# A BRIEF SURVEY OF MAJOR DISCOVERIES THAT IMPACT ON EUKARYOTIC CELLULAR PATHWAY MODELLING

***Review paper***

V.Parisi
*Dept. of Physics, Sapienza - University of Rome, (valerio.parisi@uniroma1.it)*
V. De Fonzo
*Dept. of Experimental Medicine - Sapienza University of Rome, (valeria.defonzo@gmail.com)*
F. Aluffi-Pentini
*Dept. of Mathematical Methods and Models - Sapienza University of Rome, (falpen@gmail.com)*

Keywords: Biological models, Molecular Biology, Biophysics

Abstract: An important target of molecular biology is the detailed comprehension of the dynamics of the reaction pathways regulating all cellular activities. The simulation of the biochemical reactions inside prokaryotic cells already entails great difficulties, but in eukaryotic cells, where the networks are more complex and only partially known, much greater difficulties arise. In order to infer a global behaviour from the kinetics of single reactions, a simple way to model a metabolic pathway follows the classical Michaelis-Menten approach, based on the mass action law applied to each enzymatic reaction. If, However, instead of assuming the enzymes as constant factors, appropriate attention is paid to the manifold of intermingled processes tuning the enzymes production and activity level, starting from the very regulation of genome expression, our global understanding of cellular behaviour can be substantially improved in many cases.
In this paper, among other things, we will focus on some recent researches valuable for designing upgraded models of the macromolecular crowding within cells, the promiscuity of kinases and the proteome expansion. Emphasis will be given to the schematic representation of the discussed mechanisms, providing bibliographic references for details. This should convey the essential information in a more direct and faster way, and allow grasping the message even by non experts in the field

# 1 INTRODUCTION

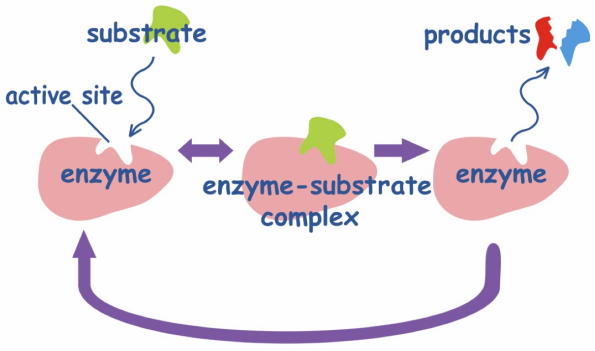
In order to decipher the overall behaviour of eukaryotic cells, a detailed analysis of metabolic pathways, both in well known networks and . in those more recently explored. For this purpose two main approaches are usually considered (Ehrenberg et al., 2009; Bruggeman and Westerhoff, 2010): the top-down approach and the bottom-up approach; the first one starts from the system properties to induce the internal processes till eventually reaching the molecular behaviour, while the second one goes in the opposite direction, starting from the molecular properties to deduce at least some of the internal processes that lead to the system behaviour. In other words one can choose to attack this problem beginning from the experimental data either of the whole systems or of the single biochemical reactions; both the approaches have proved fruitful. In this paper we shall consider only the bottom-up approach.

Since the information flow leading to each metabolic reaction originates from the genome, although modulated by the environment, often the bottom-up approaches start just, more or less indirectly, from the genome; and now that the complete genome of a number of species and a growing number of reliable data (in vitro and in vivo) about the enzyme kinetic parameters are available, it is no more unreasonable to envisage the ambitious target of building a more detailed and realistic model of the networks of biochemical reactions within a cell. Of course this entails great difficulties even in the case of simple bacteria, and much greater difficulties in the case of eukaryotic cells, where the networks are often very complex and possibly only partly known.

A classical metabolic network is a set of chemical enzyme-catalysed reactions that involve simple molecules, such as amino acids, carbohydrates and lipids. In many cases one gene encodes one enzyme that in turn catalyses one reaction, while in many other cases the behaviour is more complicated, i.e. more genes are needed for encoding one enzyme, or even one enzyme can catalyse more reactions. Moreover a protein very often undergoes post-translational modifications due to suitable enzymes. A simple approach to model a metabolic pathway follows from the classical Michaelis-Menten kinetics, based on the law of mass action applied to each reaction of the pathway, entailing a system of ordinary differential equations: the simplest example is when a constant-amount enzyme reversibly binds to a molecule of the substrate of the reaction and then catalyses the irreversible transformation of the substrate into the products of the reaction (Figure 1).

$$I)\quad E + S \longrightarrow E + P$$

$$II)\quad E + S \underset{k^d}{\overset{k^a}{\rightleftarrows}} ES \xrightarrow{k^f} E + P$$




$$III)\quad v^a = k^a[E][S]$$

$$IV)\quad v^d = k^d[ES]$$

$$V)\quad v^f = k^f[ES]$$

Figure 1: A simple reaction catalysed by an enzyme producing two products is schematised in formula I), where E indicates the enzyme, S the substrate and P the two final products. A more detailed schematisation is expressed in formula II) that exhibits the three reactions involved in I). The mass action law gives the equations III), IV) and V) for the rate of the three reactions, where v indicates the reaction rate, k a suitable constant and [] the concentration.

Since the late 1960s Savageau (Savageau, 1969; Savageau et al., 2009) proposed, for a pool of reactions, a biochemical systems theory starting from a mathematical model based on ordinary differential equations, where the kinetics of the biochemical reactions are approximated by power-law expansions instead of following rational expressions obtained from Michaelis-Menten model.

Big efforts have been produced for the simulation of the whole metabolic network of a number of bacteria, exploiting various databases sufficiently rich to give an almost complete picture. A good review is for example in a paper of Francke (Francke et al., 2005), which examines pros and cons of the different approaches.

Some of the above approaches may be incomplete if they ignore aspects that in some cases are significant; we shall examine, in Section 1, the three aspects that we consider the most important ones, i.e.: cellular compartmentalisation, non-uniform molecular distribution and macromolecular crowding. We neglect in this paper other phenomena, having an indirect influence on the metabolic network, which are originated by the three-dimensional nature of the intracellular environment. We recall two examples: the calcium waves propagating within the cell (Jaffe, 2010), whose modelling requires partial differential equations, and the rigid structures in the cytoskeleton (Ingber, 2003), which require a suitable mechanochemical approach.

Up to this point the enzymes have been implicitly assumed as constant-amount reaction ingredients. However, it is important to note that in order to design better models for the behavior of cellular pathways, it is advantageous to consider the enzymes as variable-amount ingredients, In fact, their amount is originated by the rate of their synthesis (obviously minus the rate of protein degraded or otherwise lost), which is variable, and, moreover, their production is the result of a chain of processes, each one with a peculiar dynamics. The most important are the DNA-histones interactions, the transcription, the splicing, the translation and the post-translational modifications (Figure 2).

In the next sections we focus on the following processes occurring in eukaryotic cells: genome dynamism, epigenetics, transcription control, alternative splicing, RNA interference and protein networks (Figure 2). We also provide some significant examples. Although some of such processes are known since a long time, their complete understanding and modelling are still in progress, due to recent new discoveries.

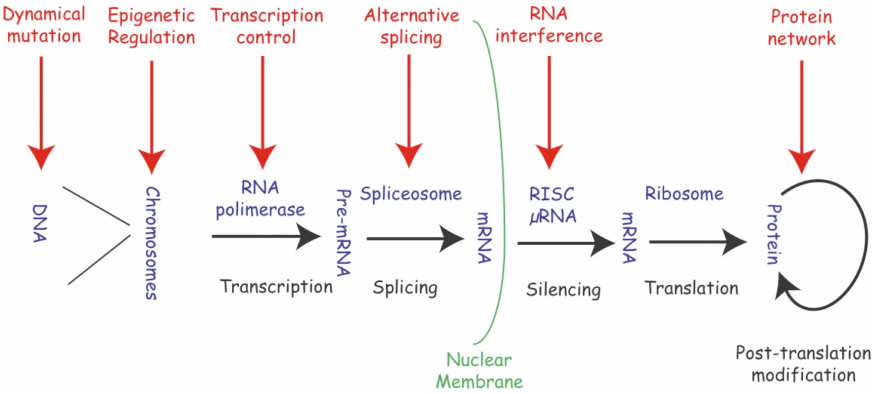



Figure 2: Chain of the processes that, starting from the DNA, leads to the production of the enzymes; the red colour indicates the critical processes considered in this paper.

## 2 MORE ON METABOLIC PATHWAYS

The greatest limits of the application of the Michaelis-Menten kinetics to biochemical networks stem from the weakness of the mass action law inside the cellular environment. The law of mass action requires that the molecules be free to move, be uniformly distributed, and spread with the traditional Fick diffusion, while for simulating the living cells we must take into account three difficulties:

a) the eukaryotic cell is largely compartmentalised because of a number of lipid membranes that constitute, inside the cytoplasm, the walls of organelles such as the endoplasmic reticulum, the cell nucleus or the Golgi apparatus (Figure 3 *LEFT*). Such membranes provide a strong obstacle to the free molecular motion from a compartment to another;

b) most molecules are non-uniformly distributed (for example the DNA-repair enzymes do not swim freely in the cell but in practice run up and down along the DNA) so that the volume of the space where they spend most

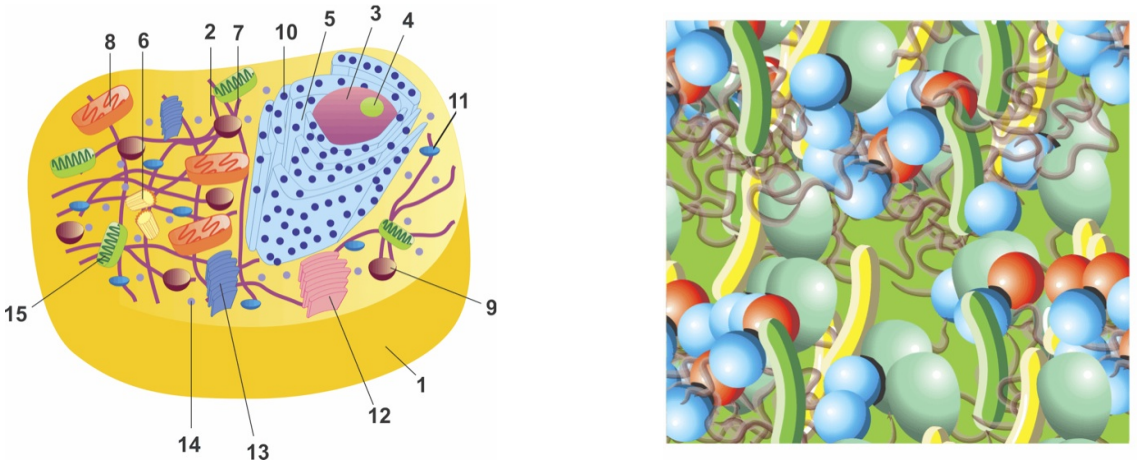
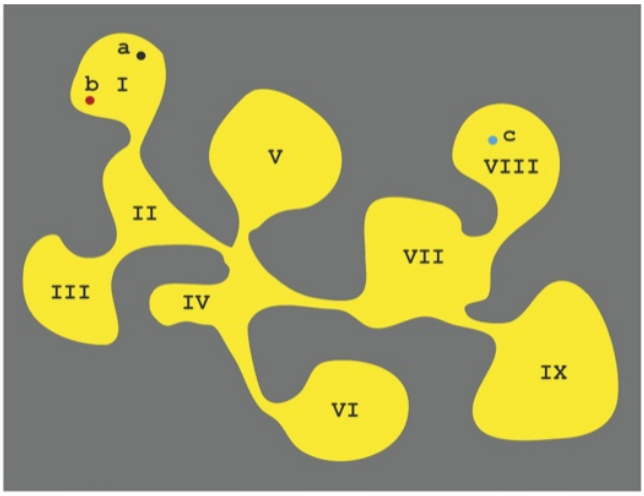



Figure 3:
*LEFT.*Example of cellular compartmentalisation: 1) Plasma membrane 2) Cytosol 3) Nucleus 4) Nucleolus 5) Rough endoplasmic reticulum 6) Centriole 7) Cytoskeleton 8) Mitochondrion 9) Vacuole 10) Ribosomes 11) Lysosome 12) Golgi apparatus 13) Smooth endoplasmic reticulum 14) Vesicle 15) Chloroplast
*RIGHT.* Schematic picture of a cellular environment crowded with macromolecules.

of the time is smaller by several orders of magnitude;




$$(1)\quad A + B + B + C \xrightarrow{k} ABBC$$

$$(2)\quad v = k[A][B]^2[C]$$

$$(3)\quad v = k[A]^{\alpha}[B]^{\beta}[C]^{\gamma}$$

Figure 4: Macromolecule crowding constrains the interaction between small molecules, subdividing the free space into a number of partially isolated regions (I - IX) indicated by yellow areas. The molecules a and b can interact very frequently, while c is isolated for a long time. The rate of a reaction, as for example (1) where one molecule of A, two of B and one of C combine, is indicated by an equation as (2) according to mass action law, while a fractional power law as (3) seems to be more adequate in a crowded environment. The square brackets indicate the chemical concentration.

c) the circulation of the molecules inside the cell is strongly hampered by the large macromolecular crowding (Figure 3 *RIGHT*), which moreover creates partially isolated spaces where small molecules are confined for very long times (Figure 4), thus causing a more frequent interaction between the molecules in the same space and a

less frequent interaction between molecules in different spaces. As a consequence, the classic mass action law with integer power exponents seems to be inadequate while the introduction of fractional power exponents seems to give more reasonable models (Schnell and Turner, 2004; Savageau, 1998).

## 3 Genome dynamism

In the classical scheme, genes are immutable and can only be transcribed. In fact, in some important cases, gene modification phenomena occur, through several dynamical genetics mechanisms (Parisi et al., 2004).

Three significant examples will be considered here.

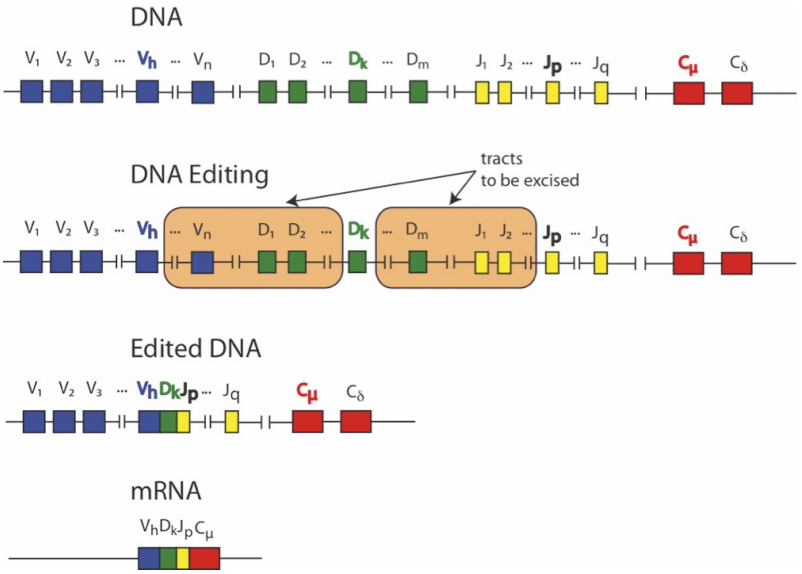



Figure 5: An example of VDJ recombination in an immunoglobulin heavy chain, where $V_h$, $D_k$, $J_p$ and $C_\mu$ are chosen.

### 3.1 V(D)J recombination

In the seventies Susumu Tonegawa discovered that the antibody diversity in the immune system is due to a genetic diversity which is not pre-existent, as it was expected, but is generated by somatic recombination: the basic mechanism is a special somatic site-specific DNA rearrangement, occurring during lymphocyte maturation. While antibody diversity lies in the variable parts of both the heavy and the light chains constituting the immunoglobulin, for the sake of simplicity we shall consider here only the heavy chain. The rearrangement acts on a genome region containing tracts belonging to three different types, named, for historical reasons, V (Variable), D (Diversity) and J (Joining) (Figure 5). In such region there is a first zone with a number of similar contiguous copies of tracts of type V, and two consecutive similar zones with tracts respectively of type D and J. To assemble a DNA tract to synthesise an immunoglobulin chain, three tracts of the above types (one for each type) are randomly selected (i.e. not genetically fixed) and joined by eliminating all intermediate tracts. Such editing-like mechanism is called VDJ recombination (Lewis, 1994).

### 3.2 Isotypic commutation

In an immunoglobulin molecule a part of the heavy chain is specific against a particular antigen, while another part of the heavy chain determines the class which the molecule belongs to (e.g. IgM, IgE, IgG, etc.): since the class determines a great number of important different features of the molecule such as the immunoglobulin localization in mucosae or in plasma, it is clear that changing the class without changing the antigen specificity, in lymphocytes maturation, is a very useful peculiarity of the immune system functionality (Manis et al., 2002).

Such peculiarity is called immunoglobulin class-switching or isotypic commutation. Max D. Cooper in the eighties discovered an unexpected DNA editing that is the base of the immunoglobulin class-switching (Figure 6 *LEFT*). In more detail, the VDJ tract of the DNA (i.e. the DNA tract coding for the heavy chain part that is specific against the antigen) is followed, on the same gene, by all the DNA tracts determining the various classes; such editing, called switch recombination, consists in eliminating the DNA between the VDJ tract and the tract determining the new class (while the following tracts, relative to other classes, remain unused).

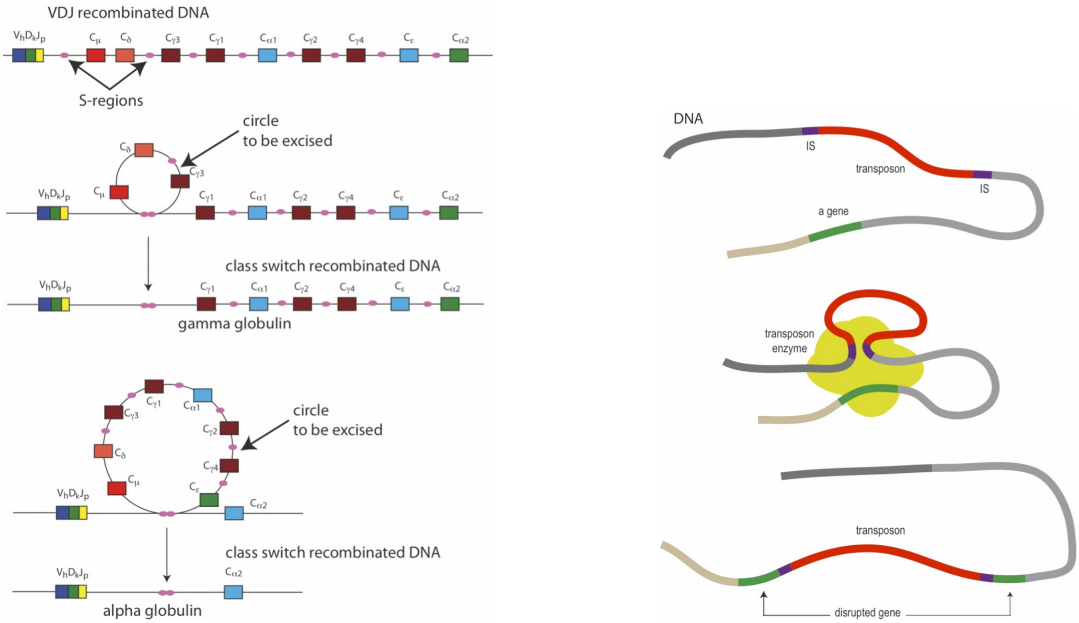



Figure 6:
*LEFT.* Immunoglobulin class-switching: two examples of recombination to obtain a gamma globulin and an alpha globulin.
*RIGHT.* An example of transposition.

### 3.3 Transposition and retrotransposition

Barbara McClintock discovered, during her studies about maize genetics in the early fifties, a mechanism, now called transposition, where some DNA sequences (the so-called jumping genes or transposons) can move to different positions along the genome (Figure 6 *RIGHT*). Later on, a somehow similar but more sophisticated mechanism has been discovered: a particular DNA tract is transcribed producing an RNA molecule, which in turn is reverse transcribed to a DNA molecule (named retrotrasposon), which is inserted elsewhere in the genome. We note that this mechanism, which has been called retrotransposition, produces a copy of the initial DNA tract, unlike a transposition, which simply displaces the involved tract (Lonnig and Saedler, 2002).

It is important to note that if transpositions (or retrotranspositions) do not involve coding genes they have negligible biological effects: if instead the insertion occurs inside a coding gene, significant effects may occur, such as deactivating the gene, or dramatically changing the transcription rate, etc., so that such mechanisms may produce remarkable perturbations to the cellular control systems.

## 4 Epigenetics

Gene expression can be modified at various levels: it can be altered by genetic mutations (such as those examined in the preceding chapter), or driven by usual transcription controls (as in the next chapter) or also by reversible chemical modifications of histones or of DNA (without affecting its sequence). We now consider this last case, called epigenetic control, which can result from a number of different mechanisms, possibly acting simultaneously on the expression of the same gene. The most common reversible modification of DNA is the addition of a methyl group to a cytosine or, less frequently, to an adenine. The methylation of the cytosine is

more frequent when the cytosine is followed by a guanine (Jaenisch and Bird, 2003). The occurrence of guanine immediately following cytosine is not uniform along the genome, but is more concentrated in some tracts called CpG islands (where the "p", for phosphorus, indicates that a "C" is immediately followed, in the strand, after a "p", by a "G") (Figure 7 *LEFT*). When methylation occurs in the neighbourhood of a promoter, it often suppresses the expression of the gene while the reverse phenomenon, the demethylation, restores the gene expression.

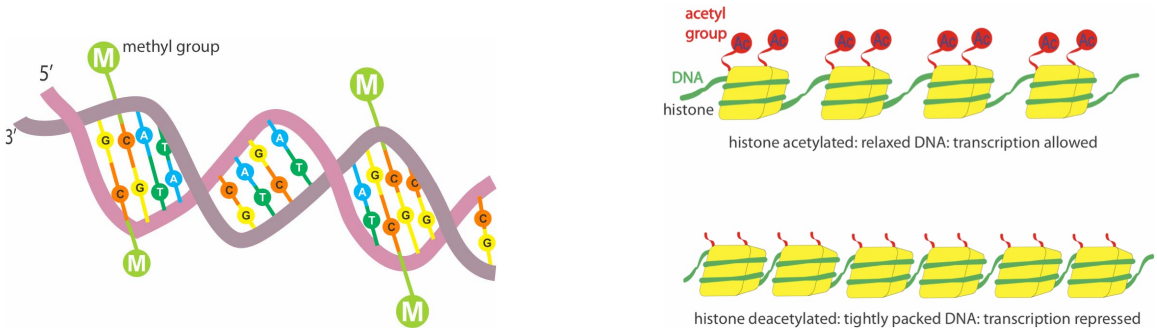



Figure 7:
*LEFT.* An example of DNA methylation.
*RIGHT.* An example of histone acetylation.

The modification of a histone (Bartova et al., 2008) is the addition of a chemical group to an amino acid of the histone: different chemical groups have been observed, the most frequent being the acetyl group and the methyl group (Figure 7 *RIGHT*). Acetylation is the most studied of these modifications. The histone acetylation, relaxing chromatin structure, enhances transcription while the histone deacetylation represses transcription. For example, acetylation of the lysine (which is chemically basic) removes histonic positive charges, reducing the affinity between histones and DNA (which is obviously acid): this makes the promoter region more accessible.

We note that although, as we saw above, the epigenetic control is originated by a number of different and often simultaneous mechanisms, their modelling can be nevertheless obtained, for example by means of a stochastic approach (Raghavan et al., 2010).

## 5 TRANSCRIPTION CONTROL

Since 1961 the standard model for transcription control was the well-known Jacob-Monod operon model, but it has been recently discovered that for the transcription of some important genes the mechanism is completely different, as we shall see below. For the sake of completeness we start to briefly describe a simplified version of the operon model proposed by Jacob and Monod in order to understand the transcription control of a cluster of genes involved in the lactose metabolism of the Escherichia coli. This simple model is well apt to illustrate the general structure of most operons. This cluster of genes belongs to a DNA region, called lac operon, that contains a first tract, the promoter, followed by a second tract, the operator, and three genes, $lacZ, lacY, lacA$, coding for the three proteins, β-galactosidase, β-galactoside permease, β-galactoside transacetylase, which are useful only when lactose is present. The control of the transcription is driven by a particular DNA binding protein called repressor (which is possibly bound to a lactose molecule) and operates, in the simplified model, as follows (Figure 8 *LEFT*). The RNA polymerase, in order to perform the transcription, must bind to the promoter and must reach, moving along the operator tract of the DNA, the region where the three genes are coded: however, if the operator is bound to the repressor, the polymerase is blocked and the transcription is repressed. The repressor is usually present in the bacterium (being continuously synthesised), but when lactose molecules are also present, the transcription is not disabled since the repressor binds to the lactose instead of binding to the operator. When instead lactose is absent, the transcription (of genes that are now useless) is repressed.

While the above model is satisfactory for a large number of genes, a very different type of control of the transcription has been recently discovered, based on the following mechanism: the transcription occurs when the DNA conformation in the promoter region is the standard double helix, but the transcription is inactive when the

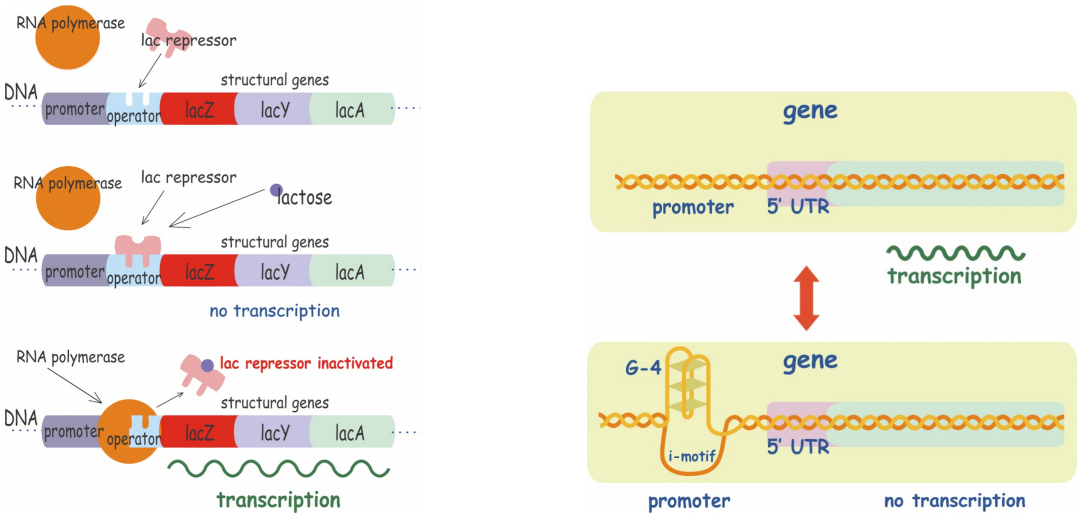



Figure 8:
*LEFT.* A simplified representation of lac operon model.
*RIGHT.* Model of G-quadruplex regulation of gene expression.

DNA is folded-back into a peculiar (non double-helix) 3D structure, called the G-quadruplex conformation that link four DNA strands stabilised by some square planar arrangements (quartets of guanines), each one made of four bases. This transcriptional control concerns a small number of crucial genes: a good example is the c Myc gene (Dang, 2010) - one of the most commonly malfunctioning genes in human cancers - whose protein, directly binding to the DNA, controls the transcription of 15% of all genes. In more detail, on the same strand of the c-Myc gene, and in the region of its promoter, there is a G-rich tract (so that the other strand is C-rich): when the two strands are normally joined, the transcription occurs, if instead the two strands separate, the G-rich tract folds into a G-quadruplex, and the transcription does not occur (Figure 8 *RIGHT*). Moreover it appears that the other strand folds into a different four-stranded structure called i-motif.

Very similar transcriptional controls, using G-quadruplex conformation, appear to occur in the same way for some genes belonging (as the c-Myc) to the category of the so-called proto-oncogen genes (Brooks and Hurley, 2010), which may mutate into cancer-causing oncogenes, but which in normal conditions control almost all cell functions, ensuring a correct cell proliferation.

## 6 Alternative splicing

In the eukaryotes the DNA transcription produces, within cell nucleus, pre-mRNA molecules, which must be suitably processed to become mature mRNA molecules and then to get out of the nucleus.

In particular, a pre-mRNA molecule first undergoes two specific chemical processes (on which we do not dwell here) at the two ends, and then a process called splicing (Figure 9 *LEFT*): an enzymatic complex (called splicesome) folds in lariat form some inner tracts of the RNA, and takes them away, so that all the remaining tracts are finally linearly joined in a mature mRNA molecule; they are called exons since they leave the nucleus while the eliminated tracts will be degraded inside the nucleus and are then called introns.

The tracts selected for the elimination, i.e. the introns, can be the same for all the molecules of a given pre-mRNA, or not; obviously in the first case the mRNA molecules produced are identical, while in the second case are different, and the process is called alternative splicing (Figure 9 *RIGHT*). This phenomenon occurs very often, is very important, and is one of the causes of the proteome expansion, i.e. the fact that in the eukaryotes the number of the proteins is much larger than the number of the genes (Nilsen and Graveley, 2010; Black, 2000). Alternative splicing mechanisms can strongly influence cellular behaviour; two of the most representative examples both deal with Drosophila Melanogaster, as shown below.

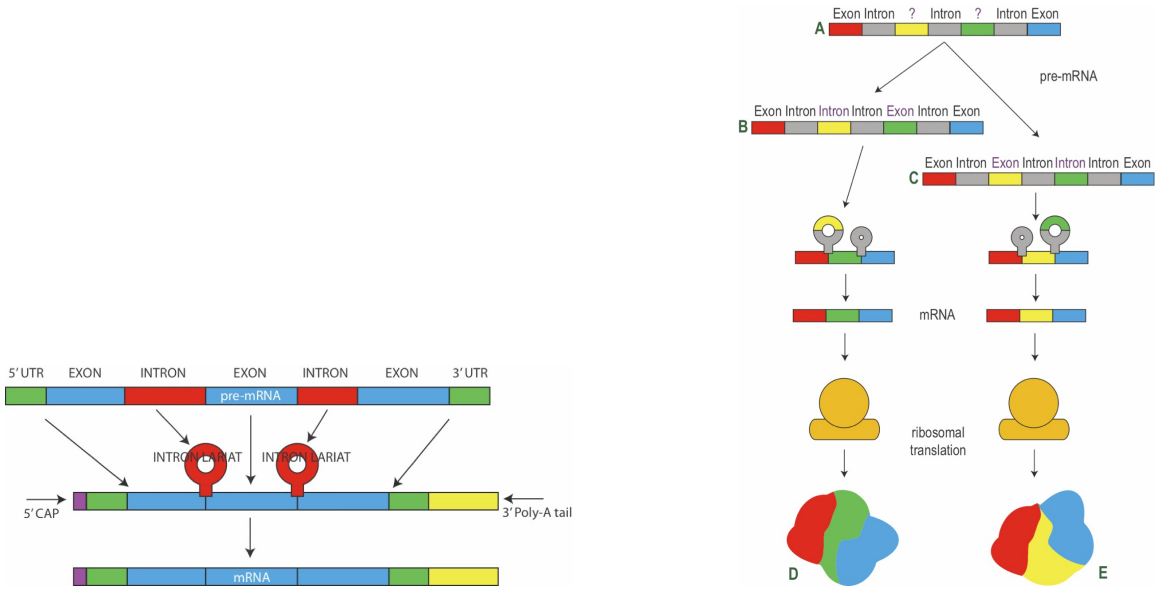



Figure 9:
*LEFT.* An example of an mRNA splicing.
*RIGHT.* An example of how an alternative splicing, starting from the same pre-mRNA molecule (A), produces two different proteins (D, E). The yellow and green tracts assume different roles (B, C) in the two splicings.

## 6.1 Sex determination

The Drosophila Melanogaster sex is genetically determined by the number of X chromosomes (two for the female and one for the male); the consequent phenotypic expression is due to the cascade interaction of three proteins, sex-lethal protein (SXL), transformer protein (TRA) and doublesex protein (DSX), each one existing in two forms, which are produced by alternative splicing processes (Figure 10) (Marin and Baker, 1998; Maniatis and Tasic, 2002).

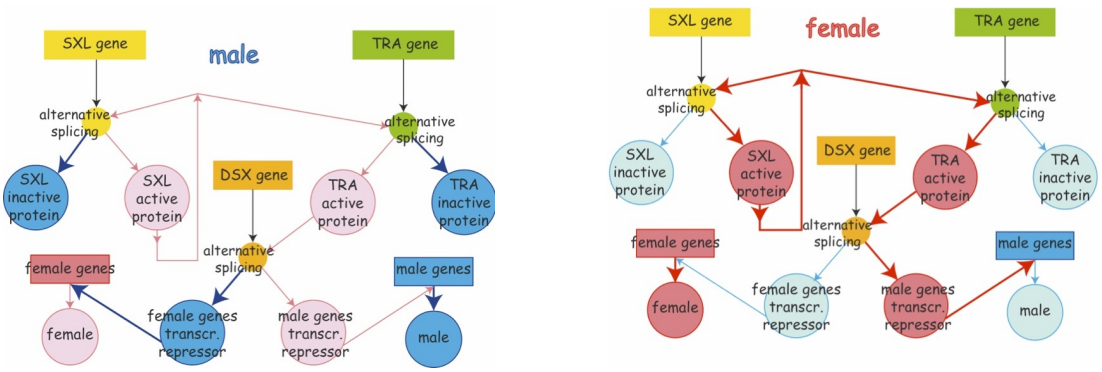



Figure 10:
Three alternative splicings determining the Drosophila Melanogaster sex. The intense colours indicate the favoured processes: red in female and blue in male.

The two forms of SXL and of TRA are active in female and inactive in male, while the two forms of DSX are both active and one represses the transcription of all the male sex-specific genes and the other one represses the transcription of all the female sex-specific genes. The production of the SXL is twice in females than in males, since the SXL gene is inside the X chromosome, so influencing the alternative splicings determining the sex of

the flies. More in detail, the abundance of active SXL in female induces the production of active SXL (in a sort of positive feedback) and of active TRA, and the abundance of active TRA leads to the DSX form that represses all male genes, while the scarcity of active SXL in male induces the production of inactive SXL and of inactive TRA and the scarcity of active TRA leads to the DSX form that represses all female genes.

## 6.2 Neural circuit assembly

In the development of the nervous system of a fly (as in fact in almost all neural networks) an important feature is the so-called self-avoidance, i.e. the tendency of branches of the same neuron to repel one another, thus preventing their connection, a wiring strategy that seems to help getting well-behaved networks. Obviously the self-avoidance process requires a preliminary self-recognition, i.e. that two branches can recognize whether they belong to the same cell, so that the repulsion process can start. A well-known example of implementation of such a strategy is based on the alternative splicing of the Dscam1 gene, as described below. The Dscam family consists of immunoglobulin cell-surface proteins expressed in the nervous systems of many different organisms; in humans such proteins seem to play an important role in neural circuit assembly: an excessive production of such proteins occurs in the Down syndrome, so that the name DSCAM comes from the acronym of Down Syndrome Cell Adhesion Molecule. While genetic recombination produces a vast repertory of antibodies (as seen above), alternative splicing produces a variety of Dscam proteins. The Dscam1 gene in the fly encodes 24 exons (Figure 11), four of them undergo alternative splicings and their assortment leads potentially to 38016 different proteins (Celotto and Graveley, 2001; Hattori et al., 2008). The self-avoidance between branches of the same neuron is based on the peculiar mechanism called homophilic repulsion, which occurs when two branches containing the same Dscam1 protein repel one another: a very rare event for different neurons because of the large number of combinations of alternative splicings.

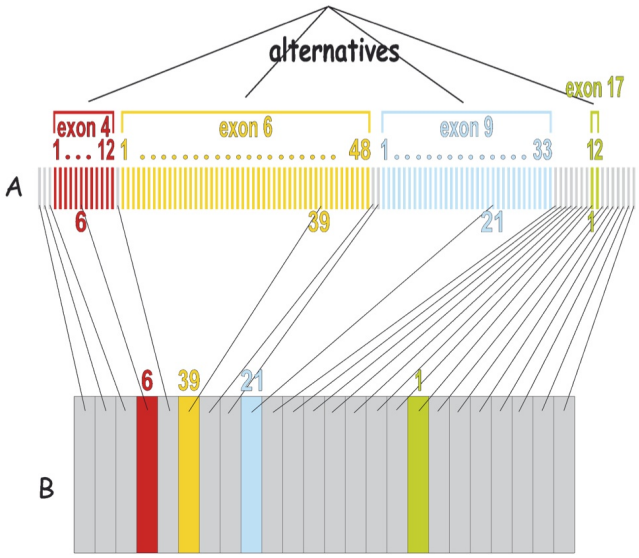



Figure 11:
An alternative splicing of the Dscam1 gene in Drosophila Melanogaster. Apart from 20 fixed grey exons that do not undergo alternative splicings, there are four exons, the 4th with 12 alternatives, the 6th with 48 alternatives, the 9th with 33 alternatives, and the 17th with 2 alternatives, where the splicing process selects only one alternative for each of them: the gene can therefore potentially produce 38016 (=12*48*33*2) different proteins. In the example from the pre-mRNA (A) is obtained a possible mRNA (B).

## 7 RNA Interference

In the early 90s some researchers attempted to obtain darker coloured petunia flowers by adding to the petunia genome supplementary copies of chalcone synthase (CHS) genes involved in violet pigment production; but quite unexpectedly they obtained instead markedly lighter colours, due to a reduced enzyme production. Only later it was discovered that an unknown mechanism prevented the translation of some molecules of the messenger RNA (Dykxhoorn et al., 2003): one of the (usually inactive) introns of the mRNA of the added CHS gene, produced a small single strand RNA fragment of an unknown molecule type (now called RNA or miRNA or microRNA) that interfered with the translation of mRNA molecules just of the CHS enzyme (Figure 12 *LEFT*). Such interference with the mRNA was the first discovered example of a mechanism, now called RNA interference (RNAi), for the cleavage (or sometimes the repression) of some specific messenger molecules.

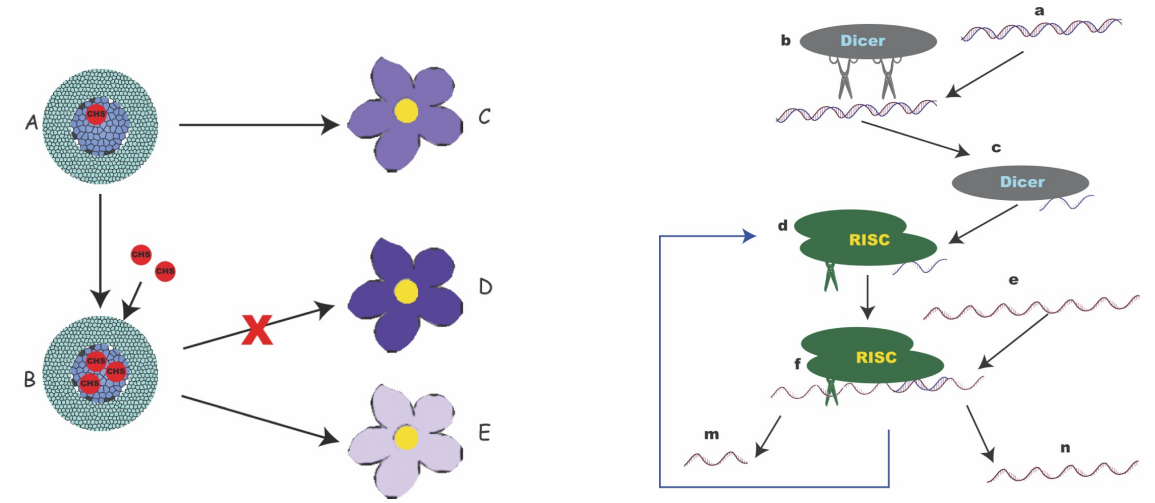



Figure 12:
*LEFT.* Unexpected result in an experiment on petunia flowers. Wild-type petunia (A) generates violet flowers (C). After the addition (B) of supplementary copies of CHS gene, lighter violet flowers (E) bloom instead of the expected darker ones (D). *RIGHT.* The pre-miRNA single strand molecule (a) is cleaved by the enzymatic complex Dicer (b) into 19-25 bp tracts of double strand RNA; after the disjunction one of the strands is just the miRNA (c). The miRNA molecule is a key player in the enzymatic complex RISC (d) and binds to a complementary tract of an mRNA molecule (e); after the binding (f) a cutting on the mRNA produces two fragments (m and n). RISC is then ready for a new process (d).

We now know that the RNAi mechanism is a powerful and fundamental process in most animal and vegetal organisms, and therefore, overlooking it, gross misunderstandings may sometimes occur, as in the above example. Such mechanism is described in Figure 12 *RIGHT*. The process starts in cytoplasm, from a short single-strand RNA molecule that forms a stem-loop and is called pre-miRNA; such molecule was originated in the nucleus by a transcription followed by suitable processes and was then exported to the cytoplasm. The pre-miRNA molecules are cleaved by an enzymatic complex, called Dicer, into 19-25 bp tracts of double strand RNA; after the disjunction one of the strands (miRNA) is kept, while the other (miRNA*) is lost. The miRNA molecule participates in an enzymatic complex called RISC (RNA-interference silencing complex) and eventually binds to a complementary tract of a mRNA molecule; after the binding another component of the RISC (called argonaute protein) can perform a cleavage on the mRNA, and the two resulting mRNA fragments are quickly degraded by normal cellular processes; the RISC is then ready to interfere with other mRNA molecules by repeating the binding and the cleavage, with the result that a particular gene is silenced.

The cleavage in the above-described RNA interference process occurs in plants for most miRNA while it does not occur in animals for most miRNA (Saumet, 1969). However also when the messenger is not cut the interference occurs somehow, since the messenger remains trapped in the RISC complex.

We note that such silencing appears well apt to be easily modelled by ordinary differential equations similar to those used for enzymatic reactions (Vohradsky et al., 2010). Besides miRNA, other small RNA molecules are

involved in similar RNA interference mechanisms, such as, for example, siRNA (which originates often from double strand molecules) (Saumet, 1969) and piRNA (Klattenhoff and W.Theurkauf, 2008); but we do not dwell on such subjects, both since the framework is still fluid, and since the main features of the mechanisms seem to be similar.

## 8 Protein networks

Normally many enzymes alter other proteins, for example by means of processes such as acetylation or glycosylation, and often the altered protein is an enzyme that activates or deactivates in turn another enzyme that in turn can activate or deactivate a third enzyme and so on, possibly forming long chains or more frequently networks. Such cascade processes normally consist only of three chemical mechanisms, i.e. hydrolysis, phosphorylation and phosphorolysis, which involve respectively three kinds of enzymes: proteases, kinases and phosphatases; in detail:

- a protease performs a cleavage of a polypeptidic molecule, hydrolyzing a specific peptide bound (Figure 13 *LEFT*);

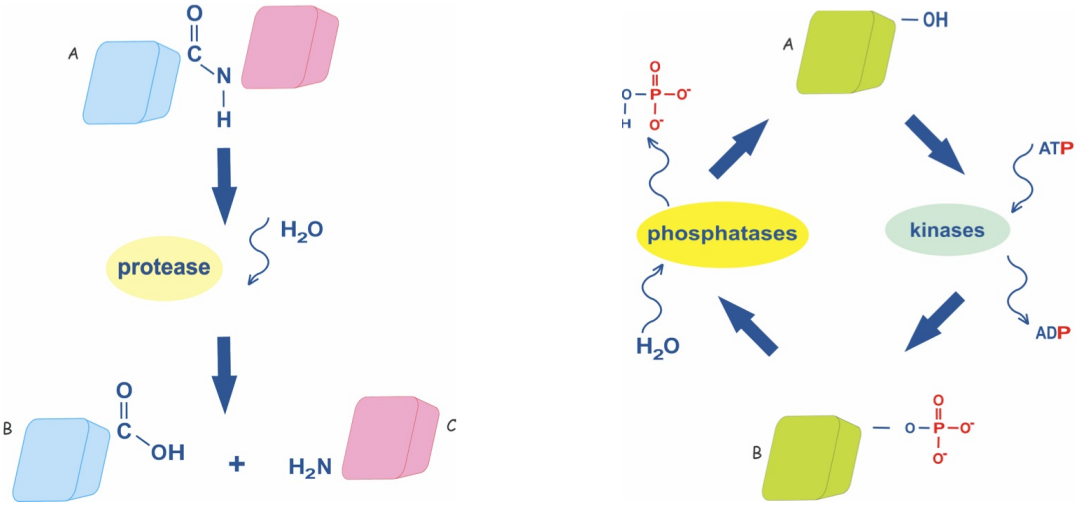



Figure 13:
*LEFT.* An example of a protease hydrolyzing a peptide bond. The protein A is split into two polypeptides B and C.
*RIGHT.* Effects of phosphorylation and phosphorolysis: A is the non-phosphorylated protein, B is the phosphorylated protein.

- a kinase performs a phosphorylation, i.e. moves a phosphate group from a high-energy phosphate-containing molecule (as ATP) to a specific amino acid (Figure 13 *RIGHT*);

### 8.1 Protease networks

Some protein networks are composed almost exclusively of proteases and are often structured as a chain (Wolan et al., 2009); a very schematic example is here represented (Figure 14): a first protease cleaves a proenzyme (here an inactive precursor of a protease), by removing the fragment that masks its active site, therefore activating a second protease, which in turn, with a similar procedure, activates a third protease, and so on. Such cascade reaction allows, even starting from few molecules, to produce a rapid and efficient response to a signal: for example, a molecule of a first protease activates hundreds molecules of a second protease, each of which activates hundreds molecules of a third protease, and so on, finally obtaining a huge number of molecules. An important example of a protease cascade pathway is given by the blood coagulation, a complicated mechanism that involves almost 13 factors, many of which are proteases.

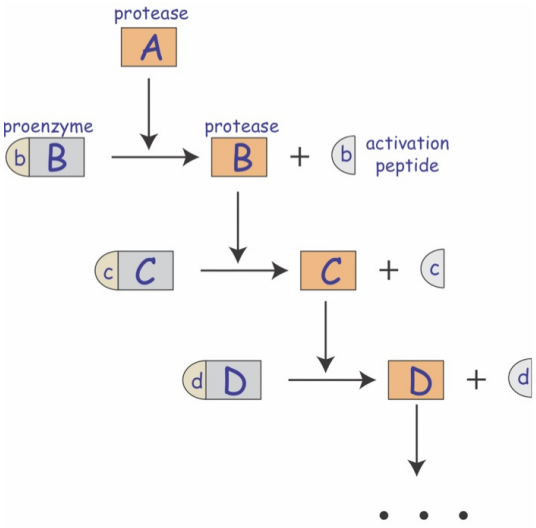



Figure 14: Example of protease chain: the protease A cleaves a proenzyme by removing the fragment b, therefore activating the protease B, which in turn, with a similar procedure, activates the protease C, and so on.

## 8.2 Kinase/phosphatase networks

Almost all kinases and phosphatases are enzymes that can be active, or not, depending on their own phosphorylation condition (Manning et al., 2002). For example, if a phosphatase is active only when is phosphorylated, some kinases activate it and some phosphatases deactivate it. In general kinases and phosphatases have the peculiarity of acting, with different efficiency, on several different substrates and of being in turn substrates of several enzymes. This implies that they can build complex networks that are similar in some aspects to neural circuits, and are very common in signal transductions; while, on the contrary, the metabolic enzymes normally catalyse only one or very few reactions.

We note that for a simply qualitative understanding of a biochemical network (for example to predict whether the addiction of a substance increases, or not, the amount of another substance), in the case of a simple metabolic network it is sufficient to check the mere existence of all the involved connections performing a somehow boolean approach (Watterson et al., 2008), while in some cases of networks of kinases and phosphatases, it could be mandatory to study also the strength of the connections (Figure 15). Important complications arise from the fact that, in order to build-up network models, the necessary parameters are usually taken from various databases with widely different characteristics, (for example with data either experimentally measured or somehow estimated from other data), and often containing some inconsistent data. In order to cope with such difficulties a natural way appears to suitably compare the values obtained from the various sources (Ranea et al., 2010) or from the various models (Orton et al., 2004), thus leading to more satisfactory modelling.

# 9 Conclusions

The detailed understanding of the dynamics of cellular pathways is of crucial importance for quantitative biology at the cellular and molecular level and one should obviously try and foresee the evolution of this field in the next future. For the sake of simplicity, we schematically consider two cascade processes: the dynamics of enzymes production and of enzymatic reactions. The latter has been studied since a long time and probably, while still requiring a substantial amount of quantitative information, it does not bode revolutions, despite conceptual improvements are still in progress. As for the dynamics of the enzyme production, in prokaryotic cells we do not expect big surprises, because the overall picture seems to be now established. In the case of eukaryotic cells, however, we do expect complications because of the gradual emerging of brand new information, which deserves

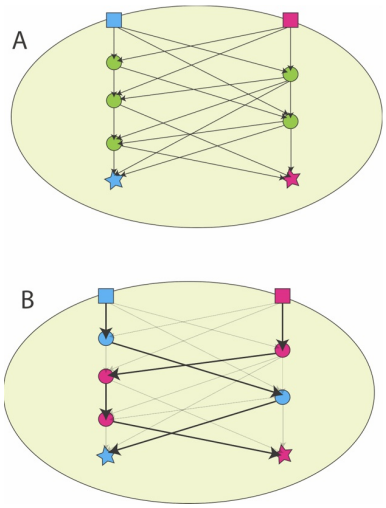



Figure 15: Hypothetic example of network of kinases where for qualitative understanding it is mandatory to study also the strength of the connections: we suppose a signal transduction model composed by two cell receptors (squares), each of them producing a specific (same colour) cellular response (stars) by means of a kinase (circles) network, where we neglect the phosphatases. The part A of the figure shows only the network, without considering the strength of the connections, so that it is impossible to understand the response specificity of the model, i.e., starting from a receptor and following all the arrows we obtain both responses. In part B, instead, we introduce the strength of the connections (showed by the thickness of the arrows), so that an easy understanding of the specificity of the response is obtained.

continuous updating of the related mechanistic models.

The good new concerns the almost parallel increase in the efficiency and power of the hardware and software tools, available for data analysis, modelling and simulation of complex biological processes . Thus, even if strictly depending upon the human and technical investment in the field, we are confident that in a not-so-far future substantial clarification of the cellular metabolism at the global level should be within our range.

# REFERENCES

Bartova, E., Krejc, J., Harnicarov, A., Galiov, G., and Kozubek, S. (2008). Histone modifications and nuclear architecture: a review. *J. Histochem. Cytochem.*, 56:711–721.

Black, D. (2000). Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology. *Cell*, 103:367–370.

Brooks, T. and Hurley, L. (2010). Targeting myc expression through g-quadruplexes. *Genes and Cancer*, 1:641–649.

Bruggeman, F. and Westerhoff, H. (2010). The nature of systems biology. *Trends Microbiol.*, 15:45–50.

Celotto, A. and Graveley, B. (2001). Alternative splicing of the drosophila dscam pre-mrna is both temporally and spatially regulated. *Genetics*, 159:599608.

Dykxhoorn, D., Novina, C., and Sharp, P. (2003). Killing the messenger: short rnas that silence gene expression. *Nat. Rev. Mol. Cell. Biol.*, 4:457467.

Ehrenberg, M., Elf, J., and Hohmann, S. (2009). Systems biology: Nobel symposium 146. *FEBS Lett.*, 583:3381.

Francke, C., Siezen, R., and Teusink, B. (2005). Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.*, 13:550558.

Hattori, D., Millard, S., WM, W. W., and Zipursky, S. (2008). Dscam-mediated cell recognition regulates neural circuit formation. *Annu. Rev. Cell Dev. Biol.*, 24:597620.

Ingber, D. (2003). Tensegrity: Ii. how structural networks influence cellular information processing networks. *J. Cell Sci*, 116:1397–1408.

Jaenisch, R. and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33:245254.

Jaffe, L. (2010). Fast calcium waves. *Cell Calcium*, 48:102113.

Klattenhoff, C. and W.Theurkauf (2008). Biogenesis and germline functions of pirnas. *Development*, 135:39.

Lewis, S. (1994). The mechanism of v(d)j joining: Lessons from molecular, immunological, and comparative analyses. *Adv Immunol.*, 56:27150.

Lonnig, W. and Saedler, H. (2002). Chromosome rearrangements and transposable elements. *Annu. Rev. Genet.*, 36:389410.

Maniatis, T. and Tasic, B. (2002). Alternative pre-mrna splicing and proteome expansion in metazoans. *Nature*, 418:236243.

Manis, J., M, M., and Alt, F. (2002). Mechanism and control of class-switch recombination. *Trends Immunol.*, 23:3139.

Manning, G., Plowman, G., Hunter, T., and Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. *Trends Biochem. Sci.*, 27:514520.

Marin, I. and Baker, B. (1998). The evolutionary dynamics of sex determination. *Science*, 281:19901994.

Nilsen, T. and Graveley, B. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463:457463.

Orton, R., Vyshemirsky, O. S. V., Gilbert, M. C. D., and Kolch, W. (2004). Computational modelling of the receptor-tyrosine-kinase-activated mapk pathway. *Biochem. J.*, 392:249261.

Parisi, V., Fonzo, V. D., and Aluffi-Pentini, F. (2004). A survey of dynamical genetics. *Riv. Biol.*, 97:223253.

Raghavan, K., Perrin, H. R. D., Goasmat, F., and Burns, J. (2010). Computational micromodel for epigenetic mechanisms. *PLoS ONE*, 5:14031.

Ranea, J., Morilla, I., Lees, J., Reid, A., Yeats, C., Clegg, A., Sanchez-Jimenez, F., and Orengo, C. (2010). Finding the dark matter in human and yeast protein network prediction and modelling. *PLoS Comput. Biol.*, 6:1000945.

Saumet, A. (1969). Anti-viral rna silencing: do we look like plants? *Retrovirology*, 3.

Savageau, M. (1969). Biochemical systems analysis: Ii. the steady-state solutions for an n-pool system using a power-law approximation. *J. Theor. Biol.*, 25:370379.

Savageau, M. (1998). Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *BioSystems*, 47:936.

Savageau, M., Coelho, P., DA, F. F. D. T., and Salvador, A. (2009). Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *PNAS*, 106:64356440.

Schnell, S. and Turner, T. (2004). Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog. Biophys. Mol. Biol.*, 85:235260.

Vohradsky, J., Panek, J., and Vomastek, T. (2010). Numerical modelling of microrna-mediated mrna decay identifies novel mechanism of microrna controlled mrna downregulation. *Nucleic Acids Res.*, 38:45794585.

Watterson, S., Marshall, S., and Ghazal, P. (2008). Logic models of pathway biology. *Drug Discov. Today*, 13:447456.

Wolan, D., Zorn, J., Gray, D., and Wells, J. (2009). Small-molecule activators of a proenzyme. *Science*, 326:853858.