
# UNIVERSITA' DELLA CALABRIA

## Dipartimento di ingegneria Informatica, Modellistica, Elettronica e Sistemistica

## Scuola di Dottorato di Ricerca

## "Archimede" in Scienze, Comunicazione e Tecnologie

## Indirizzo in: Scienze e tecnologie dei sistemi complessi

## CICLO

## XXVIII

# Gene Expression as a Digital Communication System

**Yesenia Elizabeth Cevallos Vilacrés**

**Novembre 2017**

UNIVERSITÀ DELLA CALABRIA

**UNIVERSITA' DELLA CALABRIA**

**Dipartimento di ingegneria Informatica, Modellistica, Elettronica e Sistemistica**

## Scuola di Dottorato di Ricerca

## "Archimede" in Scienze, Comunicazione e Tecnologie

**Indirizzo in**

**Scienze e tecnologie dei sistemi complessi**

**CICLO**

**XXVIII**

# Gene Expression as a Digital Communication System

**Settore Scientifico Disciplinare:** Nanocommunications at biological level. ING-INF/03, ING-INF/06

**Direttore:** Ch.mo Prof. Pietro Pantano

Firma _____

**Supervisore**: Ch.mo Prof. Salvatore Marano

Firma _____

**Dottorando:** Dottsa. Yesenia Elizabeth Cevallos Villacrés

Firma _____

To my wonderful family, especially my Mother Lolita:
Thanks for loving me; thanks for supporting me every day of my life.
All of you are my treasure.

# Acknowledgments

First, I want to thank God for permitting me to have life, health, and his lovely support as I conclude this stage of my life. Thanks, powerful God, for blessing my entire life, especially during the hard moments in which you made me understand that you are always with me in every second and in everything and then you gave me the strength to do what I have to while remembering that I must enjoy and live fully every day and gather only the best for the future. “Thanks, merciful and saint Father of my soul”.

I would like to thank my parents, brothers, and sister; my cute nephews, nieces and sister in law; and my uncles and aunts. Your love has been fundamental in providing support to my heart and my mind. Thanks for always repeating to me the words “you can..., you can...” Your presence makes my world a wonderful world.

I would also like to thank my friends, especially Ivonche, Lore and Deysi, for their sweet and enduring friendship that makes me happy. Thanks for showing me the optimistic way to see life. “I mean, a good friend really is a treasure….”

I would also like to thank Universidad Nacional del Chimborazo for providing a scholarship that gave me the chance to complete this PhD.

I would also like to thank Univesita’ della Calabria for giving me the chance to earn this PhD, which has been a discovery stage in my life. I realize that every human being is a harmonic and wonderful nanonetwork system.

To all of you, thank you from the bottom of my heart.

Yesenia

# Abstract

This PhD thesis extends upon the information theories of digital communication systems to analyse biological communications (nanocommunications) in order to accurately model biological communication as digital communication by providing an essential analysis of the analogies between both systems. As such, this work analysed gene expression from two perspectives: digital communication systems as a general perspective and internetwork systems as a specific perspective (keeping in mind that digital communication networks are a subarea of digital communication systems).

First, this work presents a novel layered network model that represents gene expression and the role of the Golgi apparatus as an internetwork router to transmit proteins to a target organ. Second, supported by the aforementioned layered network model, this work presents a digital communication system end-to-end model that represents gene expression with regard to the production of proteinaceous hormones in the endocrine system by using Shannon's theorem. In addition, each molecular process encoding biological information, from the transcription and translation of deoxyribonucleic acid (DNA) to hormone signalling, is represented by a layered network model.

These models apply the general advantages of digital internetworks and systems (i.e., performance and efficiency) to the transmission of biological information in gene expression systems. The proposed models and analysis define the duality between digital and biological communication systems, and the results herein can be used to overcome the disadvantages of both systems.

One of the most important applications of the current study is the potential use of the characteristics of both communication systems in the nano/bio-hybrid medical field (i.e., for the treatment of diseases such as cancer). Hence, the analysis presented in this study may prevent side effects by specifically enhancing the transmission of information to a suitable destination (i.e., to specific target organs), thereby facilitating the development of optimal and less expensive treatments.

# Sommario

Questa tesi di dottorato, partendo dalla teoria dell'informazione nei sistemi di comunicazione digitale, propone una sua estensione allo scopo di analizzare le comunicazioni che avvengono tra sistemi biologici (nano-comunicazioni) al fine di fornire un modello accurato della comunicazione biologica trattandola alla stregua della comunicazione digitale fornendo altresì un'analisi fondamentale delle analogie che vi sono tra i due sistemi. In particolare, questo lavoro ha analizzato l'espressione genica sotto due differenti aspetti: i) attraverso i sistemi di comunicazione digitale in una prospettiva più generale e ii) attraverso i sistemi di internetworking in una prospettiva più specifica (tenendo presente che le reti di comunicazione digitale sono un sottoinsieme dei sistemi di comunicazione digitale).

In primo luogo, questo lavoro propone un nuovo modello a rete stratificata (layered network) per descrivere l'espressione genica e, nello specifico, il ruolo dell'apparato del Golgi viene modellato come un "internetwork router" che trasmette le proteine ad un organo bersaglio. In secondo luogo, supportato dal modello a rete stratificata, questo lavoro presenta un modello di sistema di comunicazione digitale end-to-end che rappresenta l'espressione genica in relazione alla produzione di ormoni proteici nel sistema endocrino utilizzando il teorema di Shannon. Inoltre, ogni processo molecolare che codifica l'informazione biologica, dalla trascrizione e traduzione dell'acido deossiribonucleico (DNA) alla trasmissione del segnale ormonale, è rappresentato da un modello a rete stratificato.

Questi modelli offrono i vantaggi generali dei sistemi digitali e di internetworking (es. rendimento ed efficienza) alla trasmissione di informazioni biologiche nei sistemi di espressione genica. I modelli e l'analisi proposti in questo lavoro definiscono la dualità tra sistemi di comunicazione digitali e biologici, e i risultati presentati in questa tesi possono essere utili nel superare gli vantage di entrain i sistemi.

Una delle applicazioni più importanti di questo studio è il potenziale utilizzo delle caratteristiche di entrambi i sistemi di comunicazione in nanomedicina o in ambito medicale bio-ibrido (per esempio nel trattamento di malattie come il cancro). Infatti, le soluzioni presentate in questo studio potrebbero essere usate per prevenire gli effetti

collaterali migliorando in modo più specifico la trasmissione delle informazioni verso un target appropriato (es. verso specifici organi bersaglio), facilitando in questo modo lo sviluppo di trattamenti ottimizzati e meno costosi.

# Preface:

At present, nanocommunication systems or molecular communication systems are a promising part of science that can be seen as a solution for limitations in a wide array of applications, such as those in the biomedical field, industry, agriculture and military. The success of these systems is the result of the harmonic transmission of information through generations to survive, always supported by processes with energy efficiency and very high biocompatibility. Despite this successful behaviour, biological systems sometimes present serious limitations due to the incorrect processing/transmission of information, which can cause lethal diseases in species.

On the other hand, digital communication systems currently present optimal characteristics for transmitting and processing information and are a major technological development that have transformed the entire world, such as the Internet revolution. Among the better features of digital communication systems are the excellent bandwidth, very high transmission rate, compression, encryption, etc. Nevertheless, digital communication systems also present important limitations, such as their high demand for energy resources and compatibility with the environment.

Thus, science is currently focused on establishing an approach to use the best characteristics of molecular communication systems in digital communication systems and vice versa, thereby overcoming the limitations of each type of system and obtaining the benefits of each system to successfully process/transmit information. Research around the world is defining and emulating these bio-hybrid systems and establishing their applications. Hence, the literature on nanocommunication systems using the communication protocols and principles of digital transmission systems is very broad, but all of these studies follow the same paradigm i.e., establishing parallelism between both types of systems to obtain the desired communication behaviour. Particularly, modern science is focused on the digital analysis of DNA and in the way DNA is transmitted (i.e., gene expression) to biological destinations (organs, tissues or cells) because many genetic disorders could be countered from this perspective. Therefore, the main topic of many articles, journals and books is related to the study of genetic information, exploring

methods to encode and correct DNA information, schemes to modulate DNA information, mechanisms to synchronize the transmitter and receiver of DNA information, definitions of communication models to characterize noisy channels in the transmission of DNA information, definitions of propagation models to emulate the physical channels in which DNA information is sent, representation of digital sources of information (for instance, Markov sources) to emulate the transcription and translation of DNA, digital modelling of the ligand receptors in the receiver of proteins, definitions of layered networks models to represent the transmission of DNA and bio-hybrid molecular information.

This PhD thesis analyses gene expression systems from the general perspective of digital communication systems, particularly computer networks systems. These biological systems can thereby be used to the advantage of digital processing information and to address the physical limits of the transmission of information in the best way. In this manner, combining the principles of both types of systems facilitates the development of applications that can fundamentally enhance the quality of human life, specifically to search for better treatments for diseases with fewer side effects and social impacts.

This thesis is structured as follows:

Chapter one introduces basic concepts related to nanocommunications and reviews studies similar to the present work.

Chapter two defines a novel stacked-layer network model that represents gene expression and the role of the Golgi apparatus in transmitting proteins to a target organ.

Chapter three presents gene expression and protein delivery as an end-to-end digital communication system in which the entire biological process is represented by network and digital systems theories such as addressing, flow control, error control and traffic control and Shannon's theorem.

Chapter four discusses the conclusions, and a future work is mentioned.

**Rende (Italy), November 30/2017** **Yesenia Elizabeth Cevallos Villacrés**

# Contents:

# List of Figures:

# List of Tables:

# Chapter 1: Introduction

## 1.1. Definition of a Nanomachine

Nanotechnology enables the development of devices on a scale ranging from one to a few hundred nanometres. At the nanoscale, the most basic functional unit is a nanomachine, which can perform simple tasks such as computing, data storing, sensing or actuation [1-3]. A bio-nanomachine is made of biological materials (e.g., nucleic acid, protein, biological cells) or a hybrid of biological and non-biological materials [4].

## 1.2. Definition of a Nanonetwork

Communication between nanomachines will expand the capabilities and applications of individual nano-devices in terms of both the complexity and range of operation. Nanonetworks, a new network paradigm, are formed by the interconnection of nanomachines to cover larger areas, perform additional in-network processing, and fulfil complex tasks. Molecular communication is the most promising approach for the communication between nanomachines and their physio-chemical interactions composing nanonets [1]. Biological cells sense chemical concentrations with extraordinary precision. However, no cell performs this sensory task in isolation. Cells exist in communities, such as colonies, biofilms, and tissues. Experiments have shown that cells are more sensitive in groups than they are alone [5].

A molecular communication nanonetwork is a new type of distributed computer network that operates at the nanoscale [6].

## 1.3. Advantages and Applications of Nanocommunications

Currently, molecular communications or nanocommunications are one of the most important fields of research because of nature's ability to manage (through the interchange of information) resources and support the skills and evolution of living organisms over millions of years (which is also accomplished through the communication of information). Thus, nature can be considered a training centre with excellent opportunities to solve real world problems. Nature has developed numerous near-perfect

strategies, processes, and systems from the nano-to the macroscale through collaborating and evolving [7]. Therefore, the processes and systems used by living organisms could be analysed from the communications engineering paradigm to be applied principally as solutions in the medical field, where robust communication (required due to troubles in the physical layer, as in typical communication systems), fewer side effects and better quality of human life are needed for drug delivery and health monitoring [7-10]. Hence, nanotechnology has enabled the use of information theories and principles of digital systems in biological systems and vice versa; through this approach, the advantage of each type of system has been used in the other. To demonstrate such approaches, the following paragraphs cite some important engineered medical applications in nanocommunications, and chapters 2 and 3 analyse gene expression as a layered internetwork system and an end to end digital communications system, respectively.

The principal role of a communication system is to deliver information from a source to a sink. Since gene delivery systems transport genetic information encoded as DNA to living cells, such systems can be considered communication systems. Therefore, techniques developed for modelling conventional communication systems should be applicable to model gene delivery systems [11].

Farsad et al. [8], the functions of an immune system can be artificially emulated by injecting tiny artificial devices into the body, and each device is specialized for a specific task. For example, one device can be specialized to find pathogens, while another is tasked with destroying the pathogens. The functioning of these devices is very similar to that of the immune system, where each immune cell type carries a specific task. Just like the immune system, these devices must communicate and collaborate with each other to function collectively.

In Abbasi et al. [12], the authors discuss nanocommunication as a promising technology because of its ability to non-invasively access small and delicate body sites non-invasively, where conventional medical devices fall short. Hence, nano-enabled devices (like catheters and endoscopes) could reach delicate body sites, such as the spinal cord, gastrointestinal system, or the human eye. For instance, an application based on nanotechnology is “touch communication,” which uses a swarm of nanorobots as message carriers to exchange information. In “touch communication,” microbots are

applied to carry drug particles, which can be controlled and tracked (i.e., addressed) by an external macro unit (MAU) with a guiding force in real time. These microbots would survive for some time in the body, and their pathway would be the channel for exchanging information, while the loading and unloading processes correspond to the transmitting and receiving process, respectively (the receiver could be malicious cells, such as viruses or cancer). A specific application is shown in Figure 1.1, while the structure of the applied nanorobots is shown in Figure 1.2.

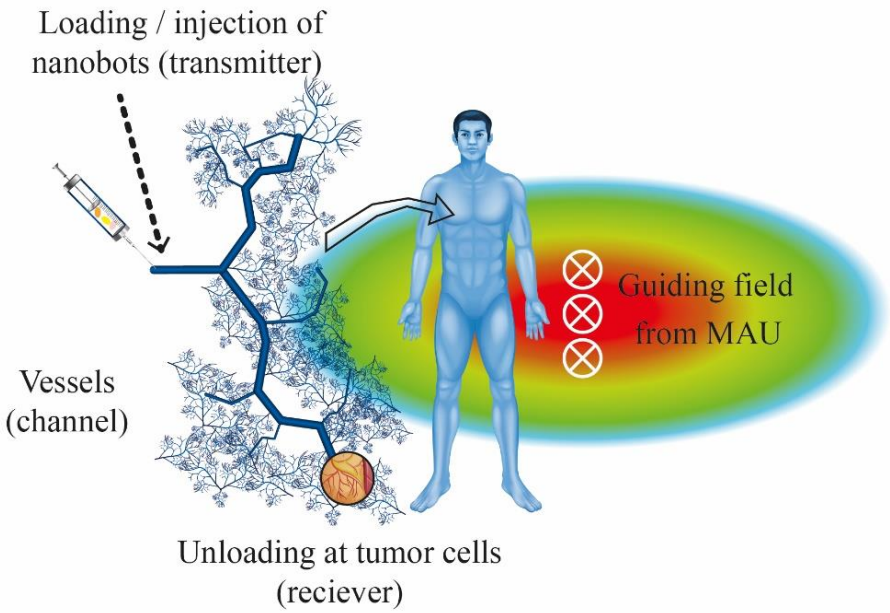


**Figure 1.1. Envisioned "touch communication" system**

Another important area of nanocommunication applications is targeted drug therapy to efficiently treat diseases. According to Nakano et al. [4], drug delivery to a target site (e.g., diseased cells or tumours) is achieved by encapsulating drug molecules in drug delivery carriers, delivering these carriers to the target sites, and releasing the drug molecules from the carriers at the target site in the body. Such drug delivery carriers are made from synthetic or natural particles (e.g., pathogens or blood cells, which are typically nano- to micrometres in size), so they can be injected into the circulatory system to propagate through the body, where they can exploit pathological conditions that appear

at a target site. Two of the most important aspects of targeted drug delivery are its ability to adjust the rate at which drugs are administered (in contrast with conventional drug delivery, wherein the rate cannot be adjusted once ingested) and addressing (i.e., simply the action of the drug in specific target sites).

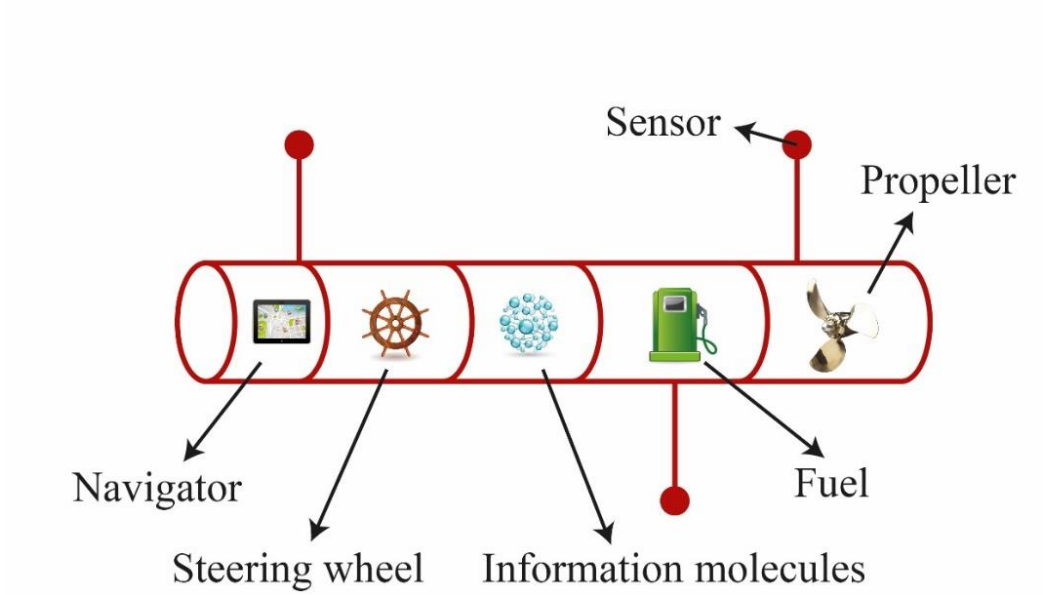


**Figure 1.2. Structure of the envisioned nanorobots**

The use of biocompatible materials may enhance targeted drug delivery and other applications, such as the use of nano-pressure sensors (due to their convenient size), which can be distributed in human eyes to detect the intraocular pressure for the early diagnosis and treatment of glaucoma to prevent vision loss. Osteoporosis in young diabetes patients could possibly be prevented through nano-devices deployed in the bones to monitor bone growth [4, 12]. Furthermore, nano-devices inside biological tissues can detect and eliminate malicious agents or cells, such as viruses or cancer cells, hence making the treatment less invasive and real time. The nanonetwork structure for accomplishing these tasks is shown in Figure 1.3 [13-15].

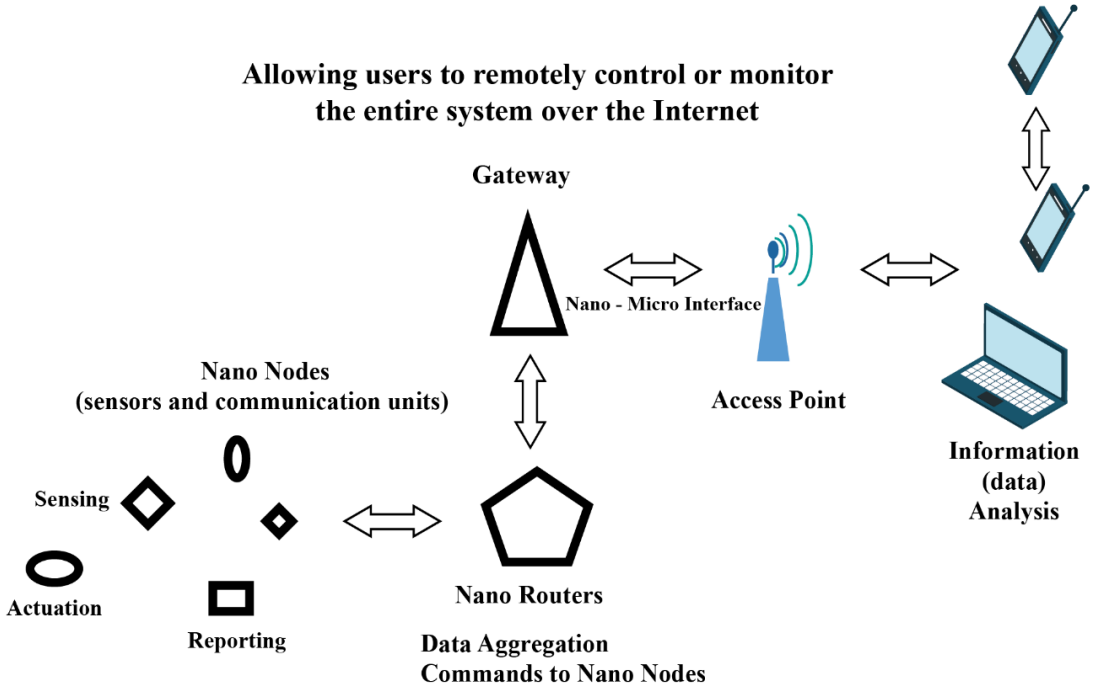


**Figure 1.3. Envisioned architecture for nano-healthcare**

The Internet of Bio-Nano-Things (IoBNT) is the paradigm through which medical devices will be connected to the Internet, thereby enabling near-real-time health services and transforming a patient's physical space into a smart space. IoBNT permits monitoring and diagnosis services with sophisticated healthcare applications that can be run inside the human body. Bolstered by breakthroughs that are being made in nanotechnology, nanonetworks are a novel networking paradigm that promises to extend medical applications beyond basic monitoring. Example applications that can be supported by nanonetworks include smart drug administration, nanoscale surgeries, and epidemic spread detection and management. The main goals of nanoscale healthcare applications are to diagnose, treat, monitor, and prevent health conditions and diseases at the molecular, cellular, or DNA (deoxyribonucleic acid) level. Current healthcare applications are implemented in two contextual domains: the patient's surrounding environment and the patient's body area (Figure 1.4) [16].

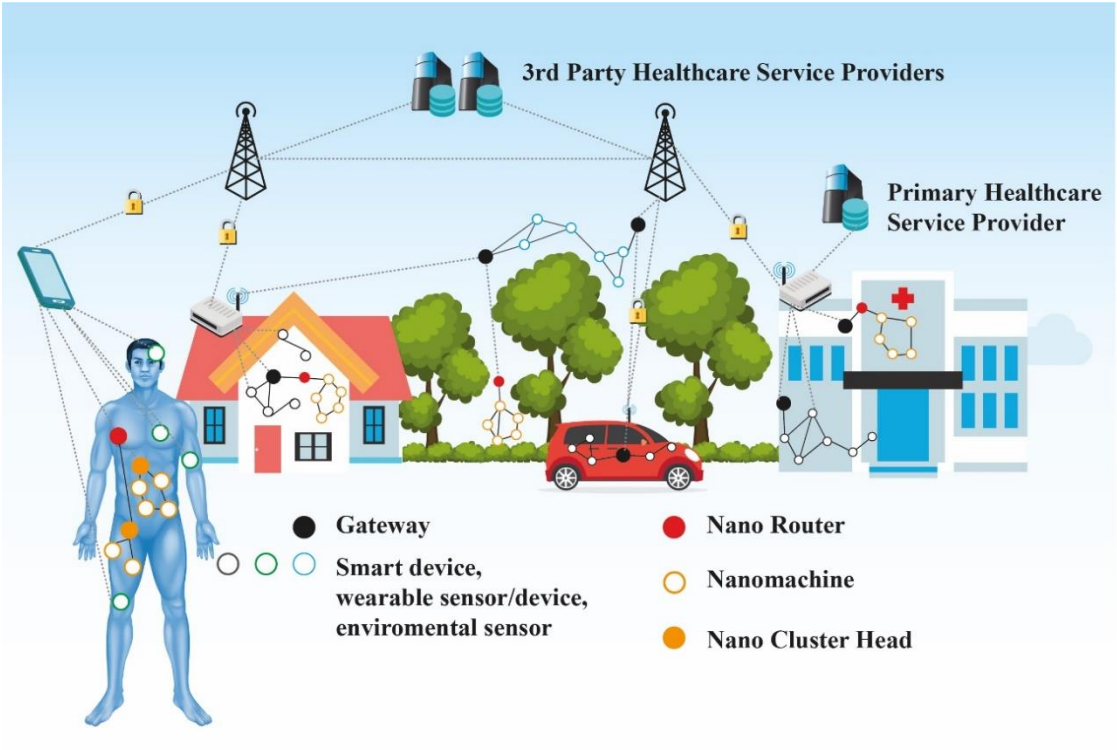


**Figure 1.4. Generic IoBNT architecture for ubiquitous healthcare**

Nanonetworks can periodically sense biological properties of the tissues or organs they monitor and send readings to gateways. Nanonetworks can also detect certain molecules, chemicals, or viruses and send alerts. In addition, detection can be more sophisticated and focus on abnormalities in the tissues or organs being monitored, such as indications of the onset of a heart attack. Nanonetworks can be used to actively perform therapeutic actions inside the body, such as administering medications and regenerative tissue engineering at the nanoscale or intracellularly. The IoBNT has a layered architecture in which the main functions in each layer are as follows [16]:

1. Application Layer. The application layer runs the healthcare applications properly for monitoring and controlling medical events in real time or nearly real time.

2. Transport Layer. The transport layer supports the reliability (related to packet loss because the body is considered a noisy environment for the transmission of information) through the dense deployment of nano-devices with the same functionality that can report the same data.

3. Network Layer. The network layer must provide routing and multi hops in the network, which is difficult due to the drift of the communication medium that establishes delay and a random position of the nano-devices. Thus, device mobility inside the body and proximity-based opportunistic routing can be suitable in these networks.
4. Medium Access Control / Physical Layer. This layer requires three main aspects to health applications: a channel capacity that guarantees reliable data delivery, an accurate channel model that accounts for the unique biological transmission medium and its associated noise, and efficient coding schemes that are resilient to errors.

Preventing cancer is one on the most important objectives in science. According to Ghavami and Lahouti [17], the overall behaviour of cancer is determined by gene expression and/or proteins. The authors note that proteomic data and collective functions of proteins are known to directly set cell function, and they propose a mechanism to prevent the growth of cancer cells with a method that distinguishes between two types of molecules. One type exists in the body in the healthy setting and the other type appears only in the presence of a certain disease. They propose an architecture for detecting abnormalities at a molecular level using sensor nanomachines over a nanocommunication channel. Once the abnormality is detected, this information is then sent through a micro communication channel to a gathering node, which makes a decision and may alarm the presence of an abnormality as necessary to the outside world.

The in vivo detection, localization, and treatment of disease at a biological scale (i.e., at the nano level) represents a major scientific advancement in science. This goal requires an in vivo communication network to link the nanomachines and enables the required collaboration. One proposal for establishing such a network is the use of chemical-exchange-based communications or molecular communication. Here, nanomachines release one or more molecules, sometimes called information or messenger molecules, by type (in the case of protein gene expression), quantity, timing, or other methods to convey a message to a receiver. Within an aqueous environment (such as a blood vessel or Brownian motion with drift) the information molecules travel towards a receiver. Note that the release of in vivo messenger molecules raises concerns about toxicity, which has

driven some molecular communications researchers to focus on transmitting a single molecule per transmission period using various types instead of quantity for signalling. Figure 1.5 schematically presents an in vivo mechanism to detect biological agents (for instance cancer cells) capable of generating molecular markers or biomarkers (i.e., the sign of the presence of a biological agent) in a medium with drift. In this mechanism, a collection of pseudo-randomly distributed nanomachines ($S_1$….$S_N$) can sense the environment and define the presence or absence of these biological agents and then release a certain molecule type to represent this decision. As Figure 1.5 indicates, the fusion centre is a collection of received molecules from the nanomachines, and it determines whether the biological agent is present [18].

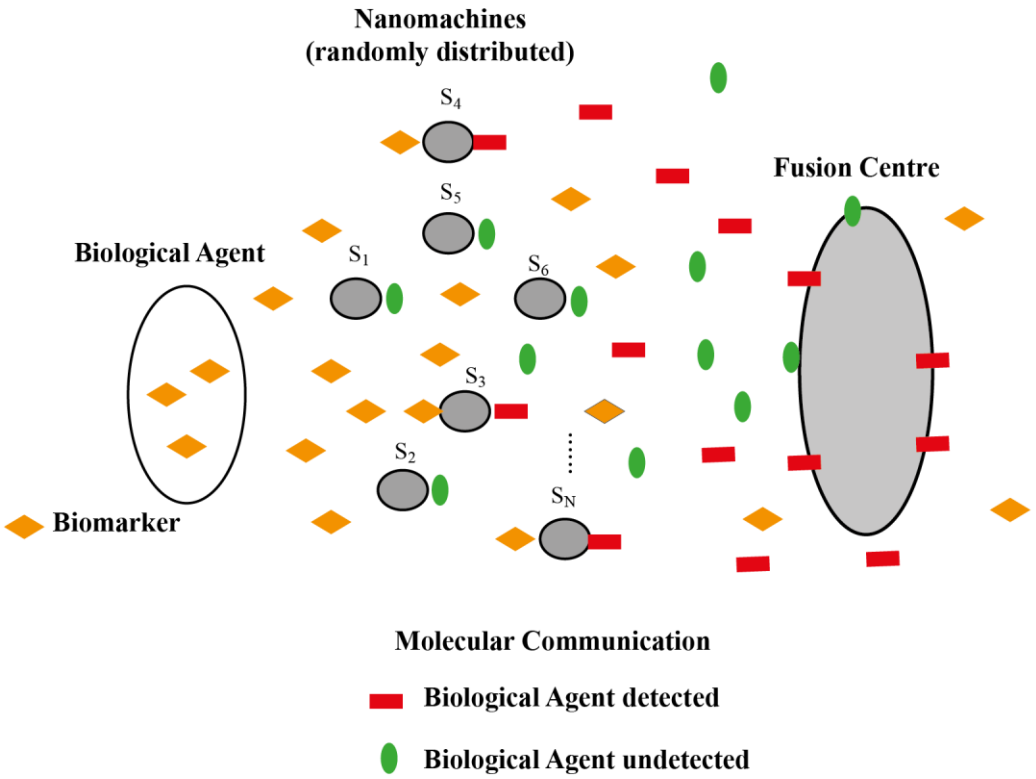


**Figure 1.5. An in vivo method to detect biological agents**

The biomolecules work collectively (i.e., as a network) on complex tasks that define the behaviour of living organisms. Examples of this cooperative work are gene regulatory networks, signal transduction networks and metabolic networks. However, the

interactions among biomolecules may affect some other biomolecules in the network, which may cause the state transition of the whole network and finally change the cellular behaviour. Therefore, controlling becomes a necessary mechanism in biomolecular networks. In Wu et al. [19], this control considers the existence of minimum steering sets that should have the most opportunities to chemically bind to existing drugs. To address this issue, this study considers drug-binding information. Specifically, the authors propose a method to identify the minimum steering sets of biomolecular networks with binding preferences such that biomolecules in the identified minimum steering sets are approved drug targets or have the strong ability to bind to existing drugs.

Biologically, the results of gene expression in eukaryotic cells are the proteins that play a fundamental role in a target organ in the body. When this target organ is in the long range of nanocommunication, the proteins are called hormones, and the endocrine system uses them to control the physiological functions in the human body. The transmission media in this long-range nanocommunication system are the blood vessels (which are considered random with a drifting medium), which include arteries, veins and capillaries. Continuously monitoring concentrations of certain parameters in blood capillaries could greatly improve the detection of a potentially critical situation in a patient and could be a fundamental way to administer drugs quickly. Therefore, in Sun et al. [20], a communication system in the blood vessel environments is proposed in which the transmitters are platelets, information is propagated through the flow of blood vessels, and the receivers are the endocrine cells (specifically soluble CD40 ligand sCD40L). At the transmitter end, the platelets secrete and release cytokines (i.e., the information molecules), which are small cell-signalling protein molecules, while at the receiver end, the cytokines are decoded by the sCD40 ligand receptors.

In Lin et al. [21], the authors defined the importance of synchronization as a fundamental parameter in nanocommunication systems (although most articles in the nanocommunications area implicitly assume synchronization between transmitter and receiver), which includes two issues. The first issue is the generation of the clock inside each nanomachine, and the second is time synchronization between the nanomachines.

The authors in [21] analysed a case with a unidirectional communications system in which the transmitter releases one molecule at a time, which is used to carry

information. The propagation occurs in an environment with drift, and the receiver absorbs and decodes the arriving molecule. The movement of molecules in the fluid environment cause a propagation delay. Two conditions in the transmission of information are assumed. The first condition establishes that the transmitter can perfectly control the molecule release time, while the receiver can perfectly measure the arrival time; the second condition establishes that the transmitter and the receiver have their own clocks. Both clocks in transmission synchronize after N rounds, and the round propagation delay is represented as a Gaussian distributed random variable, which enables calculating the offset between both clocks. Finally, the clock offset is obtained through the Newton-Raphson method. According to Lin et al. [21], an application of this type of communication could be essential, such as in cases in which engineered nanomachines require a common clock (time synchronization) to simultaneously release antibody molecules and destroy tumours in the human body.

## 1.4. References:


1. Hong L, Chen W, Liu F. Cooperative molecular communication for nanonetwork: 2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN); Jul 8; IEEE; pp. 369-70.
2. Bush SF. Interoperable nanoscale communication [future directions]. IEEE Consumer Electron Mag 2017;6: 39-47.
3. Bush SF, Paluh JL, Piro G, Rao V, Prasad RV, Eckford A. Defining communication at the bottom. IEEE Trans Mol Biol Multi-Scale Commun 2015; 1: 90-6.
4. Nakano T, Eckford AW, Haraguchi T. Molecular communication. UK: Cambridge University Press 2013. pp. 16- 20, 31, 34, 36-39, 52-67, 74.
5. Fancher S, Mugler A. Fundamental limits to collective concentration sensing in cell populations. Phys Rev Lett 2017; 118: 078101.
6. Li ZP, Zhang J, Zhang TC. Concentration aware routing protocol in molecular communication nanonetworks. Appl Mech Mater 2014; 556–562: 5024-7.
7. Raz NR, Akbarzadeh-T MR, Tafaghodi M. Bioinspired nanonetworks for targeted cancer drug delivery. IEEE Trans Nanobioscience 2015; 14: 894-906.
8. Farsad N, Yilmaz HB, Eckford A, Chae CB, Guo W. A comprehensive survey of recent advancements in molecular communication. IEEE Communications Surveys \& Tutorials 2016. 18 (3): 1887-1919.
9. Lu Y, Higgins MD, Leeson MS. Comparison of channel coding schemes for molecular communications systems. IEEE Trans Commun 2015; 63: 3991-4001.
10. Ko PY, Lee YC, Yeh PC, Lee CH, Chen KC. A new paradigm for channel coding in diffusion-based molecular communications: molecular coding distance function: 2012 IEEE Global Communications Conference (GLOBECOM); Dec 2012; IEEE; pp. 3748-53.
11. Wysocki BJ, Martin TM, Wysocki TA, Pannier AK. Modeling nonviral gene delivery as a macro-to-nano communication system. Nano Commun Netw 2013; 4: 14-22.
12. Abbasi QH, Yang K, Chopra N, *et al.* Nano-communication for biomedical applications: a review on the state-of-the-art from physical layers to novel networking concepts. IEEE Access 2016; 4: 3920-35.

**13.** Dabhi K, Maheta A. Internet of nano things-the next big thing. Int J Eng Sci 2017; 7: 10602-4.

**14.** Agarwal K, Agarwal K, Agarwal S. Evolution of internet of nano things (IoNT). Evolution 2017; 14: 2394-3386.

**15.** Nayyar A, Puri V, Le DN. Internet of nano things (IoNT): next evolutionary step in nanotechnology. Nanosci Nanotechnol 2017; 7: 4-8.

**16.** Ali NA, Abu-Elkheir M. Internet of nano-things healthcare applications: requirements, opportunities, and challenges: 2015 IEEE 11th international conference on Wireless and mobile computing, networking and communications (WiMob); Oct 2015; IEEE; pp. 9-14.

**17.** Ghavami S, Lahouti F. Abnormality detection in correlated Gaussian molecular nano-networks: design and analysis. IEEE Trans Nanobioscience 2017; 16: 189-202.

**18.** Rogers U, Koh MS. Parallel molecular distributed detection with Brownian motion. IEEE Trans Nanobioscience 2016;15: 871-80.

**19.** Wu L, Tang L, Li M, Wang J, Wu FX. Biomolecular network controllability with drug binding information. IEEE Trans Nanobioscience 2017; 16: 326-32.

**20.** Sun Y, Yang K, Liu Q. Channel capacity modelling of blood capillary-based molecular communication with blood flow drift: Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication; Sept 2017; ACM; article 19.

**21.** Lin L, Zhang J, Ma M, Yan H. Time synchronization for molecular communication with drift. IEEE Commun Lett 2017; 21: 476-9.

# Chapter 2: A Digital Communication Analysis of Gene Expression of Proteins in Biological Systems: A Layered Network Model View

**DOI 10.1007/s12559-016-9434-4**

**Cognitive Computation**

**Received: 8 January 2016 / Accepted: 11 October 2016. Published on line 21 October 2016**



**Yesenia Cevallos[1,2], Lorena Molina[1], Alex Santillán[3], Floriano De Rango[2], Ahmad Rushdi[4], Jesús B. Alonso[5]**

✉Yesenia Cevallos

ycevallos@unach.edu.ec

Lorena Molina

lmolina@unach.edu.ec

Alex Santillán

alex.santillansarmiento@plymouth.ac.uk

Floriano De Rango

derango@dimes.unical.it

Ahmad Rushdi

aarushdi@ieee.org

Jesús B. Alonso

jalonso@dsc.ulpgc.es

1. Faculty of Engineering, National University of Chimborazo, Av. Antonio José de Sucre, 060150 Riobamba, Ecuador

2. Department of Computer Engineering, Modeling, Electronics and Systems, University of Calabria, Via P. Bucci, 87036 Rende, Italy
3. Faculty of Science and Environment, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK
4. Institute for Computational Engineering and Sciences, University of Texas, Austin, TX 78712, USA
5. Department of Signaling and Communications, University of Las Palmas de Gran Canaria, 35017 Las Palmas, Spain

**Funding** This study has not explicit funding; it was made through a scholarship given to Ph.D. Student Yesenia Cevallos by National University of Chimborazo of Ecuador.

**Compliance with Ethical Standards**

**Conflict of Interest** Yesenia Cevallos, Lorena Molina, Alex Santillán, Floriano De Rango, Ahmad Rushdi and Jesús B. Alonso declare that they have no conflict of interest.

**Informed Consen**t Informed consent was not required as no human or animals were involved.

**Human and Animal Rights** This article does not contain any studies with human participants performed by any of the authors.

## 2.1. Abstract

*Background/Introduction* Biological communication is a core component of biological systems, mainly presented in the form of evolution, transmitting information from a generation to the next. Unfortunately, biological systems also include other components and functionalities that would cause unwanted information processing and/or communication problems that manifest as diseases.

*Methods* On the other hand, general communication systems, e.g. digital communications, have been well developed and analysed to yield accuracy, high performance, and efficiency. Therefore, we extend the theories of digital communication systems to analyse biological communications. However, in order to accurately model biological communication as digital ones, an analysis of the analogies between both systems is essential. In this work, we propose a novel stacked-layer network model that presents gene expression (i.e. the process by which the information carried by deoxyribonucleic acid or DNA is transformed into the appropriate proteins) and the role of the Golgi apparatus in transmitting these proteins to a target organ. This is analogous to the transmit process in digital communications where a transmitting device in some network would send digital information to a destination/receiver device in another network through a router.

*Results* The proposed stacked-layer network model exploits key networks' theories and applies them into the broad field genomic analysis, which in turn can impact our understanding and use of medical methods. For example, it would be useful in detecting a target site (e.g. tumour cells) for drug therapy, improving the targeting accuracy (ad dressing), and reducing side effects in patients from health and socio-economic perspectives.

*Conclusions* Besides improving our understanding of biological communication systems, the proposed model unleashes the true duality between digital and biological communication systems. Therefore, it could be deployed into leveraging the advantages and efficiencies of biological systems into digital communication systems as well and to further develop efficient models that would overcome the disadvantages of either system.

**Keywords**: Digital communication. Gene expression. Protein. Biological communication. Layered network model. Medical applications

## 2.2. Introduction

Biological communication has evolved over many years [1, 2] and represents a promising and ongoing area of study [1, 3–11].

Efficient biological communication is a component of biological systems (as result of their evolution transmitted to next generation), and the main characteristics of these systems include the efficient use of energy and maintenance of compatibility with the environment [12]. Unfortunately, biological systems also include certain unwanted functionalities caused by information processing and/or communication problems that manifest as diseases, some of which can be fatal.

Shannon's theorem [13] indicates that communication systems with physical limits are divided into three components: transmitter, transmission channel and receiver. So, every component of the communication system is improved to yield systems with high performance and efficiency, especially digital communications systems. Thus, the theories applied to communication systems may also be valid for biological communication systems.

Models used to analyse behaviour from common transmission theories require the establishment of bidirectional equivalences between biological and digital communication systems [14–19].

For instance, certain transmission control protocol (TCP) features are used in cellular communication to administer drugs and thus reduce the side effects and improve the treatment efficacy by releasing only drug molecules close to the target site. A connection-oriented protocol is used to detect a receiver (for example, tumour cells) and then establish and abolish communication to start and stop the treatment. A suitable transmission rate (flow control) between the transmitter and receiver is also used to avoid receiver congestion. Furthermore, the concept of reliable data transfer ensures that the target number of molecules is reached. Once the delivery of molecules is completed, a stop signal is sent from the receiver to end the connection and stop the drug flow. Flow

control is also used at the receiver. When the receiver is congested, a strategy to avoid further transmission of drug molecules is adopted [20].

Another relevant case with regard to the application of current digital communication advances in biological systems is compression because deoxyribonucleic acid (DNA) is a storage medium with high density and small size (1 g of dry DNA occupies 1 cm and has a storage potential of 455 EB of information) [21]. This enormous storage capacity could be utilized as biological database, which has been proposed by advances in informatics research [22]. Moreover, medical advances use ‘‘big data’’ to avoid disease crises, preserve successful treatments, and develop advanced treatments based on stored information [23, 24]. DNA could also be used to manage big data in medicine.

Other important aspects of digital transmission, specifically in networks, include routing principles that can be used in biological systems. With regard to internets, routing involves a mechanism to recognize changes in the network topology [25]. Therefore, new routes must be created to establish connectivity, especially when there are failures in the network. This characteristic also occurs in biological communication as chemotaxis and multicellular embryogenesis, in which identical embryonic cells undergo differentiation. Thus, the organism repairs wounds, and cells are able to regenerate [26].

The previous paragraphs explicitly suggest the use of models between biological and digital communication systems, which requires an analysis of the analogies that occur between these systems [3, 27–30]. In this article, we propose a stacked-layer network model that explains gene expression (i.e. the process by which the information carried by DNA is transformed into the appropriate proteins) and the role of the Golgi apparatus in transmitting peptide hormones to a target organ, which is similar to the digital communication process wherein a device transmitter in a network sends information to a destination device in another network through a router. The objective of this paper is to exploit network theories (e.g. independent functions of a layer into a stack, addressing, flow control, error control, and traffic control) and apply them to the treatment of diseases to reduce the side effects of drugs and improve the quality of treatment for patients from health and socio-economic perspectives. Similar proposals for the design of a stacked layers in molecular communications are presented in [31–33].

Shirazi et al. [31] proposes a protocol stack for nanonetworks, and the design consists of three layers (application, network and physical) that represent communications within short-range nanonetworks through molecular motors and calcium signalling and within long range nanonetworks using pheromones. In the first case, the communication model uses molecular rails that consist of microtubules that have the same function as wires in traditional point-to-point communication networks. In the second case, the communication model transfers calcium ions to adjacent cells through gap junctions, such as a wire connection in traditional communication networks. In the third case, information is transmitted via pheromones that can be modelled as wireless communication in traditional networks. The process of transmitting information via pheromones is similar to the processes underlying point-to-point and universal broadcast networks.

Nakano et al. [32] reports an analysis of the importance of biomedical and other molecular communication applications, and a stacked-layer design that includes a physical, data link, and network layer is proposed, which includes detailed specifications that each layer of the model must accomplish. The authors also analyse the different types of biological communications, such as pure random walk, random walk with drift, and random walk with reaction by amplifiers, to evaluate the latency (i.e. movement delay), jitter (i.e. variation in latency), and loss rate (i.e. the probability that a molecule transmitted by a biological sender is not received by the intended biological receiver).

Nakano et al. [33] provides an in-depth architectural perspective of molecular communication using a layered architecture approach, which is traditionally used in communication networks. Similar to the Open Systems Interconnection reference model and TCP/IP Internet architecture, a layered architecture (with physical, data links, network, transport and applications layer) describes issues in molecular communication as layers that are relatively independent of each other and thus facilitates the design and development of drug delivery and other applications of molecular communications in active and passive transport.

This paper is divided as follows. In section 2.3, we describe the elements of a typical digital communication system, emphasizing the transmitter and receiver components. Section 2.4 presents an analysis of the analogies between transmission of

information in digital communication systems and the protein expression, with a focus on the biological communication process required to convert information encoded in DNA into functional proteins as a digital fact. The same section discusses the similarities of routing functions between a classical router in internets and the Golgi apparatus in the cell. Besides, the method by which functional proteins (peptide hormones) are transported to a biological receiver that uses the information to perform a biological function is described. In Section 2.5, applications of the proposed analogies are discussed, and in Section 2.6, our conclusions are discussed.

## 2.3. Digital Communication Systems

Information confers awareness of an event to the receiver; therefore, the transmission of information has an implicit random origin when the information is sent to the destination at a non-deterministic moment and a user at the communication end utilizes this information [34]. Thus, communication systems have been developed to accomplish information transfer.

Digital transmission is one of the most advanced technologies for information transfer, and the characteristics that determine the potential utility of digital communication systems include storage, compression, encryption, transmission efficiency, bandwidth, and speed. The above features can be used in almost all communications because digital systems are universal; therefore, regardless of the type of communication, information can adopt a universal form that facilitates processing [35]. For example, information sent through a computer, including audio, data, text, and video, is delivered by the network using this universal form.

To perform transmission, every communication system should have three basic components: a transmitter, a communication channel, and a receiver. The transmitter includes the data terminal equipment (DTE) and data communications equipment (DCE) blocks, and the receiver at the destination end must contain DCE and DTE blocks as well (Figure 2.1). At each end, the DTE and DCE are interconnected by a physical interface that defines the electrical, mechanical, functional, and procedural characteristics between them [36].

The DTE in the transmitter contains the information source, which can be stochastic in typical communication systems, and determines the beginning of the communication process. In certain cases, the information must be altered before it can be directly delivered; thus, a transducer is utilized. Once the information presents suitable characteristics, it is sent through the physical interface to the DCE, which contains the communication controllers and produces the appropriate signal (synchronism). One of the most important functions of the DCE is to convert the information into an appropriate format for the transmission channel. In a communication system, a DCE is a modem (modulator–demodulator) or codec (coder-decoder) [36].

The communication channel is composed of the transmission medium (wired or wireless type), noise, interference, and distortion. Another undesirable effect in transmissions is attenuation that arises from the physical and electrical conditions of the transmission medium, the frequency band, and the link distance. Thus, in almost all cases, errors in communications occur in the communication channel [36].

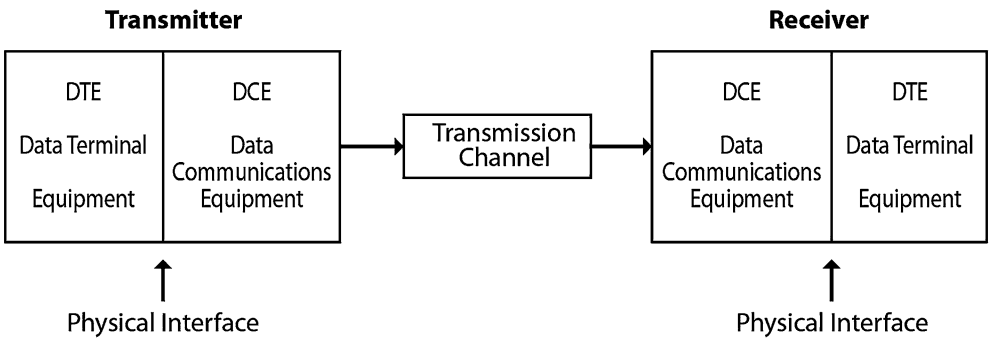


**Figure 2.1. Basic components of a communication system**

Because of the presence of the DCE in the transmitter, the information must be modified. The receiver must contain a DCE block to convert the information back to the original format. Moreover, if the information has been corrupted in the transmission medium, the DCE detects these errors and transmits the information in an appropriate way. Finally, the DTE takes this information from the physical interface and transports it to the destination [36]. This communication between the transmitter and the receiver is known as simplex (and connectionless), and the transmitter must identify the destination address (except in broadcast transmissions) before sending the information.

## 2.4. Digital Communications Analogies in Gene Expression of Protein

### 2.4.1. Comparison Between a DTE and a Nucleus

In the present article, we refer to biological communication as molecular communication, which is supported by Nakano et al. [33]. Accordingly, molecular communication presents stochastic fluctuations because of the small simple size and the limited half-life or deterioration, and it may not exhibit the same behaviour may not be observed under the same environmental conditions. Therefore, this analysis assumes that communication occurs when molecules are in a functional state.

The abstraction performed in this paper establishes that the biological DTE is represented by the nucleus (in eukaryotic organisms), which contains DNA molecules as an information source (i.e. the transmitter). The DNA contains the biological data, which are organized in nucleotide blocks called genes. These must be processed intra- or extracellularly. When we refer to extracellular communication, we assume that this type of transmission is performed over a long nanoscale distance, for example, in the endocrine system [1, 37]. We use the term ‘‘nano’’ for biological cells or molecules because they are considered to be in the ‘‘nanorange’’ [33].

Generally, a gene is defined as a set of nucleotides that stores the information required for a biological function to be performed at the destination, which can be accomplished by a protein or ribonucleic acid (RNA) [38]. In network terms, the contents of a gene may be associated with a network layer address [1, 33, 38, 39].

Therefore, an analogous process to the digital communication system described in Section 2.3 is the transfer of biological or molecular information, which starts with the transcription of DNA to RNA. One biological reason for this process is that DNA molecules cannot leave the nucleus of the cell [40], and their information must be transmitted using a different molecule. The RNA contains a copy of the information and carries it out of the nucleus [41]. This process is similar to what occurs in communication systems, in which a transducer modifies the input signal to allow for adequate processing in the other communication components.

The DNA molecule contains digital information, and it is formed of four discrete values, or four nucleotides. The nucleotides are monomers or subunits of nucleic acids (DNA and RNA), and they are formed by one nitrogenous base, a five-carbon sugar

(deoxyribose in DNA and ribose in RNA) and at least one phosphate group. The nitrogenous bases are adenine (A), thymine (T), cytosine (C), guanine (G), and uracil (U). The DNA double helix is composed of nucleotides containing the bases A, T, C, and G [42], and it maintains its structure because of the complementarity between the nitrogenous bases of each strand of the helix [43] or the affinity between adenine and thymine and between cytosine and guanine [44]. For example, if a hypothetical sequence is ATC in one strand, the complementary strand must have the TAG sequence.

The information in DNA is divided into blocks of nucleotides called genes [45, 46], which have start and termination sequences; thus, biological information is divided into data segments. In packet-switching networks, digital information is divided (to facilitate processing) into smaller units known as packets. Therefore, a packet in a digital network may be analogous to a gene in a biological communication network.

At the beginning of transcription, RNA polymerase II (RNAP II enzyme) recognizes a region on the DNA sequence called a promoter. The promoter hosts a start sequence, which is where RNAP II begins to add nucleotides to create a complementary messenger RNA (mRNA) sequence [47].

When transcription initiates, RNAP II produces a complementary single-stranded mRNA copy of one of the two DNA strands (Figure 2.2). The only difference between RNA and DNA is that RNAP II replaces thymine with uracil during this process [48, 49]. Both DNA strands do not have to be copied because they are exact complements. This concept may be comparable to the use of the digital gate NOT (or one's complement) to eliminate redundant information. Moreover, this biological process resembles digital data compression because the same amount of data is held in a smaller space.

Another element in the DNA strand is called the ''enhancer'' [50], which regulates the efficiency of transcription. This characteristic in the transcriptional process is comparable to variations of the symbol rate in multilevel digital transmissions as shown in Eq. 2.1 [51].

$$\text{symbol rate} = \frac{\text{number of symbols}}{\text{time for transmitting symbols}} \quad \text{(Bd)} \tag{2.1}$$

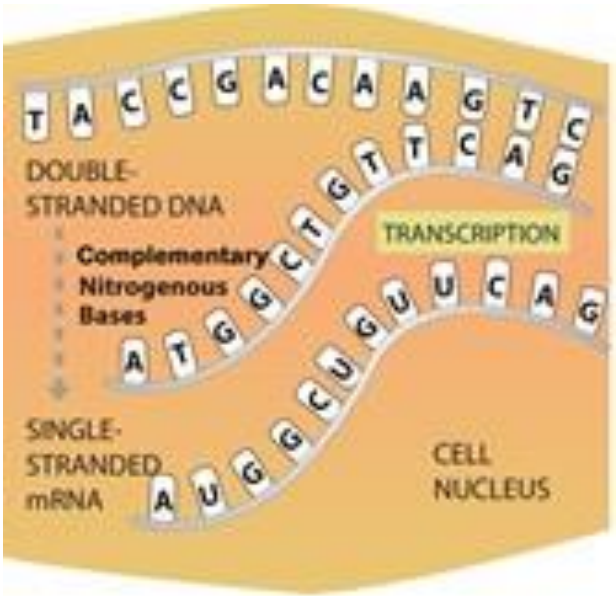


**Figure 2.2. Transcription of DNA into mRNA**

If only one symbol is sent during each transmission, then Eq. 2.1 becomes Eq. 2.2 [51]:

$$\text{symbol rate} = \frac{1}{m.\tau} \quad \text{(Bd)} \tag{2.2}$$

where $m = log_2 M$ represents the number of bits that constitute each symbol, $M$ is the number of digital levels, $\tau$ is the time required to send a bit, $m.\tau$ is the time required to send a symbol, 1 / $m.\tau$ is the bandwidth of the signal [52]. The number of symbols (as shown in Eq. 2.2) depends of the number of discrete levels, and the quantity of bits in every level may represent the amount of copied nucleotides represented by $m$ bits.

In biological or molecular communications and digital communications, the ‘‘enhancer’’ and the ‘‘variation of the symbol rate’’, respectively, are used to enhance efficiency. In biological communications, the objective may be to control the quantity of proteins through the amount of mRNA. In typical digital communications, the objective may be to expend less bandwidth because the bandwidth is less when m is greater as suggested by Eq. 2.2 [52]. Because the ‘‘enhancer’’ controls the amount of information

sent to the receiver, this objective may be understood as a flow control mechanism at the sender's end. Flow control is a fundamental task in a molecular datalink layer [33].

Transcription moves in one direction along one of the DNA strands from the 5'P to the 3'OH of the deoxyribose-phosphate backbone. This order is important because it guarantees that the gene information is copied appropriately [53]. In digital communications, the order of transmission is also fundamental. For example, in a serial communication, in which the less significant bit is generally transmitted first, 110 (7 in decimal value) and 011 (3 in decimal value) are not equivalent.

When the information is almost completely copied into single-stranded mRNA (the molecule that leaves the nucleus), an appropriate finalization sequence is also recognized by RNAP II to halt transcription [54]. The signalling to start and to stop transcription is used in the management of biological clocks that participate in the feedback mechanisms of cellular processes [33].

Three types of modifications (maturation) occur in the primary transcript molecule (called pre-mRNA) [55]:

1. Splicing. The segments in the pre-mRNA molecule that do not provide information (introns) are removed. This biological modification may be related to the digital communication concept of transmission efficiency (Eq. 2.3) [56]. Consequently, when additional bits (synchronism bits, control bits, etc.) are aggregated, transmission efficiency will decrease, whereas with fewer additional bits, such as through intron removal, the transmission efficiency will be higher.

$$n = \frac{\text{total number of information bits}}{\text{total bits in transmission}} \text{x } 100\ \% \qquad (2.3)$$

2. Capping. Post-transcriptional processing of the 5' end of the RNA occurs through the 5' cap process. At the end of transcription, the 5' end of the RNA transcript contains a free triphosphate group because this nucleotide was incorporated first in the chain. The capping process replaces the triphosphate group with another structure called the ''cap''. The cap is added by the enzyme

guanyl transferase. This enzyme catalyses the reaction between the 5' end of the RNA transcript and a guanine triphosphate molecule.

3. Polyadenylation. Post-transcriptional RNA processing at the opposite end (3' extreme) of the transcript occurs through a string of approximately 250 adenines attached to the end of the synthesized RNA chain. This string of adenine is called the ''poly-A tail''. The addition of the adenines is catalysed by the enzyme poly (A) polymerase.

The information added during capping and polyadenylation may be equivalent to delimitating data flags in digital communications systems, for example, headers and trailers that encapsulate the information in the datalink layer in protocol hierarchies in network software [25]; these flags are employed in processing and error control [33]. In eukaryotic cells, the cap and poly-A tail are added to provide stability (control and posterior processing) to mRNA molecules and avoid degradation by enzymes in the cytosol (or intracellular fluid), thus allowing the molecules to advance to the subsequent phases of biological processing. The mechanical transport of mRNA molecules through the cytosol may be analogous to the transmission of information in wired communications [1, 44] (physical layer task [33]).

Figure 2.3 and Table 2.1 summarize the analogies between biological and digital DTE (according to our communication proposal).

Table 2.2 establishes a layered network analysis of the layered tasks in typical networks and the similarities with transcription of DNA. Although biological behaviour does not map exactly to network concepts, certain functions in biological systems are similar. A layered model decomposes a large-scale system into a set of smaller units (i.e. layers) that are functionally independent and specifies the interactions among layers [33]. It is important to mention that the applications of a layered model, in biological communication (e.g. transport of drugs in disease treatment), may provide certain advantages with regard to controlling drug arrival and avoiding side effects compared with the typical transport of drugs, in which the dosage cannot be adjusted after ingestion.

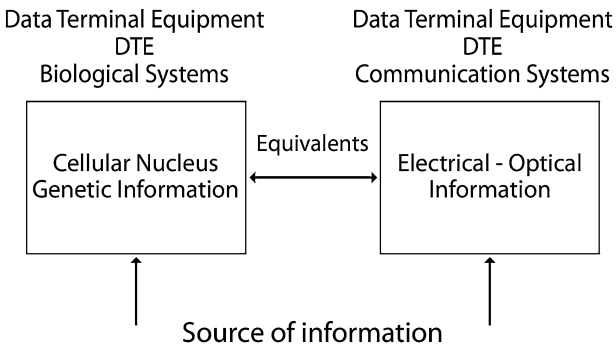


**Figure 2.3. Biological and digital DTE comparison at the transmitter end**

**Table 2.1. Summary of the biological characteristics and digital abstraction of DNA transcription**

| Biological characteristics | Digital equivalences |
|---|---|
| Presence of four nucleotides. | Digital values that represent the information. |
| DNA molecules require an RNA molecule to process biological information outside the nucleus. | Transducer. |
| DNA uses sequences of nucleotides divided into data segments. | Packets in switching networks. |
| Start and termination sequences for transcription of each gene. | Synchronism signalling. |
| Direction of DNA transcription must be 5'P–3'OH. | Transmission order of less significant bits in serial Communications. |
| Regulation in RNA production by enhancer. | Variation in the symbol rate. |
| Single DNA strand is used as a template. | Elimination of redundant information, digital data compression. |
| Removed introns. | Better transmission efficiency. |
| Cap and poly-A tail attached to the RNA molecule. | Header and trailer that encapsulate information. |

## Table 2.2. Layered model analysis of transcription of DNA

| Layered model network level | Layered model network functions in biological systems | Layered model network functions in transcription of DNA |
|---|---|---|
| Network layer | (a) *Molecular packet* It is a concept, information, communication or statement that is understood by the source and the destination (as well as the molecular router) at the molecular network layer [33].<br>(b) *Molecular network layer functionalities*<br>1. Provides functionality necessary for communication over longer distances, transmits packet without establishing a connection from a source to a destination, and provides decisions for each packet and its destination [1, 33, 39].<br>2. At the source, the network layer sends out a molecular packet [32, 33].<br>(c) *Molecular packet storage* It is a logical component at a molecular network node (e.g. sources, destinations, and molecular routers) to store molecular packets. Storage is supported by the link layer [33]. | (a), (b) Content of a specific gene may be associated with a network layer address because processing of this biological information is understood by a specific target organ (to accomplish a biological function) at the destination through intercellular communication (long distance).<br>(c) Storage of the molecular packets is necessary to process tasks into this layer. |
| Data link layer | (a) *Molecular frame* It is a concept, information, communication, or statement that is understood by the sender and the receiver at the molecular link layer [33].<br><br>(a.1) *DNA-based framing* DNA molecule that contains a message and an error correction code (i.e. molecular link layer information) may correspond to a molecular frame at the molecular link layer [33].<br>(b) *Molecular frame flow control* Allows the sender to adjust the rate of molecular frame transmission to avoid loss at the receiver [33]. | (a), (a.1) Cap and poly-A tail are added to provide stability (error control and posterior processing) to mRNA molecules, prevent degradation by enzymes and allow molecules to advance to the subsequent phases of biological processing.<br>Therefore, the information added during capping and polyadenylation may be equivalent to a header and trailer that encapsulate the information, which forms a molecular frame.<br>(b) Because the ‘‘enhancer’’ controls the quantity of proteins (i.e. the amount of information sent to the receiver), this structure may be used as a flow control mechanism at the sender end.<br><br>(c) mRNA that leaves the nucleus (with an implicit adjacent address comparable to the data link layer address for communication within a direct range of communication) will be bound by the ribosomes in the cytosol or associated with the rough endoplasmic reticulum. |

| | | |
|---|---|---|
| | (c) *Molecular link layer main task* Provides functionalities for communication within a direct range of communication [33].<br>(d) *Molecular frame storage* It is a logical component that allows the sender and receiver to store molecular frames in their respective signal molecules (i.e. information). Storage is supported by the physical layer [33]. | (d) Molecular frame storage is necessary until tasks are processed into this layer. |
| Physical layer | (a) *Molecular communication channel* It is the medium used to transmit molecules between a transmitter and receiver that communicate directly [33].<br>(b) *Molecular propagation* Allows molecules to travel over space in the environment (i.e. into the molecular communication channel) and molecules to move in a certain direction using protein motors.<br>This mode of molecule movement is referred to as active transport [1, 33, 44].<br>To use this type of communication medium, connectivity must be established in the environment a priori either artificially or through autonomous behaviour [33].<br>Compared to typical communications systems, in biological systems the information itself (signal molecules) must move through the environment at least during the molecule's half-life [1, 33].<br>(c) *Molecule storage* It is a physical component that a biological cell uses to store molecules. The storage of a molecule may simply be the molecular communication environment [33]. | (a) mRNAs that leave the nucleus are transported through the cytosol.<br>(b) Biological information must move into the cellular environment using a wired medium.<br>(c) Molecular storage is necessary until tasks are processed in this layer. |

### 2.4.2. Comparison Between DCE, Ribosomes, and Endoplasmic Reticulum

Once DNA is transcribed into a single RNA strand, the genetic information leaves the biological DTE, which requires communication with the biological DCE. In digital communication systems, connectivity from the DTE to the DCE is performed through a physical interface. In a biological system, the cytosol may represent the physical interface [57] (Figure 2.4).

The presence of DCE (codec or modem) in a conventional communication system is essential for translating information into the correct format and transmitting it through the channel. From our perspective, the ribosomes and endoplasmic reticulum (ER) represent the ‘‘biological DCE’’ because they process genetic information provide the functional structure (or format when referring to data) that is subsequently released in the biological transmission medium and ultimately arrives at the biological receiver. The genetic information is ‘‘formatted’’ during ‘‘translation’’, a process by which genetic information is converted to amino acid chains with biological functionality (inside and outside the cell). Therefore, the biological DCE codifies information via translation and provides a specific input sequence (data in mRNA) that is associated with a specific output (sequence of amino acids), and this codification process corresponds to conventional digital codification.

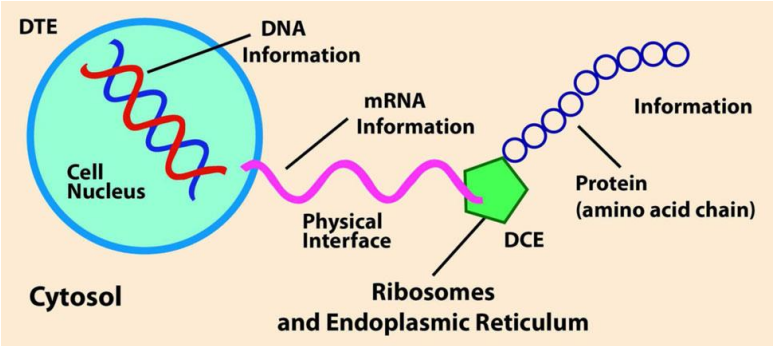


**Figure 2.4. Cytosol in a biological system is analogous to a physical interface**

During translation, the mRNA that leaves the nucleus (with an implicit adjacent address comparable to a data link layer address to facilitate communication within a direct range of communication [33]) is bound by the ribosomes in the cytosol or associated with the rough endoplasmic reticulum (RER) [58]. The movement of biological information from the nucleus to ribosomes or the ER through the cytosol, which is equivalent to the information arriving and moving into a biological communication channel, is considered a task at the physical layer [33].

Ribosomes read the information contained in the biological sequence using a codon system (a codon is a triplet of nucleotides). The codons in the mRNA can be

recognized in the ribosomes by transfer RNA (tRNA), which possesses an anticodon (the complementarity sequence of a particular codon) associated with a specific amino acid as part of its molecular structure [59]. Once again, the concept of complementarity may be equivalent to the use of the digital gate NOT (or one's complement) in each codon.

The genetic code uses three nucleotides (codon) to represent a specific amino acid [60]. The rationale behind this type of codification is that there are 20 amino acids; thus, it is necessary to sort the four nucleotides in groups of at least 3 to encode all 20 amino acids in 64 possible combinations (i.e. $4^3$= 64 because $4^2$= 16 is not sufficient to encode 20 amino acids) [61]. In digital terms, the data alphabet would consist of 64 characters. Thus, this biological communication corresponds to digital communication with multiple discrete levels using the symbol rate concept (instead of bit rate) because during the translation of each codon containing three nucleotides, a symbol (which consists of more than a bit) must be employed.

A protein is generated through the processing of specific amino acids, and the order of amino acids depends on the prior amino acid; therefore, this biological characteristic may correspond to an input in Markov's source. Because proteins consist of 20 different amino acids (or in our case, symbols), Eq. 2.4 represents the *kth*-order entropy, where $p_i$ is the probability of finding the ith amino acid and $p(i|s)$ is the conditional probability of the *ith* amino acid that occurs after the amino acid string *s* [1].

$$l_k(x) = -\sum_{i=1}^{20}\sum_{s}^{20^{k-1}} p_i p(i|s).log_2 p(i|s) \tag{2.4}$$

Briefly, protein synthesis starts in the ribosome when the first amino acid (*aa*) carried by an initiator tRNA (with the corresponding anticodon) is placed in the peptidyl site (P) of the ribosome (see Figure 2.5 [61]), which in turn is located on the position of the start codon (AUG). Then, the machinery of the ribosome allows more amino acids to be added to the first by repeating the complementarity mechanisms of the subsequent mRNA codons with specific tRNAs (Figure 2.6 [62]). The next tRNA carrying another *aa* enters the ribosome structure at the aminoacyl site (A). When two *aa* are in proximity at sites P and A, a covalent bond can form between the amino acids. After the formation

of a covalent bond, the tRNA at the P site will move to the exit site (E) and the tRNA at the A site will move to the P site, thus freeing the A site for the next tRNA (carrying a new *aa*) to repeat the process. The ribosome can advance three nucleotides along the mRNA (Figure 2.6); therefore, the addition of *aa* to the growing polypeptide can be completed [63, 64]. Once the ribosome reaches a stop codon, the ribosome releases the polypeptide chain (new protein) and the ribosome machinery disassembles [61, 62, 65].

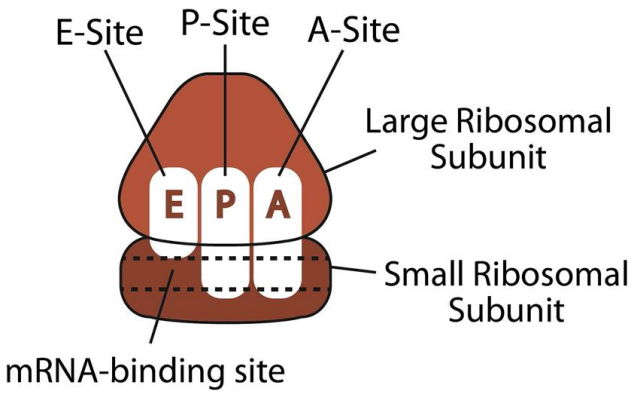


**Figure 2.5. Three sites in the ribosome structure: aminoacyl (A), peptidyl (B) and exit (E)**

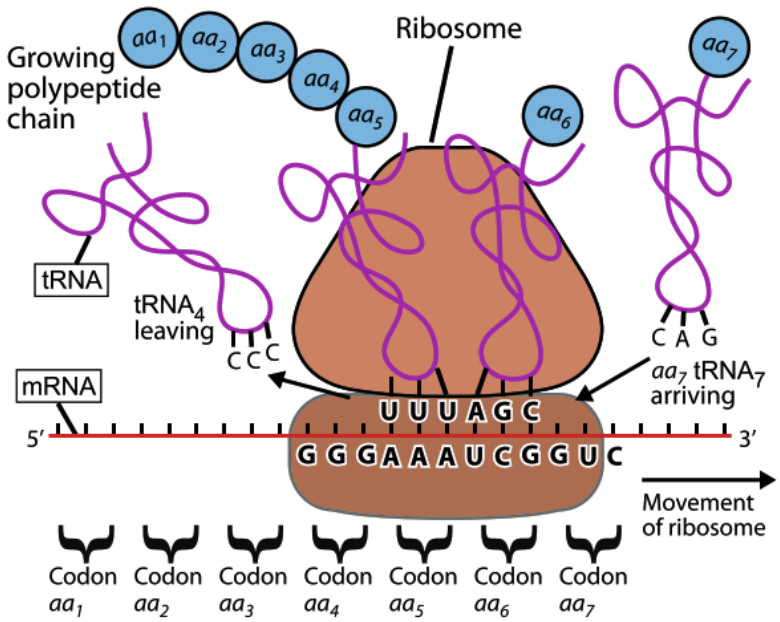


**Figure 2.6. Protein synthesis mechanisms**

In a digital communication paradigm, the ‘‘start’’ and ‘‘stop’’ codons may correspond to synchronism signals. Synchronism between the source and the destination is performed in asynchronous and synchronous transmissions through a ‘‘start bit’’ or ‘‘start flag’’, respectively. This type of information indicates that the transmitter will send the data and the receiver must take it and process it. The ‘‘stop’’ codon in biological signalling is equivalent to the ‘‘stop bit’’ (asynchronous communications) or ‘‘stop flag’’ (synchronous communications) used at the destination to indicate the end of communication. The signals to start and stop the translation of DNA introduces the management of biological clocks in cells to provide feedback during cellular processes [33].

Producing proteins through ribosomes may be equivalent to processing information in a buffer with two places to load the data (i.e. two amino acids every time). In the digital process, the first amino acid represents the first input signal into the buffer (ribosome), which is set in the upper part of the memory. When another amino acid arrives at the ribosome, it corresponds to the second input signal and is placed in site A, which can be compared to the lower part of the buffer memory. In digital communications, a new input signal must be combined (for example, using an adder) to become the first output signal. This situation may resemble the covalent bond between the ‘‘input’’ amino acid in site A and the previous amino acid in site P. The data in the lower part of the buffer are moved to the upper part, then the new input is placed in the lower part of the memory, and the process continues to obtain a sequence of outgoing signals that are represented in the biological system by the nascent amino acid sequence.

The abstraction presented in this manuscript is focused on proteins addressed to the ER because many of these proteins (e.g. peptide hormones, such as insulin) play a role outside the cell [66], where they reach the receiver and complete an end-to-end communication. When the aforementioned tagging occurs (i.e. tagging to play a role outside the cell), a signal recognition particle (SRP) is bound to the amino acid sequence (with an implicit adjacent address via molecular tagging, which is comparable to a data link layer address that facilitates communication within a direct range of communication [33]). The function of the SRP is to allow the budding protein to arrive at a channel protein in the ER that is in charge of the translocation of the protein inside the ER. Subsequently,

the SRP detaches from the protein to be recycled in the cytosol [67, 68]. Similarly, in digital systems, after the processing information and control information are used, they are discarded. Inside the ER, the proteins are folded and acquire a functional three-dimensional structure to accomplish a biological function [69, 70] (equivalent to digital information after handling by the DCE).

SRP addressing is similar to the intermediate addressing that is used (and subsequently discarded) by network protocols to reach the destination in another broadcast domain in the same internet. For example, in Figure 2.7, PC1 (in LAN1) sends information to PC5 (in LAN2), and this process requires PC1 to have four addresses (IP source or IP1, MAC source or MAC1, IP destination or IP5, MAC destination or MAC LI1). When the information arrives at the router's LAN interface, the router chooses a path and an output interface from a routing table. Then, the information is sent, and IP1 (source) and IP5 (destination) are retained. MAC1 becomes MAC LI2 and MAC LI1 becomes MAC5; therefore, the intermediate MAC addresses are discarded from one broadcast domain to another.

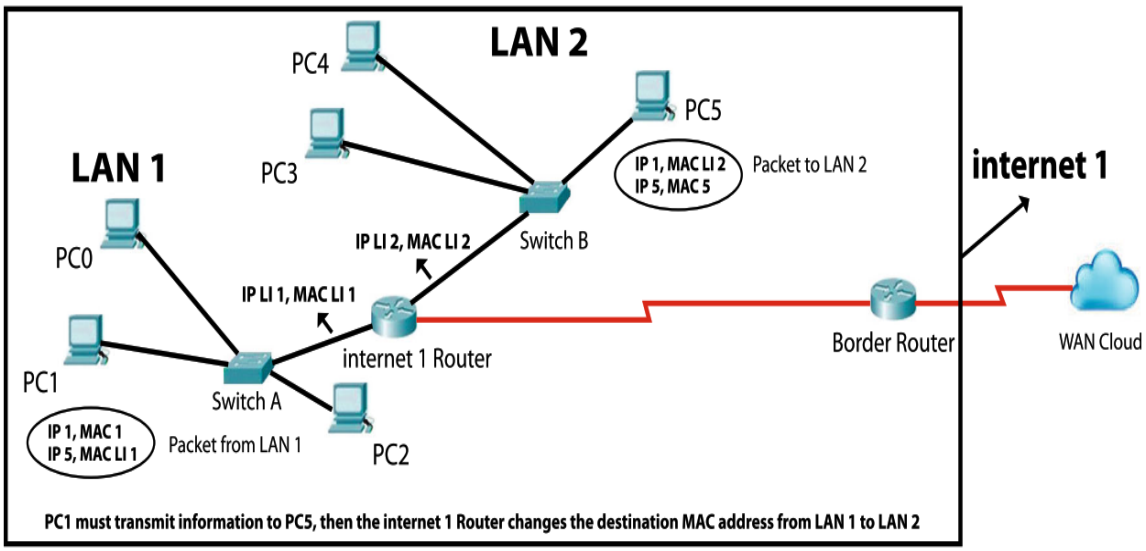


**Figure 2.7. Destination MAC addresses are changed when packets arrive at another broadcast domain**

In biological systems, information errors can arise during DNA transcription and translation, and as in conventional communication systems, errors can also arise in the

transmission medium. Errors in cellular processing and information communication are responsible for certain diseases, such as cancer, autoimmunity, and diabetes [1].

In Figure 2.8 and Table 2.3, a summary of the analogies between the biological and digital equivalents to a DCE is presented (according to our communication proposal).

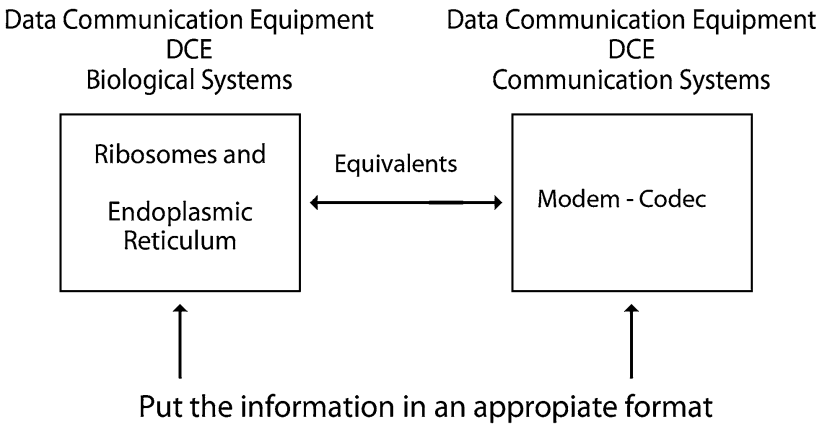


**Figure 2.8. Biological and digital DCE comparison at the transmitter end**

**Table 2.3. Summary of the biological characteristics and digital abstraction of the translation of mRNA**

| Biological characteristics | Digital equivalences |
|---|---|
| Initiation and termination codons during translation. | Synchronism signalling. |
| Specific sequence of nucleotides (input) in a codon is associated with a specific amino acid (output). | Use of a coder. |
| Each tRNA has a complementary anticodon for each codon. | Use of digital gate NOT. |
| Association of 20 amino acids with four nucleotides requires a combination of three nucleotides at one time ($4^3 = 64$). | Digital alphabet is composed of 64 characters; use of symbols (which consists of more than a bit). |
| Producing proteins through the ribosomes. | Processing information through a buffer with two locations for loading data. |
| Signal recognition particle (SRP) guides the protein to the ER. | Intermediate address used by network protocols to reach the destination. |
| SRP is recycled in the cytosol. | Intermediate addresses used by network protocols are used then discarded. |
| Protein folding. | Optimization of information after processing in a buffer. |

Table 2.4 establishes a layered network analysis of the layered tasks in typical networks and similarities in the translation of DNA. Although biological behaviour does not map exactly to network concepts, certain functions in biological systems are similar in both system types. The structure of a layered model decomposes a large-scale system

into a set of smaller units (i.e. layers) that are functionally independent of each other and specifies the interactions among layers [33].

**Table 2.4. Layered model analysis of mRNA translation of DNA**

| Layered model Network level | Layered model network functions in biological systems | Layered model network functions in translation of mRNA |
|---|---|---|
| Network layer | (a) *Molecular packet* It is a concept, information, communication or statement that is understood by the source and the destination (as well as the molecular router) at the molecular network layer [33]. | (a), (b) Protein folding continues the export process from the cell to a target organ (to accomplish a biological function) in a long distance communication, which could be understood as a network layer addressing because it may represent the information transmitted by genes at network layer. |
| | (b) *Molecular network layer functionalities* Provides functionality necessary for communication over longer distances, transmits packets without establishing a connection from a source to a destination, and provides decisions for each packet and its destination [1, 33, 39]. | The ER finishes the production of specific proteins intended for particular target organs, which may represent a network layer protocol, among the nucleus (biological DTE), and the ER (biological DCE) at the transmitter and the receiver (target organ i.e. biological DTEs/DCEs). |
| | (c) *Molecular packet storage* It is a logical component at a molecular network node (e.g. sources, destinations, and molecular routers) to store molecular packets. Storage is supported by the link layer [33]. | (c) Molecular packet storage is necessary until tasks are processed in this layer. |
| Data link layer | (a) *Molecular frame* It is a concept, information, communication or statement that is understood by the sender and the receiver at the molecular link layer. | (a) Amino acid chain may constitute a frame because when the mRNA molecules arrive at the ribosomes, these molecules must be encoded; therefore, additional biological information is provided to facilitate their processing (e.g. SRP). This possible framing at the data link layer level is performed when biological information enters the ribosomes. |
| | As in a typical network, a frame could be formed of the physical layer information and a link layer header (i.e. additional processing data containing information specific to the layer) [25, 33]. | (a.1) When the folded proteins travel to the Golgi apparatus, they are embedded in vesicles, and this format may represent a frame at the data link layer. |
| | (a.1) *Vesicle-based framing* A vesicle may naturally correspond to a molecular frame at the molecular link layer [33]. | (b) The signal recognition particle (SRP) guides the budding protein to a channel protein in the endoplasmic reticulum (ER). Subsequently, the protein travels to the Golgi apparatus. These types of addressing may be comparable to the addressing for molecular link layers (for |

| | | |
|---|---|---|
| | | communications within a direct range of communication). |
| | (b) *Molecular link layer main task* Provides functionalities for communication within a direct range of communication [33]. | (c) Molecular frame storage is necessary before processing the tasks into this layer. |
| | (c) *Molecular frame storage* It is a logical component that allows a sender and a receiver to store molecular frames in their respective signal molecules (i.e. information). Storage is supported by the physical layer [33.] | |
| Physical layer | (a) *Molecular communication channel* It is the medium used to transmit molecules between transmitters and receivers in direct communication [33]. | (a) Molecules of mRNA that leave the nucleus are transported through the cytosol to ribosomes. |
| | (b) *Signal modulation* It is a functionality at the transmitter to alter the properties of molecules to represent the information.<br>One mechanism for signal modulation is to choose one type of molecule from a set of molecule types, with each molecule type representing certain information [33]. | (b) Biological information received through the cytosol is processed in biological DCE via translation that allows for information codification by associating a specific input sequence (data in mRNA) with a specific output (sequence of amino acids); thus, codification corresponds to conventional digital codification. |
| | (c) *Molecule propagation* Functionality in which molecules travel over space in the environment (i.e. into the molecular communication channel) and move actively in a certain direction using protein motors. This mode of molecule movement is referred to as active transport [1, 33].<br>To use this type of communication medium, connectivity in the transmission must be established in the environment a priori either artificially or through the use of autonomous behaviours [33].<br>Compared to typical communications systems, in biological systems the information itself (signal molecules) must move through the environment at least during the molecule's half-life [1, 33]. | (c) Folded proteins must physically move until they reach the Golgi apparatus using a wired connection in the cytosol. |
| | (d) *Molecule storage* It is a physical component that a biological cell uses to store molecules. Molecule storage may simply be molecular communication environment at least during the molecule's useful half-life [33]. | (d) Molecular storage is necessary until tasks are processed in this layer. |

### 2.4.3. Comparison Between a Router and the Golgi Apparatus

When proteins have a functional (or quasi-functional) configuration, the ER transfers them to the Golgi apparatus (GA); therefore, the biological information must physically arrive at the GA, which is considered a physical layer task [33].

The GA consists of a series of membranous compartments (each with a different set of enzymes) that collect and dispatch proteins received from the ER [71, 72]. These tasks are analogous to the tasks performed by a router [39]. Proteins pass from the ER to the GA in vesicles that have an implicit adjacent addressing comparable to a data link layer addressing in a transmission within a direct range of communication, the vesicles may naturally correspond to a molecular frame at the molecular link layer [33].

The proteins enter the Golgi at the *cis* face. Membranous vesicles also mediate the movement of proteins through the GA. In each stage of this migration, proteins in vesicles undergo maturation to accomplish a routing function [73]. Certain sugars added to proteins could act as signals to determine the fate of the protein. Once proteins have reached the *trans* face of the Golgi, they can be addressed to other intracellular compartments or to the cell membrane to be excreted and move through the cell external environment according to Fick's first and second laws (Eqs. 2.5, 2.6) [1, 74–77]; these tasks may correspond to a physical layer [33].

$$J = -D.\nabla\emptyset + v_d.\emptyset \tag{2.5}$$

where $J$ is the flux, $D$ is the diffusion coefficient, $\phi$ is the concentration of a substance, $\nabla$ is the gradient, $v_d$ is the velocity of all particles in a distribution drift, and $v_d \,.\, \phi$ is the increase in the flux.

$$\frac{\delta\emptyset}{\delta t} = \mathrm{D}.\nabla^2\emptyset - \mathrm{v_d}\nabla\emptyset \tag{2.6}$$

The GA functions described above are similar to those performed by a router on a network [39]. As shown in the topology illustrated in Figure 2.9 (according to our communication proposal), the router determines whether the information remains inside or outside of the network. In the latter situation, routing and routed processes are required to send the information to the destination. The functions of the GA in moving proteins

forming vesicles and tagging proteins to target them to a destination may be analogous to those involved in processing the protocol data units (PDUs) in the layers of a router (i.e. encapsulating and unencapsulating information).

Table 2.5 presents a layered network analysis that compares the layered tasks in a typical network that are performed by a router and the analogous tasks in the GA. Although biological behaviour does not map exactly to network concepts, certain functions in biological systems are similar in both system types. In addition, the structure of a layered model decomposes a large-scale system into a set of smaller units (i.e. layers) that are functionally independent and specifies the interactions among layers [33].

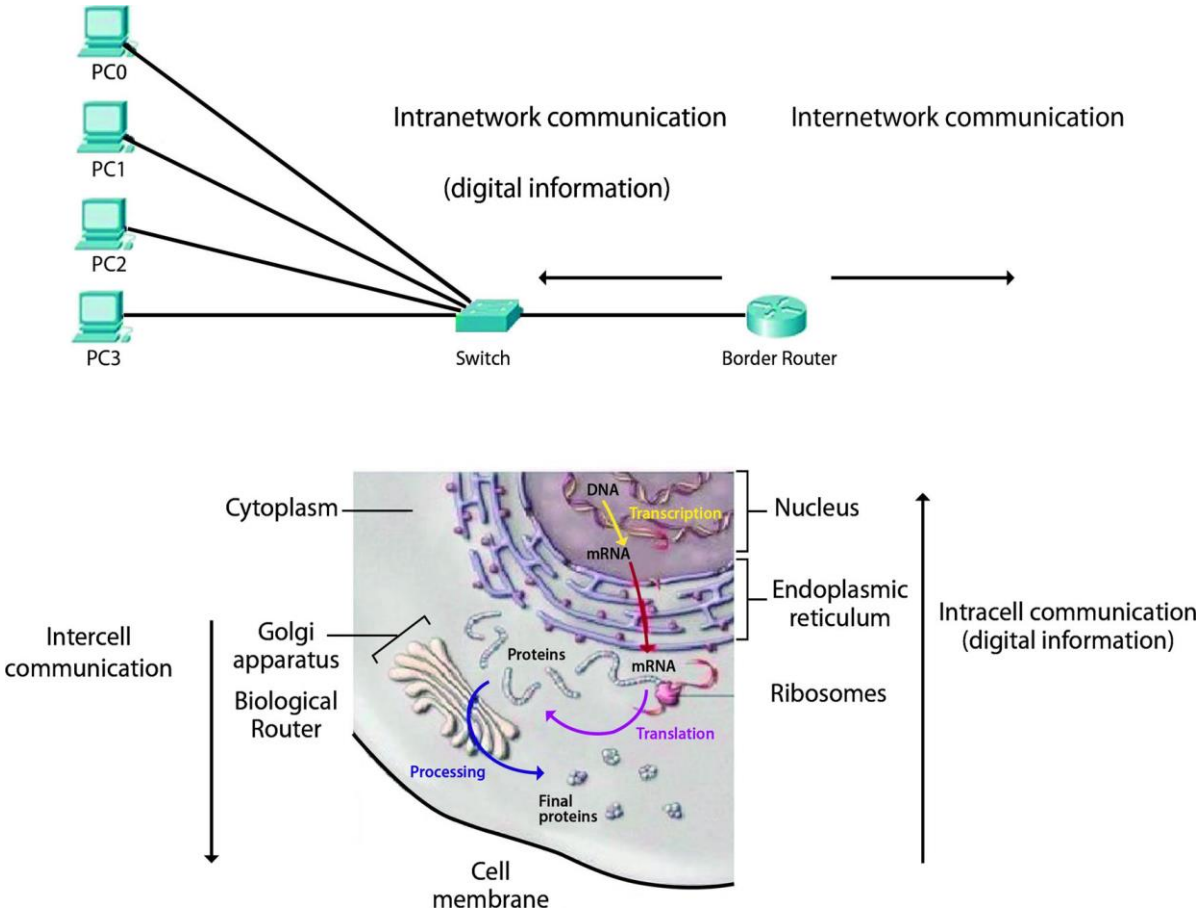


**Figure 2.9. Comparison between a router and the Golgi apparatus**

**Table 2.5. Layered model analysis comparing the Golgi apparatus with a typical router in a network**

| Layered model network level | Layered model network functions in biological systems | Layered model network functions in transmission of DNA information |
|---|---|---|
| Network layer | (a) *Molecular packet* It is a concept, information, communication or statement that is understood by the source and the destination (as well as the molecular router) at the molecular network layer[33]. | (a), (b), (c) Genetic information is provided by the nucleus, and the different types of biological information are transformed into proteins with biological functions at the receivers. The Golgi apparatus collects and dispatches these proteins to different target organs at nanodistance destinations. The functions of Golgi apparatus are necessary because over long nanolinks the information must be addressed to reach certain destinations, i.e. the adequate receivers where the biological data must be processed. These tasks permit an appropriate performance of molecular communications. Without these routing functions all senders should transmit uselessly (in intra- and intercellular communication) the information to every destination in an intent to reach the receiver.<br>The mentioned tasks are similar to these performed by a router at the network layer. |
| | (b) *Molecular router* It is a logical component of a molecular communication network. When a molecular router receives an incoming molecular packet, it directionally forwards the molecular packet towards the destination [33].<br>The Golgi apparatus manages protein traffic and acts as the primary router in cells [39]. | (d) Storage of molecular packets is necessary until the tasks are processed in this layer. |
| | (c) *Molecular network layer functionalities* 1. Provides the functionality necessary for communication over longer distances, transmits a packet without establishing a connection from a source end to a destination end, and provides decisions for each packet and its destination [1, 33, 39].<br>2. Network layer is required for communication over long distances and the transfer of molecular packets containing messages to destinations (different messages carried in molecular packets in different manners) [33] | |

| | | |
|---|---|---|
| | (d) *Molecular packet storage* It is a logical component at a molecular network node (e.g. sources, destinations, and molecular routers) to store molecular packets. Storage is supported by the link layer [33]. | |
| Data link layer | (a) *Vesicle-based framing* Vesicles may naturally correspond to a molecular frame at the molecular link layer [33]. | (a) Encapsulations in vesicles that mediate the movement of proteins through GA may correspond to a molecular frame at the molecular link layer. |
| | (b) *Molecular link layer tasks*<br>1.Provides functionalities for communication within a direct range of communication [33]. | (b)<br>1. Proteins transferred from the ER to the GA in vesicles have an implicit adjacent addressing comparable to a data link layer addressing (i.e. for communications within a direct range of communication). |
| | 2. Shared medium molecular communication link:<br>Transmitting/receiving a molecular frame on a shared link requires media access control (MAC).<br>Media access control divides a shared molecular communication link among multiple senders and transmits molecular frames from multiple senders through a shared link without causing interference between molecular frames. One mechanism for medium access control is time division multiplexing (TDM), in which different senders transmit molecular frames at different times [33]. | 2. As proteins move through the bloodstream to a target organ, the traffic of biological information from senders to receivers through this medium converts the blood into a shared medium. |
| | (c) *Molecular frame storage* It is a logical component that allows the sender and receiver to store molecular frames in their respective signal molecules (i.e. information). Storage is supported by the physical layer [33]. | (c) Molecular frame storage is necessary until tasks are processed in this layer. |
| Physical layer | (a) *Molecular communication channel* It is the medium used to transmit molecules between transmitters and receivers in direct communication [33]. | (a), (b) Biological information is transmitted by the nucleus (contained in genes) through cytosol to ribosomes and the ER to produce functional proteins (peptide hormones). Subsequently, signal molecules also arrive through the cytosol to the Golgi apparatus, which distributes them to destinations, and places them physically in the molecular communication channel towards their destination. |
| | (b) *Molecular propagation* | In our analysis the communication channel is the bloodstream. It is a |

| | | |
|---|---|---|
| | 1. Functionality in which molecules travel through space in the environment (i.e. into a molecular communication channel) [33].<br>2. Communication medium in the molecular communication environment may be represented by a concentration gradients of molecules. Molecules move actively in a certain direction [33].<br>This mode of movement is referred to as active transport, which is used by cells that secrete hormones that circulate through the bloodstream to reach distant target cells [32].<br>To use this type of communication, connectivity in transmission must be established a priori, either artificially or through autonomous behaviours [33].<br>Compared to typical communications systems, in biological systems the information itself (signal molecules) must move through the environment at least during the molecule's half-life [1, 33].<br>(c) *Molecule storage* It is a physical component that a biological cell uses to store molecules. Molecule storage may be accomplished in the communication environment [33]. | shared medium because distinct signal molecules travel through it to their destinations.<br>c) Molecular storage is necessary until tasks are processed in this layer. |

### 2.4.4. Use of Information at the Receiver

In biological systems, information is conveyed using molecules that arrive at intra-, adjacent, or intercellular destinations (long nanoscale distance, through network layer addressing [1]). We use the term ‘‘nano’’ in reference to biological cells because they are considered to be in the ‘nanorange’’ [33]. This paper analysed gene expression (transcription and translation of DNA) that produces proteins that are excreted from cells to perform a biological function. This function is executed by the endocrine system in long nanoscale distance communication [1, 37], in which network addressing is required [33]. The source end generates the information encoded in DNA (biological DTE); the ribosomes and endoplasmic reticulum (representing the biological DCE) format the data; and the Golgi apparatus (biological router) packages the proteins through the transmission medium towards its long nanoscale distance destination (network layer addressing [1]).

The method by which proteins are transported to their destinations is not conserved across all circumstances or organisms (e.g. pure random walk, random walk with drift, and random walk with reaction by amplifiers [32]). Hence, we refer to proteins excreted by the cell as peptide hormones, which travel through the bloodstream into a channel (random walk with drift [32]) to reach a target organ. This type of transmission medium may correspond to wired transmission, the behaviour of which is dependent on the physical layer [33].

As mentioned in Section 2.2, molecular systems and typical communication systems can encounter problems in processing and/or communicating information. Specifically, in molecular communication, these undesired effects include biochemical, thermal and physical noise, interference, and attenuation. The resulting damage to the signal information can cause latency (i.e. movement delay) and jitter (i.e. variation in latency) and increase the loss rate (i.e. the probability that a molecule transmitted by a biological sender is not received by the intended biological receiver) [32, 33]. Thus, for example, when a molecule moves in a fluid medium (random walk with drift) modelled as a semi-finite interval (-∞, d], the probability density function of the latency is given by Eq. 2.7 and represented in Figure 2.10 [32].

$$f(t) = \begin{cases} 0 & (t = 0) \\ \dfrac{d}{\sqrt{4\pi D t^3}} exp\left(-\dfrac{(d - vt)^2}{4Dt}\right) & (t > 0) \end{cases} \tag{2.7}$$

where $D$ is the diffusion coefficient of the molecule, $(v \geq 0)$ is the velocity of the fluid medium, and $d$ is the distance from the transmitter to the receiver. For the fluid velocity $v > 0$, the average latency is $d / v$ and the expected latency decreases in proportion to the inverse of v. The jitter is $D. d / 2.v^3$ and diminishes rapidly as $v$ increases. The loss rate can be calculated as $1 - \int_0^T f(t)dt$, which assumes that the receiver waits for the time duration $T$ [32].

Another important problem in the molecular physical layer is the loss of molecules because of their limited half-life, which leads to a loss of function [33].

In this biological scenery, as in conventional networks, an advantage of a stack of layers is the use of a fundamental task of the data link layer that transforms an imperfect channel into a line free of transmission errors or report unsolved problems to the upper layer [25].

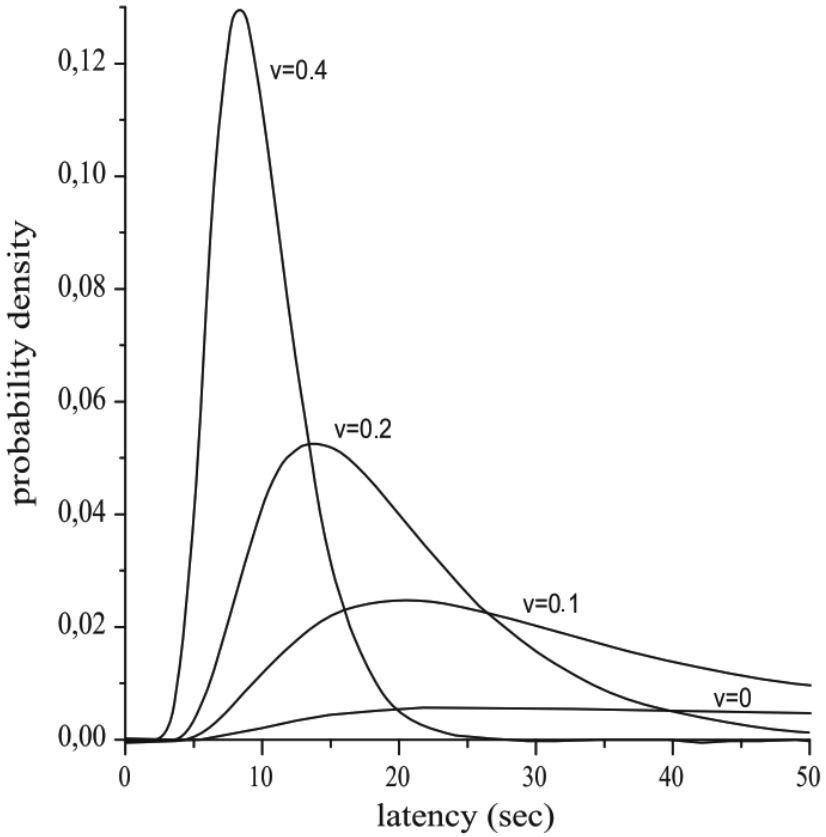


**Figure 2.10. Probability density function of the latency in a fluid medium for various fluid velocities $v = \{0, 0.1, 0.2, 0.4\}$ $\left(\frac{um}{s}\right)$, D = 0.1 $\left(\frac{um^2}{s}\right)$ and d=4 (*um*) [32]**

Subsequently, at the receiver end, the biological information must be captured through a receptor (membrane protein-chemical signal, according to the mass action law Eq. 2.8 [78]), which behaves as a transducer to decode the received signal (biological DCE). The decoded signal in turn triggers reactions inside the target cell [79–81] (biological DTE). Detection, capture, and interpretation of signal molecules by cell receptors in the receiver cell may be analogous to certain functions of the data link layer when the information arrives from the transmission channel [33].

$$L + R \leftrightarrow LR \tag{2.8}$$

where $L$ and $R$ represent a ligand and a receptor in destination cells, respectively.

This biological communication is comparable to digital communication at the receiver end, where DCE and DTE process the information.

In digital communication systems (specifically, in network communications), when information must reach a distant destination (wide area network or WAN communication), the DTE sends the information to the DCE. Next, a router modifies the information (for example, from MAC addresses in the LAN data link layer to frame relay addresses in the WAN data link layer) to permit that the data reaches the remote DCE and DTE. The biological case analysed in this paper is similar to this WAN communication (according to our communication proposal), in which routing is required to send information from the source to the destination [33]. Table 2.6 presents a layered network analysis of the layered tasks in typical networks and the use of biological information in a target organ. Although biological behaviour does not map exactly to network concepts, certain functions in biological systems are similar. The structure of a layered model decomposes a large-scale system into a set of smaller units (i.e. layers) that are functionally independent and specifies the interactions among layers [33]. Figure 2.11 resumes the biological communications described from our network communication perspective.

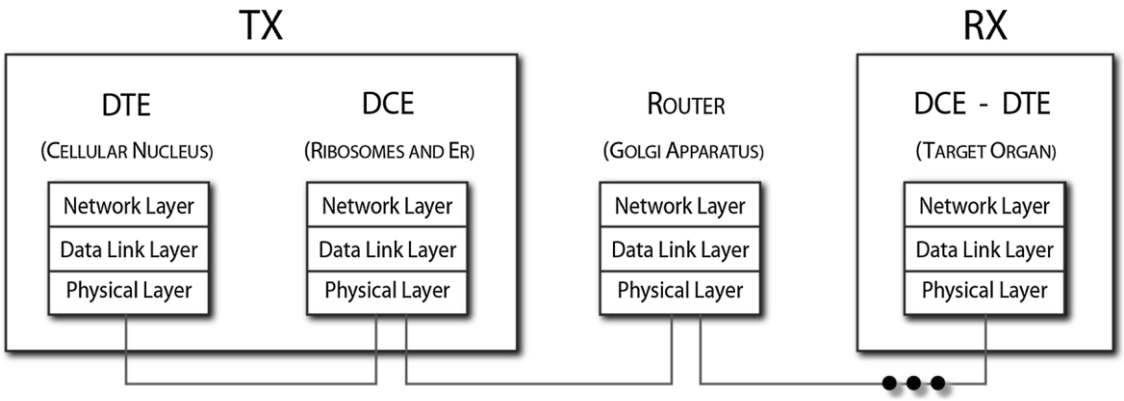


**Figure 2.11. Distribution of proteins observed in a network layered model**

**Table 2.6. Layered model analysis of the biological information at the destination**

| Layered model network level | Layered model network functions in biological systems | Layered model network functions in reception of DNA Information |
|---|---|---|
| Network layer | (a) *Molecular packet* It is a concept, information, communication or statement that is understood by the source and the destination (as well as the molecular router) at the molecular network layer [33]. | (a), (b) Information encoded in DNA is transmitted from the sender in a long nanoscale communication to the destination end (target organ). In this process, a specific type of biological information (with network layer addressing) must reach a specific destination and perform a particular biological function. Thus, identification of the destination is necessary and may be accomplished through network layer addressing. When the destination end receives the information from the sender end, the receiver ‘‘understands’’ the biological function to be performed. Hence, both ends communicate through a network layer protocol. |
| | (b) *Molecular network layer* functionalities<br>1. Provides functionality necessary for communication over longer distances, transmits packets without establishing a connection between a source and a destination, and provides decisions for each packet and its destination [1, 33, 39].<br>2. At the destination, the network layer receives a molecular packet sent from a source [33].<br>3. Receiver end uses its ‘‘specificity’’ to detect and react to specific signals [1].<br>At the receiver end, the ligand and receptor are complementary. Different ligands attach to different receptors. Specificity indicates the precision with which a signal molecule fits its complementary molecular receptor [1].<br>(c) *Molecular packet storage* It is a logical component at a molecular network node (e.g. sources, destinations, and molecular routers) that stores molecular packets. Storage is supported by the link layer [33]. | (c) Molecular packet storage is necessary until tasks are processed in this layer. |
| Data link layer | (a) Molecular frame It is a concept, information, communication or statement that is understood by the sender and the receiver at the molecular link layer [33]. | (a), (b) Decoded signal molecules contain additional information used to process, control, and interpret the molecular information at the destination end. Thus, this molecular structure and the biological processes |

| | | |
|---|---|---|
| | | may correspond to molecular frames and processes of the molecular link layer. Processes in different parts of a cell at the destination may have addressing that are comparable to data link layer addressing (i.e. for communications within a direct range of communication). |
| | As in typical network, a frame could be formed from the physical layer information and a link layer header (i.e. additional processing data) containing information specific to the layer [25, 33].<br>(b) *Molecular link layer tasks*<br>1. Provides functionalities for communication within a direct range of communication [33].<br>2. Shared medium molecular communication link: receiving a molecular frame on a shared link [33].<br>3. Molecular frame flow control: allows a sender to adjust the rate of transmission of molecular frames to avoid loss at the receiver [33].<br>4. Main task of the data link layer is to transform an imperfect channel into a line free of transmission errors [25].<br>(c) *Molecular frame storage* It is a logical component that allows the sender and receiver to store molecular frames in their respective signal molecules (i.e. information).<br>Storage is supported by the physical layer [33]. | (c) Molecular frame storage is necessary until tasks are processed in this layer. |
| Physical layer | (a) *Molecular communication channel* It is the medium used to transmit molecules between a transmitter and receiver in direct communication [33].<br>(b) *Molecular propagation* Molecules travel in the environment (i.e. into a molecular communication channel) [33].<br>Compared to typical communications systems, in biological systems the information itself (signal molecules) must move through the environment at least during the molecule's half-life [1, 33].<br>(c) *Molecule storage* It is a physical component that a biological cell uses to store molecules. Molecule storage may simply represent the molecular communication environment [33]. | (a), (b) Once the biological information arrives at its destination through the bloodstream, its processing depends on the type of signal molecules that must be captured by the receptor, and a chemical signal (ligand) must be produced to form a ligand–receptor complex<br>Subsequently, the signal molecules must be decoded into the cell to properly process the biological information.<br>(c) Molecular storage is necessary until tasks are processed in this layer. |

## 2.5. Applications

The technical principles that were analysed to formulate the proposed analogies between biological and digital systems and the design of the stacked-layer network model in this article may have many applications, including the treatment of disease.

1. *Transmission of delicate drugs.* The type of communications suggested in Sections 2.4.2, 2.4.3 and 2.4.4 which include the transmission of information encoded in DNA from the sender (with specific addressing [33]) to the receiver (target organ with a biological function) through shared medium over a long nanoscale distance, may be used to deliver delicate drugs that can be dangerous to other target organs and cause undesired and significant side effects. Depending on the protocol used to encode the information in the transmitter, only certain receivers (with the appropriate addressing) are able to decode the signal because biological processing depends on the nature of hormones [82] (Figure 2.12).

   This communication emulates a lock and key mechanism in which a receiver with the "key" can detect, read, and interpret the information. Other receivers can detect the information, but they cannot process it, thereby indicating the specificity of the process [1].

   Additionally, certain receivers can detect and decode the signal but will not do so if their address does not match the intended address. This process is similar to what occurs in Ethernet networks when a broadcast (at the data link layer level) is sent to identify an IP address: every device in the network receives the broadcast, although it is processed only if the MAC corresponds to its IP address [25].

2. *Transmission in bio-hybrid cells* Research in nanocommunications has established the use of bio-hybrid cells as sensors (getting information) and/or actuators (providing drugs) to prevent and treat diseases [8].

   The routing characteristics of DNA [1, 33, 39] (Sections 2.4.3 and 2.4.4) facilitate the transmission of information to a particular destination, where the information is then processed (e.g. in tumour cells). These properties can be utilized in cancer and pain therapies. DNA and molecular programming can discover and kill cancer cells by carrying toxic drugs to these cells through sensors and actuators [9].

A variation of the above-mentioned case is the collection of information at the destination (e.g. by a bio-hybrid collector). Enzymes produced by cells travel through the bloodstream to gatherers, where the information is stored. An analysis is then performed to determine the regular functionality of the organs or cells. Thus, the detection and prevention of diseases such as cancer may be possible [9, 83].

Thus, the design of a stacked-layer network model, such as the model proposed in this article, has the potential to utilize certain important characteristics of network theories (e.g. independent functions of layers in a stack, addressing, flow control, error control, and traffic control) in molecular communications, which may significantly improve the treatment of certain diseases [33]. For instance, a fundamental task of the data link layer is to transform an imperfect channel into a line that appears free of transmission errors. When applied to drug therapy, this functionality introduces a mechanism of control that can improve the quality of the treatment.

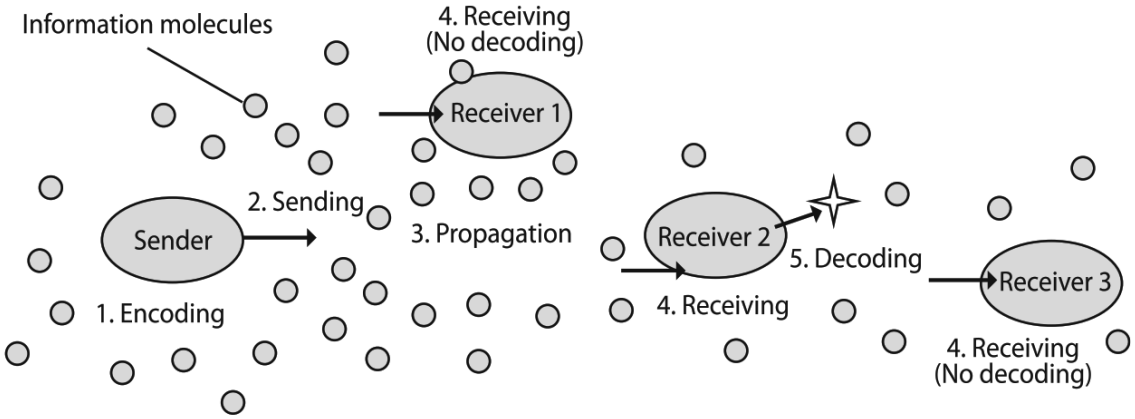


**Figure 2.12. Information in the endocrine system reaches various receivers but can be decoded only by the receiver with the appropriate address**

## 2.6. Conclusions

In this paper, we discuss the similarities between the transmission of information in conventional digital communication systems (specifically in network communications) and the expression of proteins in eukaryotic cells. The objective of this paper is to use a

network layered model to exploit network theories (e.g. independent functions of layers in a stack, addressing, flow control, error control, and traffic control) and apply these theories to the treatment of disease.

Ultimately, this model can be used to reduce the incidence of drug side effects and improve the quality of treatment for patients from health and socio-economic perspectives.

This model could also establish a foundation for designing relevant mathematical models based on digital communications characteristics. Such mathematical models may have applications in a variety of fields, such as in the medical field by detecting a target site (e.g. tumour cells) for drug therapies, improving the targeting accuracy (addressing) and reducing side effects (reducing the impact in patients). Figure 2.10 demonstrates that the fluid medium is effective for moving molecules over long distances.

A digital model of cellular behaviour can be used to apply the advantages and efficiencies of biological systems to digital communication systems and allows for the definition of duality between these systems, which can be used to overcome the disadvantages in each system.

This paper sets a challenge for the future of mathematical modelling of biological systems and digital emulations of these systems via simulations, which may lead the way towards applying network communications technology to gene expression.

## 2.7. Glossary: Cell Biology

| | |
|---|---|
| 5'-3' | The direction of nucleic acid synthesis such that the 5' phosphate of ribose or deoxyribose is joined to the 3' hydroxyl of the immediately preceding ribose or deoxyribose in a growing RNA or DNA chain. When a single strand is written, the 5' end of the molecule is conventionally written on the left and the 3' end is written on the right. |
| Active transport | Transport of a substance across a membrane that does not rely on the potential energy of a concentration gradient for the substance being transported and therefore requires an additional energy source (often ATP). |
| Amino acid | A chemical building block of proteins. There are 20 standard amino acids. A protein consists of a specific sequence of amino acids. |
| Aminoacyl (A) site | A binding site on a ribosome that accepts the incoming aminoacyl-tRNA. |
| Anticodon | A "triplet" of three nucleotides in a transfer RNA (tRNA) that is complementary to a codon in a messenger RNA (mRNA). During protein synthesis, base pairing between a codon and anticodon aligns the incoming tRNA carrying the new amino acid with the tRNA carrying the growing peptide chain. |
| ATP, adenosine triphosphate | The major source of energy for biochemical reactions in all organisms. |
| Calcium signalling | Signal transduction mechanisms where by calcium mobilization (from outside the cell or from intracellular storage pools) to the |

| | |
|---|---|
| | cytoplasm is triggered by external stimuli. Calcium signals are often seen to propagate as waves, oscillations, spikes, sparks, or puffs. The calcium acts as an intracellular messenger by activating calcium-responsive proteins. |
| Capping | Covalent modification of mRNA at the 5' end where a modified guanidine is covalently attached in a 5' – 3' linkage. |
| Cell | Fundamental structural unit of all life. The cell consists primarily of an outer plasma membrane, which separates it from the environment; the genetic material (DNA), which encodes heritable information for the maintenance of life; and the cytoplasm, a heterogeneous assemblage of ions, molecules, and fluid. |
| Cell membrane | The outer membrane of a cell, which separates it from the environment. Also called a plasma membrane or plasma lemma. |
| Chemotaxis | The movement of a cell towards or away from the source of a chemical. |
| Chromosome | A cellular structure containing genes. Excluding sperm and egg cells, humans have 46 chromosomes (23 pairs) in each cell. |
| Codon | A trinucleotide (triplet, or three-word) that specifies an amino acid or stop when parsed by the translation apparatus. The codons in mRNA molecules base pair with the anticodons of the cognate tRNAs during protein synthesis. |
| Cytoplasm | All the contents of a cell, including the plasma membrane, but not including the nucleus. |

| | |
|---|---|
| Cytoskeleton | Integrated system of molecules within eukaryotic cells which provides them with shape, internal spatial organization, motility, and may assist in communication with other cells and the environment. Red blood cells, for instance, would be spherical instead of flat if it were not for their cytoskeleton. |
| Cytosol | The semi-fluid portion of the cytoplasm, excluding the organelles. The cytosol is a concentrated solution of proteins, salts, and other molecules. |
| DNA, deoxyribonucleic acid | The substance of heredity. A long, helical, double-stranded molecule that carries the cell's genetic information. |
| Dynein | Member of a family of ATP-powered motor proteins that move towards the (-) end of microtubules by sequentially breaking and forming new bonds with microtubule proteins. |
| Embryonic stem cell | A cell found in early embryos that can renew itself and differentiate into the many cell types that are found in the human body. |
| Endocrine system | A network of glands distributed throughout the body forms the endocrine system. These glands produce hormones that are released into the circulation and distributed to distant target sites via the blood. Hormones produced by these glands act as chemical messengers to control body functions such as growth, metabolism, sexual development, and egg and sperm production. |

| | |
|---|---|
| Enhancer | A regulatory sequence in eukaryotic DNA that may be located at a great distance from the gene it controls. Binding of specific proteins to an enhancer modulates the rate of transcription of the associated gene. |
| Enzyme | A protein that speeds up a specific chemical reaction without being permanently altered or consumed. |
| ER, endoplasmic reticulum | Network of membranes in eukaryotic cells which helps in control of protein synthesis and cellular organization. |
| Eukaryote | An organism whose cells have cytoskeletons for support and their DNA contained in a nucleus, separated from the other contents of the cell. |
| Exit (E) site | The tRNA-binding site on the ribosome that binds each uncharged tRNA just prior to its release. |
| Fick's first law | An observed law stating that the rate at which one substance diffuses through another is directly proportional to the concentration gradient of the diffusing substance. |
| Fick's second law | Is used in non-steady state diffusion, i.e. when the concentration within the diffusion volume changes with respect to time. |
| GA, golgi apparatus | Eukaryotic organelle which package cell products, such as enzymes and hormones, and coordinate their transport to the outside of the cell. |
| Gap junctions | Channel formed by proteins that allows ions and other molecules to pass between adjacent cells. |

| | |
|---|---|
| Gene | A unit of heredity; a segment of DNA that contains the code for making a specific protein or RNA molecule. |
| Guanyl transferase | Guanylyl transferase is a capping enzyme complex. Guanylyl transferase is used to label either 5’ di and triphosphate ends of RNA molecules, or capped 5’ ends of RNA after chemical removal of the terminal 7-methyl-guanosine residue. |
| Hormone | A molecule that stimulates specific cellular activity; made in one part of the body and transported via the bloodstream to tissues and organs. Examples include insulin, oestrogen, and testosterone. |
| Hydrolysis | Reaction in which a covalent bond is cleaved with addition of an H from water to one product of the cleavage and of an OH from water to the other. |
| Intron | Part of a primary transcript (or the DNA from which it is transcribed). The intron is removed during RNA processing (splicing) and is not found in the mature, functional RNA. |
| Kinesin | Member of a family of motor proteins that use energy released by ATP hydrolysis to move towards the (+) end of a microtubule, transporting vesicles or particles in the process. |
| Law mass action | The law stating that the rate of any given chemical reaction is proportional to the product of the activities (or concentrations) of the reactants. |

| | |
|---|---|
| Ligand | A substance that is able to bind to and form a complex with a biomolecule to serve a biological purpose. |
| Methylation | The addition of a methyl group ($–CH_3$) to a molecule, most commonly in the context of DNA where cytosine and, less often, adenine residues can be modified in this way, sometimes resulting in a change in transcription. |
| Microtubule | Part of the cytoskeleton; a strong, hollow fibre that acts as a structural support for the cell. Microtubules also serve as tracks for transporting vesicles and give structure to flagella and cilia. |
| Motor protein | Any member of a special class of enzymes that use energy from ATP hydrolysis to walk or slide along a microfilament (myosin) or a microtubule (dynein and kinesin). |
| Myosin | The most common protein in muscle cells, responsible for the elastic and contractile properties of muscle. |
| Nucleotide | A monomer unit of nucleic acid, consisting of a purine or pyrimidine base, a sugar molecule (ribose or deoxyribose), and phosphate group (s). |
| Nucleus | Membrane-bound organelle which contains the DNA in the form of chromosomes. It is the site of DNA replication, and the site of RNA synthesis. |
| Organelle | A specialized, membrane bounded structure that has a specific function in a cell. Examples include the nucleus, Golgi, ER. |

| | |
|---|---|
| Passive transport | A kind of transport by which ions or molecules move along a concentration gradient, which means movement from an area of higher concentration to an area of lower concentration. |
| Pheromones | A chemical secreted by an animal, especially an insect, that influences the behaviour or physiology of others of the same species, as by attracting members of the opposite sex or marking the route to a food source. |
| Peptide | A natural or synthetic compound containing two or more amino acids linked together by peptide bonds. |
| Peptidyl (P) site | The binding site on a ribosome that contains the tRNA attached to the growing polypeptide chain. |
| Polarity | DNA synthesis uses 5' nucleotide triphosphates as precursor and proceeds by linking the 5' triphosphate of the incoming nucleotide to the 3'OH of the growing chain. Hence DNA strands have polarity i.e. a 5' and a 3' end. |
| Polypeptide | A linear polymer of amino acids held together by peptide linkages. |
| Poly (A) polymerase | An enzyme that adds consecutive adenosines to the 3' termini of eukaryotic mRNAs to generate poly (A) tails. |
| Promoter | A piece of genetic material that acts as a gene switch, so that a gene can become expressed in the cell. It is the region at which the RNA polymerase binds to start transcription. Most promoters are located upstream of the gene, except that some eukaryotic genes have promoters internal to the gene. |

| | |
|---|---|
| Protein | A molecule composed of amino acids lined up in a precise order determined by a gene, then folded into a specific three-dimensional shape. Proteins are responsible for countless biological functions and come in a wide range of shapes and sizes. |
| Random walk | It is no directional drift of information molecules and no chemical reaction of information molecules during propagation. Random walk is the most fundamental mechanism that molecular communication relies on to propagate a molecule. Random walk does not require any additional mechanism to propagate a molecule. |
| Random walk with chemical reactions by amplifiers | Amplifiers in the environment can increase the reliability of molecular propagation by increasing the number of propagating information molecules. Amplifiers are located in the environment and react with molecules that propagate in the environment. As a result, amplifiers produce a copy of the molecule which propagates in the environment. This class of molecular communication may be enabled by exploiting protein molecules such as those responsible for amplifying calcium ions. |
| Random walk with drift | Information molecules may undergo a directional drift which continuously propagates molecules in the direction of the drift. An example of this class of molecular communication is found in our body. Cells in the body secrete hormonal substances which circulate with the flow of the blood stream and propagate to distant target cells distributed throughout the body. |

| | |
|---|---|
| Ribosome | A molecular complex in which proteins are made. In eukaryotic cells, ribosomes either are free in the cytoplasm or are attached to the rough endoplasmic reticulum. |
| rRNA, ribonucleic acid | A molecule very similar to DNA that plays a key role in making proteins. There are three main types: messenger RNA (mRNA) is an RNA version of a gene and serves as a template for making a protein, ribosomal RNA (rRNA) is a major component of ribosomes, and transfer RNA (tRNA) transports amino acids to the ribosome and helps position them properly during protein production. |
| RNA polymerase | An enzyme that makes RNA using DNA as a template in a process called transcription. |
| RNA splicing | A process that results in the precise cutting of RNA, removal of introns to produce a fully functional RNA. |
| SRP, signal recognition particle | Protein–RNA complex that binds to signal sequences and targets polypeptide chains to the endoplasmic reticulum. |
| Start codon | The mRNA triplet (AUG) that is recognized by the ribosome as a signal for the start of translation. |
| Stop codons | One of three codons in mRNA (UAG, UGA, UAA) that function as signals for the termination of translation by ribosomes. |
| Transcription | The process of copying information from genes (made of DNA) into messenger RNA. |

| | |
|---|---|
| Translation | The process of making proteins based on genetic information encoded in messenger RNA. Translation occurs in ribosomes. |
| Vesicle | A small, membrane-bounded sac that transports substances between organelles as well as to and from the cell membrane. |

## 2.8. References


1. Bush SF. Nanoscale communication networks. Norwood: Artech House; 2010. pp. 53, 57, 67, 69, 70, 194, 195, 234.
2. Siddique N, Adeli H. Nature inspired computing: an overview and some future directions. Cogn Comput. 2015;7(6):706–14.
3. Dressler F, Akan OB. A survey on bio-inspired networking. Comput Netw. 2010;54(6):881–900.
4. Nakano T, Moore M. Molecular communication paradigm overview. J Next Gener Inf Technol. 2011;2(1):9–16.
5. Pérez ST, Vásquez JL, Travieso CM, Alonso JB. Artificial neural network in FPGA for temperature prediction. In: International conference on nonlinear speech processing. New York: Springer;2011. pp. 104–10.
6. Rushdi A, Tuqan J, Strohmer T. Map-invariant spectral analysis for the identification of DNA periodicities. EURASIP J BioinformSyst Biol. 2012;1:1–21.
7. Chahibi Y, Pierobon M, Akyildiz IF. Pharmacokinetic modeling and bio distribution estimation through the molecular communication paradigm. IEEE Trans Biomed Eng. 2015;62(10):2410–20.
8. IEEE Transactions on NanoBioscience. IEEE. http://ieeexplore. ieee.org/xpl/RecentIssue.jsp?punumber=7728.
9. Felicetti L, Femminella M, Reali G, Lio` P. Applications of molecular communications to medicine: a survey. Nano Commun Netw. 2016;7:27–45.
10. ns-3 and nanoscale simulations. 2016. https://www.nsnam.org/news/ns-3-and-nanoscale-simulations/.
11. Lee WP, Lin CH. Combining expression data and knowledge ontology for gene clustering and network reconstruction. Cogn Comput. 2016;8(2):217–27.
12. Farhan F, Mushfiq M. Biological cell and molecular communication technology: overview and challenges. Eur Sci J (ESJ). 2015;11(6):56–60.
13. Shannon C. A mathematical theory of communication. Bell Syst Tech J. 1948;27:379–423.

**14.** Burbeck S. Complexity and the evolution of computing: biological principles for managing evolving systems. Comput Syst. 2007;2004–7. http://www.evolutionofcomputing.org.

**15.** Nakano T, Suda T, Koujin T, Haraguchi T, Hiraoka Y. Molecular communication through gap junction channels: system design, experiments and modeling. In: Bio-inspired models of network, information and computing systems (Bionetics 2007). 2nd ed. IEEE; 2007. pp. 139–46.

**16.** Burbeck S, Jordan KE. An assessment of the role of computing in systems biology. IBM J Res Dev. 2006;50(6):529–43.

**17.** Ganegoda GU, Li M, Wang W, Feng Q. Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations. IEEE Trans Nanobiosci. 2015;14(2):175–83.

**18.** Raz NR, Akbarzadeh-T MR, Tafaghodi M. Bioinspired nanonetworks for targeted cancer drug delivery. IEEE Trans Nan/obiosci. 2015;14(8):894–906.

**19.** Townsend J, Keedwell E, Galton A. Artificial development of biologically plausible neural-symbolic networks. Cogn Comput. 2014;6(1):18–34.

**20.** Felicetti L, Femminella M, Reali G, Nakano T, Vasilakos AV. TCP-like molecular communications. IEEE J Sel Areas Commun. 2014;32(12):2354–67.

**21.** Shrivastava S, Badlani R. Data storage in DNA. Int J Electr Energy. 2014;2(2):119–21.

**22.** Chang WL, Vasilakos AV. DNA algorithms of implementing biomolecular databases on a biological computer. IEEE Trans Nanobiosci. 2015;14:104–11.

**23.** Worthey EA. Transformation of big data into clinically actionable knowledge: supporting the personalized medicine revolution. In: 37th annual international conference of the IEEE Engineering in Medicine and Biology Society. 2015. http://embc.embs.org/2015/keynote-worthey/.

**24.** Abdullah A, Hussain A, Khan IH. Introduction: dealing with big data-lessons from cognitive computing. Cogn Comput. 2015;7(6):635–6.

**25.** Tanenbaum AS, Wetherall DJ. Computer Networks. New York: Pearson Education; 2012. pp. 31, 43, 45, 46, 194, 216.

**26.** Loscri V, Marchal C, Mitton N, Fortino G, Vasilakos AV. Security and privacy in molecular communication and networking: opportunities and challenges. IEEE Trans Nanobiosci. 2014;13(3):199.

**27.** Choffnes ER, Relman DA, Pray L, et al. The science and applications of synthetic and systems biology: workshop summary. Washington, DC: National Academies Press; 2011.

**28.** Marchisio MA, Stelling J. Automatic design of digital synthetic gene circuits. PLoS Comput Biol. 2011;7(2):e1001083.

**29.** Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet. 2001;2(1):343–72.

**30.** Shaw IS. A study of analogies between processes in technical and biological systems. Johannesburg: University of Johannesburg; 2012.

**31.** Shirazi AZ, Mazinani SM, Eghbal SK. Protocol stack for nano networks. In: IEEE international symposium on computer, consumer and control (IS3C); 2012. pp. 849–53.

**32.** Nakano T, Moore MJ, Wei F, Vasilakos AV, Shuai J. Molecular communication and networking: opportunities and challenges. IEEE Trans Nanobiosci. 2012;11(2):135–48.

**33.** Nakano T, Suda T, Okaie Y, Moore MJ, Vasilakos AV. Molecular communication among biological nanomachines: a layered architecture and research issues. IEEE Trans Nanobiosci. 2014;13(3):169–97.

**34.** Singh RP, Sapre SD. Communication systems. New York: Tata McGraw-Hill Education; 2008. pp 364.

**35.** Abu-Rgheff MA. Introduction to CDMA wireless communications. Cambridge: Academic Press; 2007. pp. 51.

**36.** Morris DJ. Communication for command and control systems: international series on systems and control, vol. 5. Amsterdam: Elsevier; 2014. pp. 98–100.

**37.** Akyildiz I, Pierobon M, Balasubramaniam S, Koucheryavy Y. The internet of bio-nano things. IEEE Commun Mag. 2015;53(3):35.

**38.** Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO,et al. What is a gene, post-ENCODE? History and updated definition. Genome Res. 2007;17(6):669–81.

**39.** Tipton HF, Krause M. Information security management handbook. 6th ed. v. 2. Boca Raton: CRC Press; 2008. https://books.google.it/books?id=EqpjYH_Z6MQC.

**40.** Bettelheim F, Brown W, Campbell M, Farrell S, Torres O. Introduction to general, organic and biochemistry. Boston: Cengage Learning; 2013. pp. 614-640.

**41.** Lieberman M, Marks AD. Marks' basic medical biochemistry: a clinical approach. Philadelphia: Lippincott Williams & Wilkins; 2009. pp. 206.

**42.** Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al. Molecular biology of the cell. 500 tips series. New York: Garland Science; 2014. pp. 334–60. https://books.google.co.uk/books?id=1ZUDoQEACAAJ.

**43.** Farsad N, Yilmaz HB, Eckford A, Chae CB, Guo W. A comprehensive survey of recent advancements in molecular communication. New York: Garland Science; 2014.

**44.** Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, et al. Essential cell biology. New York: Garland Science; 2013.pp. 185–7, 383–407, 487–513.

**45.** Goodman SR. Medical cell biology. Amsterdam: Elsevier; 2007. pp. 175. https://books.google.es/books?id=tRbCHk9easQC.

**46.** Starr C, McMillan B. Human biology. Belmont: Cengage Learning; 2013. pp. 412–3.

**47.** Zhang C, Liu H, Zheng K, Hao Y, Tan Z. DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. Nucleic Acids Res.2014;136(4):1381–90.

**48.** Jorde LB, Carey JC, Bamshad MJ. Medical genetics. Amsterdam: Elsevier; 2015. pp. 6–11. https://books.google.com.ec/books?id=FrRgCgAAQBAJ.

**49.** Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson genetics in medicine. Amsterdam: Elsevier; 2015. pp. 23. https://books.google.com.ec/books?id=4yV1CQAAQBAJ.

**50.** Solomon E, Martin C, Martin D, Berg L. Biology. Stamford: Cengage Learning; 2014. pp. 305–7. https://books.google.es/book s?id=Z0rAAgAAQBAJ.

**51.** Anttalainen T. Introduction to telecommunications network engineering. Artech House telecommunications library. Norwood: Artech House; 2003. pp. 144–8. https://books.google.es/ books?id=S0_Pvzr-TeEC.

**52.** Laskar J, Chakraborty S, Pham AV, Tantzeris MM. Advanced integrated communication microsystems, vol. 174. Hoboken: Wiley; 2009. pp. 10.

**53.** Snape A, Papachristodoulou D, Elliott WH, Elliott DC. Biochemistry and molecular biology. Oxford: Oxford University Press; 2014. pp. 377.

**54.** Snustad DP, Simmons MJ. Principles of genetics, binder ready version. Hoboken: Wiley; 2015. pp. 267–74. https://books.google.com.ec/books?id=NBB0CgAAQBAJ.

**55.** Reece JB, Meyers N, Urry LA, Cain ML, Wasserman SA, Minorsky PV, et al. Campbell biology Australian and New Zealand version. Melbourne: Pearson Higher Education AU; 2014. pp. 340–4. https://books.google.es/books?id=5t6aBQAAQBAJ.

**56.** FitzGerald J, Dennis A. Business data communications and networking. Hoboken: Wiley; 2009. pp. 136. https://books.google.es/ books?id=EKW1_wcuOpgC.

**57.** Hillis DM. Principles of life. Sunderland: Palgrave Macmillan; 2011. pp. 60-65, 187-206.

**58.** Holmes R, et al. Salters-Nuffield advanced biology: AS student book. Oxford: Heinemann; 2005. pp. 229.

**59.** Fitzgerald-Hayes M, Reichsman F. DNA and biotechnology. Cambridge: Academic Press; 2010. pp. 52-74.

**60.** Del Vecchio D, Murray RM. Biomolecular feedback systems. Princeton: Princeton University Press; 2014. pp. 48.

**61.** Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. How cells read the genome: from DNA to protein. New York: Garland Science; 2014.

**62.** Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Nucleic acids, the genetic code, and the synthesis of macromolecules. New York: W.H. Freeman; 2000.

**63.** Ignatova Z, Martinez-Perez I, Zimmermann KH. DNA computing models. Berlin: Springer; 2008.

**64.** Drlica KS, Perlin DS. Antibiotic resistance: understanding and responding to an emerging crisis. New York: Pearson Education; 2011. pp. 217. https://books.google.co.uk/books?id=YUppJBKF wPwC.

**65.** Lewin B. Genes 9. v. 9. Burlington: Jones & Bartlett Learning; 2008. pp. 110. https://books.google.co.uk/books?id=BrNpbPkzoxAC.

**66.** Jackson ARW, Jackson JM. Environmental science: the natural environment and human impact. New York: Pearson Education; 2000. pp. 139–40. https://books.google.co.uk/books?id=WLYC dqaHj1MC.

**67.** Russell PJ. Biology: the dynamic science, vol. 1. Belmont: Cengage Learning; 2008. pp. 323. https://books.google.co.uk/ books?id=VYMFAAAAQBAJ.

**68.** Plopper G, Plopper RPIG. Principles of cell biology. 2nd ed. Navigate advantage access. Burlington: Jones & Bartlett Learning, LLC; 2014. pp. 277–85. https://books.google.es/books?id=z_ ZGBQAAQBAJ.

**69.** Stabler LB, Metz M, Wilkes A. Kaplan AP biology 2016. Kaplan test prep. New York: Kaplan Publishing; 2015. pp. 124. https:// books.google.es/books?id=DFgRCgAAQBAJ.

**70.** Dubitzky W, Southgate J, Fuß H. Understanding the dynamics of biological systems: lessons learned from integrative systems biology. New York: Springer; 2011. pp. 1. https://books.google.es/ books?id=MNtTKe0S4ngC.

**71.** Cracraft J, Donoghue MJ. Assembling the tree of life. Oxford: Oxford University Press; 2004. pp. 60.

**72.** Bolsover SR, Shephard EA, White HA, Hyams JS. Cell biology: a short course. Hoboken: Wiley; 2011. pp. 223–31. https://books. google.es/books?id=Kt_hL1stQQkC.

**73.** Surgeons AAO, Elling B, Elling KM, Rothenberg MA. Paramedic: anatomy & physiology. Sudbury: Jones & Bartlett Learning; 2012. pp. 22. https://books.google.es/books?id=-LhaI-gOUgC.

**74.** Beckerman M. Molecular and cellular signaling. New York: Springer; 2006. pp. 6–9.

**75.** Seth G, Hossler P, Yee JC, Hu S. Engineering cells for cell culture bioprocessing—physiological fundamentals. In: Hu WS, editor. Cell culture engineering, vol. 101. New York: Springer; 2006. pp. 119–64.

**76.** Sircar S. Principles of medical physiology. New York: Thieme; 2008. pp. 30. https://books.google.es/books?id=zFl7y5xqHj4C.

**77.** Pusztai A, Bardocz S. Lectins: biomedical perspectives. London: Taylor & Francis; 2005. pp. 168. https://books.google.es/books? id=qK40HAP6bGgC.

**78.** Demchenko AP. Introduction to fluorescence sensing. New York: Springer; 2008.

**79.** Starck JM, Ricklefs RE. Avian growth and development: evolution within the altricial-precocial spectrum, vol. 8. Oxford: Oxford University Press; 1998. pp. 174.

**80.** Lister AL, Van Der Kraak GJ. Endocrine disruption: why is it so complicated? Water Qual Res J Can. 2001;36(2):175–90.

**81.** Roper MG, Guillo C. New technologies in affinity assays to explore biological communication. Anal Bioanal Chem. 2009;393(2):459–65.

**82.** Nakano T, Moore M, Enomoto A, Suda T. Molecular communication technology as a biological ICT. In: Sawai H, editor. Biological functions for information and communication technologies, vol. 320. Berlin: Springer; 2011. pp. 49–86.

**83.** Du W, Cao Z, Wang Y, Zhou F, Pang W, Chen X, et al. Specific biomarkers: detection of cancer biomarkers through high throughput transcriptomes data. Cogn Comput. 2015;7(6):652–66.

# Chapter 3: Modeling Gene Expression and Protein Delivery as an End-to-end Digital Communication System

**Yesenia Cevallos[1,2], Lorena Molina[1], Guillermo Machado[1,3], Tadashi Nakano[4], Ahmad Rushdi[5], Jesús B. Alonso[6], Floriano De Rango[2]**

✉Yesenia Cevallos
ycevallos@unach.edu.ec
Lorena Molina
lmolina@unach.edu.ec
Guillermo Machado
gmachado@unach.edu.ec
Tadashi Nakano
tadasi.nakano@fbs.osakau.ac.jp
Ahmad Rushdi
aarushdi@ieee.org
Jesús B. Alonso
jalonso@dsc.ulpgc.es
Floriano De Rango
derango@dimes.unical.it

1. Faculty of Engineering, National University of Chimborazo, Av. Antonio José de Sucre, 060150 Riobamba, Ecuador
2. Department of Computer Engineering, Modeling, Electronics and Systems, University of Calabria, Pietro Bucci Street, Building 41C, 87036 Rende (CS), Italy
3. Faculty of Public Health, Polytechnic High School of Chimborazo, 060150 Riobamba, Ecuador

4. Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita 565-0871, Japan
5. Department of Electrical and Computer Engineering. University of California, Davis, CA 95616, USA
6. Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria, Office D-102, Block B, Building of Electronic and Telecommunications, Department of Signalling and Communications, Campus Tafira, 35017, Las Palmas, Spain

**Funding** This study has not explicit funding; it was made through a scholarship given to Ph.D. Student Yesenia Cevallos by National University of Chimborazo of Ecuador.

**Compliance with Ethical Standards**

**Conflict of Interest** Yesenia Cevallos, Lorena Molina, Guillermo Machado, Tadashi Nakano, Ahmad Rushdi and Jesús B. Alonso, Floriano De Rango declare that they have no conflict of interest.

**Informed Consen**t Informed consent was not required as no human or animals were involved.

**Human and Animal Rights** This article does not contain any studies with human participants performed by any of the authors.

## 3.1. Abstract

In this article, we analyse gene expression from a digital communication systems perspective. Specifically, we use network theories, such as addressing, flow control, error control, traffic control, and Shannon's theorem, to design an end-to-end digital communication system that represents gene expression. In particular, in this digital communication system, we define layered network models that represent the transcription and translation of deoxyribonucleic acid (DNA) and the end-to-end transmission of proteins to a target organ. These models take advantage of the key features of digital communication systems (i.e., performance and efficiency) to the transmission of biological information in gene expression systems. The proposed analysis and design establish the duality between digital and biological communication systems, and our results could be used to overcome the challenges that both communication systems face. One of the most important applications of this study is the development of communication systems for medical purposes (i.e., in the treatment of diseases, such as cancer). The analysis presented in this study may, for instance, enable the avoidance of side effects through the transmission of biological information to a suitable destination (i.e., to specific target organs) in order to achieve optimal treatments.

**Keywords**: Gene expression, Proteins, Digital communication, Biological communication, DNA, RNA, Medical applications

## 3.2. Introduction

Communication theories have established an important parallel between digital communication systems and biological systems [1]. Thus, the analysis of the behaviour of these systems based on common transmission theories requires the establishment of bidirectional equivalences between the systems, which are discussed below.

Undoubtedly, biological systems can be viewed as a communication system [2]. They use communication components such as the sender, channel and receiver to achieve information transmission. At the sender, a physiological process is required for the generation or storage of information-carrying molecules. Subsequently, the information-carrying molecules are released into a channel through which the information-carrying

molecules are transported to the receiver due to a propagation mechanism. At the receiver, a detector must be present to measure a certain property or properties of the information-carrying molecules. This property could be the presence or absence of information-carrying molecules, their concentration, time of arrival, type of molecule, or any other measurable parameter [3].

The above-mentioned model can be used to describe the transmission of malicious information related to human diseases. For example, viruses, such as human immunodeficiency virus (HIV), information-carrying molecules in this case, propagate through the entire body, causing infection. Fortunately, treatments for diseases and drug transmission systems are able to provide information that results in the destruction of malicious cells. In addition, advanced therapies use gene delivery vectors containing information encoded in DNA to repair abnormal or variant genes. Although the information and/or the mechanisms to delivery information may differ among biological systems, the communication model is the same. Communication theories can be thus used as a foundation for the design of a layered protocol platform that combines the advantages of internetworks and computational science to evaluate and model HIV transmission [4].

The sharing of resources and information is one of the most important functionalities of computer networks. In addition, the management (i.e., monitoring and control) of computer networks is typically required to maintain their appropriate functioning and ensure their adequate performance [5]. Similar to the network issues discussed in [6], biomolecules function cooperatively in biological networks to detect malfunctions in diseases and identify drug targets. Network control is used to steer the information transmission among various nodes and to change the state of the network while identifying drug targets and providing insight into experimental designs for drug discovery.

The propagation of molecules in environments, such as air, water and blood plasma, via simplified multilayer diffusion channels in the respiratory system is analysed in [7]. The modelling of the channels in the time domain (in terms of delay and capacity) is conducted according to Brownian motion and Fick´s laws and the analysis of the channel in the frequency domain (as in conventional channels) using Fourier transforms. This and many other studies in molecular communication [8] focus on the

characterization of diffusion channels under the physical constraints specific to molecular communication; these constrains are similar to those present in digital communication systems and are even more important in biological communication systems given that these channels are used to deliver information-carrying molecules that play a vital role in the organs of the human body or the administration of drugs for the treatment of a disease [7].

In [9], a biological system is modelled as a digital communication system. In this biological communication system, nucleotide bases in DNA serve as transmitters, DNA serves as the communication medium, and the living organisms serve as the receivers. In this biological communication system, errors are hypothesized to occur in the transmission channel and produce mutations in the DNA. Then, [9] presents the development of robust algorithms that operate in one of two regions, i.e., non-protein-coding DNA or protein-coding DNA, as a mechanism to prevent mutations and avoid lethal consequences, such as diseases in biological receivers.

In [10], gene expression or, equivalently, the flow of biological information is presented from an information communication theory perspective; this system is modelled as a digital communication system. Due to errors in the biological environment (i.e., channel), the functions of cells, organisms or species may be severely affected. However, similarly to digital communication systems, error control and correcting codes can be used to overcome the errors in the transmission medium. In particular, [10] proposes the use of linear and convolutional codes to detect and correct errors in a digital representation of gene expression.

The intracellular transmission of genetic information via a biological communication system is represented as a digital communication system in [11]. In this biological communication system, the identification, reproduction and mathematical classification of nucleotide sequences are necessary to subsequently identify, reproduce and mathematically classify DNA sequences. The errors in the system are efficiently corrected using G linear codes. According to [11], the methodology used to characterize this biological communication system could also be used to prevent mutations in DNA, produce new drugs and create genetic improvements.

Due to the similarities between the transmission of information in communication

systems and the transmission of drugs in the human body, in [12, 13, 14, 15], several transmission platforms and models used to deliver drugs only to specific targets to minimizes the side effects, are discussed. These specialized transmission systems, which are called drug delivery systems, could be used in the medical treatment of cancer, HIV, genetic diseases, gastrointestinal diseases, and neurological diseases. The routes of transmission (i.e., administration) of drugs into the body include oral ingestion, injection, and skin application [12].

The above-mentioned studies have applied the key features of each system to the other. In biological systems, the evolutionary process used by living organisms over billions of years [3] and characteristics, such as energy efficiency, biocompatibility, the ability to work in an aqueous medium and pervasiveness, are the most important feature [16]. In digital systems, performance and efficiency are the two important features [1].

Nevertheless, both systems have common challenges. Attenuation, delay, interference, noise and loss of information could result in fatal communication problems. For instance, failure in biological systems could result in the development of lethal diseases, such as cancer, in human beings [17, 18].

However, biological communication (or molecular communication) is a promising scientific field, and its application to medical treatment could greatly improve the quality of human life; for instance, conventional therapies could non-invasively access small and delicate body sites [19].

In this article, we analyse gene expression from a digital communication systems perspective. Specifically, we use network theories, such as addressing, flow control, error control, traffic control and Shannon's theorem, to design layered network models that represent the transcription and translation of DNA and the end-to-end transmission of proteins to a target organ. These models take advantage of the key features of digital communication systems for the transmission of biological information in gene expression systems. Additionally, these models help us improve our understanding of biological communication systems. Further, by analysing the models, we establish the duality between digital and biological communication systems and provide solutions to overcome the challenges that both systems face. One of the most important applications of this analysis could be the development of communication systems for medical purposes (i.e.,

the treatment of diseases, such as cancer). This analysis may, for instance, enable the avoidance of side effects through the transmission of biological information to the suitable destination (i.e., a target organ) in order to achieve optimal treatments.

This paper is organized as follows. In Section 3.3, we describe the elements of the communication architecture in digital systems and biological systems, emphasizing the processing/transmission of information between the sender and receiver. Because both systems behave very similarly and their components are analogous, in Section 3.4, we present an analysis of gene expression and protein delivery based on conventional digital communication systems, i.e., from Shannon's perspective. This section is divided into 6 subsections. In the first subsection, we describe the cell nucleus as biological data terminal equipment (DTE). In the second subsection, we define the ribosomes and endoplasmic reticulum (ER) as biological data communication equipment (DCE). At the end of this subsection, we design a layered network model that represents the transcription and translation of DNA. In the third subsection, we define the Golgi apparatus as an internet (border) router. At the end of this subsection, we design a layered network model that represents this biological fact. In the fourth subsection, we represent gene expression in a binary format in which each codon and each amino acid is represented by a digital code. Subsequently, in the fifth subsection, we use Shannon's theorem to digitally transmit biological information through a fluid with drift and determine the capacity $C$ (in bits per second), i.e., the highest rate of information transmission from a transmitter to a receiver with respect to a given number of transmitted molecules. In the sixth subsection, the digital receiver uses Markov chains to emulate ligand-receptor systems. At the end of this subsection, we design a layered network model that represents the end-to-end transmission of proteins to a target organ. In Section 3.5, we discuss the possible applications of our study in the medical field. The conclusions are presented in Section 3.6.

## 3.3. Communication system architecture

### 3.3.1. Digital communication system components

Information confers awareness of an event to the receiver. The transmission of information has an implicit random origin from which information is sent to a destination

at a non-deterministic moment, and a user at the communication end (i.e., a receiver) utilizes this information [1, 20]. Digital communication systems have been developed to accomplish reliable information transmission.

Digital communication is one of the most advanced information transmission technologies, and the characteristics that determine the potential usefulness of digital communication systems include storage, compression, encryption, transmission efficiency, bandwidth and speed. These features can be applied to nearly all forms of communication because digital systems are universal; therefore, regardless of the type of communication, information can adopt a universal form to facilitate its processing [1, 21]. For example, information sent through a computer, including audio, data, text and video, is delivered by a network using this universal form.

To achieve transmission, digital communication systems should have the following three basic components (according to Shannon's communication model): a transmitter, a communication channel and a receiver. The transmitter end includes the data terminal equipment (DTE) and data communication equipment (DCE) blocks, and the receiver at the destination end must also contain DCE and DTE blocks (Figure 3.1 [1]). At each end, the DTE and DCE are interconnected by a physical interface that defines their electrical, mechanical, functional and procedural characteristics [1, 22].

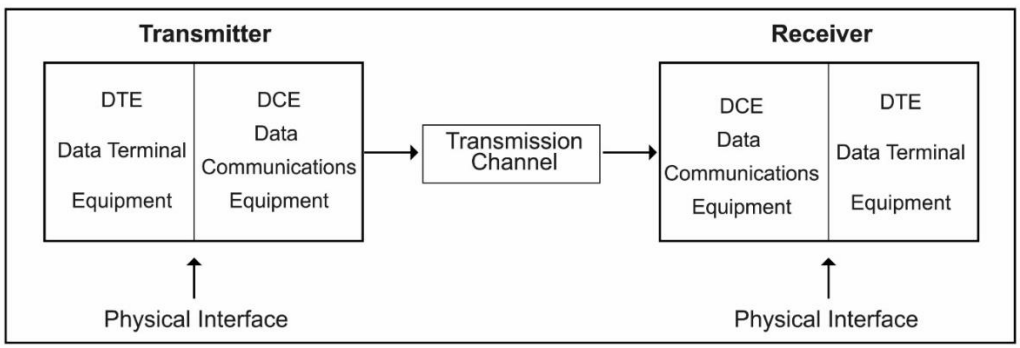


**Figure 3.1. Basic components of a communication system**

The DTE in the transmitter contains the information source, which is often stochastic in typical communication systems, and determines the beginning of the

communication process. In certain cases, the information must be altered before it can be delivered; thus, a transducer is utilized. Once the information presents suitable characteristics, it is sent through the physical interface to the DCE, which contains the communication controllers and produces the appropriate signal (synchronism). One of the most important functions of the DCE is the conversion of information into a format that is appropriate for the transmission channel. In communication systems, the DCE is a modem (modulator–demodulator) or codec (coder-decoder) [1, 22].

The communication channel comprises the transmission medium (wired or wireless type), noise, interference and distortion. Attenuation, which arises from the physical and electrical conditions in the transmission medium, the frequency band and the link distance, is an undesirable effect in transmissions. Thus, in nearly all cases, errors in communication occur in the communication channel [1, 22].

Due to the presence of the DCE in the transmitter, the information must be modified. The receiver must contain a DCE block to convert the information back to its original format. Moreover, if the information has been corrupted in the transmission medium, the DCE detects these errors and transmits the information appropriately. Finally, the DTE collects the information from the physical interface and transports it to its destination [1, 22]. This communication between the transmitter and receiver is known as simplex (and connectionless), and, except for in broadcast transmissions, the transmitter must identify the destination address before sending the information.

### 3.3.2. Biological system components

Living organisms can perform specific tasks and functions through the exchange of biological data (through molecular communication) from sources that release information-carrying molecules into the environment (medium) in which the information-carrying molecules propagate to the receiver. The form in which the biological information is presented, the mechanisms used to transport the information into the medium, and the way in which this biological information is sensed by the receiver depend on the biological functions of the cells, tissues or organ; however, the following five processes are commonly found in biological communication systems (molecular communication systems (Figure 3.2) [2]:

1. Encoding. Encoding is the process by which a biological transmitter transforms information into informational molecules with specific biochemical characteristics that can be detected by the biological receiver. Information can be encoded in various forms by the informational molecules; for instance, information could be encoded in different types of molecules by to a particular nucleotide sequence in the DNA.
2. Sending. Sending is the process by which a biological transmitter releases informational molecules into the environment. For instance, if the biological transmitter is a cell, the sending process could include budding vesicles or the catalysis of a chemical reaction that produces informational molecules elsewhere.
3. Propagation. Propagation is the process by which informational molecules travel from a biological transmitter through the environment to the biological receiver. An informational molecule may passively diffuse through the environment (as shown in the upper panel of Figure 3.2) or propagate actively (e.g., via molecular motors that guide the molecules as shown in the lower panel of Figure 3.2). Molecular communications display stochastic behaviour that can be quantified as noise in the propagation medium.
4. Receiving. Receiving is the process by which a biological receiver captures the informational molecules propagated in the environment. Surface receptors in the receiver can bind a specific type of informational molecule and induce reactions on the surface that cause reactions within the biological receiver.
5. Decoding. Decoding is the process by which a biological receiver, upon capturing the informational molecules, chemically reacts with the molecules. This process allows the biological receiver to accomplish a function in a cell, tissue or organ.

As stated in the above paragraphs, an obvious parallelism exists between digital communication systems and biological communication systems. Also, based on the previously mentioned observations and studies discussed in Introduction, biological systems can be analysed from the perspective of Shannon's communication model (Figure 3.3 [3]). Hence, in the following sections in this

article, we use this paradigm to analyse biological transmission in gene expression systems from a digital communication systems perspective.

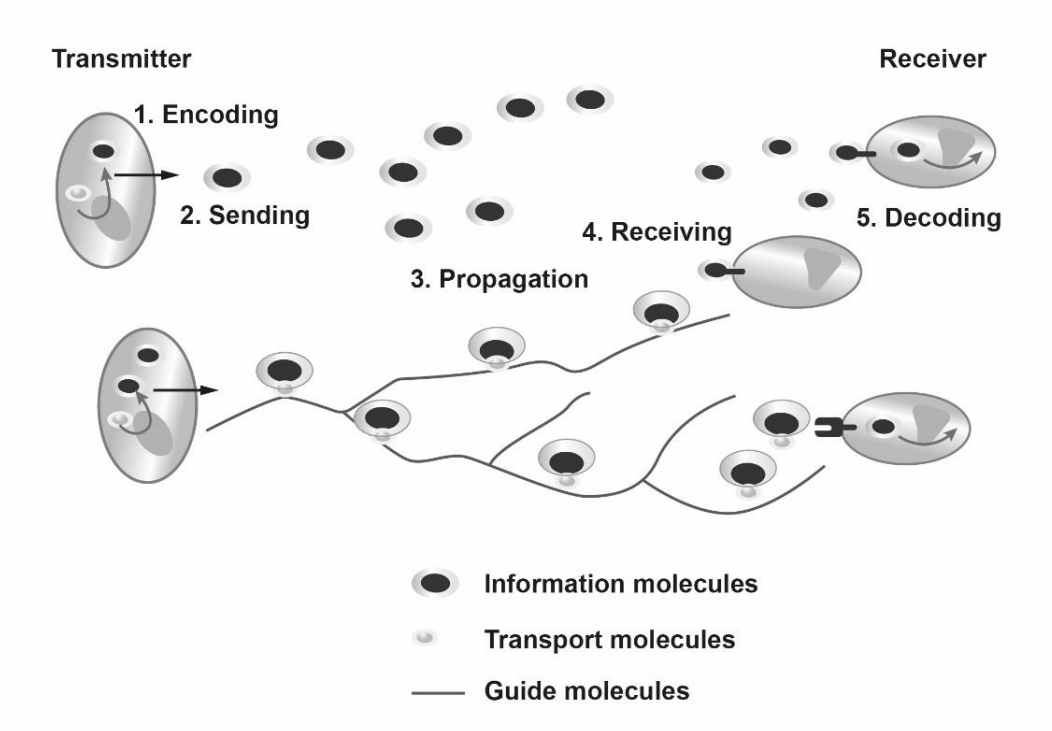


**Figure 3.2. Model of molecular communication**

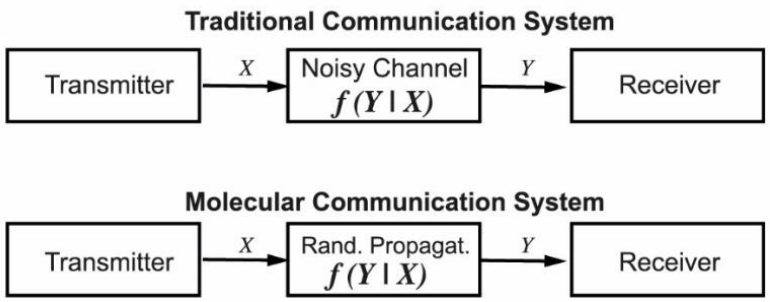


**Figure 3.3. Shannon's communication model applied to digital and molecular communication systems [3]**

## 3.4. Gene expression and protein delivery as a digital communication system

In this section, we extend the architecture (Figure 3.1) to describe gene expression in cells and protein delivery to a target organ. In doing so, we use the models of biological communication (Figures 3.2 and 3.3) to describe key processes.

### 3.4.1. The cell nucleus as biological DTE

We hypothesize that the biological DTE is represented by the nucleus (in eukaryotic organisms), which contains DNA molecules as the information source (i.e., the transmitter) [1]. DNA contains biological data that are organized in nucleotide blocks called genes. The biological data must be processed intracellularly or extracellularly. We define extracellular communication as a type of transmission that occurs over a long distance (i.e., distance longer than the size of a cell), such as transmission in the endocrine system [1, 23, 24].

A gene is typically defined as a set of nucleotides that stores the information required for a biological function to be performed at a destination; this function can be accomplished by a protein or ribonucleic acid (RNA) [25]. In network terms, the contents of a gene may be associated with a network layer address [1]. Therefore, the transmission of biological or molecular information begins with the transcription of DNA into RNA and is analogous to the digital communication system process described in Section 3.3. One biological reason for this process is that DNA molecules cannot leave the nucleus of the cell [26]; their information must therefore be transmitted using a different molecule. RNA contains a copy of the information and carries the information out of the nucleus [27]. This process is similar to the process that occurs in communication systems in which a transducer modifies an input signal to allow it to be processed by other components of the communication system [1].

Within DNA molecules, information is encoded by four discrete values or four nucleotides, i.e., DNA molecules contain *digital* information. The nucleotides are monomers or subunits of nucleic acids (DNA and RNA) comprising one nitrogenous base, a five-carbon sugar (deoxyribose in DNA and ribose in RNA) and at least one phosphate group. The nitrogenous bases include adenine (A), thymine (T), cytosine (C), guanine (G) and uracil (U). The DNA double helix comprises nucleotides containing the

bases A, T, C and G [28] and maintains its structure due to the complementarity between the nitrogenous bases of each strand of the helix [29], i.e., the affinity of adenine to thymine and that of cytosine to guanine [30]. For example, if the hypothetical sequence in one strand is ATC, the complementary strand must have the sequence TAG [1].

The information in DNA is divided into blocks of nucleotides called genes [31, 32] that possess start and termination sequences; thus, biological information is divided into data segments. In packet-switching networks, digital information is divided into smaller units known as packets to facilitate processing. Therefore, a packet in a digital network may be analogous to a gene in a biological communication network [1].

At the beginning of transcription, the molecular motor RNA polymerase II (the RNAP II enzyme) recognizes a region of the DNA sequence called the promoter region [33]. The promoter hosts the start sequence, which is the location at which RNAP II begins to add nucleotides to create a complementary messenger RNA (mRNA) sequence [34].

During the initiation of transcription, RNAP II produces a complementary single-stranded mRNA copy of one of the two DNA strands (Figure 3.4 [1]). The only difference between RNA and DNA is that RNAP II uses uracil instead of thymine during this process [35, 36]. There is no need to copy both DNA strands because the strands are exact complements. This concept may be comparable to the use of the digital gate NOT (or one's complement) to eliminate redundant information [1]. Moreover, this biological process resembles digital data compression because the same amount of data is contained in a smaller space [1]. In fact, the maximum compression of the genetic code for the optimal use of its nucleotides is given by [37]:

$$\frac{H\ (pv)}{log_2 a} \tag{3.1}$$

where *H(pv)* is the entropy of the probability vector *pv* with respect to a specific nucleotide, and *a* is the number of letters in an alphabet; in this case, the alphabet consists of the four nucleotides mentioned above.

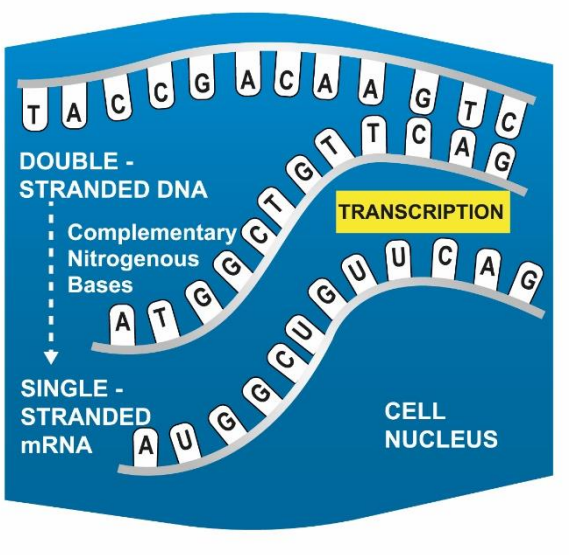


**Figure 3.4. Transcription of DNA into mRNA**

The "enhancer" is another element of the DNA strand [38] that controls the quantity of protein produced according to the amount of mRNA. Because the "enhancer" controls the amount of information sent to the receiver, this process may be understood as a flow control at the sender end [1]. Datalink layers are responsible for flow control to ensure that a rapid sender cannot swamp a slow receiver with more messages than can be processed [39, 40].

Transcription proceeds unidirectionally along one of the DNA strands from the 5´P to the 3´OH of the deoxyribose phosphate backbone. This order is important to ensure that the genetic information is copied appropriately [41]. In digital communications, the order of transmission is also fundamental [1]. For example, in a serial communication in which the less significant bit is generally transmitted first, 001 (1 in decimal value) and 100 (4 in decimal value) are not equivalent.

When the information is almost completely copied into single-stranded mRNA (the molecule that leaves the nucleus), an appropriate finalization sequence is recognized by RNAP II, and transcription is halted [42]. The signalling to start and stop transcription is used in the management of biological clocks that participate in the feedback mechanisms of cellular processes [43].

The following three types of modifications (maturation) occur in the primary transcript molecule (i.e., pre-mRNA) [44]:

1. Splicing. Segments in the pre-mRNA molecule that do not provide information (introns) are removed.
2. Capping. Post-transcriptional processing of the 5´ end of the RNA occurs through the 5´ cap process. At the end of transcription, the 5´ end of the RNA transcript contains a free triphosphate group because this nucleotide was incorporated first into the chain. The capping process replaces the triphosphate group with another structure called the "cap". The cap is added by the enzyme guanylyl transferase. This enzyme catalyses the reaction between the 5´ end of the RNA transcript and a guanine triphosphate molecule.
3. Polyadenylation. Post-transcriptional RNA processing at the opposite end (3´ extreme) of the transcript occurs through a string of approximately 250 adenines that are attached to the end of the synthesized RNA chain. This string of adenines is called the "poly-A tail". The addition of the adenines is catalysed by the enzyme poly (A) polymerase.

The information added during capping and polyadenylation may be equivalent to the delimiting data flags used in digital communication systems, such as headers and trailers that encapsulate the information in the datalink layer in protocol hierarchies in network software (Figure 3.5) [1, 2, 39]. These flags are used for processing and error control [43]. In eukaryotic cells, the cap and poly-A tail are added to provide stability (control and posterior processing) to the mRNA molecules and prevent the degradation of the mRNA by enzymes in the cytosol or intracellular fluid, thus allowing the molecules to advance to the subsequent phases of biological processing. The mechanical transport of mRNA molecules through the cytosol may be analogous to the transmission of information in wired communications (physical layer task) [1, 23, 30, 43].

Figure 3.6 [1] illustrates the analogies between biological and digital DTE (according to our communication proposal).

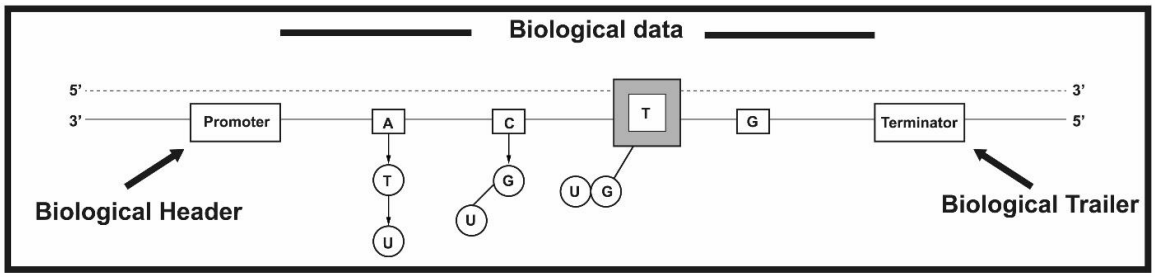


**Figure 3.5. Biological frame in the transcription of DNA**

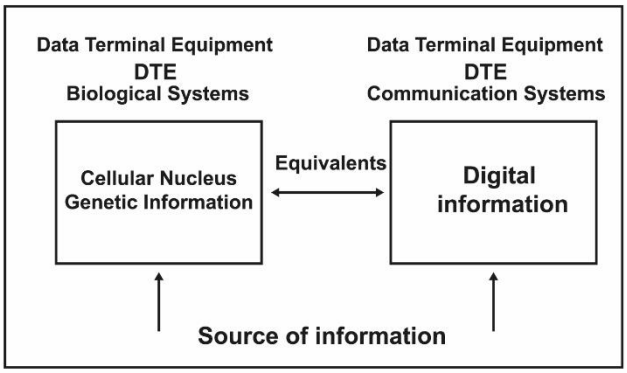


**Figure 3.6. Biological and digital DTE at the transmitter end**

### 3.4.2. The ribosomes and endoplasmic reticulum (ER) as biological DCE

Once DNA is transcribed into a single RNA strand, the genetic information leaves the biological DTE, which requires communication with the biological DCE. In digital communication systems, the connectivity between the DTE and DCE occurs through a physical interface. In a biological system, the cytosol may represent the physical interface [1] (Figure 3.7 [1]).

In a digital communication system, the DCE (codec or modem) is essential for translating the information into the correct format and transmitting the information through a channel. According to our perspective, the ribosomes and ER represent the "biological DCE" because they process genetic information and provide the functional structure (or "format" when referring to data) that is subsequently released into the biological transmission medium and ultimately arrives at the biological receiver [1]. The genetic information is "formatted" during "translation", which is a process by which genetic information is converted into amino acid chains with biological functionality

inside and outside the cell. Therefore, the biological DCE codifies information via translation and provides a specific input sequence (data in mRNA) that is associated with a specific output (sequence of amino acids); this codification process corresponds to conventional digital codification [1].

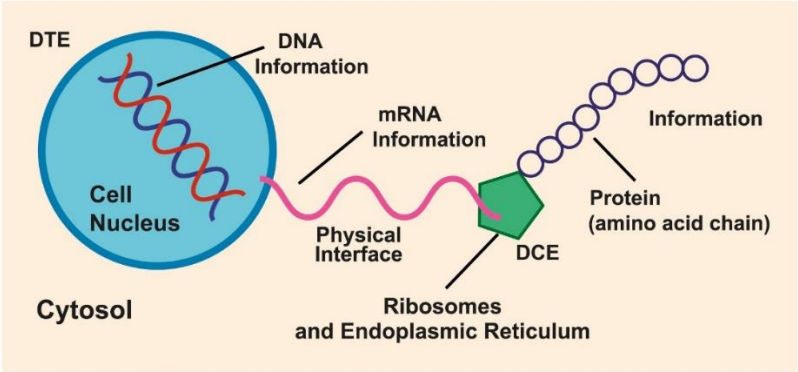


**Figure 3.7. The cytosol in a biological system is analogous to a physical interface in a digital system**

During translation, the mRNA that leaves the nucleus has an implicit adjacent address that is comparable to a data link layer address to facilitate communication within a direct range of communication [1, 43]; thus, the mRNA is bound by the ribosomes in the cytosol or those associated with the rough ER (RER) [45]. The movement of biological information from the nucleus to the ribosomes or ER through the cytosol, which is equivalent to information arriving and moving into a biological communication channel, is considered a task at the physical layer [1, 43].

Ribosomes, which are structures that serve as molecular motors, read the information contained in the biological sequence using a codon system (a codon is a triplet of nucleotides) [33]. In the ribosome, the codons in the mRNA are recognized by transfer RNAs (tRNAs) that possess an anticodon (a sequence complementary to a particular codon) associated with a unique amino acid that binds specifically to the molecular structure of that tRNA [46]. Once again, the concept of complementarity may be equivalent to the use of the digital gate NOT (or one's complement) in each codon [1].

Protein synthesis occurs in the ribosome through the signalling of tRNAs that indicate to the ribosomes the beginning (start codon) and end (stop codon) of the process to ensure the proper reading of the biological information [30]. From a digital perspective, the analysis of amino acid interactions that form proteins is essential for understanding the evolutionary relationships among organisms, the development of new drugs and the production of synthetic proteins [12].

In a digital communication paradigm, the "start" and "stop" codons may correspond to synchronism signals. Synchronism between the source and destination is performed in synchronous transmissions through a "start flag". In this type of information transmission, the transmitter sends the data, and the receiver must collect and process the data. The "stop" codon in biological signalling may be equivalent to the "stop flag" in synchronous communications that is used at the destination to indicate the end of communication. We emphasize synchronous transmissions (and not asynchronous transmissions) because the transmission of a large amount of information is accomplished through synchronous communications; for instance, 455 EB of data can be encoded in 1 gram of single-stranded DNA [39, 47]. The signals to start and stop the translation of DNA allow the biological clocks present in cells to provide feedback during cellular processes [1, 43].

Because proteins (i.e., polypeptide chains) are generated by the processing of specific amino acids, the order of the amino acids depends on the prior amino acid; therefore, this biological characteristic may correspond to input in Markov's source. Proteins consist of 20 different amino acids; therefore, Eq. 3.2 represents the *k*th-order entropy, where $p_i$ is the probability of finding the *i*th amino acid, and $p(i|s)$ is the conditional probability of the *i*th amino acid to occur after the amino acid string *s* [1, 23].

$$I_k(x) = -\sum_{i=1}^{20} \sum_{s}^{20^{k-1}} p_i p(i|s).log_2 p(i|s) \tag{3.2}$$

If all codons (amino acids) occur with an equal probability, the number of possible sequences (*nps*) in polypeptide chains of length *N* can be calculated using Eq. 3.3.

$$nps = 20^N \tag{3.3}$$

Because many amino acids do not have the same probability, Eq. 3.3 must be modified. Thus, if we consider a long sequence of *N* symbols selected from an alphabet of either codons or amino acids, we can determine the probability of that sequence using Eq. 3.4 [37].

$$P = \prod_i p(i)^{N.p(i)} \tag{3.4}$$

where *Np(i)* is the probability of the *ith* symbol [37].

In the previous equation, if we consider the logarithm at both sides,

$$log_2 P = N. \sum_i p(i). log_2 p(i) = -N. H \tag{3.5}$$

Where

$$H = -\sum_i p(i). log_2 p(i) \tag{3.6}$$

and *H* is the information or Shannon's entropy of the probability space containing the events *i* [37].

Accordingly, the probability of obtaining a long sequence of *N* independent symbols or events from a finite alphabet is [37]:

$$P = 2^{-N.H} \tag{3.7}$$

The number of sequences of length *N* is nearly [37]

$$2^{N.H} \tag{3.8}$$

The analysis presented in this article focuses on proteins processed in the ER because many proteins (e.g., peptide hormones, such as insulin) play a role outside the cell [48], where they reach the receiver and complete an end-to-end communication.

When the aforementioned tagging occurs (i.e., tagging to play a role outside the cell), a signal recognition particle (SRP) is bound to the amino acid sequence, thereby providing it with an implicit adjacent address via molecular tagging. This molecular tag is comparable to a data link layer address that facilitates communication within a direct range of communication [1, 43]). The function of the SRP is to allow the nascent protein to arrive at a channel protein in the ER that oversees the translocation of the protein within the ER. Subsequently, the SRP detaches from the protein and is recycled in the cytosol [28]. Similarly, in digital systems, after the processing information and control information are used, they are discarded. Inside the ER, the proteins are folded and acquire the functional three-dimensional structure necessary for them to accomplish their specific biological functions [28] (equivalent to digital information after processing by the DCE) [1].

In biological systems, information errors can occur during DNA transcription and translation; similarly, in conventional communication systems, errors can occur in the transmission medium. Errors in cellular processing and information communication are responsible for many medical disorders, such as cancer, autoimmunity, and diabetes [18, 23, 49]. Figure 3.8 [1] illustrates the analogies between the biological and digital equivalents of DCE. Supported by our previous studies, i.e., [1], Figure 3.9 presents a layered network model of the transcription and translation of DNA.

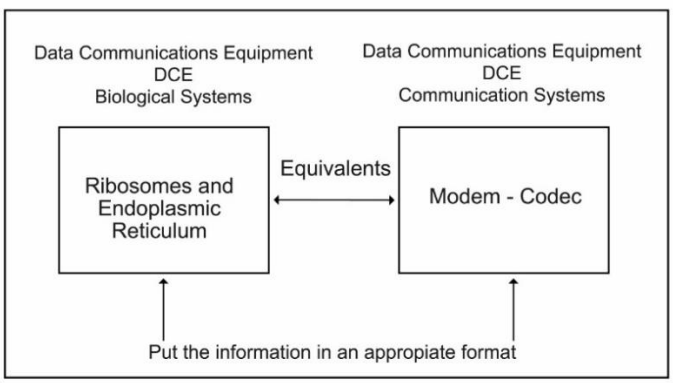


**Figure 3.8. Biological and digital DCE at the transmitter end**

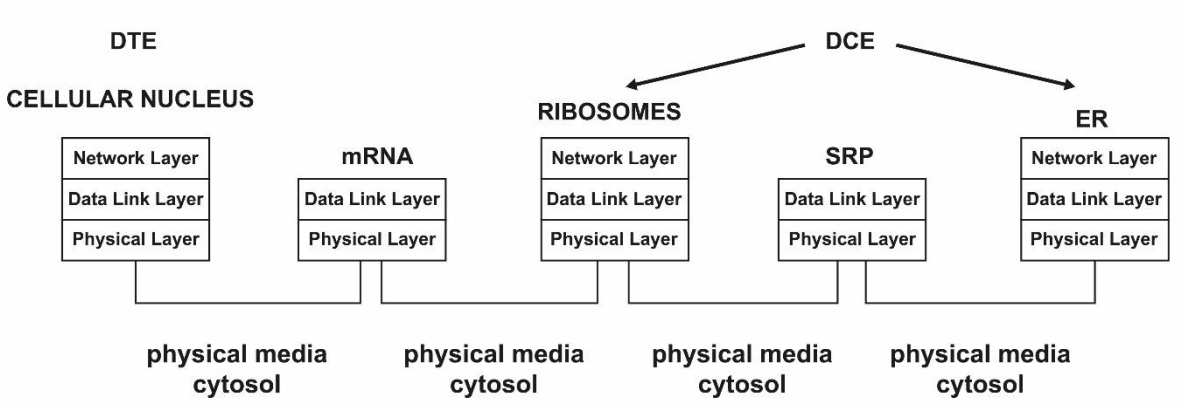


**Figure 3.9. DTE-DCE layered network model of the transcription and translation of DNA**

Figure 3.9 presents a layered network model of gene expression from the typical network perspective. The structure of a layered model decomposes a large-scale system into a set of smaller units (i.e., layers) that are functionally independent of each other and specifies the interactions among the layers. Hence, an advantage of using a stack of layers is the fundamental use of the data link layer to transform an imperfect channel into a line free of transmission errors or report unsolved problems to the upper layer [39]. Therefore, the application of such a model to biological systems (e.g., drug delivery) could provide high reliability [1].

### 3.4.3. The Golgi apparatus as an internet (border) router

When proteins are functional, the RER transfers the proteins via molecular motors to the GA. Because each protein contains an implicit adjacent address via molecular tagging, which is comparable to a data link layer address that facilitates communication within a direct range of communication [1, 43]) [2], the proteins are routed to the appropriate intercellular destinations; however, the GA determines whether the proteins remain inside the cell [28]. During this process, the proteins, along with their information content, move from the RER to the GA, where the information is deposited into vesicles that bind the cis GA face. Then, new vesicles containing the protein information are generated, and other cellular components required for processing the proteins are added. The new vesicles deposit their contents into the medial GA face, and, again, new vesicles

containing the protein and the elements necessary for further processing are formed. Finally, the vesicles reach the trans GA face, where a process identical to the previously defined process occurs; thus, the proteins are inserted into new vesicles but are directed to the endoplasmic membrane to be secreted outside the cell.

The functions of the GA are similar to those performed by a router in a network as shown in the topology presented in Figure 3.10 [1]. The router determines whether the information remains inside the network or leaves the network; thus, routing and routed processes are indispensable for sending information to the appropriate destination. Furthermore, the actions of depositing proteins, forming vesicles and attaching information to specify a protein destination are analogous to those required in the processing of Protocol Data Units (PDUs) in the layers of a router [1]. Figure 3.11 represents a layered network model of the communication of information from the DTE to the GA.

Figure 3.11 presents a layered network model from the typical networks perspective representing gene expression as an intranet with its corresponding internet (border) router. The structure of a layered model decomposes a large-scale system into a set of smaller units (i.e., layers) that are functionally independent of each other and specifies the interactions among the layers. Hence, an advantage of using a stack of layers is the fundamental use the data link layer to transform an imperfect channel into a line free of transmission errors or report unsolved problems to the upper layer [39]. Therefore, the application of such a model to biological systems (e.g., drug delivery) could provide high reliability [1].

The usefulness of the digital theories that have been applied to the current biological analysis could be very important for the treatment of diseases because the proteins transmitted from the sender end, which correspond to the regulation of gene expression, define the development of a multicellular organisms, such as humans, and may cause the development of pathological abnormalities, such as cleft palate or cancer [50]. Thus, the regulation of gene expression could be considered a flow control mechanism in digital communications.

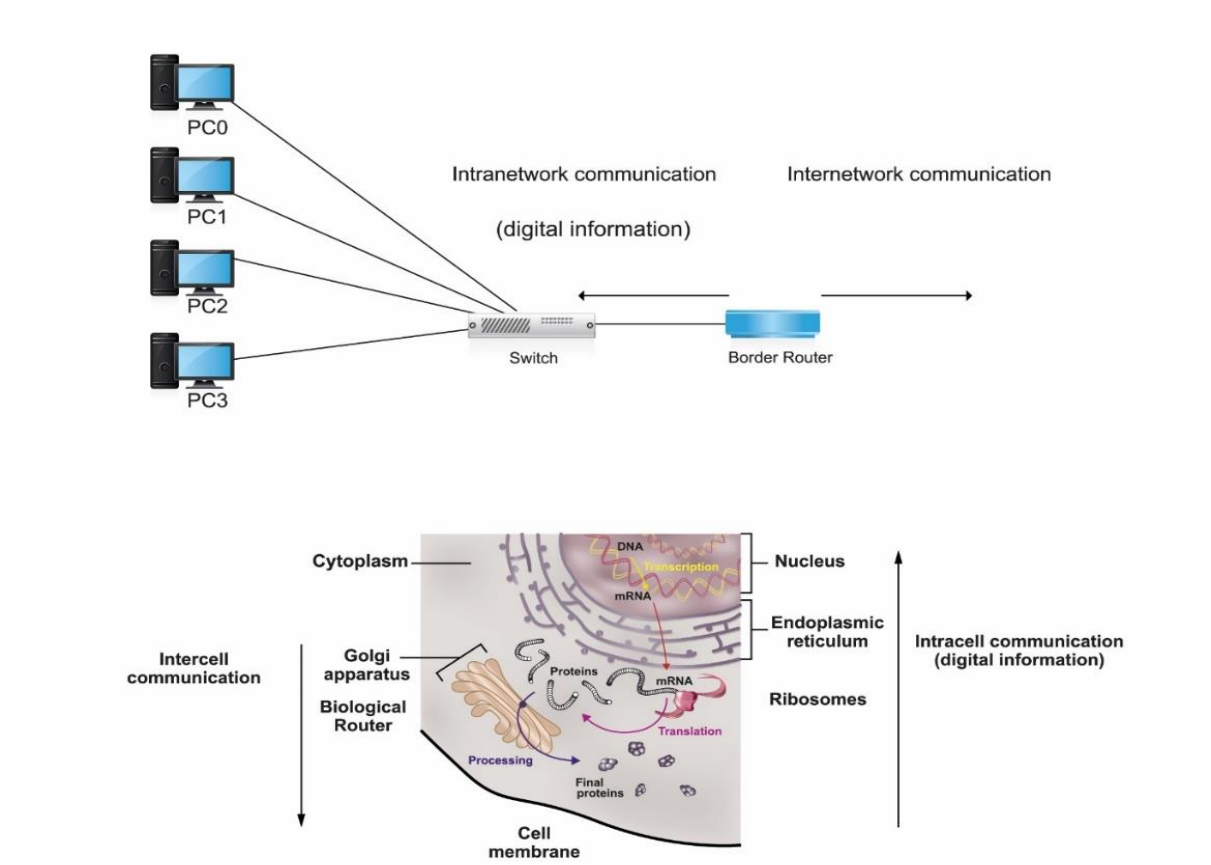


**Figure 3.10. The functions of the GA are similar to the tasks of an internet router**

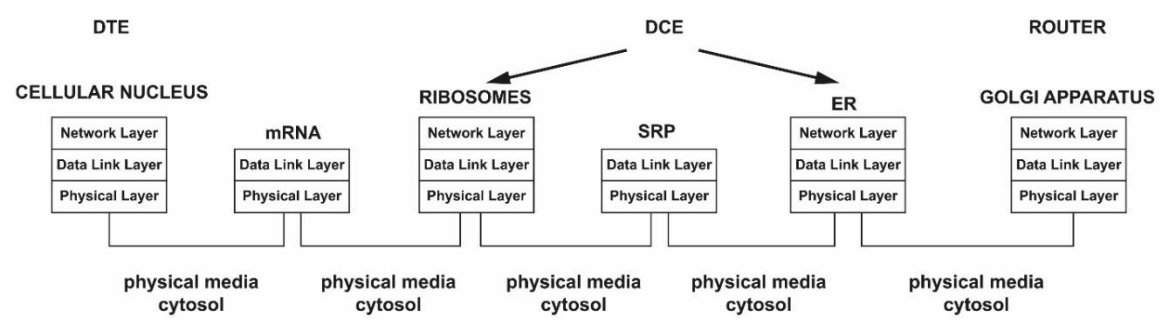


**Figure 3.11. Layered network communication model of the communication of information from the DTE to the GA**

### 3.4.4. Binary representation of gene expression

Genetic information is encoded by a four-letter alphabet in which each letter corresponds to a nucleotide base. In contrast, the amino acid alphabet consists of 20 separate letters, one for each amino acid. Hence, according to information theory, the quantity of information carried by a single nucleotide base is 2 bits $\left(-log_2(1/4)\right)$, and the quantity of information required to unambiguously specify one amino acid from a set of 20 is 4.3 bits $\left(-log_2(1/20)\right)$. Therefore, the use of 6 bits to specify an amino acid represents an excess of information; however, this excess information capacity may explain the redundancy in the genetic code [51]. In digital communication systems, redundancy provides an essential mechanism for controlling the information and detecting errors; Similarly, the genetic code contains redundant information to reduce errors in gene expression [13, 39, 52].

Various combinations can be used to encode a nucleotide base with 2 bits; we use the method chosen in [53]. Using this method, the digital encoding most clearly distinguishes pyrimidine bases (PY, uracil or cytosine) from purine bases (R, adenine or guanine). The number of heterocyclic rings in the pyrimidine and purine bases differs, and the mutations that preserve this classification, i.e., transitions, are more frequent but less damaging than transversion mutations, in which the classification of the base is changed. The binary identifiers are selected to reflect the molecular similarities exhibited by the nucleotides. Thus, each nucleotide base is encoded according to Table 3.1. In this coding system, the first bit is 0 for the PY bases (two heterocyclic rings) and 1 for the R bases (one heterocyclic ring). The second bit is 0 for the "weak" bases that form 2 hydrogen bonds with each other during Watson-Crick pairing (W=U or A) and 1 for the "strong" bases that form 3 hydrogen bonds (S=C or G). Therefore, the digital encoding of the nucleotide bases is as follows: U 00, C 01, A 10, and G 11 [53].

The standard amino acid correspondence (Tables 3.2 and 3.3) entries are recast as 6-bit binary messages, i.e., the adequate correspondence of the basic unit of mRNA (codons) to protein translation is represented using 6 bits. Due to the clustering of amino acids with similar physicochemical properties (the most important amino acid properties for proper protein folding and function are size, hydropathy, and charge), individual bit

positions correlate with specific properties. The above-mentioned classification system considers the most "determinative" bits first and prioritizes the same nucleotide molecular features prioritized by nature; for instance, this organization replaces thymine with uracil. In addition, Tables 3.2 and 3.3 indicate the amino acid corresponding to each codon, reflecting the long-established notion that the second letter of each codon conveys the most information about the intended amino acid, followed by the first letter. Thus, related amino acids tend to be grouped into the same vertical column. The third letter of a codon is often degenerate because it does not change the identity of the encoded amino acid. Thus, to prioritize the most significant bits, we used a classification system that reorders the nucleotides in a codon in the order 2, 1, 3. An example of the binary representation is shown using the codon **AUG**, which codes for the amino acid methionine (Table 3.2). The 6-bit index of the codon is determined by concatenating the 2-bit identifiers from the second nucleotide ($U_{00}$), the first nucleotide ($A_{10}$), and the third nucleotide ($G_{11}$) in order, yielding **001011**. This method is equivalent to the following series of questions: Is the second base a purine? (0 for No, 1 for Yes). Is the second base strong? (0 for No, 1 for Yes). These questions are repeated for the first base of the codon and then for the third base of the codon [53].

**Table 3.1. Nucleotide bases represented by 2 bits**

| | **PY** **(pyrimidine: first bit is represented as 0)** | **R** **(purine: first bit is represented as 1)** |
|---|---|---|
| **W** (Weak: second bit is represented as 0) | $U_{00}$ | $A_{10}$ |
| **S** (Strong: second bit is represented as 1) | $C_{01}$ | $G_{11}$ |

**Table 3.2. Method used to determine the 6-bit index corresponding to each codon. In this case, the codon is AUG, which represents methionine and corresponds to 001011**

| **Bit 1** | **Bit 2** | **Bit 3** | **Bit 4** | **Bit 5** | **Bit 6** |
|---|---|---|---|---|---|
| **Is the second base R?** | Is the second base S? | Is the first base R? | Is the first base S? | Is the third base R? | Is the third base S? |
| **NO** | NO | YES | NO | YES | YES |
| **0** | 0 | 1 | 0 | 1 | 1 |
| **$U_{00}$** | | $A_{10}$ | | $G_{11}$ | |
| **A** | | U | | G | |
| **Methionine** | | | | | |

Following this method, each of the 64 codons is assigned a unique 6-bit index that places the most important information first. A complete amino acid correspondence table generated using this method is provided in Table 3.3 [53].

To transmit biological information associated with gene expression from a digital communication perspective, i.e., to emulate the transcription of DNA in the biological DTE discussed in Section 3.4.1, we use the method outlined in the previous paragraphs to associate each nucleotide base with 2 bits. Subsequently, to emulate the translation of DNA by initially processing the specific triplets of nucleotides discussed in Section 3.4.2, we use a codon selector (Figure 3.12) to obtain six bits that represent each specific codon. Then, to complete the translation process, it is necessary to associate these 6 bits with the digital equivalence of an amino acid (i.e., an entry in Table 3.3); thus, we used a digital decoder of 6 bits to 64 bits (Figure 3.12). As previously discussed (Sections 3.3.3 and 3.4.2), after biological information is processed in biological DCE, the information travels to the GA, from which it is transmitted to the transmission channel (i.e., the bloodstream). Hence, we used serial communication to transmit each decoded bit individually to an internet (border) router, and from this router, the information was transmitted to the transmission channel, which was also performed using serial communication (Figure 3.13).

## Table 3.3. Equivalences of amino acids and bits

| First Base | Second Base | | | | | | | | | | | | Third Base |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | U | | | C | | | A | | | G | | | |
| U | UUU | 000000 | \|Phe | UCU | 010000 | Ser | UAU | 100000 | Tyr | UGU | 110000 | Cys | U |
| | UUC | 000001 | Phe | UCC | 010001 | Ser | UAC | 100001 | Tyr | UGC | 110001 | Cys | C |
| | UUA | 000010 | Leu | UCA | 010010 | Ser | UAA | 100010 | Stp | UGA | 110010 | Stp | A |
| | UUG | 000011 | Leu | UCG | 010011 | Ser | UAG | 100011 | Stp | UGG | 110011 | Trp | G |
| C | CUU | 000100 | Leu | CCU | 010100 | Pro | CAU | 100100 | His | CGU | 110100 | Arg | U |
| | CUC | 000101 | Leu | CCC | 010101 | Pro | CAC | 100101 | His | CGC | 110101 | Arg | C |
| | CUA | 000110 | Leu | CCA | 010110 | Pro | CAA | 100110 | Gln | CGA | 110110 | Arg | A |
| | CUG | 000111 | Leu | CCG | 010111 | Pro | CAG | 100111 | Gln | CGG | 110111 | Arg | G |
| A | AUU | 001000 | Ile | ACU | 011000 | Thr | AAU | 101000 | Asn | AGU | 111000 | Ser | U |
| | AUC | 001001 | Ile | ACC | 011001 | Thr | AAC | 101001 | Asn | AGC | 111001 | Ser | C |
| | AUA | 001010 | Ile | ACA | 011010 | Thr | AAA | 101010 | Lys | AGA | 111010 | Arg | A |
| | AUG | 001011 | Met | ACG | 011011 | Thr | AAG | 101011 | Lys | AGG | 111011 | Arg | G |
| G | GUU | 001100 | Val | GCU | 011100 | Ala | GAU | 101100 | Asp | GGU | 111100 | Gly | U |
| | GUC | 001101 | Val | GCC | 011101 | Ala | GAC | 101101 | Asp | GGC | 111101 | Gly | C |
| | GUA | 001110 | Val | GCA | 011110 | Ala | GAA | 101110 | Glu | GGA | 111110 | Gly | A |
| | GUG | 0001111 | Val | GCG | 011111 | Ala | GAG | 101111 | Glu | GGG | 111111 | Gly | G |

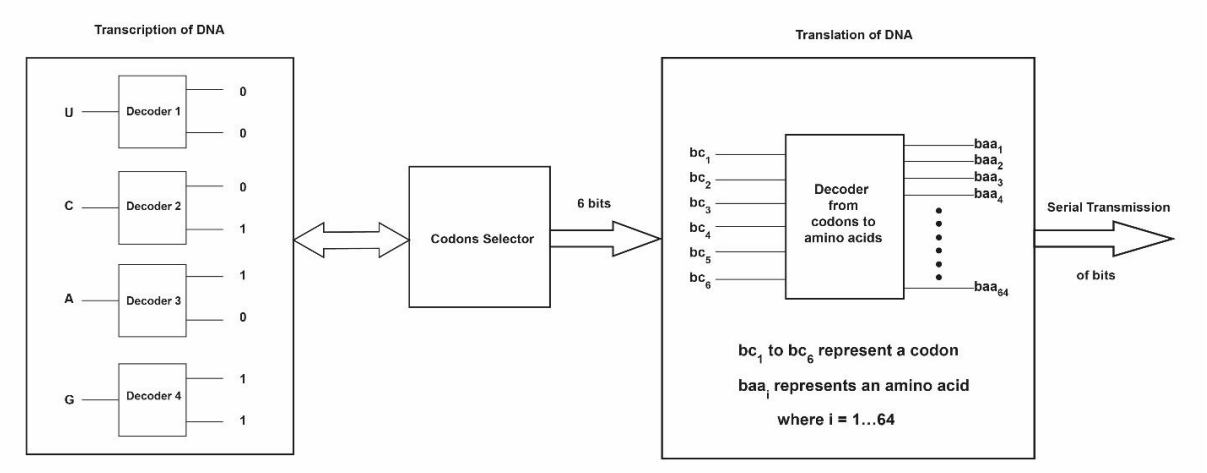


**Figure 3.12. Gene expression represented in bits**

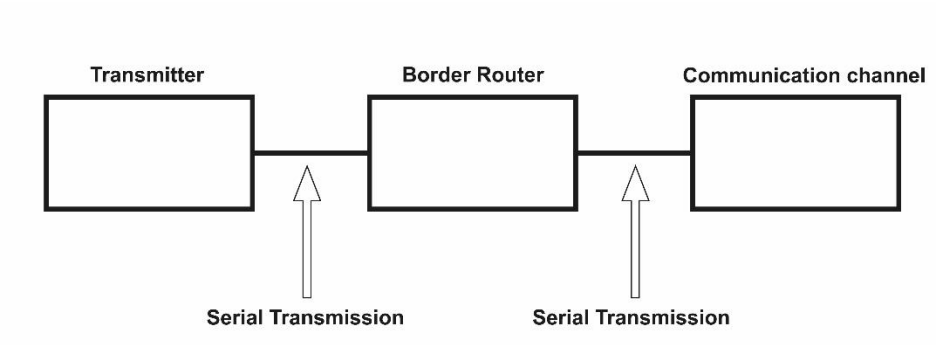


**Figure 3.13. Block diagram representing the serial transmission of information in gene expression**

### 3.4.5. Protein delivery through a communication channel according to Shannon's theorem

The mode of protein delivery to different destinations in the body is not the same in all cases and depends on the specific requirements of the system involved (e.g., the endocrine system). Hence, we consider cases in which the proteins secreted by the cell (e.g., hormones of a proteinaceous nature) move through the bloodstream (physical transmission medium-passive transport) to a target organ (address destination). This type of molecular communication is referred to as intercellular communication (i.e., distances in the range of mm to m) [1, 24]. Thus, Eq. 3.9 [54] and Figure 3.14 [2, 8, 55, 56, 57] characterize the movement of molecules in a fluid medium with drift (e.g., the bloodstream).

$$f(t) = \sqrt{\frac{\lambda}{2\pi t^3}}\, exp\left(-\lambda \frac{(t-\mu)^2}{2\mu^2 t}\right), for\ time\ (t) > 0 \tag{3.9}$$

where the mean is $\mu = d/v$, the shape parameter is $\lambda = d^2/2D$, the velocity of the fluid medium is $v \geq 0$, the diffusion coefficient is $D$, and the distance from the transmitter to the receiver is $d$.

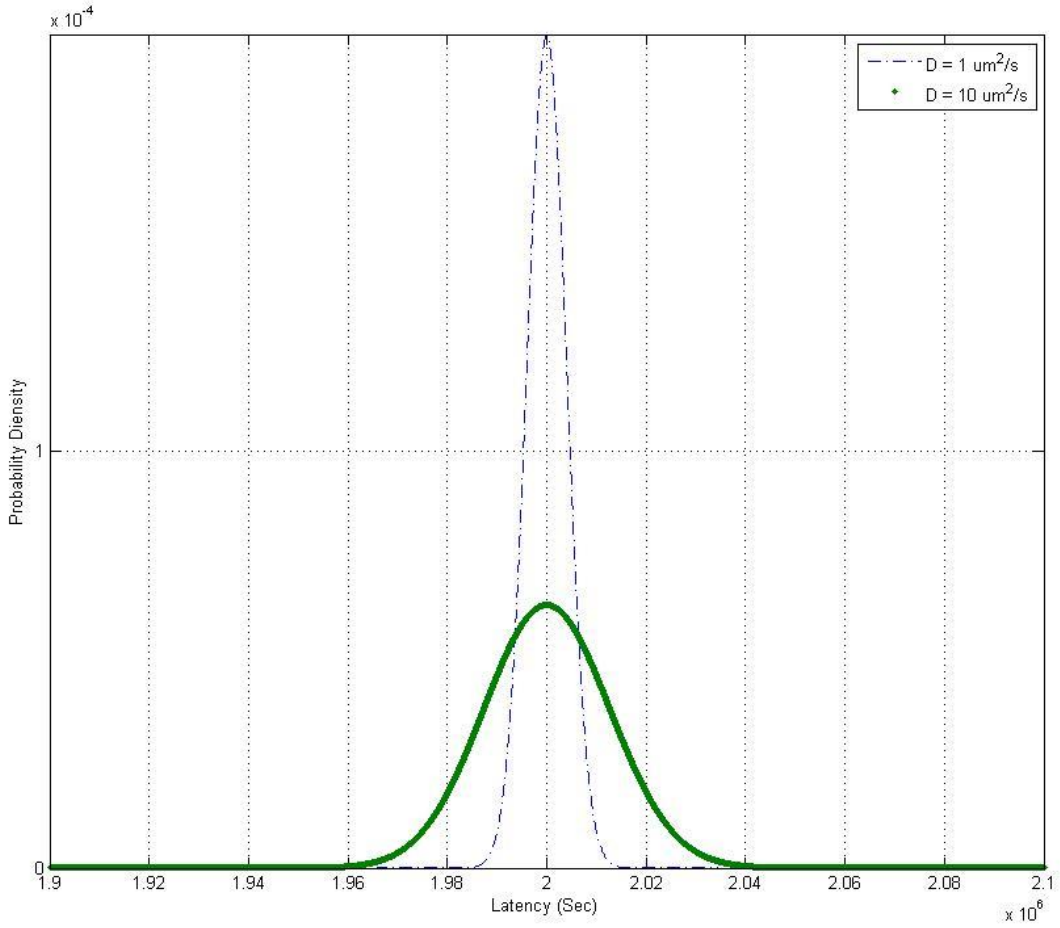


**Figure 3.14. Probability density function of the latency in a fluid medium at a velocity v = 5 cm/s, d = 1 m, $at\ D = 1\ \mu m^2/s$ ; $and\ v = 5\ cm/s, d = 1\ m\ and\ D = 10\ \mu m^2/s$**

As mentioned in Section 3.3, molecular systems and typical communication systems can encounter problems during the transmission of information via communication channels. Specifically, in molecular communication, these problems include biochemical, thermal and physical noise (due to the stochastic nature [54]); interference, which can be controlled by an appropriate transmission rate; and attenuation, which depends on the distance travelled and the physical characteristics of the fluid medium [1, 2]. The resulting damage to the signal information can cause latency (i.e., movement delay), which is expressed as $d/v$, and jitter (i.e., variation in latency) which mathematical expression is $D.d/2.v^3$, and the loss rate (i.e., the probability that a

molecule transmitted by a biological sender is not received by the intended biological receiver) can increase. The mathematical expression for the loss rate is $1 - \int_0^T f(t)dt$, which assumes that the receiver waits for the time duration $T$ [1,8].

In this biological scenario, similarly to conventional digital communication systems (e.g., network communication systems), digital communication theories could be used to overcome the transmission troubles in the channel. Therefore, applying the previously mentioned digital techniques to gene expression could convert the proposal in this article into a real solution for the analysed biological case. Hence, an advantage of using a stack of layers is the fundamental use of the data link layer to transform an imperfect channel into a line free of transmission errors or report unsolved problems to the upper layer [39]; therefore, the application of such a technique to biological systems (e.g., drug delivery) could provide high reliability [1].

Thus, to counter the errors in a noisy channel, we use Shannon's theorem to determine the capacity $C$ (in bits per second), i.e., the highest rate of information transmission from a transmitter to a receiver with respect to a given number of transmitted molecules [7, 13, 58]:

$$C = \max\{I(X;Y)\} \tag{3.10}$$

where $I(X; Y)$ is the Mutual Information (MI), and $X$ and $Y$ are the transmitted signal and the received signal, respectively (see Figure 3.3).

Figure 3.15 [59] shows the use of the Blahut-Arimoto algorithm to maximize the MI. Using this method, the transmitter is either allowed to send one molecule per slot, which is consistent with the serial transmission of information described in Section 3.4.4, or is not allowed to transmit information.

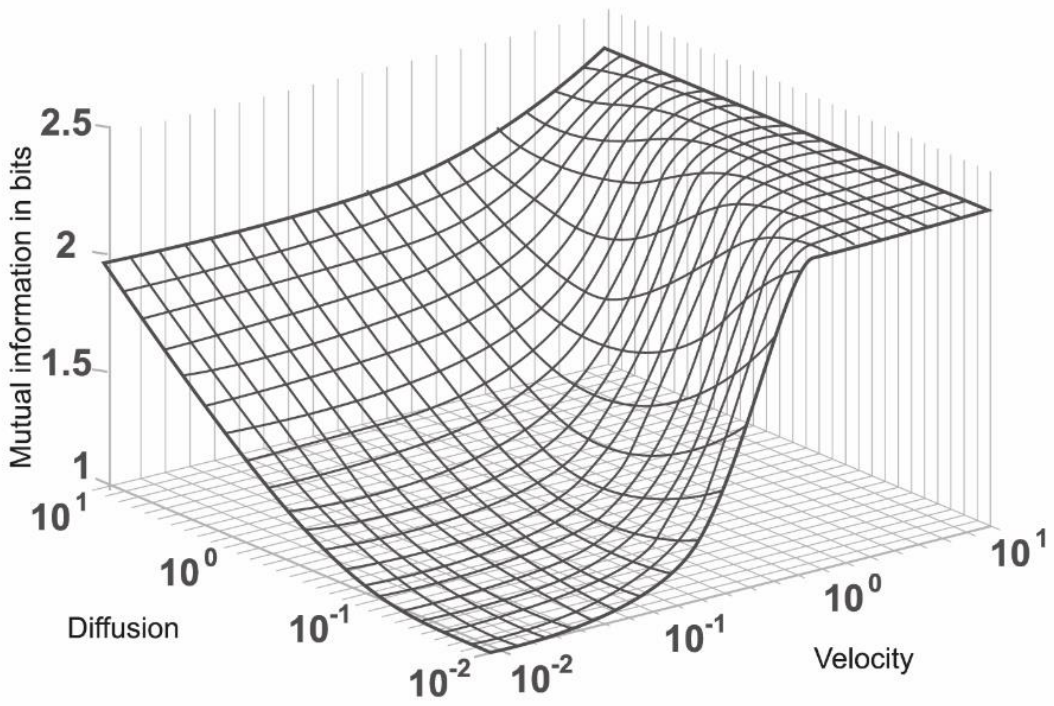


**Figure 3.15. Grid plot of the MI at various velocities and diffusion constants [56]**

Figure 3.15 shows the MI in bits with respect to the diffusion coefficient, which depends on the biochemical and physical properties of the fluid medium, and velocity (i.e., the drift in the transmission medium). The velocity-diffusion plot can be roughly divided into the following three regions [59]:

1. A diffusion-dominated region in which the MI is relatively insensitive to the velocity.
2. A high-velocity region in which the MI is insensitive to the diffusion coefficient.
3. An intermediate region in which the MI is highly sensitive to both the velocity and diffusion coefficient of the medium.

Therefore, the MI increases as the velocity increases and reaches a maximum value of $log_2(N)$, where $N$ represents the number of slots available for transmission. If a

transmitter is NOT allowed to transmit a molecule in any slot, the maximum MI is $log_2(N+1)$. At high velocities, the diffusion coefficient of the MI is insensitive to changes in the diffusion coefficient.

The diffusion coefficient is a measure of the uncertainty in the propagation time. Hence, we expect the MI to be lower with a high diffusion coefficient, which is the case at high velocities. However, surprisingly, a higher diffusion coefficient results in a higher MI at low velocities because at low velocities, the diffusion in the medium enhances the propagation of the molecule from the transmitter to the receiver. Thus, to improve the MI at low velocity and low diffusion coefficients, it is necessary to release multiple molecules at one time.

Using this method, transmitting various molecules (identical or different) simultaneously is possible due to the lineal channel properties [55, 60, 61, 62].

From a network perspective, the channel capacity is comparable to a flow control mechanism.

### 3.4.6. Digital processing of biological information at its destination

A target cell, tissue or organ performs a physiological function in the human body due to the communication of biological information from a transmitter to a receiver through the bloodstream (in the example used in our investigation). To send this information, the transmitter uses the data contained in the DNA molecules to identify the target cell, tissue or organ at the destination. Subsequently, the processing of information by a receiver depends on the type of proteinaceous hormone involved. Here, we briefly describe a case in which this processing is performed through ligands and their receptors.

Ligands acting as signals attach to receptors on the surface of a receiver. The receptors amplify and integrate the signals with input from other receptors and transmit the resulting signals to the destination cell. Thus, the communication architecture in the receiver possesses the following four features that characterize signal transduction [23]:

- Specificity. At the receiver end, the specificity determines the affinity between ligands and receptors. Specific ligands attach to specific (or complementary) receptors. Specificity indicates the precision with which a signal molecule fits

its complementary molecular receptor. From a network perspective, specificity could be understood as an addressing mechanism. This process is similar to the process that occurs in Ethernet networks when a broadcast (at the data link layer level) is sent to identify an Internet Protocol (IP) address; in this case, all devices in the network receive the broadcast, but the broadcast is processed only if the MAC corresponds to its IP address [1, 39].

- Amplification. Amplification is the increase in the strength of the signal. During amplification, the response to the first signal activates a second signal, the response to the second signal causes the activation of a third signal, etc. From a network perspective, this cell function is similar to the function of internetwork devices (i.e., switches, routers, etc.) that amplify the received signal before it is processed [63].
- Desensitization. Desensitization is the ability to remove the signal once received. The sensitivity of receptor systems is the function by which the signal is attenuated after a certain threshold period (i.e., the receiver becomes insensitive to the signal). After this threshold period, the receiver becomes sensitive again. This process may be comparable to the network case mentioned previously because when a computer receives a broadcast with an IP that does not match its own, the computer ignores the information, but if a new broadcast is sent, the computer processes it [1, 39].
- Integration. Integration is the ability to properly interpret and incorporate the received signal with other signals. Integration is defined as the ability of a system to receive multiple signals and produce a unified response appropriate to the needs of the cell or organism. In Sections 3.4.4 and 3.4.5, we examined how the transmission of biological data (represented by bits) is sent into the channel bit by bit separately due to the linear properties of the transmission channel. In the mentioned case, all internetwork device receivers are capable of processing various input signals simultaneously [39]. Thus, the simultaneous processing of received signals at the destination in digital communication systems and biological systems may be analogous.

Similarly, to digital communication systems, the receiving of biological

information through receptors can be described by discrete states, i.e., bound (*B*) or unbound (*U*). In the *U* state, the receptor waits for the arrival of molecular information; once the information arrives, the receptor enters the *B* state. In the *B* state, the receptor cannot receive other molecular information (rendering it insensitive to the signal), and a certain amount of processing time is required before the receptor returns to the *U* state [64].

This binding process can be represented by a discrete-time Markov chain using steps of length $\Delta t$ and a state transition probability matrix (at the $i$th step) as follows [64]:

$$P_i = \begin{bmatrix} 1 - \Delta t \,.\, r_{UB}.\, c(i\Delta t) & \Delta t.\, r_{UB}.\, c(i\Delta t) \\ \Delta t \,.\, r_{BU} & 1 - \Delta t \,.\, r_{BU} \end{bmatrix} \tag{3.11}$$

where $r_{UB}.\, c(i\Delta t)$ is the transition rate from *U* to *B*, which is proportional to the ligand concentration $c(i\Delta t)$, and the transition rate from *B* to *U* is $r_{BU}$, which is independent of the ligand concentration.

In Eq. 3.11, if $c_i = c(i\Delta t)$, $\alpha_{c_i} = \Delta t\; r_{UB}\; c(i\Delta t)$ and $\beta = \Delta t\; r_{BU}$, $P_i$ becomes

$$P_i = \begin{bmatrix} 1 - \alpha_{c_i} & \alpha_{c_i} \\ \beta & 1 - \beta \end{bmatrix} \tag{3.12}$$

where $\alpha_{c_i}$ is the transition probability from *U* to *B*, which is sensitive to the input ligand concentration $c_i$, and $\beta$ is the probability of transition from *B* to *U*, which is transition-insensitive. If we consider the extreme ligand concentration from the minimum allowed concentration $c_i = L$ to the maximum allowed concentration $c_i = H$, the signal transduction process is a time-inhomogeneous Markov chain according to either Eq. 3.13 or Eq. 3.14. The choice of Eq. 3.13 or Eq. 3.14 is determined by the ligand concentration [64].

$$P_H = \begin{bmatrix} 1 - \alpha_H & \alpha_H \\ \beta & 1 - \beta \end{bmatrix} \tag{3.13}$$

$$P_L = \begin{bmatrix} 1-\alpha_L & \alpha_L \\ \beta & 1-\beta \end{bmatrix} \quad (3.14)$$

In the described biological process, the membrane protein behaves as a transducer that decodes the received signal (DCE), triggering several reactions inside the target cell, tissue or organ of the body (DTE) [65]. This behaviour is comparable to the function performed at the receiver end in digital communication systems to process information that will be profitable at the destination.

As previously mentioned and according to the internetwork communication characteristics discussed throughout this article, the gene expression of proteins could be considered an internet in which the intracellular communication at the transmitter end emulates a LAN network in which the distance from the ER to the GA ($d_{LAN}$ in Figure 3.16) is 2 μm [2]. We do not consider in $d_{LAN}$ the distance between the ribosomes and ER because the ribosomes that participate in the translation process are found in the cytosol and RER; therefore, this distance is not fixed. However, if this distance could be considered $d_{LAN}$, it would still represent a biological LAN cover distance because it represents an intracellular distance [2].

Alternatively, the transmission of proteinaceous hormones could be defined as a WAN network (Figure 3.16) because it covers a distance $d_{WAN}$ (in the range of mm to m) [24, 55].

The receiver end also emulates a LAN network with a cover distance of up to the size of a cell (approximately 100 μm) [2].

Similarly, to typical networks, the biological WAN covers a larger distance than the biological LANs [66, 67]; In addition, similarly to conventional networks, the biological LANs are interconnected by routers [68] (Figure 3.16). Hence, at the transmitter end, the Golgi apparatus acts a biological router (Section 3.4.3). We consider that the receiver end, i.e., the target organ, also behaves as a router because among a set of receivers that are likely destinations connected by a common transmission channel (i.e., the bloodstream), the receivers for which the information is intended (i.e., with the suitable addressing) react to received information. Once again, the described case may be equivalent to the case of Ethernet networks in which a broadcast is sent through the transmission channel (with a bus logical topology) to each computer in the network, but

only the computer with the appropriate IP address processes the received communication [1, 69]. Therefore, the end-to-end addressing could be comparable to network addressing at the network layer [1].

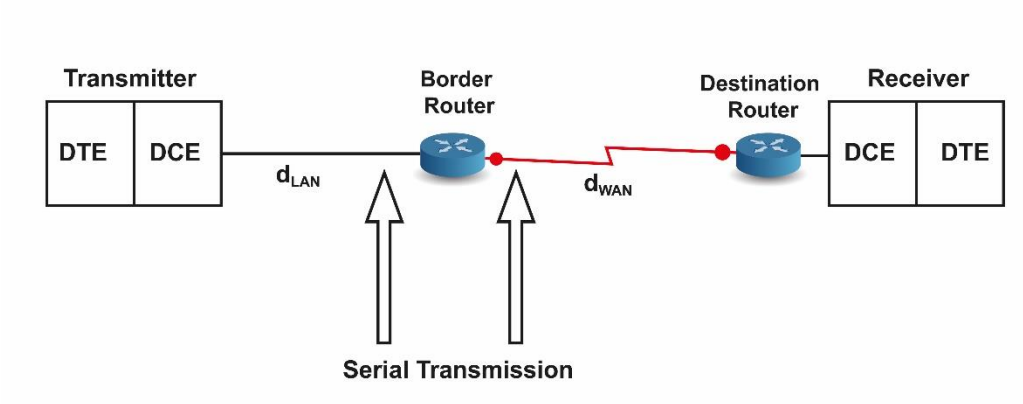


**Figure 3.16. Internetwork equivalence of gene expression**

Once the target cell receives the biological data, the information is communicated to other organelles using an implicit adjacent address that is comparable to the data link layer address used to facilitate communication within a direct range of communication [1, 43]. The received biological message is physically transmitted to the target cell. Figure 3.17 presents an end-to-end layered network model of the gene expression of proteins [1].

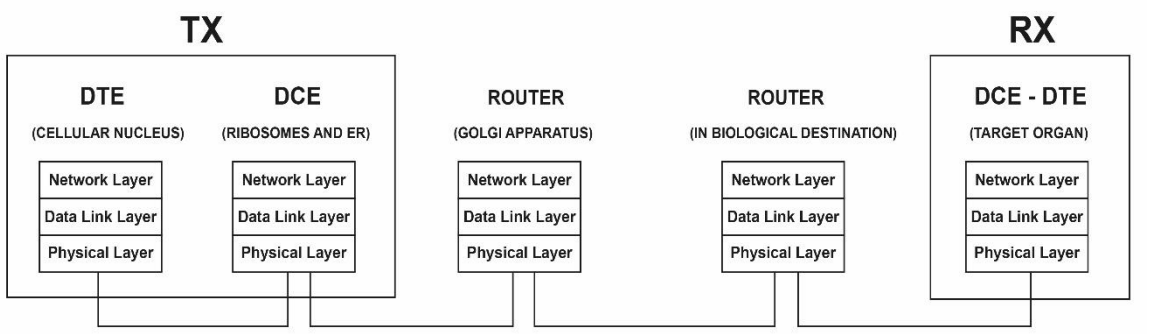


**Figure 3.17. Gene expression as a layered network model**

Figure 3.17 presents a layered network model from the typical network perspective

of gene expression as an internet. The structure of a layered model decomposes a large-scale system into a set of smaller units (i.e., layers) that are functionally independent of each other and specifies the interactions among the layers. One advantage of a stack of layers is the fundamental use of the data link layer to transform an imperfect channel into a line that is free of transmission errors or report unsolved problems to the upper layer [39]; therefore, the application of such a model to biological systems (e.g., drug delivery) is highly reliable [1].

### 3.5. Applications

1. The analysis of gene expression as a digital communication system specifically from a network perspective (in Sections 3.3 and 3.4) could be applied to bio-hybrid medical treatments that inject delicate and expensive drugs into the human body. In these cases, Eq. 3.10 describes the highest quantity of the drug (i.e., C) according to the bloodstream velocity ($v$) and the physical conditions of the blood (D) (Figure 3.15). The N slots in Eq. 3.10 could be considered periods during which the injected particles (i.e., bits) are transmitted. Similarly, in typical communications, if the transmission rate is $\leq C$, there is a high probability that the information will be received adequately at the destination (target organ) through a noisy channel (such as the bloodstream).

   Notably, the proposed applications may provide certain advantages in controlling drug arrival over the typical transport of drugs, in which the dosage cannot be adjusted after ingestion. Thus, characteristics, such as addressing and flow control, could be adjusted to deliver drugs to their desired locations at a controlled rate and dosage (depending on their toxicity level) while minimizing the effects of the drug on the healthy parts of the body. Hence, patients could be more comfortable, the risks related to medical errors could be reduced, and the cost-effectiveness of these drugs relative to that of conventional drugs could be increased. Moreover, the proposed application could benefit from applications of nanotechnology, which has enabled the development of microneedles, which are extremely small needles that enter the skin without stimulating the nerves [12, 70].

2. Hormone therapy management consists of the administration and adequate transmission of precise amounts of hormones to avoid considerable side effects during the treatment of diseases, such as cancer [13]. Thus, the mechanisms, such as addressing, flow control, error control and traffic control and Shannon's theorem that were used to design the layered network model of the transmission of proteinaceous hormones in the endocrine system to a target organ (Section 3.4) may be used in hormone therapy because these network theories could act only in certain target organs (i.e., destination addressing), provide the appropriate dose (i.e., flow control, error control and traffic control and Shannon's theorem (expressed in Eq. 3.10)) and avoid side effects due to a higher targeting accuracy (i.e., addressing).
3. The proposed analogies between biological systems (in regard to the gene expression of proteins) and digital communication systems (specifically, internetwork systems) could be used in bio-hybrid applications in which the benefits and characteristics of both systems are applied. For instance, the authors in [55] define a molecular packet as having a structure that is very similar to an Internet packet [5], i.e., the molecular packet is composed of a header and a payload (Figure 3.18). The basic idea proposed in [55] is inspired by the Internet Protocol (IP) in which an IP datagram is divided into smaller fragments to ensure that the fragments can pass through underlying communication links. Thus, a large molecular message is segmented into a set of smaller packets to facilitate their diffusion in a transmission drift channel to reach a biological destination.

   The application of the network theories proposed in [55] is comparable to that discussed in Section 3.4.1, in which we mentioned that the biological information in DNA is divided into data segments (smaller units) during transcription to facilitate processing; this process is analogous to the process that occurs in packet-switching networks. Furthermore, the linear properties of the channel (Section 3.4.5) are also used in [55] to transmit various molecular packets. At the destination, the receiver processes a set of fragments and reassembles the molecular message [55] using features that are comparable to the integration properties of the receiver discussed in Section 3.4.6.

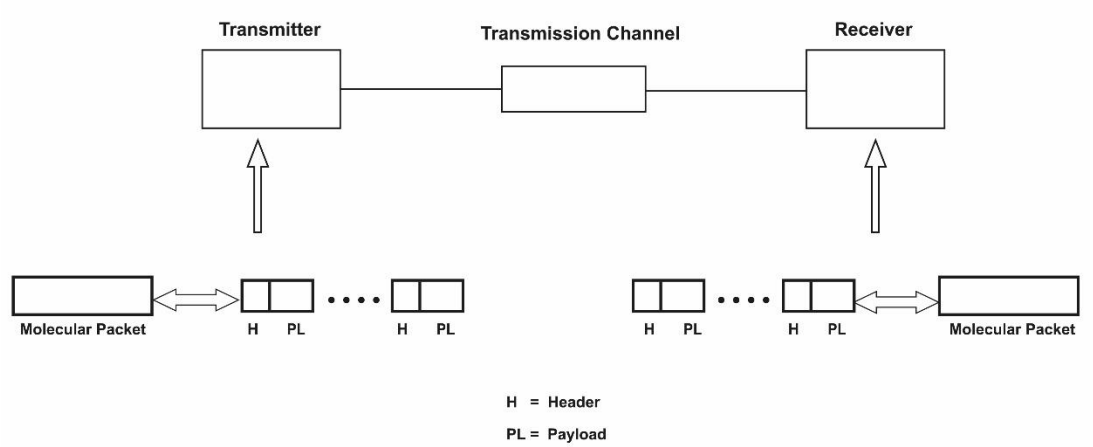


**Figure 3.18. Molecular packet fragmentation and reassembly**

## 3.6. Conclusions

We analysed gene expression from a digital communication systems perspective. Specifically, we used network theories, such as addressing, flow control, error control, traffic control and Shannon's theorem, to design layered network models that represent the transcription and translation of DNA and the end-to-end transmission of proteins to a target organ. These models apply the advantageous features of digital communication systems to the transmission of biological information in gene expression systems. Additionally, these models help us improve our understanding of biological communication systems. Further, the proposed analysis establishes the duality between digital and biological communication systems and proposes solutions to overcome the challenges that both systems face. One of the most important applications of this analysis could be the development of communication systems for medical purposes (i.e., in the treatment of diseases, such as cancer). The above-mentioned analysis may, for instance, avoid side effects through the transmission of information only to a suitable destination (i.e., to specific target organs), in order to achieve optimal treatments. In general, such an approach improves the quality of the end-to-end communication in both types of systems because the failures in the transmission media that occur in biological and digital communication systems drastically degrade the information that is transmitted. One way to overcome these challenges may be to use a layered network model in which the

fundamental task of the data link layer is to transform an imperfect channel into a line free of transmission errors or report unsolved problems to the upper layer (or to the practitioner who controls the medical process); therefore, the application of these methods to biological systems (e.g., drug delivery) could be highly reliable.

### 3.7. List of abbreviations

| | |
|---|---|
| A | Adenine |
| B | Bounded sites |
| *C* | Capacity of a noisy channel |
| C | Cytosine |
| DCE | Data Communication Equipment |
| DNA | Deoxyribonucleic Acid |
| DTE | Data Terminal Equipment |
| ER | Endoplasmic Reticulum |
| GA | Golgi Apparatus |
| G | Guanine |
| HIV | Human Immunodeficiency Virus |
| IoBNT | Internet of Bio-Nano Things |
| ISI | Inter Symbol Interference |
| MI | Mutual Information |
| mRNA | messenger Ribonucleic Acid |
| PDUs | Protocol Data Units |
| PY | Pyrimidine bases |
| R | Purine bases |
| RNA | Ribonucleic Acid |
| RNAP II | RNA Polymerase II |
| RER | Rough Endoplasmic Reticulum |
| S | Strong bases |
| SRP | Signal Recognition Particle |
| T | Thymine |
| tRNA | transfer Ribonucleic Acid |
| U | Uracil |
| *U* | Unbounded sites |
| W | Weak bases |
| *X* | transmitted signal |

| | |
|---|---|
| $Y$ | received signal |

### 3.8. References


1. Cevallos Y, Molina L, Santillán A, De Rango F, Rushdi A, Alonso JB. A digital communication analysis of gene expression of proteins in biological systems: A layered network model view. Cognit Comput 2017; 9(1): 43-67.
2. Nakano T, Eckford AW, Haraguchi T. Molecular communication. New York, NY: Cambridge University Press 2013; pp. 16- 20, 31, 34, 36-39, 52-67, 74.
3. Farsad N, Yilmaz HB, Eckford A, Chae CB, Guo W. A comprehensive survey of recent advancements in molecular communication. IEEE Communications Surveys & Tutorials 2016; 18(3): 1887-919.
4. Dong M, Li W, Xu X. Evaluation and modeling of HIV based on communication theory in biological systems. Infect Genet Evol 2016; 46: 241-7.
5. Comer, D. Computer Networks and Internets, Global Edition. Pearson Education Limited. 2016; pp. 119-124, 359-362, 567-576.
6. Wu L, Shen Y, Li M, Wu FX. Network output controllability-based method for drug target identification. IEEE Transactions on NanoBioscience 2015; 14(2): 184-91.
7. Mustam SM, Syed-Yusof SK, Zubair S. Capacity and delay spread in multilayer diffusion-based molecular communication (DBMC) channel. IEEE Transactions on NanoBioscience 2016; 15(7): 599-612.
8. Nakano T, Moore MJ, Wei F, Vasilakos AV, Shuai J. Molecular communication and networking: opportunities and challenges. IEEE Transactions on NanoBioscience 2012; 11(2): 135-48.
9. Haughton D, Balado F. BioCode: Two biologically compatible algorithms for embedding data in non-coding and coding regions of DNA. BMC Bioinf 2013; 14: 121.
10. May EE. Towards a biological coding theory discipline. New Thesis 2004; 1(1): 19-38.
11. Faria LC, Rocha AS, Palazzo R. Transmission of intra-cellular genetic information: A system proposal. J Theo Biol 2014; 358: 208-31.
12. Chahibi, Youssef. Molecular communication for drug delivery systems: A survey. Nano Communication Networks. Elsevier. 2017.

**13.** Akan, Ozgur B and Ramezani, Hamideh and Khan, Tooba and Abbasi, Naveed A and Kuscu, Murat. Fundamentals of molecular information and communication science. Proceedings of the IEEE. IEEE. 2017;105(2):306-318.

**14.** Iwasaki, Satoru and Nakano, Tadashi. Graph-based Modeling of Mobile Molecular Communication Systems. IEEE Communications Letters. 2017. Available from: http://ieeexplore.ieee.org/document/8080203/.

**15.** Nakano, Tadashi and Okaie, Yutaka and Kobayashi, Shouhei and Koujin, Takako and Chan, Chen-Hao and Hsu, et al. Performance Evaluation of Leader--Follower-Based Mobile Molecular Communication Networks for Target Detection Applications. IEEE Transactions on Communications. 2017;65(2):663-676. Available from: http://ieeexplore.ieee.org/document/7742377/.

**16.** Raut P, Sarwade N. Study of environmental effects on the connectivity of molecular communication based Internet of nano things: Proceedings of the International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET); 2016 Mar 23-25; Chennai: IEEE 2016.

**17.** Acton QA. Cancer gene therapy: New insights for the healthcare professional: 2012 edition: Scholarlybrief, Atlanta, Georgia: Scholarly Editions 2012; pp. 91.

**18.** Ravi D, Wong C, Deligianni F, *et al.* Deep learning for health informatics. IEEE Journal of Biomedical and Health Informatics 2017; 21(1): 4-21.

**19.** Abbasi QH, Yang K, Chopra N, *et al.* Nano-communication for biomedical applications: A review on the state-of-the-art from physical layers to novel networking concepts. IEEE Access 2016; 4: 3920-35.

**20.** Singh RP, Sapre SD. Communication systems. New York: Tata McGraw-Hill Education 2008; pp. 364.

**21.** Abu-Rgheff MA. Introduction to CDMA wireless communications. Cambridge: Academic Press 2007; pp. 51

**22.** Morris DJ. Communication for command and control systems: International series on systems and control. Amsterdam: Elsevier Science 2014; pp. 98–100.

**23.** Bush SF. Nanoscale communication networks. Norwood: Artech House 2010; pp. 53, 57, 67, 69, 70, 194, 195, 234.

**24.** Akyildiz IF, Pierobon M, Balasubramaniam S, Koucheryavy Y. The internet of bio-nano things. IEEE Communications Magazine 2015; 53(3): 32-40.

**25.** Shrivastava S, Badlani R. Data storage in DNA. International Journal of Electrical Energy 2014; 2(2): 119-24.

**26.** Bettelheim F, Brown W, Campbell M, Farrell S, Torres O. Introduction to general, organic and biochemistry. Boston: Cengage Learning 2015; pp. 614-640.

**27.** Lieberman M, Marks AD. Marks' basic medical biochemistry: A clinical approach. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins 2009; pp.206.

**28.** Alberts B, Johnson A, Lewis J, *et al.* Molecular biology of the cell. New York: Garland Science 2014; pp. 334-60.

**29.** Farsad N, Yilmaz HB, Eckford A, Chae CB, Guo W. A comprehensive survey of recent advancements in molecular communication. IEEE Communications Surveys \& Tutorials 2016. 18 (3): 1887-1919.

**30.** Alberts B, Bray D, Hopkin K, *et al.* Essential cell biology. New York: Garland Science 2013; pp. 185–7, 383–407, 487–513.

**31.** Goodman SR. Medical cell biology. Amsterdam: Elsevier; 2007. pp. 175. https://books.google.es/books?id=tRbCHk9easQC.

**32.** Starr C, McMillan B. Human biology. Belmont: Cengage Learning 2013; pp. 412-3.

**33.** Sneppen, K. Models of Life: Dynamics and Regulation in Biological Systems. Cambridge University Press. 2014; pp. 24.

**34.** Zhang C, Liu HH, Zheng KW, Hao YH, Tan Z. DNA G-quadruplex formation in response to remote downstream transcription activity: Long-range sensing and signal transducing in DNA double helix. Nucleic Acids Res 2013; 41(14): 7144-52.

**35.** Jorde LB, Carey JC, Bamshad MJ. Medical genetics. Amsterdam: Elsevier; 2015. pp. 6–11. https://books.google.com.ec/books?id=FrRgCgAAQBAJ.

**36.** Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson genetics in medicine. Amsterdam: Elsevier Health Sciences 2015; pp. 23.

**37.** Yockey, H.P. Information Theory, Evolution, and the Origin of Life. Cambridge University Press. 2005; pp 26-28, 108-109.

**38.** Solomon E, Martin C, Martin DW, Berg LR. Biology. Stamford: Cengage Learning 2014; pp. 305-7.

**39.** Tanenbaum AS, Wetherall DJ. Computer Networks. New York: Pearson Education 2012; pp. 31, 43, 45, 46, 194, 197, 200, 202, 216, 257-298, 340, 356-358.

**40.** Nakano, Tadashi and Okaie, Yutaka and Vasilakos, Athanasios V. Transmission rate control for molecular communication among biological nanomachines. IEEE Journal on Selected Areas in Communications. 2013;31(12):835-846.

**41.** Papachristodoulou D, Snape A, Elliott WH, Elliott DC. Biochemistry and molecular biology. Oxford: Oxford University Press 2014; pp. 377.

**42.** Snustad DP, Simmons MJ. Principles of genetics, binder ready version. Hoboken: Wiley 2015; pp. 267-74. https://books.google.com.ec/books?id=NBB0CgAAQBAJ.

**43.** Nakano T, Suda T, Okaie Y, Moore MJ, Vasilakos AV. Molecular communication among biological nanomachines: A layered architecture and research issues. IEEE Transactions on NanoBioscience 2014; 13(3): 169-97Reece JB, Meyers N, Urry LA, Cain ML, Wasserman SA, Minorsky PV. Campbell biology Australian and New Zealand edition. Melbourne: Pearson Higher Education AU 2014; pp. 340-4.

**44.** Reece JB, Meyers N, Urry LA, Cain ML, Wasserman SA, Minorsky PV. Campbell biology Australian and New Zealand edition. Melbourne: Pearson Higher Education AU 2014; pp. 340-4. https://books.google.es/books?id=5t6aBQAAQBAJ.

**45.** Holmes R. Salters-nuffield advanced biology as student book. Oxford: Heinemann 2005; pp. 229.

**46.** Fitzgerald-Hayes M, Reichsman F. DNA and biotechnology. Cambridge: Academic Press 2009; pp. 52-74.

**47.** De Silva, Pavani Yashodha and Ganegoda, Gamage Upeksha. New Trends of Digital Data Storage in DNA. BioMed Research International Journal. Hindawi Publishing Corporation. 2016.

**48.** Jackson ARW, Jackson JM. Environmental science: The natural environment and human impact. New York: Pearson Education 2000; pp. 139-40. https://books.google.co.uk/books?id=WLYCdqaHj1MC.

**49.** Wei ZS, Yang JY, Shen HB, Yu DJ. A cascade random forests algorithm for predicting protein-protein interaction sites. IEEE Transactions on NanoBioscience 2015; 14(7): 746-60.

**50.** Lodish B, Kaiser K, Bretscher P, Amon, Scott. Molecular cell biology. New York, NY: WH Freeman and Company 2013; pp. 279-281.

**51.** Konieczny, L. and Roterman-Konieczna, I. and Spòlnik, P. Systems Biology: Functional Strategies of Living Organisms. Springer International Publishing. 2013; pp. 17-18.

**52.** Krebs, J.E. and Goldstein, E.S. and Kilpatrick, S.T. Lewin's GENES XII. Jones & Bartlett Learning 2017; pp. 114, 127, 142.

**53.** Nemzer, Louis R. A binary representation of the genetic code. Biosystems Journal. Elsevier. 2017; 155: 10-19.

**54.** Haselmayr W, Aejaz SMH, Asyhari AT, Springer A, Guo W. On the impact of transposition errors in diffusion-based channels. 2017. arXiv preprint arXiv:1701.02971.

**55.** Furubayashi, Taro and Nakano, Tadashi and Eckford, Andrew and Okaie, Yutaka and Yomo, Tetsuya. Packet fragmentation and reassembly in molecular communication. IEEE transactions on nanobioscience Journal. IEEE. 2016;15(3): 284-288.

**56.** Bronzino, J.D. and Peterson, D.R. Tissue Engineering and Artificial Organs. The Biomedical Engineering Handbook, Fourth Edition. CRC Press. 2016; pp 66-69.

**57.** Mustam, Saizalmursidi Md and Syed-Yusof, Sharifah K and Zubair, Suleiman. Capacity and Delay Spread in Multilayer Diffusion-Based Molecular Communication (DBMC) Channel. IEEE transactions on nanobioscience Journal. 2016; 15(7):599-612.

**58.** Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948; 27(3): 379-423, 623–56.

**59.** Raut P, Sarwade N. Establishing a molecular communication channel for nano networks. International Journal of VLSI Design & Communication Systems 2013; 4(2): 27.

**60.** Shah Mohammadian, Hoda and Messier, Geoffrey G and Magierowski, Sebastian. Nano-machine molecular communication over a moving propagation medium. Journal Nano Communication Networks 2013;4(3): 142-153.

**61.** Mosayebi, Reza and Arjmandi, Hamidreza and Gohari, Amin and Nasiri-Kenari, Masoumeh and Mitra, Urbashi. Receivers for diffusion-based molecular

communication: Exploiting memory and sampling rate. IEEE Journal on Selected Areas in Communications. 2014; 32(12):2368-2380.

**62.** Eckford, Andrew W. Nanoscale communication with Brownian motion. Information Sciences and Systems, 2007 Mar 160-165. CISS'07. 41st Annual.

**63.** Uhlig, T. and Sellmaier, F. and Schmidhuber, M. Spacecraft Operations. Springer Vienna. 2014; pp 313.

**64.** Eckford AW, Thomas PJ. Information theory of intercellular signal transduction: Proceedings of the 49th Asilomar Conference on Signals, Systems and Computers; 2015 Nov 8-11 119-122; Pacific Grove, CA: IEEE 2015.

**65.** Lister AL, Van Der Kraak GJ. Endocrine disruption: Why is it so complicated? Water Qual Res J Can 2001; 36(2): 175–90.

**66.** Wu, C.H. and Irwin, J.D. Introduction to Computer Networks and Cybersecurity. CRC Press. 2016; pp. 269.

**67.** Bidgoli, H. MIS. Cengage Learning. 2015; pp. 122.

**68.** Eric Coll, M.E. Telecom 101: Fourth Edition - 2016. High-Quality Reference Book and Study Guide Covering All Major Topics, Up To Date To 2016... in Plain English. 2016; pp. 215.

**69.** Lammle, T. CCENT ICND1 Study Guide: Exam 100-105. Wiley. 2016; pp. 27-29.

**70.** Chahibi, Youssef and Akyildiz, Ian F. Molecular communication noise and capacity analysis for particulate drug delivery systems. IEEE Transactions on Communications Journal. IEEE. 2014; 62(11): 3891-3903.

# Chapter 4: Conclusions

1. Because biological systems and digital systems have fundamentally the same components (i.e., transmitter, communication channel and receiver) for the transmission of information, their communication behaviours are similar. To delineate these specific similarities, it is possible to define equivalence models that overcome the limitations of each system and apply the best communication strategies of one system in the other system, as was performed in this work. By using internetwork system theories, this work presented a layered network model that represents gene expression. Using theories of digital communication systems (including Shannon's theorem) and the aforementioned layered network model, this work presents an end-to-end digital communication system model that represents the gene expression of proteins.

2. Bio-hybrid nanocommunication systems currently represent an ideal solution for medical therapies due to their biocompatibility, which avoids the immune response (which can reject the drugs) and does not stimulate nerves, thus providing an effective treatment. In addition, their small size makes them suitable for accessing delicate body sites.

3. The information management methods that are characteristic of internetwork systems and digital communication systems confer more reliability with regard to the transmission/processing of information at each stage of the transcription and translation of DNA and in the communication of information to a target organ. Thus, these methods are appropriate for the treatment of genetic disorders.

4. The Addressing of information used in internetwork systems and digital communication systems applied to gene expression systems is fundamental to directing information and detecting a suitable destination (i.e., target organ). In this way, it is possible to avoid side effects that could be lethal, especially when

administering specific drugs. This is a very important aspect that must be kept in mind during the treatment of diseases due to the physical, psychological and social repercussions of patients.

5. Gene expression systems are one of the most important systems in biological systems because they define the development of a multicellular organism. The processing of information in a biological receiver establishes this development. Hence, when a target organ cannot process the biological data (due an overwhelming condition during the reception of information), the development of pathological abnormalities may occur. Therefore, a typical mechanism in internetwork systems and digital communication systems for regulation regarding the transmission/reception of information is the flow control of data from the transmitter to the receiver, which also provides a solution for avoiding these abnormalities.

6. A future study will present a nanonetwork simulation of the layered network model proposed in this thesis according to the IEEE 1906.1 draft standard.