is the primary genetic molecule in most living systems in the sense that it is heritable and encodes 'informa-

tion' that is necessary to synthesize biomolecules that embody the structure and behaviour of biological systems.

One of the underlying themes of Molecular Biology is that the physical structure of molecules relate to their biological functions within biological systems. The structure of DNA was theorized by James Watson and Francis Crick in the 1950's using crystallographic data collected by Rosalind Franklin [5, 6]. There is a complicated history around the discovery of DNA, partly outlined by James Watson's memoir of the events leading up to it [6]. Unfortunately Rosalind Franklin did not receive equal credit in the discovery of the structure of DNA to James Watson and Francis Crick [7, 8].

A representation of the structure of a short segment of DNA is given in Figure 2. The structure is composed of two strands which are described as *antiparallel* due to having reverse orientations in a specific sense. Explicitly, a DNA strand has an orientation due to where the phosphodiester bonds attach to ribose sugars at either the $3'$ or $5'$ carbons of the ribose sugars. Phosphodiester bonds are a pair of ester bonds between a phosphate and some other structures. See [9] for a detailed definition and mechanism of synthesis of ester bonds. Labelling the carbon atoms of a sugar molecule such as ribose with labels such as $3'$ or $5'$ is a common convention in biochemistry [4]. In brief, one can count the number of carbons away from the oxygen starting from a particular carbon known as the α carbon. The α carbon is defined in relation to the adjacent oxygen atom and where it would sit in its straight chain form. A more detailed explanation of this labelling convention can be found in [4], however it suffices to say here that the $3'$ carbon is three bonds away from the ring oxygen and the $5'$ carbon is five bonds away from the ring oxygen. While the $5'$ carbon is not part of the pentagon ring, it is a part of ribose when it is a free molecule apart from DNA. DNA is particularly stable in a double-stranded configuration because of hydrogen bonds between its nitrogenous bases, and it is the geometric properties of how these bases are oriented in space that lead to the complementarity of specific (nitrogenous) base pairs (A with T and G with C).



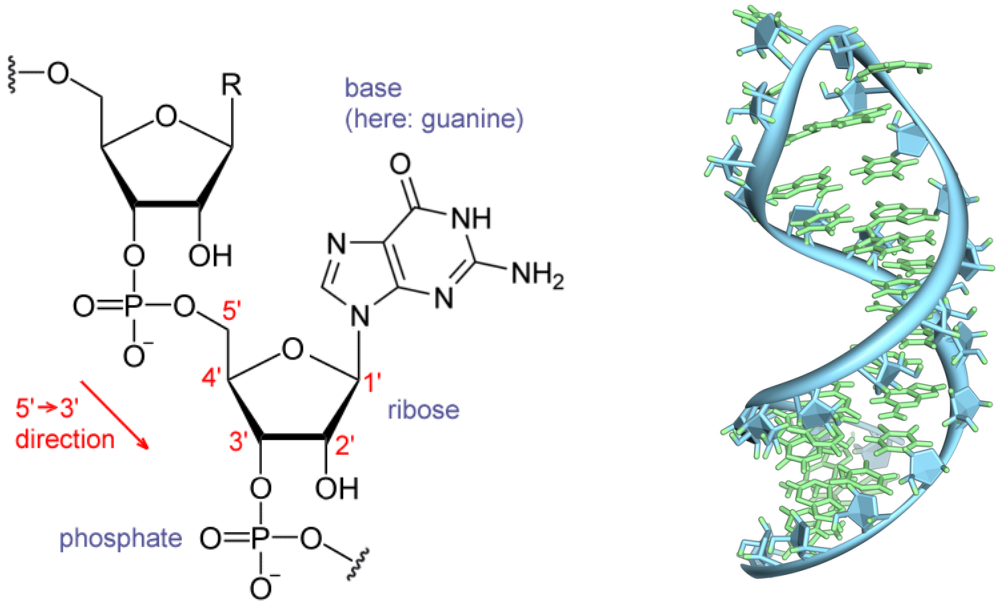
(Image Credit: Madeleine Price Ball CC-BY-SA 3.0/CC0)

Figure 2: Cartoon model of the structure of double-stranded DNA. The yellow 'P' symbol represents the phosphorus atom at the center of each phosphate group, which are participating in phosphodiester bonds. The sequence of alternating phosphate and ribose sugars is called the *phosphate-deoxyribose backbone* or *phosphate-sugar backbone of DNA*. The orange heptagons represent the ribose sugar units, which complete the sugar-phosphate backbone via phosphodiester bonds. The green, red, purple, and blue structures represent different nitrogenous bases found in DNA: adenine, cytosine, thymine, and guanine. A nitrogenous base attached together with a ribose sugar is called a *nucleoside*. A nitrogenous base together with a ribose and a phosphate is called a *nucleotide*. The two strands are considered *antiparallel*, which can be recognized from the orientation of the pentose sugars (i.e. deoxyribose) on the left strand vs the right strand. This is formally indicated by the notations $3'$ and $5'$ which pertain to the bonding and orientation of atoms to the pentose ring.

2.1.2 RNA

Ribonucleic acid (RNA) is structurally very similar to DNA, and in some viruses it serves as the primary genetic material [2]. RNA is an important type of macromolecule because it can catalyze chemical reactions and is integral to the processes that lead from DNA to protein. While DNA has the nitrogenous bases adenine, thymine, guanine and cytosine, RNA's nitrogenous bases include adenine, uracil, guanine, and cytosine. Like thymine, uracil is complementary to adenine through hydrogen bonding. Panel (a) of Figure 3 shows the key structural difference between RNA and DNA that leads RNA to be more reactive (and less stable) and enables its catalytic activity. Specifically at the $2'$ carbon of pentose sugar (i.e. five-carbon atom sugar) is a hydrogen (H) in DNA, whereas RNA has a hydroxyl group (-OH). The absence of this hydroxyl group in DNA is why it is referred to as 'deoxy'-ribonucleic acid. While double-stranded RNA is found in specific types of viruses, panel (b) of Figure 3 shows a rather typical situation of a single strand of RNA self-associating into a structure commonly referred to as a *hairpin loop*.

(a) (b)



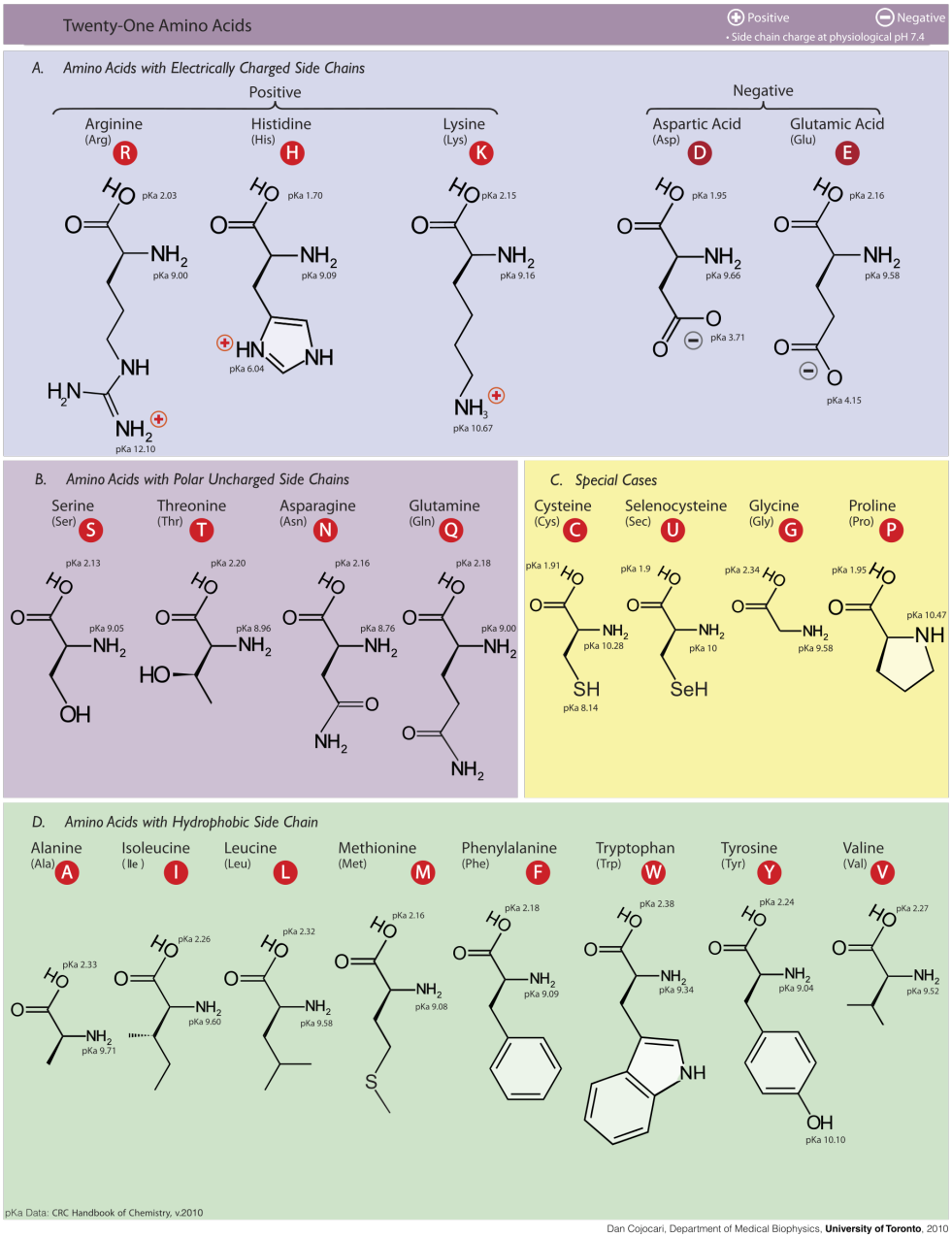
(Image Credit: Username Narayanese CC-BY-SA 3.0) (Public Domain)

Figure 3: (a) Skeletal formula illustrating binding among a nitrogenous base (guanine), and the ribose sugars and phosphates that make up the sugar-phosphate backbone of RNA. The orientation of these bonds gives us the notion of a $3' \to 5'$ direction to an RNA sequence. This also illustrates one of the key structural differences between DNA and RNA: a hydroxyl group (-OH) at the $2'$ carbon. (b) Ribbon and stick cartoon of the 3D structure of a single-stranded RNA molecule in a hairpin conformation.

2.1.3 Protein

Among DNA, RNA, and protein, it is protein that is more complicated in terms of the chemical substituents that go into it. Similar to the previously discussed macromolecules,

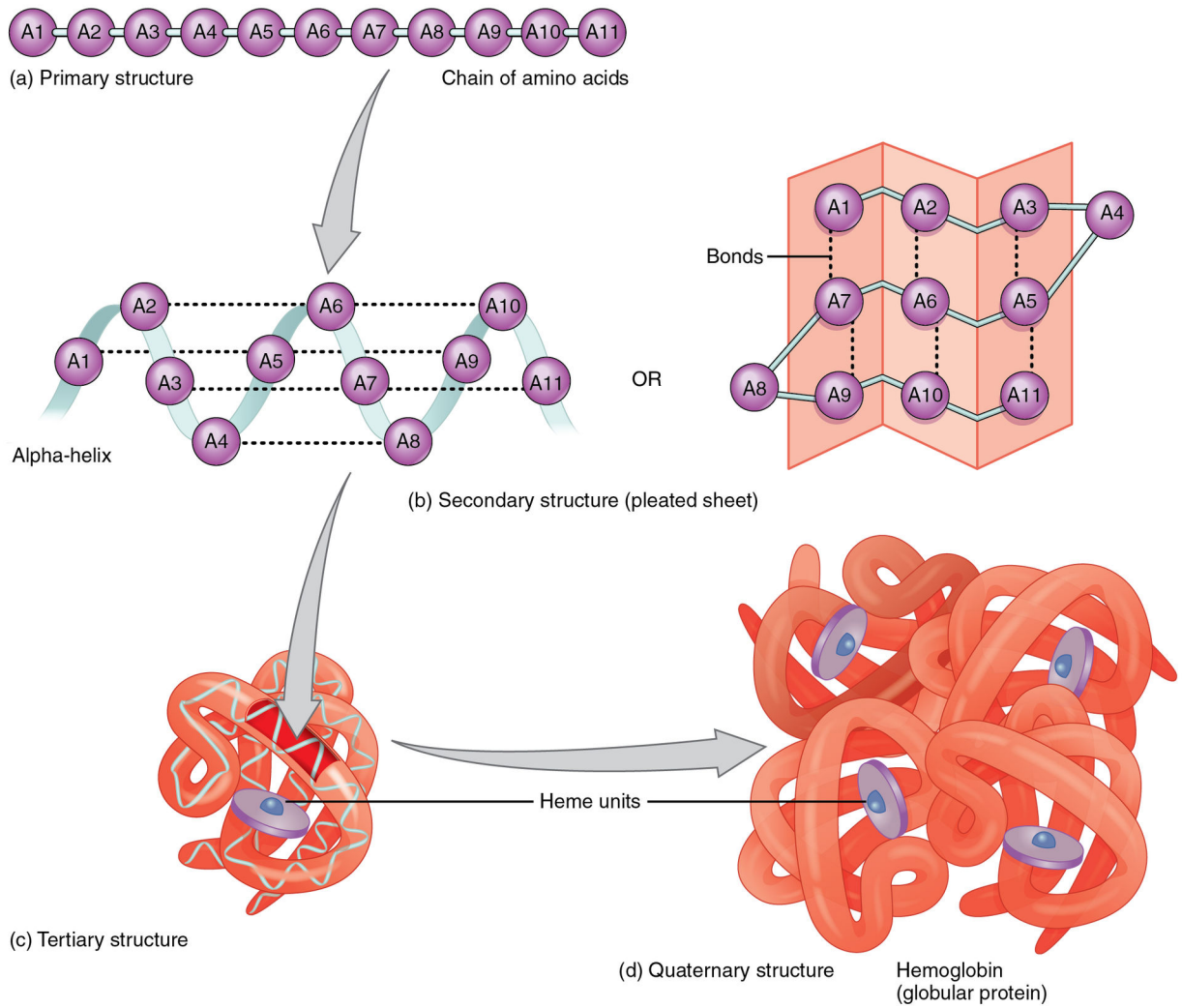
proteins in their simplest form can be thought of as sequences of individual units (monomers) that are put together to make a sequence (called a *polymer*). Simple proteins composed of sequences of amino acids, which are illustrated in Figure 4, and are joined to together with a type of covalent bond called a *peptide bond*. A sequence of amino acids is often called a *polypeptide*. It is partly due to the diversity of physical and chemical properties of amino acids including positive/negative charge, polarity or non-polarity, steric hinderance (i.e. constraints in conformation in space), buffering capacity, and hydrogen bonding that proteins themselves have a great diversity of properties.



(Image Credit: Dan Cojocari CC-BY-SA 3.0)

Figure 4: Classification and structure of the common twenty amino acids that are incorporated into protein. Panel A shows the amino acids with either negatively and positively charged side chains. Panel B shows the amino acids with polar uncharged side chains. Panel C shows amino acids with sides chains that are unique among the twenty amino acids including a thiol group in cysteine, a selenium atom in selenocysteine, a lack of chirality in the α carbon of glycine, and proline which is not technically an amino acid due to lacking a primary amine at the α carbon. Panel D shows amino acids that have non-polar (hydrophobic) side chains.

While various combinations of the amino acids in Figure 4 can be imagined, an important aspect of proteins is the way in which those amino acids go together. Like many polymers, proteins are composed of sequences of monomers (i.e. amino acids). Figure 5 illustrates how proteins can be organized into four levels of structure. The primary structure is simply the sequence of amino acids, shown in part (a) of Figure 5, which leads to a clear domain of application for Combinatorics. Since amino acids can repeat, we can claim that there are 20^n possible sequences of amino acids provided that none of those options are excluded due to physical constraints. Secondary structure of proteins describe the 3D structures of local segments of a sequence, shown in part (b) of Figure 5, which can include motif structures such as α-helices and β-pleated sheets. The dashed lines in part (b) of Figure 5 show bonds, specifically hydrogen bonds, which strongly influence the local structure of a protein. The tertiary structure of a protein describes the global 3D structure of a single sequence of amino acids. This is shown in part (c) of Figure 5. The important distinction between secondary and tertiary structure is to what degree the global vs local structure is being considered. The last structural level of proteins is the quaternary structure, shown in part (d) of Figure 5. The quaternary structure accounts for the 3D structure of not only a single peptide sequence, but also the structure of complexes. A *complex* generally refers to collections of proteins (and sometimes other types of molecules) connected by hydrogen bonding and/or disulfide bridges. Disulfide bridges are covalent bonds between thiol groups which are the -SH groups found in cysteines. Hydrogen bonds are weaker interactions than covalent disulfide bridges. Hydrogen bonds can provide flexibility, but can also be disrupted by changes in pH, salt concentrations, or temperature which can affect the functioning of the protein.



(Image Credit: OpenStax College CC-BY-SA 3.0/4.0)

Figure 5: Cartoon diagram of the order of protein structure: (a) primary, (b) secondary, (c) tertiary, and (d) quaternary. The cartoon uses hemoglobin as an example of a protein that exhibits all four levels of structure, as well as the existence of heme groups that are not found in all proteins.

Now that we have briefly described DNA, RNA, and protein, we can move onto to the bigger picture of information flow in a cell: The Central Dogma.

2.2 The Central Dogma

Macromolecules can be related to each other through chemical reactions in direct and indirect ways. In the case of DNA, RNA, and protein there is an abstraction called the *Central Dogma of Molecular Biology* (CDMB). In a standard treatment of the subject, the CDMB is often a statement that the flow of genetic information goes from DNA to RNA, and from RNA to protein. Sometimes DNA replication is included in this notion, but even the DNA $\rightarrow$ RNA $\rightarrow$ protein is not historically accurate. What Francis Crick had actually proposed as

the central dogma is that once genetic information in DNA had been encoded to protein, there was no way to reverse the flow of information [10]. Even so, this modified view of the CDMB is a useful abstract representation which has been visualized in Figure 6 along with the processes of reverse transcription and RNA replication that were not known to Francis Crick when he developed this concept. As shown in Figure 6, Francis Crick was essentially correct insofar as that there still doesn't exist a known biological process for going back from genetic information encoded in the form of protein sequences.

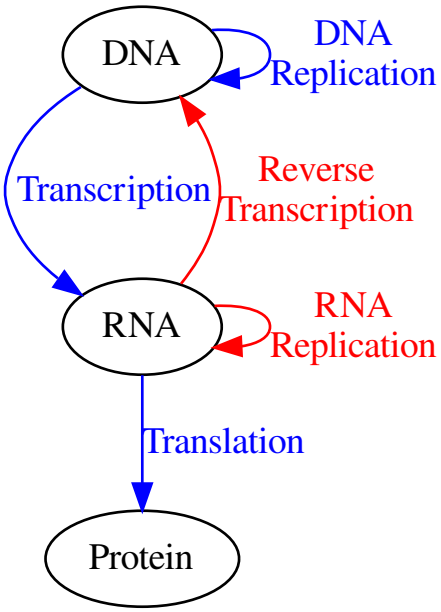
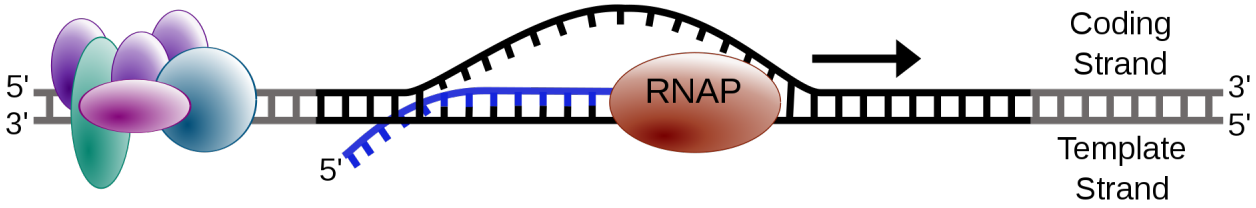


Figure 6: Directed acyclic graph illustrating the (modified) Central Dogma of a Molecular Biology. The nodes represent the three common types of "information-carrying" macromolecules: DNA, RNA, and protein. The arcs (directed edges) are labelled by the molecular processes that convert from one type of macromolecule to another. Those arcs and arc labels coloured blue represent processes that are part of the central dogma set out by James Watson, while those in red represent processes that were discovered later.

We will refer the reader to [4] on the topics of DNA replication, reverse transcription, and RNA replication. However, to understand how genes in the form of segments of DNA encode proteins it is particularly apt to give a brief description of the processes of transcription (DNA $\rightarrow$ RNA) and translation (RNA $\rightarrow$ Protein).

2.2.1 Transcription

Transcription is a process that takes place at genes in a cell, which for our purposes is some location in the DNA of an organism with specialized sequences. In terms of input-to-output, information encoded in a DNA sequence is copied into a newly synthesized RNA molecule. Such RNA molecules may have regulatory or catalytic roles in the cell, but in the CDMB one can think of them as an intermediate step towards producing protein. Transcription as a process is broadly described in three phases that we will briefly summarize.



(Public Domain)

Figure 7: Cartoon model of transcription occurring along a segment of DNA, producing a strand of messenger RNA through the catalytic action of RNA polymerase (RNAP). The cluster of circles and ellipses to the left of the diagram represent various initiation factors that help transcription get started at a specific location. RNAP reads the template strand in the $3' \to 5'$ direction while 'writing' messenger RNA in the $5' \to 3'$ direction. The coding strand is not 'read' by RNAP, however the messenger RNA transcript will have a corresponding sequence notwithstanding the hydroxyl difference in pentose sugars and presence of uridine instead of thymine.

The first phase is called *initiation*, which involves RNA polymerase (RNAP) along with a collection of transcription factors (i.e. other molecules related to stability and gene regulation) binding to the promoter sequence of the gene. The promoter sequence of a gene is a subsequence of DNA near the start of the gene whose content allows for RNAP to bind. Thus the physical availability of a promoter for binding is a form of gene regulation that molecular machines can modulate. Sometimes a collection of genes are controlled by a single promoter, and in such cases these collections are termed an *operon*. A later section will discuss the *lac* operon, which is a classic example from Molecular Biology of a gene regulatory network.

The second phase is *elongation*, in which RNAP reads the template strand in the $3' \to 5'$ direction while synthesizing messenger RNA in the $5' \to 3'$ direction (See Figure 7). The resulting RNA will have the same nitrogenous bases as the DNA coding strand except where a thymine exists in the coding strand there will instead by a uracil. Bose thymine and uracil are complementary to adenine through hydrogen bonding.

The third and final stage of transcription is the *termination* phase in which RNAP dissociates from the template strand and stops the production of the RNA molecule. There are multiple mechanisms in which termination of transcription occurs, which are not discussed

here.

We will describe translation momentarily, but it warrants mentioning that RNA produced through transcription does not always get immediately translated into protein. This is sometimes because the RNA molecule is not protein-encoding, but also because the RNA itself can require additional processing before the biologically-relevant protein product can result from translation. A few such processes include RNA splicing, the addition of a 7-methylguanosine to the $5'$ end, and polyadenylation [11]. Such a RNA molecule that has been processed for translation is called *messenger RNA* (mRNA).

2.2.2 Translation

Translation (Figure 8) is a process that involves the construction of proteins from amino acids according to a sort of code. Since proteins are constructed from sequences of amino acids, a related question is how this ordering comes to be. A codon is a subsequence of three nucleotides that corresponds to an amino acid. Given an RNA sequence, there are three possible sliding windows in which to consider codons which are called *reading frames*. The reading frame used in the context of translation is set by a special type of codon called the *start codon*.[1] This mapping is illustrated in Panel (b) of Figure 8.[2] While there exist exceptions, most organisms have this same mapping from subsequences of mRNA to amino acids. There are three codons that do not correspond to amino acids, and rather are involved in the stopping (i.e. termination) of translation.

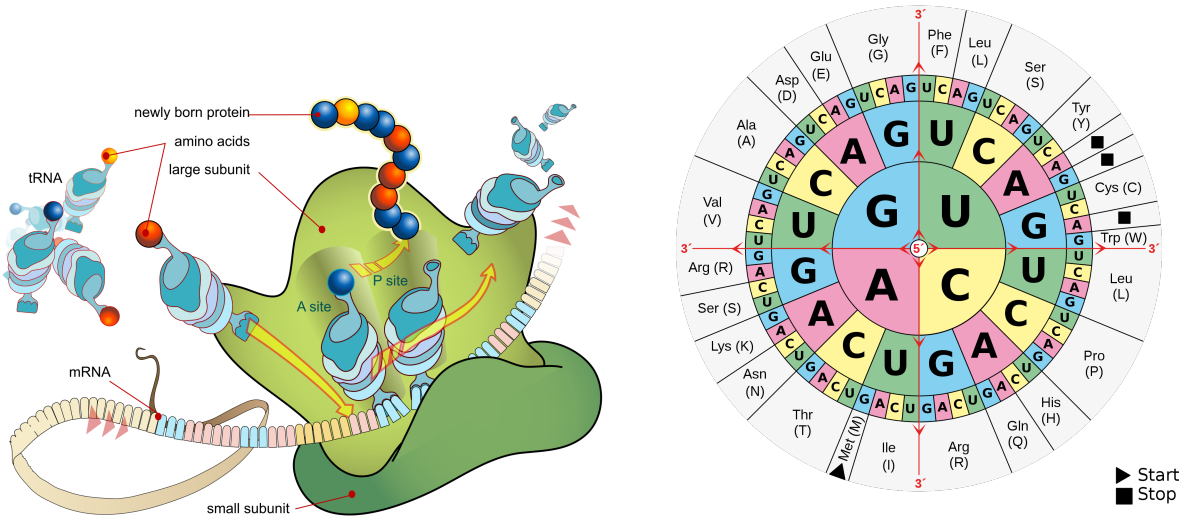
In terms of physical entities, the following must be present in order for translation to occur:

- Ribosome
- mRNA
- transfer RNAs (tRNAs)
- nucleotide triphosphates (ATP and GTP)

[1] Multiple overlapping reading frames are possible, but we do not consider them here.

[2] Notice that some codons map to the same amino acid. This is due to the *Wobble Effect* that is further explained in [12].

(a) (b)



(Public Domain) (Public Domain)

Figure 8: (a) Cartoon model of translation. (b) Circular RNA codon table.

The ribosome itself is a molecular machine composed of both RNA and protein, and is generally subdivided into a large and small subunit which correspond respectively to the larger light green and smaller dark green blobs in Panel (a) of Figure 8. One model of how the ribosome functions is to think of it reading in an mRNA molecule like a tape, moving from codon-to-codon according to a ratcheting-like motion [13].

We have explained the origin of mRNA already, but it warrants emphasizing that its sequence content represents "instructions" from a gene on how to assemble the sequence of a protein. In the abstract this follows the mapping shown in Panel (b) of Figure 8, but this mapping must be physically embodied by a real system.

In the most immediate sense, the physical embodiment of the codon-to-amino-acid code comes from the transfer RNA molecules. Transfer RNA molecules have (1) an amino acid bonded to them and (2) a subsequence of RNA that is complementary to a specific codon through hydrogen bonding interactions. These complementary triples are called *anticodons*. We will not explain why certain amino acids get bounded to specific RNA molecules with a specific anticodon, but it is an important question in evolutionary biochemistry.

The last physical component we listed above are the nucleotide triphosphates. They are the molecules whose hydrolysis (i.e. breakdown with the use of water) provides the energy required for the synthesis of protein to take place. Two examples of nucleotide triphosphates that are used in translation are adenosine triphosphate and guanosine triphosphate.

Up until now we have described translation in terms of the components that are involved, or as an abstract map, but since translation is a process it is important to briefly summarize

it in terms of discernible steps. As with transcription, translation is also described by the phases *initiation*, *elongation*, and *termination*. In initiation the ribosome assembles around an mRNA molecule with the first tRNA hydrogen-bonded to the a start codon. The elongation phase is a sort of ratcheting process in which the mRNA molecule is fed into the ribosome in a discrete way. This is done in a way that is consistent with the reading frame set by the start codon, and involves tRNAs moving into and out of the A site and P site of the ribosome (Figure 8). This incorprating, shifting, and releasing of tRNA coincides with adding amino acids to a growing polypeptide chain which must itself be repeatedly repositioned repeatedly through covalent bonding. All of this requires considerable energy in the for of ATP and GTP. The last stage of translation, namely termination, involves the ribosome reaching any one of three *stop codons* shown in Figure 8(b).

2.3 Summary

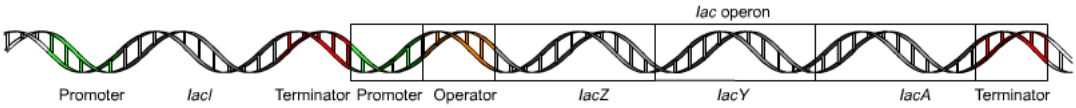
In this section we have introduced the big-picture view of how information flows from DNA to RNA, and then from RNA to protein. This is essential to understanding how genes are expressed, and it is through gene expression that gene regulatory networks function. While many details were exluded for brevity, it should prepare the reader with a better big picture for the sections that are to come. The next section will introduce the *lac* operon that exemplifies the notion of a gene regulatory network.

3 The *lac* Operon

The closest natural phenomena to genetic circuits are gene regulatory networks, and have only been possible to study with the development of methodologies in biochemistry and molecular biology within the last century. Gene regulatory circuits are broadly conceived of as network models whose nodes represent genes and whose edges represent 'interactions' between genes [14]. While such networks were once frequently inferred from transcriptional profiles generated from microarray analysis [15], other technologies such as RNA-Seq have emerged that provide similar information with some useful tradeoffs [16].[3] The main biological functions of gene regulatory networks are two-fold: (1) keep essential molecular functions stable and (2) adapt to the environment. Housekeeping genes perform the former role to keep the organism in homeostasis [17], and various other genes are orchestrated together in complex patterns of interaction to allow the organism to change its biochemical makeup in response to biotic and abiotic environmental stimuli. This broad conception of gene networks is useful for appreciating the potential complexity and abstract function of genetic circuits, however a concrete example of a specific gene regulatory network will now be provided to clarify 'how' such networks work.

[3]Remember that *RNA* is an abbreviation of *ribonucleic acid*, which is a similar type of molecule to DNA. The main structural difference between RNA and DNA is an additional alcohol group on the pentose sugars of the pentose-phosphate backbone. Messenger RNA (mRNA) is an intermediate macromolecule that confers protein structure specified by the DNA primary structure. Non-zero concentrations of mRNA indicate gene expression, and changes in mRNA concentrations indicate differential expression. RNA-Seq is an analytical technique that aims to quantify the number of mRNA transcripts for each gene.

Among the earliest (1961) gene regulatory networks to be studied is the *lac* operon, which controls only a small aspect of enteric bacterial catabolism of sugars [18]. The term "*lac operon*" is short for "*lactose operon*", which is a fitting label due to its functional role being the regulated metabolism of the lactose. An operon is a subsequence of DNA containing one or more genes that are under the shared control of a single promoter, which is itself a subsequence of DNA important in intiating the expression of genes [4].



(Public Domain)

Figure 9: Diagram of the *lac* operon. The collection of regions enclosed in rectangles consist of the entire *lac* operon, with each rectangle indicating the following elements: (1) promoter, (2) operator, (3) *lacZ*, (4) *lacY*, (5) *lacA*, and (6) terminator. Upstream (left) of the *lac* operon is the operon for the *lacI* gene along with its own promoter and terminator.

Figure 9 illustrates the structure of the *lac* operon. Because the *lac* operon contains genes that encode for multiple distinct proteins, its mRNA transcript is labelled as "*polycistronic*".

The first subsequence of the *lac* operon is its promoter. While it is RNA polymerase that directly catalyzes the synthesis of nascent mRNA[4], the initiation of transcription is facilitated by the binding of various transcription factors. One of these transcription factors is an inititiation factor, called a "*σ-factor*", that stabilizes RNA polymerase binding at an initiation sequence.

The second subsequence of the *lac* operon is the operator, which is a site for a repressor to bind. The binding of a repressor prevents the transcription of the operon, which is a product of the *lacI* gene that resides in a monocistronic operon upstream from the *lac* operon. By default *lacI* is expressed, but the binding of allolactose to the repressor protein itself will reduce the association strength that the repressor has to the operator. Allolactose is similar to lactose itself, with a $\beta 1-6$ glycosidic linkage instead of a $\beta 1-4$ glycosidic linkage as found in lactose.[5] While the environment may have lactose rather than allolactose present, some portion of lactose is converted to allolactose. Thus allolactose is a signal of the presence of lactose for some bacteria, with allolactose activating the *lac* operon through negative induction.

The third subsequence of the *lac* operon is *lacZ*, which is a gene encoding a β-galactosidase. A β-galactosidase is an enzyme that cleaves (i.e. separates) lactose into D-galactose and D-glucose. Figure 10 illustrates how numerous monosaccharides are interrelated through group transfer and interconversion, including glucose and galactose. Glucose can be broken down

[4]Nascent mRNA is newly-synthesized messenger RNA that has not undergone any modifications such as splicing, $5'$ cap addition, or poly(A) tail addition.

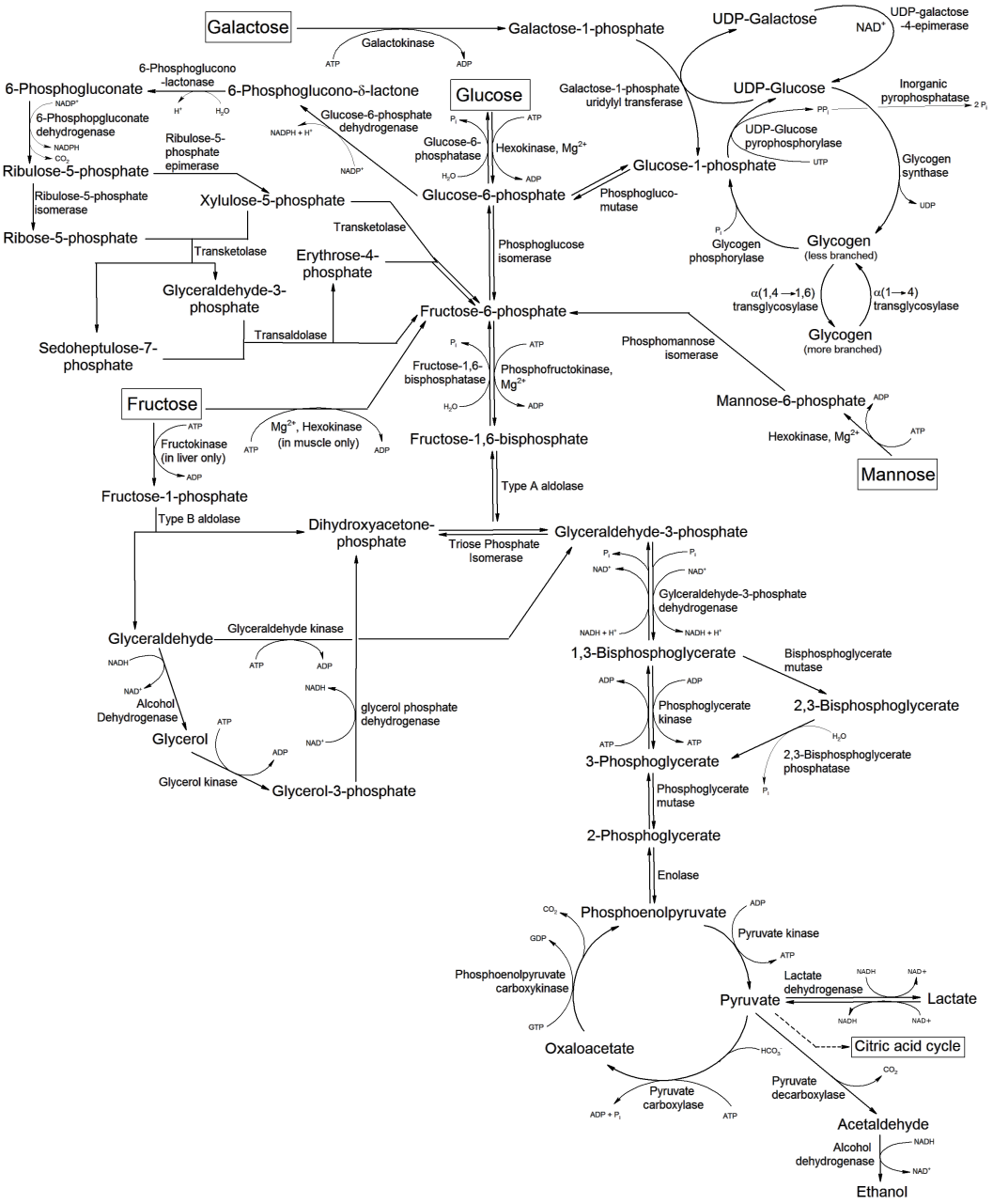
[5]A *glycosidic linkage* is a covalent bond between a carbohydrate and some other molecule that may or may not be a sugar. In the case of allolactose, this is a covalent bond between D-glucose and D-galactose.

in the glycolysis pathway to produce ATP, and galactose can be converted to glucose 6-phosphate which also enters into the glycolysis pathway [11]. A given galactose molecule is phosphorylated by galactokinase using adenosine triphosphate as the phosphate donor, which results in galactose 1-phosphate. Galactose 1-phosphate is then converted to UDP-galactose in a reaction catalyzed by galactose 1-phosphate uridylyl transferase which exchanges the UDP on UDP-glucose for a phosphate on a galactose 1-phosphate.[6] The UDP-galactose is oxidized by NAD^+ in a reaction catalyzed by UDP-galactose-5-epimerase, which results in a molecule of UDP-glucose.[7] The aforementioned uridylyl transferase reaction also involves replacing the uridylyl group on UDP-glucose with a phosphate, producing glucose 1-phosphate. Glucose 1-phosphate is then converted to glucose 6-phosphate in a reaction catalyzed by phosphoglucomutase. Glucose 6-phosphate continue through the glycolysis pathway to produce ATP and ATP-equivalents in the form of NADH.[8]

[6] *UDP* on its own is an abbreviation for *uridine diphosphate*, but attached to another molecule as a functional group it is referred to as a *uridylyl group*.

[7] *NAD* or *nicotinamide adenine dinucleotide* is an important coenzyme in biological systems because it acts as an electron carrier to deliver electrons needed in reduction reactions.

[8] NADH is the most common form of reduced NAD^+.



(Image Credit : Username LHcheM under a CC BY-SA 3.0 license.)

Figure 10: Metabolic pathway of some common monosaccharides, including interconversion of galactose and glucose. The shortest path of reactions in this diagram starting at glucose and ending with pyruvate is the canonical glycolysis pathway. Interconversion of various intermediates of the glycolysis pathway with common sugars such as fructose, galactose, and mannose illustrate how organisms are adaptive to availability of different energy and carbon sources.

Much of catabolism of nutrients happens within the cell. With cell membranes being relatively impermeable to all but small and non-polar compounds, organisms have adapta-

tions for importing nutrients [11]. The fourth subsequence of the *lac* operon is the *lacY* gene, which encodes the β-galactoside permease protein [19]. Permeases are transmembrane transport proteins that allow specific molecules to pass from one side of a membrane to another. Amongst other molecules, this protein will import lactose (which is polar and medium-sized) into the cell.

The fifth subsequence of the *lac* operon is the *lacA* gene which codes for the protein β-galactoside transacetylase. Transacetylases are enzymes that transfer acyl groups (R-C(=O)OH), which often involves the esterification reaction of an acid. In this case, the acyl group on acetyl-coenzyme A (a common cofactor in biological systems that acts as a good leaving group in organic reactions). Unfortunately this reaction's biological role is not well understood, inspite of the *lac* operon being heavily studied [20, 21].

The last subsequence of the *lac* operon is the terminator. Practically all genes have these because they signal for the RNA polymerase to stop transcribing the DNA. This allows specific collections of genes to be coexpressed in response to specific environmental conditions, such as the extracellular presence of lactose.

These subsequences only make up a part of the regulation of lactose catabolism. The lactose repressor provides one mechanism for negatively inducing the operon as a function of allolactose. A second mechanism involves responding to low glucose concentrations. When glucose concentrations are low, the concentration of cyclic adenosine monophosphate (cAMP; an endogenous signal molecule indicating an organisms energy state) increases [11]. When cAMP binds to catabolite activator protein (CAP), coded for by an enzyme outside the *lac* operon, this sets about an increased production of β-galactosidase [22]. Since β-galactosidase cleaves lactose into D-glucose and D-galactose, it is responsible for the catabolic but <u>not</u> transportation of lactose. These two regulatory mechanisms can work in parallel.

The canonical example of the *lac* operon gave an illustrative example of how regulatory structures behave can encode a biological function. In genetic circuit design the function of a collection of genes may not have a biological or evolutionary teleology, but rather an arbitrary one.

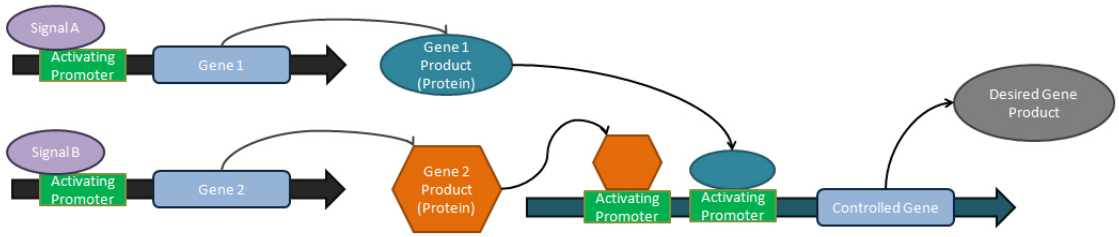
4 Genetic Circuit Design

With a better understanding of the underlying biochemistry from the last section, we now return to the subject of genetic circuit design. We will further clarify the analogy between circuits and gene regulation, and discuss the motivation and goals of the field.

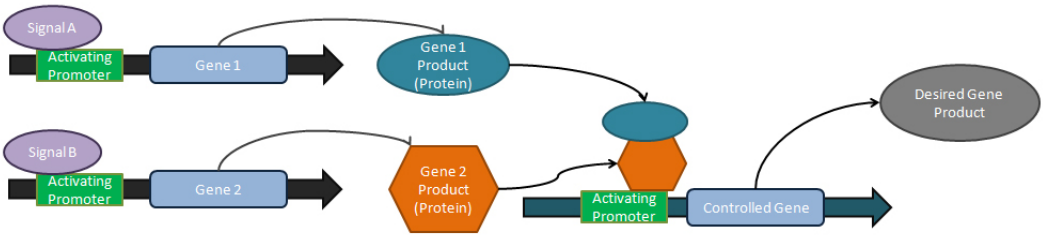
The analogy between of biological circuits to electrical circuits is twofold. The first is that a system can be designed to have a desired behaviour by configuring various components. In the case of electrical circuits this includes conductors, semiconductors, transistors, capacitors, resistors, power sources and various signal inputs and outputs. A biological circuit is implemented in terms of various molecules, cofactors, genes within DNA, and the complex

apparatuses for transcription, splicing, translation, post-translational modifications, and vesicle transport.

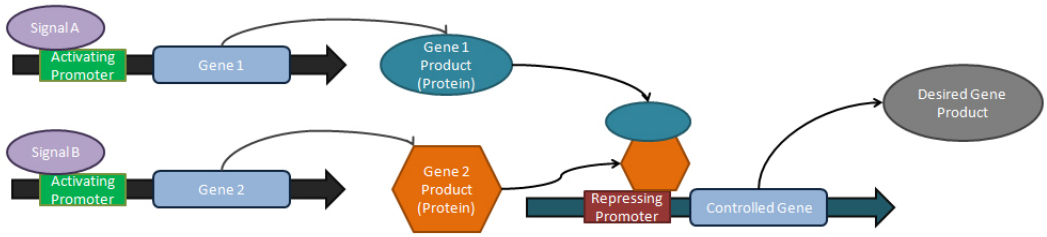
The second aspect of the analogy is the encoding of logic gates in terms of the components (See Figure 11 for three examples). While this is not inherently necessary, it allows design patterns and processes generally used for non-biological systems to be readily repurposed to biological systems. There are two possible unary logic gates (i.e. one input), and sixteen binary gates (i.e. two inputs), which can be implemented using molecular components [12]. It is through the composition of these logic functions that more complex functions can be represented, and thus encode more complicated behaviour [1].



(a)(Public Domain)



(b)(Public Domain)



(c)(Public Domain)

Figure 11: Schematic models of hypothetical implementations of logic gates using gene transcription and translation with regulation at activating or deactivating promoters. From top-to-bottom: (a) OR gate, (b) AND gate, and (c) NAND gate.

Analogies can be very useful in both the design and motivation of systems, however the details of implementation cannot be ignored. In biological circuits the response of one variable to another will not be a logic gate because of physical properties of the components and how they interact. In particular, the components of biological circuits are molecules which

are frequently transported by diffusion within the cell.[9] Since chemical diffusion is a result of the random thermal motion of ions and molecules in a fluid, there are substantial time delays between the synthesis and utilization of a molecule due to uncertainty in the molecules motion. Even once a molecule reaches its target, the interaction with its target may not be ideal. For example, transcription factors binding to a promoter do not simply bind once and wait for an RNA polymerase to show up. Rather, transcription factors will bind and unbind randomly with a frequency that depends on their binding affinity to the promoter [11].

A common modelling approach is to hypothesize a system of differential equations that govern the concentration-dependent interaction of components with respect to time [23]. There is no guarantee that a given system will fit observed data, however a combination of parametrization and heuristics can assist in posing models.[10] The law of mass action is not strictly a law of nature, but rather an often-correct principle which states the rate of a chemical reaction is directly proportional to the product of the activities or concentrations of the reactants.[11] These systems often yield sigmoidal (s-shaped) curves as solutions, which give a continuous and monotonic approximation to the notion of transitioning a gate from 'off' to 'on' or *vice versa*.

For genetic circuits in particular, detailed knowledge of the genome of the model organisms should be known in considerable detail to understand what behaviour to expect from the system under different environmental conditions. While humans were not the first organisms to have a full consensus genome constructed, the Human Genome Project that finished in 2001 involved developing technologies, principles, and approaches for reverse-engineering the genetic structure of complex organisms [24]. Craig Venter, one of the directors of the Human Genome Project, also designed and implemented the first 'minimal genome' bacteria where as many genetic components were removed as possible [25]. These are really important developments in understanding the genetics of organisms, but mapping out the genes of an organism is not sufficient in designing new systems. An understanding of the network of interactions between parts must be obtained, which has led to projects like the BioGrid database that attempts to collate and curate all biochemical interactions [26, 27].

The main goal of genetic circuit design to develop knowledge, processes, and tools to designing biological systems with desired behaviour. Such systems may allow the development of new drugs or treatments, improved food production, bioremediation of contaminated ecosystems, and highly-parallelized computation.

5 Advancements

This section will discuss some of the acheivements of genetic circuit design.

[9] Sometimes molecules are transported by vessicles or transport proteins.

[10] The paramatrization is often proposed as a direct proportionality.

[11] An activity of a molecule is a function of concentration, charge, and size of a molecule. For sufficiently large concentrations, the concentration alone can be used instead of the activity.

One of the earlier accomplishments in genetic circuit design was that of the toggle switch. A toggle switch is a device that can hold a one of two discrete states, and be manipulated to change from one of those states to the other. The first genetic toggle switch was controlled by two repressible promotors, which promotors that are default 'on' until bound by a repressor protein [28]. This work demonstrated that biological circuits can encode discrete memory, and behave like a finite state machine.

The next notable publication came in the same year (2000) as the toggle switch. This second study demonstrated the design of an oscillatory genetic circuit. Instead of toggling between two discrete states, the genetic circuit programmed periodic protein concentrations [29]. Because the feedback system came about by the use of repressors, this oscillating system was termed a *repressilator*. This work demonstrated that genetic circuits can produce periodic functions in the form of dynamic chemical equilibria.

Periodic protein concentrations implies that the system does not stabilize, but rather it oscillates through a collection of states. Another paper from 2000 demonstrated that negative feedback loops in the circuit structure provided stability to the target concentrations, and therefore approached a stable steady state [30]. When such feedbacks that produce stability occur in an organisms physiology, the stablilized state is referred to as *homeostasis*. Such a regulatory system that self-corrects is called *autoregulation*. As components are added to a system, their resulting pattern of interactions can result in chaos or nonholonomic properties.[12] This study demonstrated that inspite of random thermal motion and other processes adding deviations from the fixed point, negative feedback can bring the system back toward the fixed point.

In general it is not obvious if a system will exhibit stable, chaotic, or oscillatory behavior. Tyson *et al.* 2008 made a qualitative analysis of systems of differential equations representing genetic circuits and found mathematical principles to discern between the two outcomes [31]. Toggle switches provide a system with memory, oscillators can track time, and autoregulators provide stability and noise reduction. One might imagine that a clock device could be developed that acted as a stable clock that recorded time. This is what Dalchau *et al.* 2018 conceived of, and was able to produce a working example [32]. These works demonstrate that inspite of complex dynamics and interactions, it is possible to engineer genetic circuit systems with a specific and stable functionality.

The aforementioned studies are oriented around the first principles and proof of concepts that involve the implementation of genetic circuits. However, they make little use of the analogy of logic gates that provide a framework for scalable design of complex functions. Earlier works focused on producing a single working circuit, such as AND and NOR gates [33, 34], but by 2013 all of the canonical 2-inputs gates had been implemented [35].

Xie *et al* 2011 is of particular note in developing a wetware[13] implementation using

[12] *Chaos* is a high sensitivity in the outcome of a system to its initial conditions. A related concept to chaos is that of a *nonholonomic system*, whose final state depends on its history.

[13] The term *wetware* is an adapted term from the terms *hardware* and *software* found in Computer Science.

interferring RNA that was capable of detecting whether a biological cell was normal or cancerous [36]. The type of cancer cells used in this study were HeLa cells. This cell line is famously from Henrietta Lacks whose family had purportedly not given permission to use her cells for research. Contamination of cell lines involving HeLa cells has been an ongoing issue [37], but it does not detract from Xie *et al* 2011 work on a particular culture. The circuit was a classifier that could discriminate between the two type of cells, and would signal only the cancer cells to undergo apoptosis. *Apoptosis* is a synonym for *programmed cell death*, which can be thought of as a biochemical process endogenous to a cell that involves its self destruction. To validate their system they put their classifier circuit into a normal cell line and a HeLa cancer cell line, and observed that no cell culture growth characteristics changed in the normal cells while the HeLa cells showed cell mortality. This work demonstrated that (1) the gate logic does not have to entirely rely on transcriptional control and (2) that genetic circuits might have real-world applications in fighting uncontrolled cell growth.

6 Challenges

While genetic circuit design has come a long way in only a couple of decades, there are ongoing challenges in the field to be overcome. We will briefly outline some of those challenges here.

The first challenge comes back to the nature of diffusion and chemistry. Chemical diffusion implies that where the molecules go is uncontrolled within a given cellular compartment, and difficult to control between cellular compartments. If components of biological circuits didn't have any interactions than the desired interactions, then this would be more of an inefficiency than a general challenge. In practice chemicals in biological systems undergo numerous interactions, some are covalent and others not. This means that components do not independently affect the behaviour of the cell, and may result in further inefficiencies or even the circuit having undesired functions. This lack of independence of the components is called *non-orthogonality* in the literature.

The second challenge to genetic circuit design is metabolic load.[14] Metabolic load is the amount of resources in the forms of energy, carbon, and other nutrients that the organism is using to run all of its biological and synthetic processes [11]. Macromolecules such as DNA, RNA, and proteins in particular require a lot of metabolic resources. As genetic and biological circuits tend to be implemented in terms of these molecules, in addition to other molecules, running a designed circuit tends to require a lot of resources. The limits on available resources lead to selection pressures that ensure natural organisms are (locally) optimized to be efficient with those resources [38, 39, 40]. The modifications to an organism's metabolic flux due to the addition of a synthetic circuit are not necessarily within the feasible resources that can be spared, or in the weaker case the organism is simply less likely to be able to reproduce. This selective pressure can lead to cultures evolving to no longer

Wetware in this case corresponds to the physical objects in a biological system that are physical embodiments of the 'software'.

[14] This is sometimes also called *metabolic burden*.

have the original circuit, or have mechanisms that deactivate the circuit.

The most promising candidate solution in the literature for dealing with non-orthogonality of parts and metabolic load is to distribute the parts of a circuit across multiple strains of microorganism [41]. In order for a given logic circuit to be implemented across multiple organisms, it must be implemented with cell-to-cell communication. In other words, signals must be allowed to travel between cells. By isolated interfering components into different cells, the risk of unexpected behaviour might be lowered. By taking a large circuit that would otherwise be implemented in each individual of a population, and instead implementing smaller sub-circuits into different strains would potentially imply lower metabolic load for each individual.

7 Conclusions

In this review we briefly introduced the biological basis of genetic circuits, and the analogy that motivates them. The main takeaways from these sections are (1) regulatory control can be achieved through gene transcription, and (2) known chemical interactions can be approximately represented by logic gates.

We then provided a summary of the major acheivements of the field since its inception approximately twenty years ago. Lastly we discussed some challenges of the field. The classic accomplishments of this field were development of toggle switches, clocks, and autoregulation implemented by genetic circuits. However, it is important to note that genetic circuit design is not just studying the construction of abstract logic gates. The example of Xie *et al* 2011 illustrates that genetic circuits can be integrated with cellular structures to serve a desired biological function such as cancer detection.

Future work should more comprehensively collate and categorize the genetic circuits that have been implemented, including the wetware implementation and the logical design. Future work should also further discuss ways of dealing with the challenges associated with genetic circuits, and recent trends of the literature.

References

[1] Jennifer A N Brophy and Christopher A Voigt. Principles of genetic circuit design. *Nature Methods*, 11(5):508–520, April 2014.

[2] Bruce Alberts. *Molecular biology of the cell*. Garland Science, Taylor and Francis Group, New York, NY, 2015.

[3] Claude Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October 1948.

[4] Michael Cos David Nelson and Albert Lehninger. *Lehninger principles of biochemistry*. W.H. Freeman and Company ; Macmillan Higher Education, New York, NY : Houndmills, Basingstoke, seventh edition edition, 2017. OCLC: ocn986827885.

[5] James Watson and Francis Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. 171(4356):737–738, April 1953.

[6] James Watson. *The double helix : a personal account of the discovery of the structure of DNA*. Touchstone Book, published by Simon & Schuster, New York, 2001.

[7] Brenda Maddox. The double helix and the 'wronged heroine'. *Nature*, 421(6921):407–408, January 2003.

[8] Rosalind franklin was so much more than the 'wronged heroine' of DNA. *Nature*, 583(7817):492–492, July 2020.

[9] L. G. Wade. *Organic chemistry*. Pearson, Boston, 2013.

[10] Matthew Cobb. 60 years ago, francis crick changed the logic of biology. 15(9):e2003243, September 2017.

[11] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt. Genetic circuit design automation. *Science*, 352(6281):aac7341–aac7341, March 2016.

[12] Seth G. Abels and Emil F. Khisamutdinov. Nucleic acid computing and its potential to transform silicon-based technology. *DNA and RNA Nanotechnology*, 2(1), January 2015.

[13] Wen Zhang, Jack A. Dunkle, and Jamie H. D. Cate. Structures of the ribosome in intermediate states of ratcheting. 325(5943):1014–1017, August 2009.

[14] Albert-László Barabási and Zoltán N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, February 2004.

[15] T. I. Lee. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298(5594):799–804, October 2002.

[16] Federico M. Giorgi, Cristian Del Fabbro, and Francesco Licausi. Comparative study of RNA-seq- and microarray-derived coexpression networks in arabidopsis thaliana. *Bioinformatics*, 29(6):717–724, February 2013.

[17] Atul J. Butte, Victor J. Dzau, and Susan B. Glueck. Further defining housekeeping, or "maintenance," genes focus on "a compendium of gene expression in normal human tissues". *Physiological Genomics*, 7(2):95–96, December 2001.

[18] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318 – 356, 1961.

[19] A. R. Tarlov and E. P. Kennedy. The beta-galactoside permease system and the metabolism of phospholipids in escherichia coli. *J Biol Chem*, 240:49–53, Jan 1965.

[20] Xing-Guo Wang, Laurence R Olsen, and Steven L Roderick. Structure of the lac operon galactoside acetyltransferase. *Structure*, 10(4):581–588, April 2002.

[21] Steven L. Roderick. The lac operon galactoside acetyltransferase. *Comptes Rendus Biologies*, 328(6):568–575, June 2005.

[22] T.Philip Malan, Annie Kolb, Henri Buc, and William R. McClure. Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the p1 promoter. *Journal of Molecular Biology*, 180(4):881–909, December 1984.

[23] Jafar Kazemi and Sadjad Ozgoli. Intracellular regulatory control of gene expression process. *CoRR*, abs/1606.04163, 2016.

[24] J. Craig Venter, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith, Mark Yandell, Cheryl A. Evans, Robert A. Holt, Jeannine D. Gocayne, Peter Amanatides, Richard M. Ballew, Daniel H. Huson, Jennifer Russo Wortman, Qing Zhang, Chinnappa D. Kodira, Xiangqun H. Zheng, Lin Chen, Marian Skupski, Gangadharan Subramanian, Paul D. Thomas, Jinghui Zhang, George L. Gabor Miklos, Catherine Nelson, Samuel Broder, Andrew G. Clark, Joe Nadeau, Victor A. McKusick, Norton Zinder, Arnold J. Levine, Richard J. Roberts, Mel Simon, Carolyn Slayman, Michael Hunkapiller, Randall Bolanos, Arthur Delcher, Ian Dew, Daniel Fasulo, Michael Flanigan, Liliana Florea, Aaron Halpern, Sridhar Hannenhalli, Saul Kravitz, Samuel Levy, Clark Mobarry, Knut Reinert, Karin Remington, Jane Abu-Threideh, Ellen Beasley, Kendra Biddick, Vivien Bonazzi, Rhonda Brandon, Michele Cargill, Ishwar Chandramouliswaran, Rosane Charlab, Kabir Chaturvedi, Zuoming Deng, Valentina Di Francesco, Patrick Dunn, Karen Eilbeck, Carlos Evangelista, Andrei E. Gabrielian, Weiniu Gan, Wangmao Ge, Fangcheng Gong, Zhiping Gu, Ping Guan, Thomas J. Heiman, Maureen E. Higgins, Rui-Ru Ji, Zhaoxi Ke, Karen A. Ketchum, Zhongwu Lai, Yiding Lei, Zhenya Li, Jiayin Li, Yong Liang, Xiaoying Lin, Fu Lu, Gennady V. Merkulov, Natalia Milshina, Helen M. Moore, Ashwinikumar K Naik, Vaibhav A. Narayan, Beena Neelam, Deborah Nusskern, Douglas B. Rusch, Steven Salzberg, Wei Shao, Bixiong Shue, Jingtao Sun, Zhen Yuan Wang, Aihui Wang, Xin Wang, Jian Wang, Ming-Hui Wei, Ron Wides, Chunlin Xiao, Chunhua Yan, Alison Yao, Jane Ye, Ming Zhan, Weiqing Zhang, Hongyu Zhang, Qi Zhao, Liansheng Zheng, Fei Zhong, Wenyan Zhong, Shiaoping C. Zhu, Shaying Zhao, Dennis Gilbert, Suzanna Baumhueter, Gene Spier, Christine Carter, Anibal Cravchik, Trevor Woodage, Feroze Ali, Huijin An, Aderonke Awe, Danita Baldwin, Holly Baden, Mary Barnstead, Ian Barrow, Karen Beeson, Dana Busam, Amy Carver, Angela Center, Ming Lai Cheng, Liz Curry, Steve Danaher, Lionel Davenport, Raymond Desilets, Susanne Dietz, Kristina Dodson, Lisa Doup, Steven Ferriera, Neha Garg, Andres Gluecksmann, Brit Hart, Jason Haynes, Charles Haynes, Cheryl Heiner, Suzanne Hladun, Damon Hostin, Jarrett Houck, Timothy Howland, Chinyere Ibegwam, Jeffery Johnson, Francis Kalush, Lesley Kline, Shashi Koduru, Amy Love, Felecia Mann, David May, Steven McCawley, Tina McIntosh, Ivy McMullen, Mee Moy, Linda Moy, Brian Murphy, Keith Nelson, Cynthia Pfannkoch, Eric Pratts, Vinita Puri, Hina Qureshi, Matthew Reardon, Robert Rodriguez, Yu-Hui Rogers, Deanna Romblad, Bob Ruhfel, Richard Scott, Cynthia Sitter, Michelle Smallwood, Erin Stewart, Renee Strong, Ellen Suh, Reginald Thomas, Ni Ni Tint, Sukyee Tse, Claire Vech, Gary Wang, Jeremy Wetter, Sherita Williams,

Monica Williams, Sandra Windsor, Emily Winn-Deen, Keriellen Wolfe, Jayshree Zaveri, Karena Zaveri, Josep F. Abril, Roderic Guigó, Michael J. Campbell, Kimmen V. Sjolander, Brian Karlak, Anish Kejariwal, Huaiyu Mi, Betty Lazareva, Thomas Hatton, Apurva Narechania, Karen Diemer, Anushya Muruganujan, Nan Guo, Shinji Sato, Vineet Bafna, Sorin Istrail, Ross Lippert, Russell Schwartz, Brian Walenz, Shibu Yooseph, David Allen, Anand Basu, James Baxendale, Louis Blick, Marcelo Caminha, John Carnes-Stine, Parris Caulk, Yen-Hui Chiang, My Coyne, Carl Dahlke, Anne Deslattes Mays, Maria Dombroski, Michael Donnelly, Dale Ely, Shiva Esparham, Carl Fosler, Harold Gire, Stephen Glanowski, Kenneth Glasser, Anna Glodek, Mark Gorokhov, Ken Graham, Barry Gropman, Michael Harris, Jeremy Heil, Scott Henderson, Jeffrey Hoover, Donald Jennings, Catherine Jordan, James Jordan, John Kasha, Leonid Kagan, Cheryl Kraft, Alexander Levitsky, Mark Lewis, Xiangjun Liu, John Lopez, Daniel Ma, William Majoros, Joe McDaniel, Sean Murphy, Matthew Newman, Trung Nguyen, Ngoc Nguyen, Marc Nodell, Sue Pan, Jim Peck, Marshall Peterson, William Rowe, Robert Sanders, John Scott, Michael Simpson, Thomas Smith, Arlan Sprague, Timothy Stockwell, Russell Turner, Eli Venter, Mei Wang, Meiyuan Wen, David Wu, Mitchell Wu, Ashley Xia, Ali Zandieh, and Xiaohong Zhu. The sequence of the human genome. *Science*, 291(5507):1304–1351, 2001.

[25] Clyde A. Hutchison, Scott N. Peterson, Steven R. Gill, Robin T. Cline, Owen White, Claire M. Fraser, Hamilton O. Smith, and J. Craig Venter. Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, 286(5447):2165–2169, 1999.

[26] C. Stark. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34(90001):D535–D539, January 2006.

[27] Rose Oughtred, Chris Stark, Bobby-Joe Breitkreutz, Jennifer Rust, Lorrie Boucher, Christie Chang, Nadine Kolas, Lara O'Donnell, Genie Leung, Rochelle McAdam, Frederick Zhang, Sonam Dolma, Andrew Willems, Jasmin Coulombe-Huntington, Andrew Chatr-aryamontri, Kara Dolinski, and Mike Tyers. The BioGRID interaction database: 2019 update. *Nucleic Acids Research*, 47(D1):D529–D541, November 2018.

[28] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342, Jan 2000.

[29] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, Jan 2000.

[30] Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, Jun 2000.

[31] John J Tyson, Reka Albert, Albert Goldbeter, Peter Ruoff, and Jill Sible. Biological switches and clocks. *Journal of The Royal Society Interface*, 5(suppl_1), June 2008.

[32] Neil Dalchau, Gregory Szép, Rosa Hernansaiz-Ballesteros, Chris P. Barnes, Luca Cardelli, Andrew Phillips, and Attila Csikász-Nagy. Computing with biological switches and clocks. *Natural Computing*, 17(4):761–779, Dec 2018.

[33] J Christopher Anderson, Christopher A Voigt, and Adam P Arkin. Environmental signal integration by a modular AND gate. *Molecular Systems Biology*, 3(1):133, January 2007.

[34] Alvin Tamsir, Jeffrey J. Tabor, and Christopher A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, 469(7329):212–215, December 2010.

[35] Piro Siuti, John Yazbek, and Timothy K Lu. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, 31(5):448–452, February 2013.

[36] Zhen Xie, Liliana Wroblewska, Laura Prochazka, Ron Weiss, and Yaakov Benenson. Multi-input rnai-based logic circuit for identification of specific cancer cells. *Science*, 333(6047):1307–1311, 2011.

[37] John R. Masters. HeLa cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer*, 2(4):315–319, April 2002.

[38] Charles Darwin. *On the Origin of Species by Means of Natural Selection*. Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.

[39] A. J. Lotka. Natural selection as a physical principle. *Proceedings of the National Academy of Sciences*, 8(6):151–154, June 1922.

[40] A. J. Lotka. Contribution to the energetics of evolution. *Proceedings of the National Academy of Sciences*, 8(6):147–151, June 1922.

[41] Behzad D. Karkaria, Neythen J. Treloar, Chris P. Barnes, and Alex J. H. Fedorec. From microbial communities to distributed computing systems. *Frontiers in Bioengineering and Biotechnology*, 8:834, 2020.