Microbial Circuits: A New Paradigm in Computing

Abstract: *The essence of the classical computation paradigm is the processing of strings of 0s and 1s. The basic components constituting the entire mass of classical computers are logic gates fabricated from semiconductor material. In this model a high voltage at the input of a gate amounts to a 1 and a low voltage to a 0. Broadly speaking, we may state that the presence of a certain stimulus, which may be of any form signifies a 1 and the absence of this stimulus a 0. Efforts have been continually made to develop new methods of computation by varying the nature of these stimuli. This study intends to show how DNA – the "molecule of life", may function as a logic gate. An attempt is made to demonstrate how existing genetic regulatory mechanisms of repression and transcription may be harnessed to engineer in vivo digital circuits. We propose to conduct an experiment to discuss the feasibility of DNA working as an inverter. This experiment is based on the four basic processes of molecular biology namely, transcription – the formation of RNA from DNA, repression – deactivation of a gene by a repressor protein, the central dogma – the unidirectional flow of information from DNA to RNA to protein and the "one gene – one polypeptide" hypothesis of Beadle and Tatum. The presence of a repressor is considered as a logic 1 and its absence a 0. Logic signals are represented by the synthesis rate of DNA binding protein as stated in the "Operon Mode" given by Jacob and Monod in 1961. The results obtained will be analyzed with respect to the feasibility and robustness of this model of the DNA inverter. We also propose methods to combine microbial gates to form more complex circuits.*

Keywords: *DNA inverter, Microbial circuits, Lac Operon, Gene Expression, Set Valued Logic*

Saif Ali
Sikander Hayat
Department of Computer Engineering
Faculty of Engineering and Technology
Jamia Millia Islamia, Delhi

1. INTRODUCTION

Biological cells possess important characteristics, such as energy efficiency, miniature scale and self-reproduction that make them suitable for use as logic elements for computational circuits. The microbial circuit design process transforms the logic circuit into a genetic network whose chemical activity in vivo implements the computation specified by the logic circuit. In this paper we discuss the computationally ability that is inherently present in the molecule of life - DNA. The paper will also discuss the expediency of a computer system based on such logic gates. The experimental results and the advantages of such a system over a conventional computer system are also discussed.

2. OBJECTIVES

- To demonstrate that protein synthesis is a process which follows logical NOT configuration.
- To discuss the possibility of eliciting NAND functionality from the trp Operon.
- To outline the process of microbial circuit design.

3. BIOLOGICAL CONCEPTS USED

3.1 Regulation of gene expression

Of the estimated 100,000 genes in the human genome only a fraction of these are expressed at any given time. Some gene products are present in large amounts in cells such as RUBISCO in plants and photosynthetic bacteria. Other gene products occur in smaller amounts; for instance, a cell may have enzymes required for lactose utilization[1]. The intracellular concentration of gene product is driven by at least seven processes working as a sequential circuit and each having potential points of regulation. These are:

1. Synthesis of primary RNA transcript.
2. Posttranscriptional processing of mRNA.
3. Protein synthesis (translation).
4. mRNA degradation.
5. Posttranslational modification of proteins.
6. Protein degradation.
7. Protein targeting and transport.

3.2 Principles of gene regulation

Genes for products that are required at all times are expressed at a constant level in every cell. Genes for the enzymes of central metabolic pathways such as (citric acid cycle), fall in this denomination and are referred to as *housekeeping gene*. The unvarying expression of such a gene is called *constitutive gene expression*[1].

The cellular levels of some gene products rise and fall in response to molecular signals; this is **regulated gene expression**. Gene products that increase in concentration under particular molecular circumstances are referred to as **inducible;** the process of increasing their expression is called induction. For example, in the event of exhaustion of glucose in the culture medium the bacteria use lactose as a carbon source of energy. This requires induction of lactose utilizing genes.

Another kind of gene regulation is *repression*. For example, ample supplies of *tryptophan* lead to repression of genes for the enzymes that catalyze *tryptophan* biosynthesis[5].

Transcription is mediated and regulated by protein-DNA interactions, especially those involving the protein components of RNA polymerase.

3.2.1 How the activity of RNA polymerase is regulated

RNA polymerase bind to DNA and initiate transcription at sites called promoters, found near points at which RNA synthesis begins on DNA template the regulation of transcription initiation often entails changes in how RNA polymerase interacts with a promoter. The nucleotide sequences of promoters vary only slightly. These slight variations influence the binding affinity of RNA polymerase and thus the frequency of transcription initiation. Much of this variation is due to differences in promoter sequence.

Although housekeeping genes are expressed constitutively, the cellular concentrations of the proteins they encode varies widely. For these genes, the RNA polymerase promoter interaction strongly influences the rate of transcription initiation; differences in promoter sequence allow the cell to synthesis the appropriate level of each housekeeping gene product.

Along with the above process, *regulatory proteins* further modulate the basal rate of transcription initiation at promoters of non-housekeeping genes .These proteins often work by enhancing or interfering with the interaction between RNA polymerase and the promoter[1].

3.3 Proteins that bind to or near promoters regulate transcription initiation

At least three types of proteins regulate transcription initiation by RNA polymerase. *Specificity factors* alter the specificity of RNA polymerase for a given set of promoters. *Repressors* impede access of RNA polymerase to the promoter. *Activators* enhance the RNA-promoter interaction.

Repressors bind to specific sites on the DNA .in prokaryotic cells, such binding sites called OPERATORS, are generally near a promoter. RNA polymerase binding, or its movement along the DNA after binding, is blocked when the repressor is present. Regulation brought about by means of a repressor protein that blocks transcription is referred to as **negative regulation.**

Repressor binding to DNA is regulated by a molecular signal called **effector** usually a protein that binds to the repressor and causes conformational change. The interaction between repressor and signal molecule either increases or decreases transcription.

Activators provide a molecular counter part to repressors; they bind to DNA and enhance the activity of RNA polymerase at a promoter; this is **positive regulation**. Activator binding sites are often adjacent to promoters that are bound weakly or not at all by RNA polymerase alone, such that little transcription occurs in the absence of the activator. Some activators bind to DNA sites called enhancers that are distant from the promoter, influencing the rate of transcription at a promoter that may be located thousands of base pairs downstream. Some activators are normally bound to DNA enhancing transcription until binding of signal molecule.in other cases triggers dissosiation the activator bind to DNA only after interaction with a signal molecule. Signal molecule can therefore increase or decrease transcription depending on how they affect the activator[1, 5, 6].

3.4 OPERON Model (François Jacob, Jacques Monod, 1960)

Bacteria have a simple general mechanism for coordinating the regulation of genes that encode products involved in a set of related processes: these genes are clustered on the chromosome and are transcribed together. Most prokaryotic mRNA s are polycistronic –multiple genes on as ingle transcript-and the single promoter that initiates transcription of the cluster is the site of regulation for expression of all the genes in the cluster. The gene cluster and promoter, plus additional sequences that function together in gene regulation, are called an *OPERON*[6].

3.4.1 Example : Lac Operon

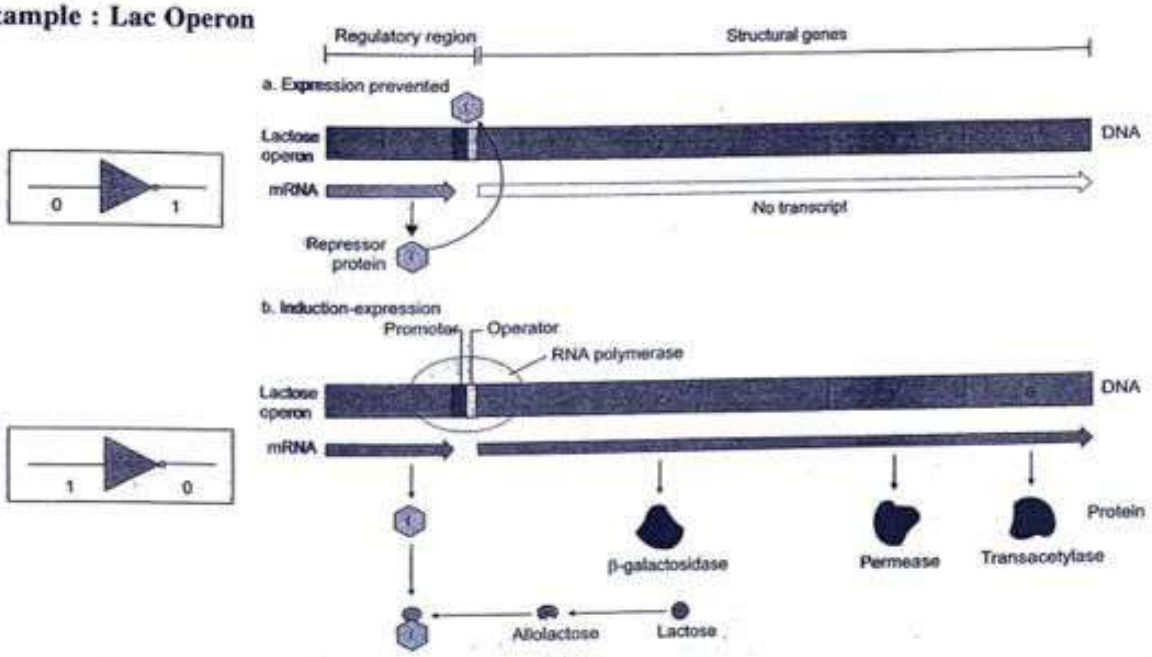


Fig. 1

Table 1. Truth Table of DNA Inverter

Input (Repressor Protein)	Operator	Output (Gene Expression)
Absent - Logic 0	ON	Expressed Logic 1
Present Logic 1	OFF	Not Expressed Logic 0

Conclusion 1

- **The expression of gene, based on presence or absence of repressor is similar to that of an inverter**

When the sugar lactose is added to the cultures of *E.coli*, it induces three enzymes necessary to break down the lactose into glucose and galactose. These three genes b-galactosidase, permease and transacetylase are referred to as *structural genes* and are regulated as a unit by a single switch referred to as *operator*. The genes are expressed or not expressed depending on whether the operator switch is ON or OFF. This operator is switched ON or OFF by a protein referred to as repressor. When this protein binds to the operator and blocks it, the switch is turned OFF and the three genes are not expressed. The repressor is coded by another gene, which is called the 'I' gene, and is referred to as *regulatory gene* and its product as regulatory protein.

Inducer: It's a chemical which inactivates the repressor and ceases its action of joining with the operator region.

3.4.2 Example: Tryptophan Operon

This operon is similar in concept to Lac O only difference being the feedback action of the final proteins on further gene regulation. Two new terms are introduced here.

- **Co Repressor**: It's a non-protein compound which may come from outside or from metabolism within the cell. Its concentration controls transcription in the cell. If concentration of a co repressor rises, aporepressor - co repressor complex is formed, the complex joins the operator gene and further production of co repressor stops. *In tryptophan OPERON, the end product, i.e tryptophan acts as co repressor.*
- **Apo repressor:** It's a protein coded by the regulator gene. It combines with co repressor, and apo repressor –co repressor complex joins the operator gene, which is turned OFF, and enzyme synthesis ends[1].

Table 2. Truth Table of DNA NAND gate

Input A (Apo repressor)	Input B (Co repressor)	Output (Gene Expression)
Absent - Logic 0	Absent - Logic 0	Gene Expressed Logic 1
Absent - Logic 0	Present - Logic 1	Gene Expressed Logic 1
Present - Logic 1	Absent - Logic 0	Gene Expressed Logic 1
Present - Logic 1	Present - Logic 1	Not Expressed Logic 0

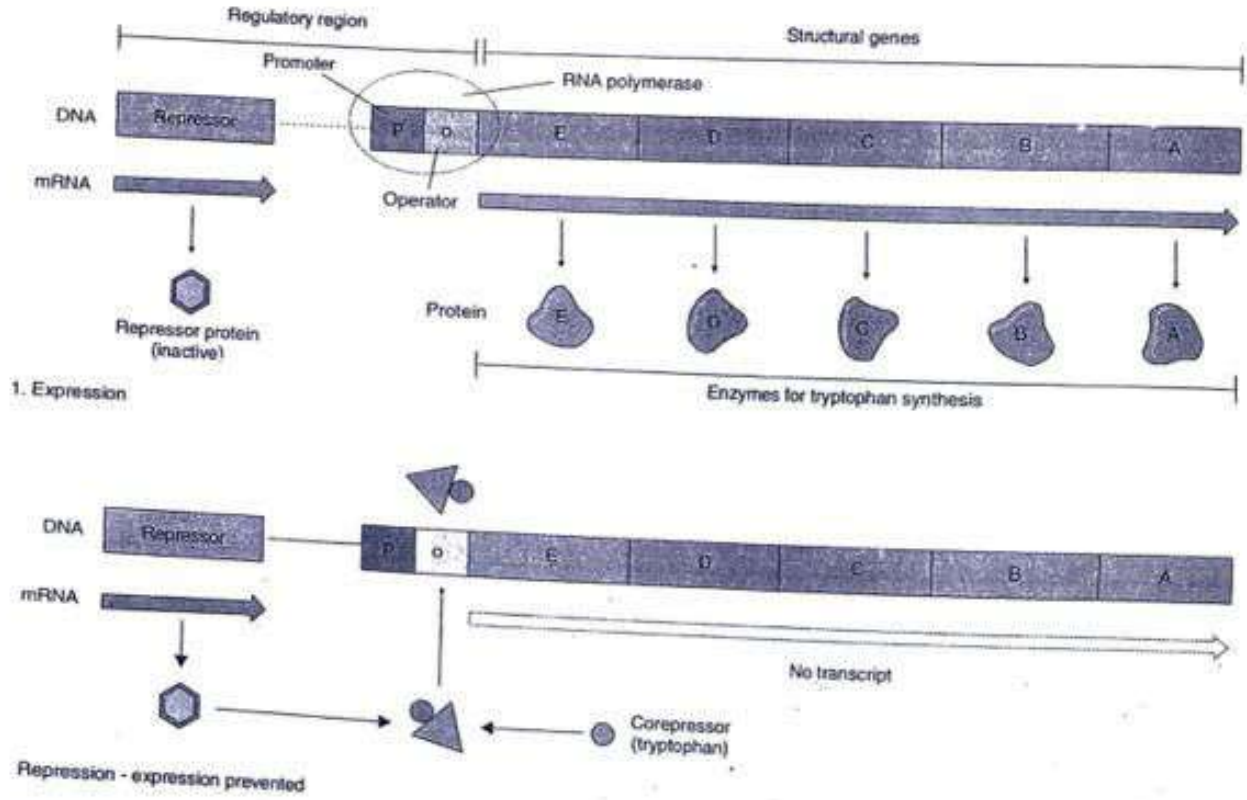


Fig. 2

Conclusion 2

- **The expression of gene, based on presence or absence of the combination of apo repressor and co repressor is similar to that of a NAND gate.**

4. THE BASICS OF *IN-VIVO* DIGITAL GATES

The approach for engineering in vivo digital circuits harnesses existing genetic regulatory mechanisms of repression and transcription.

The three basic elements of any electronic circuit are:

- The basic computational element (inverter in our case).
- Connecting wires.
- Storage element.

4.1 Computation: Analysis of a DNA inverter

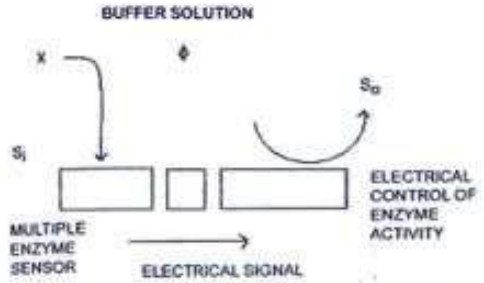
Let ϕA denote a signal defined as the synthesis rate of a DNA binding protein. Hence *the synthesis rate of the input protein A is the input signal to be inverted.* Let ρA denote the strength of repression, defined as the concentration of operator that is bound by repressor. To reduce the noise and increase the reliability of the system, we use cooperative binding of input proteins, which then bind to the operator and repress the gene. Also, continuous synthesis of input protein is required to replenish the used proteins in order to maintain repression.

For low values of ϕA, the amount of repression increases only slightly as more input protein is synthesized because concentrations are too low for significant dimerisation. At higher levels of ϕA (when input proteins dimerize easily), cooperative binding results in non-linear increase in the repression activity. Finally, at saturation levels of input protein when the operator is mostly bounded, the curve reaches an asymptotic boundary. The purpose for this stage is to provide gain: as a result of this stage, the analog signal approximates its digital meaning. *Thus the relation between ϕA and ρA denoted by C is sigmoidal.* The next stage involves transcription of the structural gene by the RNA polymerase, and results in the inversion of the signal. Let ΨZ denote the rate of mRNA synthesis for the output protein Z. Then, in the steady state relation Υ(transcription stage) between ΨZ and ρA, ΨZ decreases monotonically as ρA increases. With no repression, the transcription process proceeds at maximum speed (**max ϕZ**). Any increase in the repression activity results in a decrease in transcription

5.2.2 Generation function

The bio-device should produce specific molecules in response to the electrical signal. We can realize such a system to regulate the generation of specific molecules by controlling the activity of the enzyme electrically.

Many research groups have reported the electrochemical behavior of enzyme molecules immobilized in the conductive polymer matrix. Based on this idea, we can incorporate the enzyme molecule in conductive polymer membranes[4].

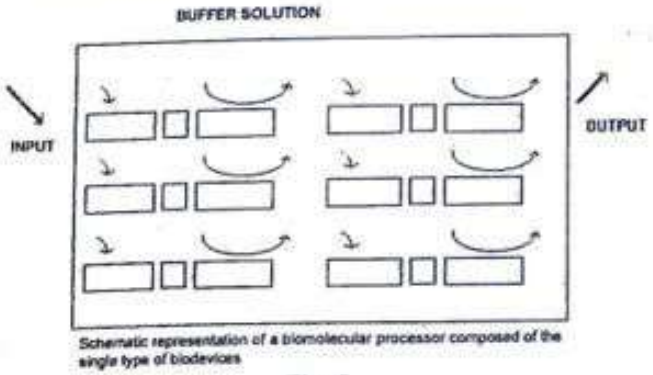


Biodevice based on the electrical control of enzyme activity. The Biodevice regulates the generation of specific molecules by controlling enzyme activity.

Fig. 4

For the generation function, the enzymes that catalyses the reaction to produce the set of molecules (S_o) are immobilized in the conductive polymer membrane. An electrical signal applied to the membrane causes the changes in electrochemical dynamics and corresponding changes in catalytic activities of the enzymes.

When the condition ($X \cap S_i = \phi$) is satisfied, an electrical signal from the multipleenzyme sensor activates the enzymes in the conductive polymer matrix to produce the set of molecules S_o. Otherwise, the signal inhibits the enzyme action, so the biodevice produces no information.



Schematic representation of a biomolecular processor composed of the single type of biodevices

Fig. 5

6. THE CHALLENGES FACED IN SUCCESSFUL IMPLEMENTATION

Stability of bio-molecules

Long-term stability of immobilized bio-molecules is a problem. For example, ®glucose-oxidase has a lifetime of only a few months. Self-assembly and self-repair are particularly important issues in bio-molecular computing.

Speed limit

Typical Bio-devices are slower by a magnitude of 10^4 than electronic devices. It is important to determine the degree of parallelism that makes up for this delay.

Iterative computations and "cleanup" problem

A bio-molecular computer performs data-driven computation, which means that molecules are passed directly and asynchronously as 'tokens'. If we can use reversible enzyme reactions, iterative computations can be made possible but the slow data rate will decrease the efficiency. If only irreversible reactions are available, continuous supply of buffer solution is required to clean up the solution.

In contrast, bio-molecular computers seem particularly well suited for massively parallel processing applications.

REFERENCES

[1] Lehninger D. Principles of Biochemistry, 3rd edition, pages 1072-1080 Macmillan Press/Worth Publishers 2000.

[2] Ron Weiss," Programming Colonies of Biological cells", Massachusetts Institute of Technology, Department of Electrical engineering and Computer science".

[3] Yuminaka Y., Aoki T. and Higuchi T., "Design of a Set Logic Network Based on Frequency Multiplexing and its Applications to Image Processing". Proc. IEEE 21st symposium 1991. Multiple Valued Logic, IEEE CS Press, California.

[4] Y.Yuminaka,T. Aoki, and T.Higuchi,Tohoku University: "Interconnection Free Biomolecular Computing", 1992s COMPUTER,IEEE Computer society.

[5] Lewin B., Genes VII, Oxford University Press, Oxford 1999.

[6] Lim D., WCB McGraw Hill, Second Edition, 1998.

❐

varieties of substrate molecules represent SVL logic values. Since, typical enzyme molecule might consist of 300 amino acids chosen from 20 types that commonly occur in living systems, the system can also use the enormous number of potential enzyme shapes for discriminating logic values.

SVL is a switching algebra for computing based on *mutiplexable* information carriers.

Let L be the set of logic values {0, 1...r-1}. An SVL function of n variables denoted by **F (X1, X2, ..., Xn)** is a mapping

$$F(2^L)^n \longrightarrow 2^L$$

Where, 2^L is the power set of L.

Example

Table 3

SVL Function $F(X_1, X_2) = X1 - X2 = X_1 \cap X_2^-$

X2 / X1	ϕ	{0}	{1}	{0, 1}
ϕ	ϕ	ϕ	ϕ	ϕ
{0}	{0}	ϕ	{0}	ϕ
{1}	{1}	{1}	ϕ	ϕ
{0, 1}	{0, 1}	{1}	{0}	ϕ

5.1.2 Model of bio-molecular switching device[4]

Let L be the set of all the substrates and that one kind of substrate represents one logic value. Therefore, L is considered to be the set of logic values.

The logic values thus represented by substrates can be transmitted simultaneously in solution. This simultaneous transmission is interpreted algebraically as "logic -value multiplexing. Furthermore, an enzyme-based biosensor can exactly discriminate the molecular information. Such a sensor consists both of detector enzymes that are immobilized on a membrane and of a transducer that produces an electric signal in response to enzyme action. Constructing logical systems first requires consideration of how to realize the primitive logical operations that define arbitrary logical functions.

$\cup$ "Union" can be obtained by simply mixing the corresponding substrates in the solution. To realize product type functions like "$\cap$" with sum type functions " $\cup$", we must design a basic building block as an inhibitor. This device generates molecules if it does not detect specific molecules.

Let X be the set of input substrates. A bio-molecular switching device is defined as:

$$\mathbf{BIO\ (X;\ S_i,\ S_o) = \{So\ if\ X \cap Si = \phi\ \}}$$

$$\{\ \phi\ \mathbf{if}\ X \cap Si\ != \phi\ \}$$

Where S_i and S_o are sets of substrates inherent in each device, and ϕ denotes empty set[3].

5.2 Implementation of bio-devices: Building a circuit

The function of the Bio-devices model can be realized in many ways, one of them being consideration of "selection" and "generation" switching functions.

- *Selection* is the function of detecting specific substrates in solution using enzyme –based biosensors.
- *Generation* is the function of producing specific substrates in response to the electric signal from the bio-device's *selection function.*

5.2.1 Selection function

Enzyme-based biosensors can be used for the realization of the Selection Function. The sensor consists of a detector enzyme immobilized on a membrane and transducer that produces a signal, usually electric, in response to effects of the enzyme action.

The Bio-device's BIO (X; S_i, S_o) should detect the *set –theoretic condition*

$$(X \cap S_i\ != \phi)$$

where X denotes the set of substrates in solution and S_i denotes the set of substrates inherent in each device. This implies that the multiple enzymes – based biosensor, which uses many kinds of detector enzymes to respond to any element of the substrate set S_i, plays an important role in the selection function.

For example, consider the three types of biosensors on which the sets of enzymes, {glucose oxidase}, {alcohol oxidase} and {glucose oxidase, alcohol oxidase} are immobilized (i.e., S_i={glucose}, S_i={alcohol} and S_i={glucose, alcohol}, respectively, for the three sensors). The responses of the two single-enzyme sensors are selective for their substrates, i.e., glucose and alcohol respectively. On the other hand the two enzyme sensors respond to both glucose and alcohol simultaneously.

On injection of four different substrate solutions i.e. (X=ϕ, {glucose}, {alcohol}, {alcohol, glucose}) the sensor system has digital responses, if and only if the condition

$X \cap S_i\ != \phi$ is satisfied. Obviously, this corresponds to the selective function required for the proposed Bio-devices.

activity, and hence inversion is attained.

In the last stage, rRNA translates mRNA product into output protein. The rate of this process is defined as the out put signal ϕZ of the inverter. Let P (translation stage) denoting the mapping between ϕZ and ΨZ in general, increases until an asymptotic boundary is reached. This boundary is defined by the protein synthesis capabilities of the cell, as well as other factors such as mRNA stability[2].

Inversion: $\phi Z = \Pi \circ \Upsilon \circ C (\phi A)$

4.1.1 Equilibrium behavior

Theoretical way to show the transfer curve mapping ϕA to ϕZ is to use a set of simulations.

Let each simulation include a different steady state rate of input protein synthesis, and record the level of the output protein if the system settles into a steady state. In this case, the definition of a steady state is some number of simulation time steps where all the state variables do not fluctuate by more than a small threshold.

4.2 Connections: Analysis of a ring oscillator

We can also theoretically analyze the connections of a microbial circuit by building a ring oscillator. A ring oscillator is a simple circuit that consists of three inverters connected in a series loop. The simulation results must show expected oscillations in protein concentrations, as well as a phase shift between the values.

4.3 Storage: Analysis of an RS - latch

The RS latch is a component for persistently maintaining a data bit. It consists of two cross-coupled NAND gates, with inputs $\overline{S}$ and $\overline{R}$ for setting and resetting the complementary output values A and B. The inverters with inputs $\overline{B}$ and $\overline{R}$ and common output A constitute one of the NAND gates, while the inverters with inputs A and S and the common output B constitute the other NAND gate[2].

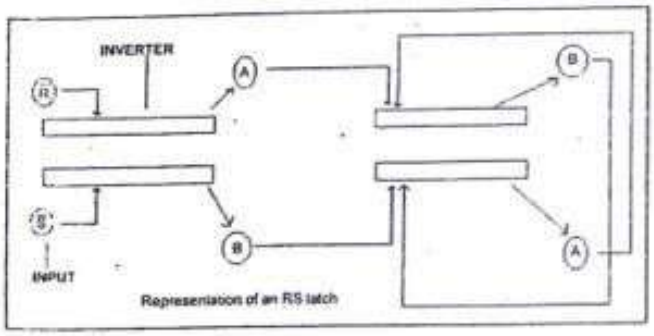


Fig. 3

5. INTERCONNECTION-FREE MICROBIAL CIRCUIT DESIGN

After we are able to successfully make logic NOT and NAND gates, the next obvious question to be answered is: How to combine these to get more complex circuits?

5.1 Combination of elements to get complex circuits

Any hypothetical microbial circuit has to be based on the following facts:

- The specificity of enzyme (produced using the output protein)* to their target gene.
- Different rate of synthesis of different input binding proteins.

To reduce losses and noise along connecting wires, it is proposed to make microbial circuits interconnection-free. Thus, whatever the complexity of our circuit, it has to be able to differentiate between different input and output signals and reliably harness the existing genetic regulatory mechanism of repression and transcription to give the desired output.

The disadvantages of a conventional system, based on traditional VLSI architecture are increased chip area, power consumption and noise. And a promising candidate for breaking through these difficulties is a **'Bio-molecular Computer'** – based on the dynamics of bio-molecular activities, rather than on electronic switching.

Such a model is based on the specificity of enzymes in their choice of reactants, called *substrates*. It represents the parallel distribution of logical information in the varieties of substrate molecules and uses enzyme specificity for parallel selection.

In a bio-molecular processor based on this model, the specified substrates are *broadcast* in solution/cell from sources. They diffuse with random molecular motion and carry information by their presence or absence in the solution*.

At the specified destination, enzyme –based biosensors selectively detect the released substrates (enzyme), which automatically triggers a specific bio-molecular element in the solution. Therefore the data transfer is in this sense, 'inter connection free' [4].

5.1.1 Brief overview of SVL3 and its applications

This system uses SVL (set-valued logic), a special class of multiple valued logic arithmetic system.

Let's assume that we can use a large number of enzymes and their substrates in our system and that the