tripleC cognition
communication
co-operation

Intracellular management of information: from DNA to proteins

Juan M. Lara

Instituto de Neurociencias de Castilla y León, c/ Pintor Fernando Gallego 1, 37007 Salamanca. Email: juan.lara@incyl.org. Web: http://www.incyl.org

Abstract: *The living features of cells constitute an information flow from a central database, the nuclear DNA, to molecular effectors, proteins, which are synthesised in the cytoplasm. This flow comprises two changes in information units: transcription and translation. Transcription is carried out in the nucleus and consists in the transduction of specific instructions from the DNA to an intermediary, the RNA, which in its canonical form is a messenger (mRNA) that takes information from the cell nucleus. In the cytoplasm, information in the messenger gives the commands for a specialised molecular complex to build a specific protein. After regulation of the form, proteins are integrated as structural and/or functional units in the complex network of biological functions in the cell.*

Keywords: Cell; nucleotide; codon; amino acid; protein

Life is an efficient flow of information which, through sequential transductions, contains the entropy in levels apt for the functionality, growth, reproduction and evolution of the cells. Thus, in accordance with the general scientific consensus, but, nonetheless, debated, the cell is the minimum unit of life.
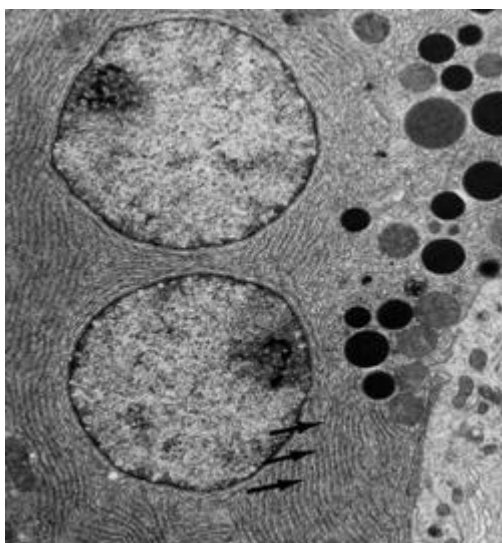

Figure 1: Electron microscopy micrograph showing two secretory cells with similar nuclear and cytoplasmic content. Arrows indicate structures specialised in synthesising the dense granules' content (zymogens).

To sketch how I interpret this flow of information I will take, as a model, a pair of eukaryotic cells (with true nucleus), specifically two adult and fully functional cells of the exocrine pancreas of any mammal, those which produce a mixture of digestive enzymes denominated zymogen (Cross and Mercer, 1993) (Figure 1).

Specialised formation is unnecessary to appreciate that both cells are very similar, almost identical. They have similar sizes, their nuclei are very similar in size and appearance, the same as the remainder of the cell (the cytoplasm) which, moreover, is laden with dark structures that are accumulations of zymogen prepared to be secreted to the exterior. I have chosen these cells because I believe they demonstrate the objective efficiency and similar results of the flow of biological information in two independent, although related, cells: both cells are somatic from the same animal, thus their genetic heritage is identical. Furthermore, probably both cells originate from the division of a single cell, which

makes the synchronicity in their development and their relationship with the medium similar. They are "twin" cells which on developing a programme of similar instructions generate almost identical forms and functions, or, equally, starting from the same genetic potential, develop this through a sequence of transductions to distinct supports that result in equivalent structures, forms and functions. In the same animal, any other somatic cell (e.g. muscular, adipose or neural), including other cells of the same organ, in spite of having the same "book of instructions", interprets it in such a way that its form, structure and function makes it completely different (Figure 2).
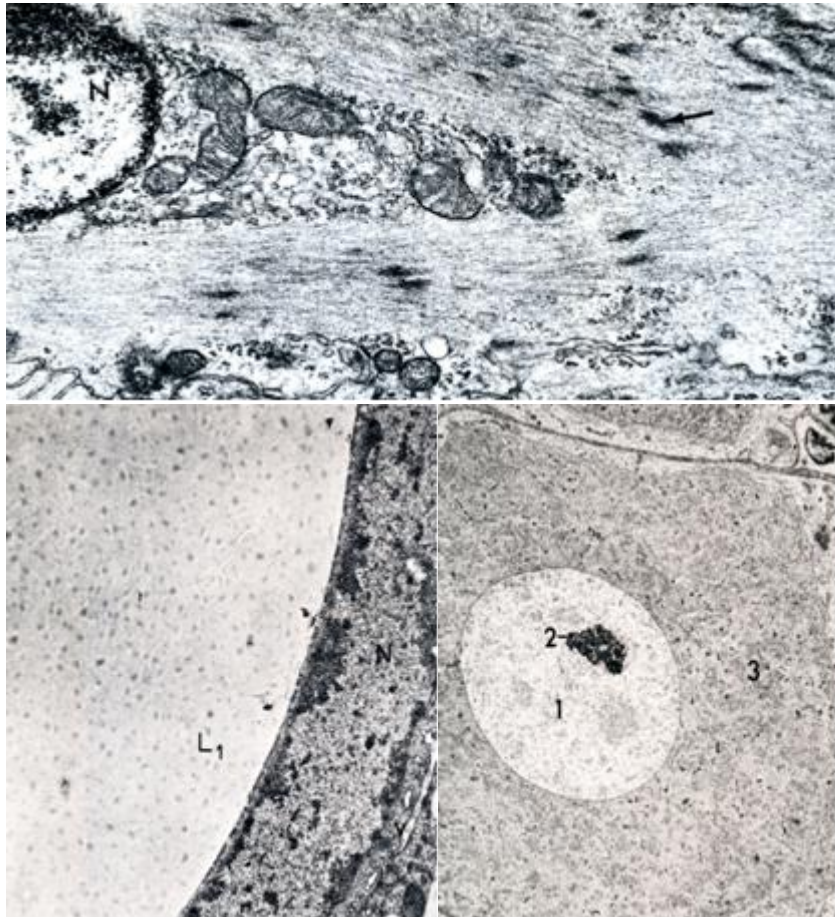

Figure 2: Electron microscopy micrographs showing a muscle cell (above), an adipocyte or a fatty cell (inferior left corner) and a neuron (inferior right corner) (Weiss and Roy, 1982) from the same species as the cells in figure 1. In the muscle cell, N indicates the nucleus and the arrow in the cytoplasm indicates dense structures involved in contraction. In the adipocyte, N indicates the nucleus and L1 the main lipid droplet that almost occupies the entire cytoplasm. In the neuron, number 1 shows a very electron-lucent nucleus, where an electron-dense structure (2) stands out, which is the nucleolus.

As a first approximation, it seems scarcely probable that two cells, so similar are so by chance, more so by the ordered execution, and in parallel, of a series of instructions in a complex flow of information.

Although "we usually believe that everything unknown can be explained from what is already known" (Chudnovsky et al., 2008), it is evident that we lack both the knowledge and the tools to describe, quantify and understand in its details this flow of information within the cells. Nevertheless, I will sketch what we know. I shall deal with the processing of information within a cell (intracellular) considering its two basic compartments, the nucleus and cytoplasm.

1. The intracellular information

The eukaryotic cell, the most complex of those known, is the structural and functional

base of yeasts, fungi, plants and animals. It is composed of two principal compartments: the nucleus and cytoplasm. The separation of both parts of the cell is marked by a physical frontier, the nuclear envelope, which separates, spatially and temporally different stages of the flow of information which makes the life of the cell possible. The most internal component is the nucleus, which not only contains the principal data base of the cell, the nuclear DNA, but also contains and equipment of elements that administer this information, regulating functionally and temporally their activation or blockage. It also activates or blocks, in accordance with a programme contained in the DNA itself, the duplication of the information (replication) and the transduction of the content of specific

zones of the DNA (genes) to a messenger (mRNA) in a process denominated transcription (Figure 3). Still within the nucleus, the information contained in this new support can mature through regulated alternative processes which permit the generation of diverse messengers from the transcription of a single gene (Figure 3).

The messenger is the support of information which, once mature, reaches the cytoplasm crossing the nuclear envelope by very selective complex molecules (pore complexes). Once in the cytoplasm the messenger is submitted to another transduction with the participation of complex molecular machinery, the ribosomes, which translate the information of the messenger to proteins (Figure 3).
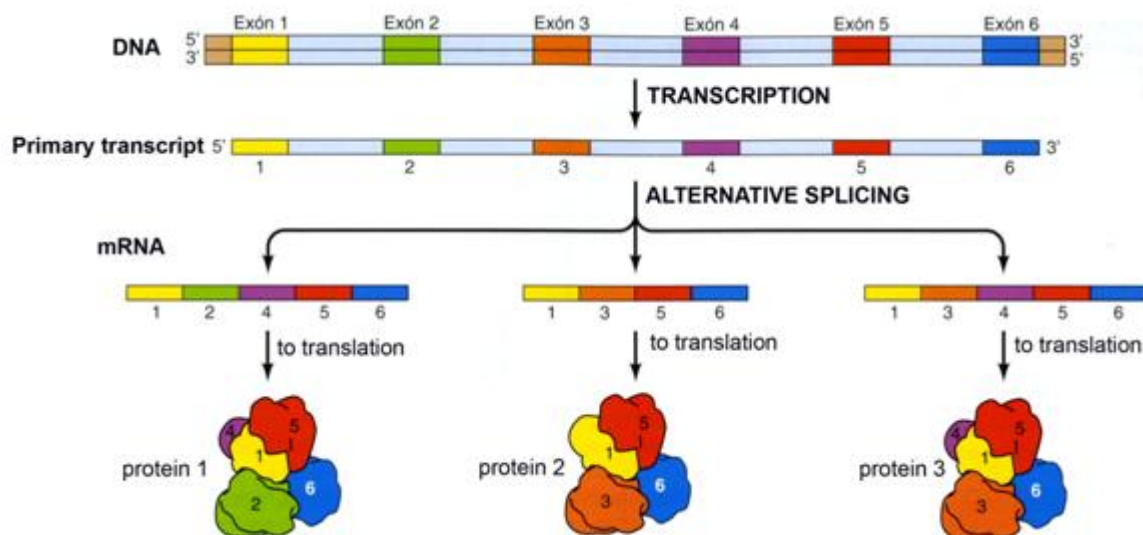

Figure 3: Scheme showing the two main information transduction processes in the eukaryotic cell (from Cooper and Hausman, 2008). Transcription takes place in the nucleus and transduction is done in the cytoplasm. In the space-time period between the two processes, the information that the messenger contains could be modified by an alternative processing that, starting from the same immature messenger, can generate various mature messengers that will be translated into different proteins.

This new support contains diverse types of codification, as well as being the executor of specific instructions (for example determined enzymatic activity), it has information about when, where and until when to execute these instructions. Thus because of the characteristics of the sequence itself of the components of the proteins, the amino acids, or through the participation of external regulators (chaperones and chaperonins) the new proteins acquire a spatial organisation

that mediates the activation or non-activation of specific portions (domains) so that they perform determined functions, which can also be regulated by agents external to the protein in question. Together with functional potentialities, the sequence of amino acids and their spatial organisation contain codes of localisation for many proteins, which facilitate its ordered distribution in the cell. What is more, each protein contains little known information which, in combination with

external agents, determines when they cease to be functional and when they are recycled.

Finally, probably the most obscure step of the flow of information in the cytoplasm is the integration of each protein in the complex functional mesh, in which one protein is usually involved in several pathways or cascades of biochemical reactions, occasionally of opposed significance, at least apparently. The regulation of this molecular network seems to be based on subtle structural modifications of the proteins originated by the interaction with other proteins or with other components of the cellular medium (e.g. ions).

It seems evident that the life of the cell and, therefore, life in a wider sense, requires that the metabolic pathways are executed correctly, when and where they are necessary. We discuss this because these exquisite coordinated chains of reactions are occurring in billions of cells in every single one of us.

1.1. Processing of information in the cellular nucleus

Since the nineteenth century there has been sufficient evidence to consider that the cellular nucleus in some way contained hereditary information, but eight decades passed from when nuclein was isolated (Miescher, 1871) until the structure of DNA was known and the genetic code began to be understood (Watson and Crick, 1953). The principal difficulty stemmed from the fact that DNA is, in its entirety, an enormous molecule ($2.9x10^9$ pairs of nucleotides, according to Lodish et al., 2007), but formed by only four components (nucleotides: adenine, cytosine, guanine and thymine) which are repeated thousands of millions of times in an apparently random manner. It was not easy to know how the information was coded, nor how it was transmitted to the daughter cells. In the second half of the twentieth century it was deduced that the complementary pairs of nucleotides (adenine-thymine, guanine-cytosine) (Figure 4) allows the exact replication of a DNA molecule, and if we consider the nucleotides as letters (by convention A, C, G, T) we have a limited alphabet available.

The support of the basis of genetic data is, therefore, a very simple system. Four letters whose order, far from being random, compose an immense library contained in a specific number of tomes (chromosomes[1]) characteristic of each eukaryote species. At this level, the unit of information is the nucleotide which, as a letter of our alphabet or each of the symbols of a binary code, is objectively real, but lacks information *per se* (Figure 5).

Only one part of the DNA is susceptible to being translated to a second support (and not in a simultaneous manner), whereas the greater part of the sequence of nucleotides participates in the administration of this information regulating the duplication of the DNA (replication) or determining, among other things, which portions (genes) are transcribed to other nucleic acid (RNA) (Figure 5) and when (Lewin, 2008).

In addition, in the strict sense, not all the sequence of nucleotides of a gene is transcribed and of this some portions will be eliminated (introns) and only some others (exons) can form part of the second basic support of information (mRNA or messenger RNA) (Figure 3). Moreover, still in the nucleus, some of these exons can be eliminated by a complex process of maturation by cut and splice from which alternative mRNA molecules can result (Figure 3).

---

[1] DNA associates to different kind of proteins that conduct its packing in a way that it can be properly stored in the cell nucleus. Furthermore, chromatin packing functions as a regulator of the accessibility to the information from each zone of the DNA. "*Librum apertum, ut scientiarum arcana reseres…clausum, ut aedem prout oporteat in intimo pectore custodias*" (Universidad de Salamanca, 2004).
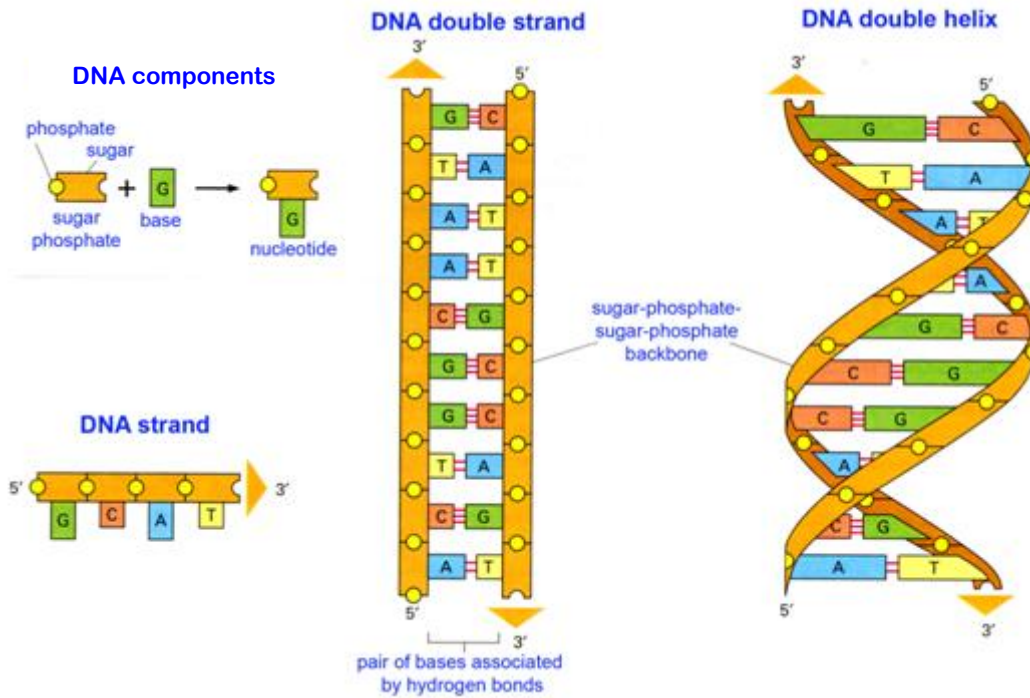
Figure 4: Scheme showing the organisation of the DNA (from Alberts et al., 2002). A nucleotide is the association of a phosphate-sugar and a base (cytosine, thymine, adenine or guanine). Each nucleotide can establish a head-tail association (5'-3') with any other nucleotide, generating a DNA sequence. A lineal DNA sequence can associate (or generate by replication) to another complementary sequence. In the cell nucleus, the canonical organization of the DNA is the double helix structure described by Watson and Crick.

Figure 5: The nucleotide gene sequence that codifies to the "Disable 1b" protein in the zebrafish is shown. Due to space reasons, only the canonical exons' sequence is displayed. The sequences labelled in red are the ones transcribed to mRNA taking part of the mature messenger, but they do not codify to amino acids.

I consider mRNA as the second basic support of the cellular information, being very similar in its composition to DNA, from which it originates by complementary transcription from one of the strands of DNA using uracil (U) as complementary of adenine. More relevant is that mRNA is a simple chain of nucleotides (not double like DNA), which makes it more fragile and, therefore, less adequate as a store for information in the long term. Moreover, in mRNA the unit of information is the trinucleotide of consecutive nucleotides (codon) considered in direction of reading (5'-3') from the first sequence AUG (start codon for translation) (Figure 6); nonetheless, the sequences of nucleotides that do not form part of the coding sequence are important for the modulation of the information contained in the corresponding molecule of mRNA. What is more, for the interpretation of the flow of information it is important to bear in mind that the DNA is repeatedly transcribed. That is, many copies of mRNA are generated from a single sequence of DNA.

This messenger, as its name indicates, is an intermediary between the administration of the information which occurs in the nucleus and that which takes place in the cytoplasm. It is formed and matures in the nucleus and is translated in the cytoplasm. The spatial and temporal separation between the transcription (transduction from DNA to RNA) and the translation (transduction of mRNA to protein) is considered essential for understanding the cellular evolution.

1.2.   Processing of information in the cytoplasm

The mRNA passes from the nucleus to the cytoplasm with a coherent orientation with the direction of reading, in such a way that it can begin to be read (translated) even when part of the linear molecule is still in the nucleus (Watson et al., 2006). The machinery of translation, constituted by complex molecules (ribosomes), functions as precise templates of assembly of amino acids in an order determined by the sequence of codons of the messenger (mRNA).

The amino acids are the components of the proteins, that is, a protein is a chain of amino acids joined to each other by a particular type of bond (peptide bond). Therefore, in the flow of cellular information the translation is a transduction of codons to amino acids, with which the support of information passes from mRNA (polynucleotide) to protein (polypeptide).

Distinct from the passage of information from DNA to mRNA, in this case the transformation of the unit of information is very marked, from a codon (trinucleotide) to a completely different type of molecule, an amino acid. Furthermore, as there are 4 nucleotides that can constitute codons, there will be 4^3 (64) codons possible to code all the amino acids of the living cell. Surprisingly, although there are hundreds of thousands of proteins, their structural bases, the amino acids, only number 20 in actual cells. Of the 64 possible codons, one functions as the start of reading (AUG) and is the only one which codes the amino acid Methionine (MET), three others are interpreted as the signal of termination of translation (UAA, UAG and UGA), thus there remain 60 codons for 19 amino acids (Fig. 6).

The assembly of the nucleotides is performed in the sequence marked by the messenger, the translation is always carried out in the same direction (from 5' to 3') and the amino acids, on associating through peptide bonds, organise themselves in a chain in which the first amino acid is always MET and denominated amino terminal. In contrast, the last amino acid of the chain is denominated carboxyl terminal. Thus the chain of amino acids that constitutes a protein is an orientated linear sequence and is denominated primary structure (Figure 7).

On leaving the ribosome the proteins which are being synthesised integrate into the cytoplasm, a complex fluid whose physical-chemical conditions determine the spatial arrangement of the emergent sequence of amino acids. If other factors, which I shall mention later, do not participate, the conditions of the cytoplasm and the sequence of the amino acids determine the second level of hierarchical structure of the proteins, the secondary structure, which results from the folding of localised parts of the amino acid chain (Figure 7). In the

cytoplasm a sequence of amino acids tends to assume a structure denominated random coil, but if hydrogen bonds which stabilise the structure are formed between some residues (lateral chains of amino acids), parts of the chain of amino acids fold.

Figure 6: The codification of amino acids through codons is shown in the left panel. In the right panel a list of amino acids and their abbreviations is shown.

Figure 7: Primary (A) and secondary (B) structure of proteins. The primary structure of a protein is the chain of amino acids linked by peptide bonds (A). The secondary structure of a protein is the folding of certain areas of a protein sequence. The scheme shows two of the more frequent secondary structures, alpha helix (B) and beta sheet.

A superior hierarchical structural level is the tertiary structure of the proteins, which is the global arrangement of the chain of amino acids (Figure 8). In this case the structure is stabilised by hydrophobic interactions and hydrogen bonds which, together, maintain the elements of the secondary structure compact. Nonetheless, these stabilising forces are weak, thus the tertiary structure undergoes continual fluctuations which are relevant in the functional regulation of the proteins.
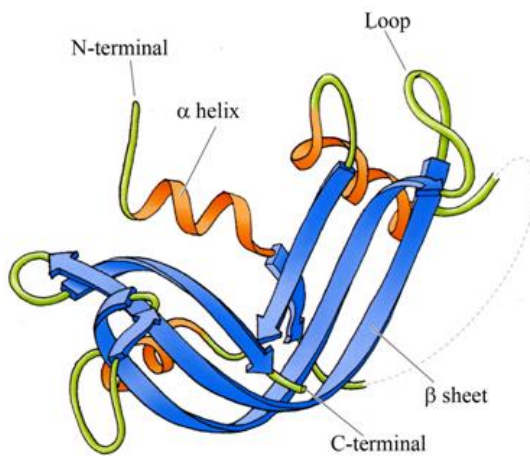

Figure 8: Scheme showing the tertiary structure of a protein (ribonuclease) containing several areas of the amino acid sequence with the secondary structure of "alpha helix" and "beta sheet" linked by loop regions.

In addition, in the context of the information, the proteins have distinct levels of organisation. The primary structure, linear, contains, with great frequency, sequences of amino acids that the molecular machinery of the cytoplasm interprets as signals of the destiny of the nascent protein (e.g. those which are to integrate in specific structures such as the peroxisomes, mitochondria or endoplasmatic reticulum). They mark zones of the linear sequence so that other molecules associate to it which prevent its folding (chaperones) and maintain them in close arrangement to the primary until they reach a specific destination (on occasions, an organelle, on occasions, complexes specialised in regulating their adequate

folding) or modify the interactions between amino acids altering the conditions of folding (Karp, 2008). Once the tertiary structure is reached, the molecular machinery of the cell can restructure the three dimensional arrangement of the chain of amino acids. Moreover, the three dimensional arrangement of a protein can be conditioned by the characteristics of the medium in which it is (e.g. pH, or levels of concentration of specific ions), or as a result of the function of other proteins.

The importance of the three dimensional structure of the proteins lies in the fact that the functionality of the molecule depends on the spatial organisation of sequences of amino acids or the interaction of groups of these sequences (Karp, 2008). Basically, the activation or deactivation of one or more structural and functional characteristics of the molecule depends on its arrangement. Therefore, slight changes in it act as systems of regulation.

The machinery of translation in its entirety, the ribosomes, reads, simultaneously, thousands of copies of various thousands of types, which determines that thousands of copies of thousands of different proteins are generated. This copious molecular population is not random, neither in the type of protein, nor the quantity of each, nor in its location, once it has reached its functional arrangement. What is more, the types of proteins and the quantities of each are regulated by other proteins and molecular complexes which deactivate and recycle them (Karp, 2008; Lodish et al., 2007).

The result of the former is that in each moment of the life of the cell each type of specific protein is integrated into the global functioning (Alberts et al., 2002; Figure 9) and its production, activity and degradation is controlled by the nuclear DNA. This supposes the existence of fine mechanisms of retro-alimentation that regulate the nuclear activity based, in turn, on proteins and RNA's which function, at the same time, as sensors of cytoplasmic activity and regulators of nuclear processes.
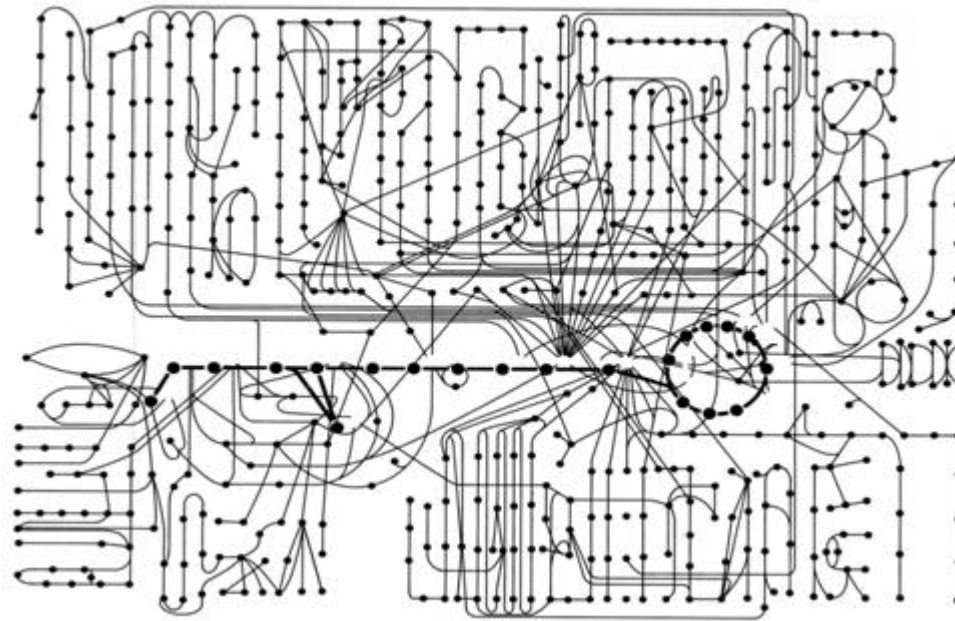
Figure 9: Some metabolic routes and their relationships are shown (Alberts et al., 2002). The scheme represents approximately 500 common metabolic reactions. Each molecule from a metabolic route is represented with a black circle.

2. Conclusion

The cell is the simplest organisation displaying permanent living features. Its functionality is based on the information flow from the central instructions book, the DNA, to the effector units, the proteins.

In the eukaryotic cells information suffers two transductions, first in the nucleus, transcription, and then in the cytoplasm, translation. In these cells both processes are spatially and temporally separated, which leads to message maturation.

The amount, activity, localisation and turnover of the proteins are regulated by both internal and external factors and among them DNA activity should be emphasised. In turn, DNA activity is regulated by many proteins, generated at the cytoplasm from DNA information.

An isolated eukaryotic cell can be considered the result of a self-regulated information flow.

References

Alberts, B.; Johnson, A.; Lewis, J. ; Raff, M. ; Roberts, K, ; and Walter, F. (2002). *Biología molecular de la célula*. Barcelona: Omega.

Cooper G.M. and Hausman R.E., (2008). *La célula*. Madrid: Marbán.

Croos, P.C. and Mercer K.L. (1993). *Cell and Tissue ultrastructure . A functional Perspective*. New York: Freeman.

Chudnovsky, E., Jejada, J. and Punset, E. (2008). *El templo de la ciencia. Los científicos y sus creencias*. Barcelona: Destino.

Karp, G. (2008). *Cell and Molecular Biology*. Hoboken, NJ: Wesley & sons

Lewin, B. (2008). *Genes IX*. Sudbury, MA: Jones & Barlett

Lodish, H.; Berk, A.; Kaiser, C.A.; Krieger, M.; Scott, M.P.; Bretscher, A.; Ploegh, H. and Matsudaira, P. (2007). *Molecular Cell Biology*. New York: Freeman.

Miescher, F. (1871). Über die Chemische Zusammensetzung der Eiterzellen. *Medicinische-chemische Untersuchungen, 4,* 441-460

Universidad de Salamanca. (2004). Ceremonial para la Investidura de Nuevos Doctores. Universidad de Salamanca
        [Online] <http://campus.usal.es/gabinete/protocolo/Ceremonial_Nuevos_doctores.pdf> [consulted: 21/05/2009]

Watson, J.D.; Baker, T.A.; Bell, S.P.; Gann, A.; Levine, M. and Losick, R. (2006). *Biología molecular del gen.* Madrid:
        Panamericana.

Watson, J.D. and Crick, F.H.C. (1953). Molecular structure of nucleic acids. A structure for deoxyribosenucleic acids.
        Nature, *137*, 737-738.

Weiss, L. and Greep, R.O. (1982). *Histología.* Barcelona: El Ateneo.

About the Author

Juan M. Lara

Studied for B.Sc. and Doctorate in Biological Sciences at the University of Seville, Spain. At present is Professor of Cell Biology at the University of Salamanca and researcher of the Institute of Neurosciences of Castille and Leon (Salamanca). His research is focussed on the understanding of the structural and molecular modifications in the spontaneous regeneration of the nervous system in fish.