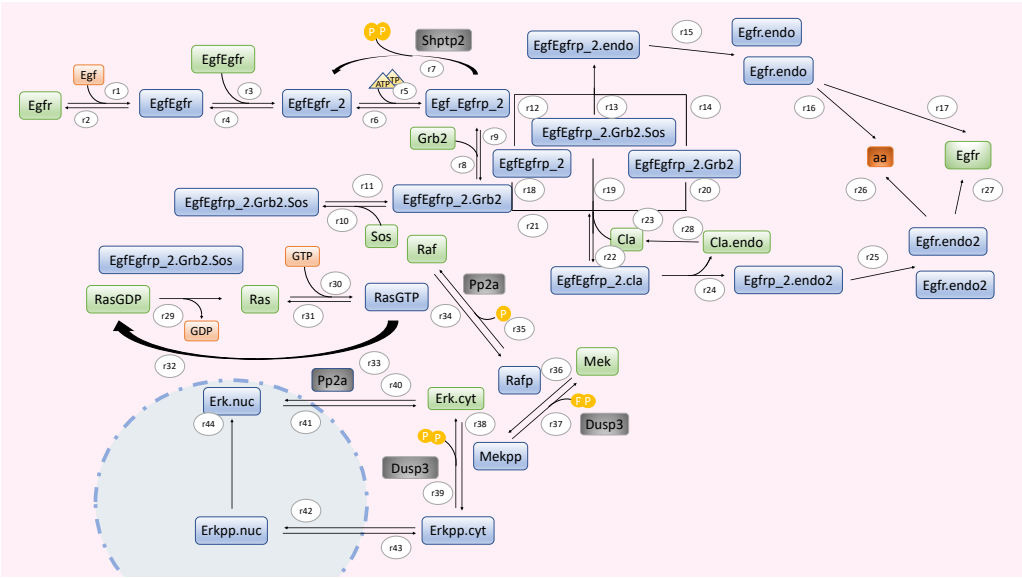
# UNIVERSITY OF NAPLES FEDERICO II

## DOCTORATE IN
## MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

## XXXIII CYCLE

**Rocco Caggiano**

# *In silico* SIMULATION OF TUMOUR CELL PROLIFERATION AND MOVEMENT BASED ON BIOCHEMICAL MODELS OF MAPK CASCADE




**Year 2021**

**UNIVERSITY OF NAPLES FEDERICO II**

**DOCTORATE IN**
**MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY**

**XXXIII CYCLE**

# *In silico* SIMULATION OF TUMOUR CELL PROLIFERATION AND MOVEMENT BASED ON BIOCHEMICAL MODELS OF MAPK CASCADE

Tutor
Prof. Giovanni Paolella

Candidate
Rocco Caggiano

**Year 2021**

Table of contents

Abstract

Systems biology allows analytical investigation of intracellular dynamics, analyzing complex processes and taking into account the interactions among the various subsystems. In this study, biochemical models describing the behavior of regulatory molecular networks were created and interfaced with a simulation system able to reproduce motility and proliferation of eukaryotic cell cultures. The primary focus was on MAPK cascades, particularly Erk1/2 activation by growth factors and mitogens such as EGF through tyrosine kinase receptors (RTKs) as Egfr, which represent a fundamental signal transduction and regulatory network affecting many cellular processes, including proliferation, motility, differentiation and survival. Erk1/2 predicted levels were related to reactions representing the progression of the cell cycle and used to modulate cell growth in a cell simulator.
The biochemical model was built starting from literature data and a database of estimated protein concentrations representative of different cell types and experimental conditions and may be run for prolonged time frames and in various experimental conditions, including a vast array of cell lines. A software tool developed on purpose is able to run the model and interface with the cell simulator.

# 1. Background

## 1.1 Systems biology's approach to the study of molecular pathways

### *1.1.1 Systems Biology*

Systems biology has evolved as a scientific discipline in which computational and mathematical modeling is used to study biological systems (Vidal 2009). This discipline focuses the attention on the study of the complex metabolic pathways of an organism, or the networks that govern intracellular signaling. It studies molecular interactions at different levels, enabling the identification and description of the subcellular machinery that makes functional operational units in cells, tissues and organ systems resulting in physiological behaviors (Ideker 2001, Kirschner 2005) using both experimental and computational frameworks to answer biological questions. Omics technology provides a platform to extract knowledge using bioinformatics, statistical methods and network analysis. The dynamical models of certain components in these networks must be verified by low-throughput, high-fidelity and single cell experiments that provide new strategies to improve and optimize the dynamical models.

Unlike the classical molecular biology, characterized by a qualitative scientific approach, aimed at verifying the validity of a (verbal) model through the formulation and verification of hypotheses, and studying the complexity of biological phenomena analyzing subset question, often through in vitro experiments, systems biology, prefers a holistic analytical approach, studying the phenomenon in its globality, and prefers to involve quantitative analysis (Kell 2004). Methodologies can be applied either in a “bottom-up approach”, that puts small functional units together to make a system or in a “top-down approach”, that starts from the global view of the system and then tries to study smaller subsystems (Batchelor 2009, Purvis 2013).

Despite the different approaches, systems biology is highly dependent on the extensive biological information that has been acquired through molecular biology, and systems biology studies often require confirmation of their results and hypotheses using a reductionist approach. There are two complementary different approaches to systems biology (Jewett 2008). In the first, different types of high-throughput generated data, often referred to as omics data, are analyzed in an integrative way; it is useful for mapping cellular functions at the genome scale. In the second, detailed models for specific processes, such as enzymatic reactions, are assembled into a model describing the system being studied, enabling detailed timescale resolution of the impact of individual components on overall system properties. Both approaches allow the identification of how the interactions between the many different cellular

components give rise to biological functions that cannot be identified through a reductionist approach.

The first systems biology approaches date back to the 1950s and 1960s, with Denis Noble's computer model describing the function of the heart pacemaker being a landmark study (Noble 1960). In the late 1960s a research driven by Arnold Fredrickson and Henry Tsuchiya at the University of Minnesota, aimed to improve industrial fermentation process designs, developed an extensive modeling framework for cellular growth models (Fredrickson 1970). In the 1970s biochemists and biophysicists had the idea of exploiting mathematical modeling to study complex biological processes, such as signaling pathways and metabolic pathways. In 1974 Reinhart Heinrich at Humboldt University and Tom Rapoport described a model that allowed the quantification of flux control in metabolic pathways (Heinrich 1974). A few months earlier, Kacser and her colleagues published a similar mathematical model with the same objective (Kacser 1973), and these two frameworks have now been combined into what is referred to as metabolic control analysis (MCA). In the following years the MCA was developed significantly, in particular by Hans Westerhoff and Douglas Kell (Kell 1989), and today it represents a valid tool useful for the study of cellular processes and for classroom illustration of how fluxes are controlled in metabolic pathways.

One of the most important keys needed to develop a successful model is to have high quality experimental data. In recent decades, our knowledge of the foundation of living organisms in terms of various components of cells, tissues and organ systems has been greatly expanded due to advances in technologies for high-throughput measurements such as genomics, transcriptomics, proteomics and metabolomics. In genetics and genomics, entire genomes of many organisms have been sequenced and the gene expression profiles comprehensively mapped. In biochemistry, mass spectrometry-based protein surveys have provided extensive lists of proteins and protein complexes, while molecular and cell biology have provided information on how proteins are organized to orchestrate the functions of subcellular systems such as cell organelles and cellular machinery components. The availability of such multiscale information has catalyzed the formation of systems biology as a discipline in biomedical sciences (Tavassoly 2018). The development of this knowledge led Ideker and coworkers to analyze these in the context of annotated biological networks, as illustrated in a study of the galactose regulation of yeast (Ideker 2001). This led to the coining of the term "systems biology" by Leroy Hood (Ideker 2001) and Hiroaki Kitano (Kitano 2002), who independently established the first institutes for systems biology in Seattle and Tokyo, respectively, in 2000.

Today, systems biology finds wide application in basic studies of

biology, engineering cells for the production of valuable chemicals, and understanding the molecular mechanisms underlying complex human diseases (Nielsen 2017). As an example, systems biology is used to study the evolution of a cancer cell from a normal cell. This involves interactions among molecular components at the cell level. At the same time, systems biology can be used to integrate the interactions among cancer cells and the evolution of tumors. It is also capable of describing the interaction of different tissues such as blood vessels, tumors and the immune system to shed light on the complex phenomena of cancer at the organ level (Anderson 1998, Kreeger 2010).

### *1.1.2 Mathematical Modelling*

The center of systems biology is mathematical modelling, by which a biological system can be translated into a mathematical model for subsequent computer simulation and analysis. In order to be useful to biological questions, such models must accurately reproduce the biological system under study, and be able to make predictions about its behavior. Thus, while the basis of a model is the topological representation of its components and their links, it is the description of the biological system's dynamic behavior which equips the model with predictive power (Tomita 1999). However, it is important to underline that a model is not a real and exact representation of the biological system, but only a simplified description that could assist the analysis. A mathematical model can be used to generate new insights, make testable predictions, identify gaps in our biochemical knowledge, test conditions that may be difficult to obtain in the laboratory and help to identify what is right and wrong with our hypotheses (Orton 2005).

It is now extremely clear and ascertained that the eukaryotic signal transduction pathways are extremely complex. Pathways that were once described as simple and linear are now described in an extremely branched way, and modules once conceived as autonomous are known to crosstalk (Schlessinger 2000). This led to ideating new approaches and methodologies to study these complex communication networks, and an exhaustive answer came precisely from the use of quantitative mathematical models, to better understand the behavior of cellular signaling networks.

Analyzing the models of biological systems available in literature so far, a possible common workflow could be hypothesized and described as composed of five main steps (Orton 2005):

1. The first step involves the selection of the biological system to model (e.g., the MAPK cascade), especially identifying the biological question the model aims to answer (e.g., which concentration of EGF is sufficient to activate Erk1/2?).

2. The second step involves the definition of the model. Essentially, this involves drawing a detailed topological chart of the system that shows all the species (e.g., proteins) involved, what reactions they can participate in and where. Each reaction is defined according to its kinetic nature. It is also necessary to define the kinetic parameters (such as rate constants and initial species concentrations) characteristic of the reaction. The search for detailed kinetic parameters takes place through the consultation of the scientific literature and numerous databases available today, while to estimate the missing kinetic parameters it may be necessary to resort to laboratory experiments or computational techniques (Mendes 1998). It is important to keep in mind that many biological processes are very complex and not yet well understood. Therefore, defining a model often involves making simplifying assumptions that reduce complex and poorly understood processes into simpler ones that can still represent the biological processes well enough to explain the observed data.
3. The next step involves translating these kinetic reactions into a set of differential equations that describe how the concentration of each species changes over time. This set of differential equations is then simulated (or solved) over a period of time. Several software tools are now available that allow the generation and simulation of models based on differential equations.
4. The simulation typically returns a table of data or a curve showing how each species' concentration changes over time. The results obtained must now be validated against other available experimental data. If the model results agree with the other experimental data, the model itself can be further analyzed and expanded, otherwise it must be revised. This is a 'debugging' loop involving model definition, simulation and validation, where the model is refined in order to obtain a behavior which conforms with experimental observations. This validation step is extremely important in order to generate a model that produces reliable predictions.
5. After the model has been validated, it can be used, and the simulation results interpreted. What is commonly performed is a sensitivity analysis, in which the systems' reactions to the variation of some crucial parameters are evaluated. For example, it is evaluated how the peak height or the duration of the phosphorylated Erk signal or its nuclear at translocation at different activator pathway concentrations. Sensitivity analysis can also be used to assess how robust the system is. A robust system can absorb fairly large disturbances and continue to perform reasonably well. If a system variable has low sensitivity with

respect to a parameter, it is robust to alterations in that parameter. The structural strength of a model can also be analyzed by monitoring how it behaves when parts of it are removed. This can be useful, because there is often redundancy in biological systems where multiple pathways are available for the production or activation of a certain protein.

### *1.1.3 Ordinary differential equations (ODEs)*

A biochemical process has a dynamic that can be studied by translating its components (nodes and edges) into a set of ordinary differential equations (ODEs). The technical definition of a differential equation is an equation involving one or more unknown functions and their derivatives. For example, a differential equation describes how a property of interest, such as the concentration of a substrate, changes over time; this is usually expressed by describing how the rate of change of the concentration is related to the concentration at that moment.

Simulation uses numerical methods to solve the differential equations and approximate the change in concentration of all the species in a system over time. There are many advanced methods of numerical integration, such as the Euler's method, Runge-Kutta, Rosenbrock and Richardson extrapolation. After a first step, which solve the differential equation, the chosen method is then repeated for another step size using the result from the previous step as the new starting point. Differential equations can easily be used to model large biological systems involving many more species and reactions, as well as more enzyme reactions.

Chemical reactions can be classified as irreversible, when the equilibrium is completely displaced either on the part of the reactants or of the products, and as reversible, where instead the reactants and the products coexist in chemical equilibrium. Almost all chemical reactions are reversible. This means that reagents can be obtained from the products and vice versa. This situation is called dynamic equilibrium: it can be considered reached when the direct and reverse reaction rates are equal. It is called "dynamic" because, even when at equilibrium, direct and reverse reactions continue to occur, although with no change of the species concentrations. Since at equilibrium the concentrations of the components are constant, their ratio will also be constant; this means that if the concentration of one specie is changed, the others' automatically change to conserve the ratio. This mathematical constraint is expressed by the law of mass action. The following reversible reaction

$$A \rightleftharpoons B$$

describes the conversion of A into B. It is a simple, non-catalyzed reaction, which can be described by the kinetics of mass action. The reaction rate constant k represent the specific speed of the reaction, measured at 1 M reagent concentration. The reaction proceeds at speed v=k[A]. In this case the reaction speed depends by the concentration of A: the higher the value of [A], the higher the speed, and therefore the faster A will be consumed, and the faster B will be produced. This type of reaction is also referred to as a first order reaction, and k is measured in $s^{-1}$. In a second order reaction, where two equal or different species react to form a product, on the other hand, the reaction rate is directly proportional to the product of the concentrations of the two reacting substrates, or to the square of the concentration of only one of the reactants. In this case k is measured in L $mol^{-1}$ $s^{-1}$.

In the biochemical model, the mass action kinetics can be used to model reactions such as those which physiologically inactivate species, proteins degradation/synthesis, or complex/decomplex reactions. Most other reactions need to be catalyzed to take place at interesting rates. Enzymes are proteins that convert specific reactants into products while remaining basically unchanged. A classical scheme for an enzyme catalysed reaction:

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

For this kind of reaction, the rate of production depends nonlinearly on the concentration of the substrate. This reaction can be described by the Michaelis–Menten kinetics, one of the best-known models of enzyme kinetics (Srinivasan 2021). This model relates the concentration of substrate to the reaction rate.

$$v = \frac{d[P]}{dt} = V_{max}\frac{[S]}{K_M + [S]}$$

It introduces a saturating behaviour in the dynamics, plotting the rate (extracted from initial velocity phase) as a function of substrate concentration and fitting that to a single-site binding hyperbola. The $V_{max}$ is reached only when all the enzyme present in solution is bound to the substrate; it depends on the concentration of the enzyme in solution according to the relationship $V_{max} = k_{cat}$ [E]T, where $k_{cat}$ is the catalytic constant, or turnover number, an intrinsic kinetic property of the enzyme that expresses the number of substrate molecules converted into product by an enzyme molecule in the unit of time, when the enzyme is saturated. The Michaelis-Menten constant, $K_M$, represents the substrate concentration necessary for the reaction to have a speed equal to half the maximum speed (or semi-maximal speed). Application of this model to

several turnover enzymes reactions requires several approximations and assumptions (Cleland 1967, Segel 2013):

1. the enzyme is a catalyst and is not consumed in the reaction, and rapidly reacts with the substrate to form the enzyme-substrate complex;
2. only a single substrate and a single enzyme-substrate complex are involved, and the enzyme-substrate complex breaks down directly to form free enzyme and product;
3. enzyme, substrate and enzyme-substrate complex are at equilibrium, that is, the rate at which ES dissociates to E and S is much faster than the rate at which ES breaks down to form E+P.

In the field of biochemical modelling, the Michaelis-Menten equation helps to model reactions such as phosphorylation/dephosphorylation reactions.

Many studies on the kinetics of enzymes have been made, and many of the collected informations are collected in the BRENDA database (http://www.brenda-enzymes.org). BRENDA (Braunschweiger Enzymdatenbank) is the most important database that collects information on enzymes, such as:

1. the reactions they catalyze, their specificity, structure, thermal stability, optimal catalysis conditions in terms of pH;
2. the organisms from which the enzymes were isolated and purified, the specific activity, the $k_M$, inhibitors, all the bibliographic references relating to the data presented and links to other databases.

An important part of BRENDA is represented by more than 110,000 enzyme ligands entries. These include not only metabolites of primary metabolism, co-substrates or cofactors but also enzyme inhibitors or metal ions. Each entry includes its chemical structure. The database is maintained by the Institute of Biochemistry of the University of Cologne. The data it contains are extracted directly from the scientific literature; the database is handled manually, although the informations are extracted by the means of computer tools.

### *1.1.4 The Systems Biology Markup Language (SBML)*

When modeling a system composed of several biochemical species, each represented by a differential equation, the ODE based methods require the simultaneous solving of all the differential equations representing the system.
Software implementing these methods generally have a graphical interface that allows the user to enter the biochemical reactions and kinetic constants, which are used by the instrument to automatically generate and simulate the model. In certain cases, advanced features allowing to view and analyze the models and visualize the simulation results are available. Over the years, various simulation environments have been created, such as Gepasi, E-CELL, MATLAB, COPASI, NEURON, and STEPS; each of them use its own template, so it is very difficult to reuse the same template in different

environment. For a model to be effective, it must ensure some degree of interoperability, so that it can run in different execution environments. For this purpose, it was necessary to achieve a standardization of the different model descriptions available. To solve this problem, the Software Platforms for Systems Biology forum was created, which included several representatives from the teams developing the software packages.

The forum decided at the first meeting to develop a simple, XML-based language for representing and exchanging models between simulation/analysis tools: the Systems Biology Markup Language (SBML) (Hucka 2003). An XML (eXtensible Markup Language) format was chosen as it was increasingly widespread and accepted as a standard data language for bioinformatics (Achard 2001). The base definition, SBML Level 1, is the result of analyzing common features in representation languages used by several simulators. Subsequent releases of SMBL have added several features and structures to level 1: the current version of SMBL is level 3.

An SMBL model is nothing more than a serie of one or more lists, each of which contains the elements that make up the model. In these lists information on compartment, species, reactions, parameters, unit definition and rules of the model can be found. A software package can read a model expressed in SBML and translate it into its own internal format for model analysis.

It has the classic characteristics of an XML file, where each element consists of a matched pair of start/end tags enclosed by '<' and '>' characters. The first line contains a particular sequence of characters (beginning with '<? xml') declaring the rest of the data stream as conforming to the XML encoding standard. The sbml element, starting from line 2, defines the SMBL model and the version with which it was created, allowing it to be distinguished from others. Inside it is the model subelement, which has a single attribute that specifies the name of the model, and a series of different subelements, each one acting as a container for all the elements of which it is made up (Hucka 2003).

Over the years a large database that collects all the mathematical models created in the SBML format has been created. It's named Biomodels, available at http://www.ebi.ac.uk/biomodels/ . It is divided in two main branches:

1. Literature-based models composed of models derived from literature. Prior to be added to the database, models are curated and annotated. In this way, all the model components are cross-referenced to external database information. The models are evaluated for compliance with MIRIAM guidelines to ensure reproducibility of simulation results and supply of adequate provenance informations.

2. Path2Models, containing automatically generated models of pathway resources such as KEGG, MetaCyc and BioCarta. They can be supplemented with kinetic information fetched from dedicated databases or predicted from the pathway structure. These models have been classified on the basis of the resource from which they were generated: there are 'metabolic' models, referring to quantitative, kinetic metabolic pathways, 'non metabolic' models, referring to qualitative, logical non-metabolic pathways, and 'whole-genome metabolism', referring to 'genome-scale metabolic network reconstructions'.

BioModels currently hosts more than 140000 models, including over 1200 models directly derived from literature (Chelliah 2015).

*1.1.5 Starting Protein Concentrations*

To simulate this type of reactions it is necessary to know the initial values (at time zero) of the species involved in the model. Protein measurements are key data for the development of a mechanistic models (Erickson 2019). Techniques for evaluating the proteome of a cell line are becoming increasingly efficient, faster, and affordable. Analytical approaches based on mass spectrometry (MS) demonstrated the ability to evaluate the protein profile in a comprehensive fashion, even of multiple mammalian cell lines in parallel (Geiger 2012). In addition, other proteomic analysis methods have been developed which eliminate the need for references, simplifying protocols and making them suitable for large-scale proteomics approaches (Wisnievski 2014). A recent study conducted a proteomic analysis on several human mammary epithelial cell lines (HMEC), fibroblasts and tumor cell lines in parallel, using ultra-sensitive targeted proteomic approach. The aim of the study was to quantify the different protein species involved in the Erk cascade. The analysis revealed that most of these proteins have a similar level of expression between different cell lines (Shi 2016). Through approaches like this, proteomics reinforces the reliability of models applied to different cell lines. Today, phosphoproteomic techniques are becoming increasingly important, as protein phosphorylation is important in many cellular regulatory processes and in signal transduction pathways. These innovative methodologies were applied to study serine-threonine family kinases, identifying proteins with decreased phosphorylation following addition of Egfr, PI3K, mTOR, and Mek inhibitors (Moritz 2010).

Other interesting information useful to construct models, can come from the analysis of the cellular transcriptome. The next-generation RNA sequencing (RNA-seq) technologies have become a powerful tool to study the presence and quantity of RNA molecules in biological samples and have revolutionized transcriptomic studies. RNA-seq data are highly reproducible

for both technical and biological replicates and the quantification of expression levels is highly accurate.

In hosting laboratory a standardization method that calculates intracellular protein concentrations starting from quantitative data derived from both mass spectrometry-based proteomics experiments and RNA-seq-based experiments was developed (see material and methods). In this way, the integration of experimentally determined data of a different nature can offer a solution in the search for the missing parameters necessary for the construction of effective models.

## 1.2 The Erk1/2 cascade

Cell motility plays an important role in many physiologic and pathologic processes, such as tissue remodeling, metastatic tumor cell migration, stem cell mobilization and homing. The main challenge in the field of cell motility study is to develop a comprehensive physical description of how and why cells move. To this purpose it is necessary to find new ways to model the biological properties of cells. This goal can be achieved through the development of new experimental techniques aimed to extract physical information from living systems, and to define, on the basis of them, new theoretical models, through which better understanding this process.

The mitogen-activated protein kinases (MAPKs) have been considered one of the crucial regulators of cell motility. Several studies have highlighted that Erk signaling also regulates cell motility through the phosphorylation of various components that are involved in cell motility machinery (Tanimura 2017). In addition to cell motility, Erk signaling also plays critical roles in the regulation of various cellular functions, including proliferation, differentiation, and survival in response to extracellular signals. Consistently, with the diverse roles of Erk in cellular regulation, its constitutive activation results in tumour formation and progression (Yarden 2001), and thus inhibition of Erk signaling is believed to be a promising strategy for cancer therapy (Sebolt-Leopold 2004).

### *1.2.1 The cascade*

The Extracellular signal-regulated protein kinases 1 and 2 (Erk1/2) signaling pathway represents the prototypical and most extensively studied subfamily of MAPKs. It refers to a module of three kinases that sequentially activate each other by phosphorylation in response to a wide range of extracellular stimuli, such as cytokines, growth factors, cellular stress and cell adhesion. The pathway employs one of the more generic signaling patterns found in biological signal transduction: a loop formed by a kinase phosphorylating a target protein and an opposite phosphatase which is in

charge of dephosphorylating the same target. Therefore, the pathway plays a central role in many key cellular processes, such as proliferation, cell differentiation and survival, and in the cellular adaptation to chemical and physical stress.

The cascade starts by the interaction of an extracellular stimulus with a multimolecular complex of receptors such as receptor tyrosine kinases (RTKs) and G protein-coupled receptors (GPCRs). The epidermal growth factor receptor (Egfr) belongs to a widely expressed tyrosine kinase receptor family, frequently mutated or overexpressed in cancer. On the extracellular membrane four different monomers, also named Erbb1-4, can be expressed, and can form both homo- and heterodimers. Two members of this family, i.e. ERBB2 and ERBB3, are non-autonomous: the former lacks the capacity to interact with the ligand, while the latter shows a defective kinase activity. However, by heterodimerization with other members of the family, they can evoke strong intracellular signals (Citri 2006). In the absence of their ligand, an equilibrium between Egfr inactive monomers and inactive dimers exists (Chung 2010, Jura 2009). EGF binding stabilizes receptor conformations that expose an extracellular dimerization interface, triggering the accumulation of active Egfr dimers (Ferguson 2003, Ogiso 2002). After Egfr dimerization, the tyrosine kinase domain of one Egfr moiety phosphorylates several Tyr residues in the partner moiety.

The extracellular domain of Egfr can adopt two conformations, closed and extended, the latter being dimerization-competent. EGF binding stabilizes the extended conformation, thus favoring dimer formation, and allows the Egfr kinase domain to interact with its Tyr substrates (Zhang 2006). The kinase activity is antagonized by phosphatase acting at the membrane already at the very early stages of Egfr signalling. Young (2019) found that PTP1B acts by Egfr dephosphorylation in order to finely tuning its signalling pathway (Young 2019). Liang (2018) investigated whether the formation of Egfr active dimers in the absence of EGF could be sufficient to the downstream activation of Ras by developing a synthetic Egfr analogue peptide. Interestingly, they found that the presence of EGF is necessary to the conformational changes, like oligomerization and reorganization on the cell surface, needed for an efficient signaling through the Sos-Ras-MAPK pathway (Liang 2018).

The cellular responses to different concentration of a stimulating specie can be expressed in both linear and nonlinear modalities. While the former is an almost obvious model of dose-response intracellular signalling, the latter shows a threshold-controlled mechanism. An example of such modalities concerns the EGF-dependent internalization of Egfr. The activated receptor is internalized through both clathrin mediated endocytosis (CME) and non-clathrin endocytosis (NCE). The first is a saturable mechanism, active at all

concentrations of EGF, coupled to the recycling of the receptors on the cell surface and thus to sustainment of signalling. NCE is observed only at high concentrations of EGF and is largely devoted to commit the receptor to lysosomal degradation; thus, the intense activation of NCE above a certain EGF threshold, might regulate the net signalling output, in response to increasing EGF concentration, in a nonlinear way (Sigismund 2013). Investigating this field, Sigismund at al. demonstrated how the ubiquitination of the Egfr at the plasma membrane is threshold controlled. This threshold controls the modality of Egfr internalization at the PM, and thereby enables cells to translate quantitative inputs (EGF concentrations) into qualitatively different internalization mechanisms. At low EGF concentrations the Egfr is scarcely ubiquitinated and internalized primarily through CME. At high EGF concentrations, Egfr is mostly endocytosed through NCE, as the receptor becomes ubiquitinated. Thus, the ubiquitination threshold controls receptor fate and the balance between maintenance versus attenuation of Egfr signalling (Capuani 2015).

Phosphorylated tyrosines serve as binding sites for adapters proteins that contain a Src homology 2 (Sh2) domain, which is expressed in Grb2 (growth factor receptor-bound protein 2); with its SH3 domain inteacts with the guanine nucleotide exchange factor Sos1/2, the best characterized route of Ras activation that occurs at the plasma membrane (Cargnello 2011).

The Ras proteins superfamily includes three members: H-Ras, K-Ras, and N-Ras. Ras mainly localizes on the cytoplasmatic surface of the plasma membrane (Hancock 2003). In the inactive state, Ras binds GDP. After the cells are stimulated, Ras receives activating inputs from guanine nucleotide exchange factors (GEFs, such as Sos1/2) and deactivating inputs from GTPase activating proteins (GAPs). GEFs facilitate Ras conformational change from the inactive GDP-bound form to the active GTP-bound form by promoting nucleotide release. As intracellular GTP/GDP ratios are estimated to be greater than 10 under nutrient replete conditions, a free Ras G domain is more likely to bind GTP than GDP. GAPs (e.g., Rasa1, NF1) prompt the reverse transition by accelerating GTP hydrolysis. Thus, Ras GAPs have the potential to function as tumor suppressors, and loss-of-function mutations affecting Ras GAPs have the potential to be oncogenic (Erickson 2019).

It is important to focus the attention on two important aspects to better understand the importance of these proteins in the overall cascade:

1. The influence of Sos dynamics in MAPK signaling is underscored by the relatively low Sos protein copy number in the Egfr-MAPK pathway. While the absolute abundance of most core proteins was between 50,000 and 70,000 copies per cell, adaptors concentrations like Sos were found at far

lower amounts (2000 to 5000 copies per cell). MAPK signaling showed saturation in all cells between 3000 and 10,000 occupied EGFRs, consistent with the idea that adaptors limit signalling. Shi (2016) shows how the relative stoichiometry of core MAPK pathway proteins is very similar across different cell types, with cell-specific differences mostly restricted to variable amounts of feedback regulators and receptors. The low abundance of adaptors relative to Egfr could be responsible for previous observations that only a fraction of total cell surface Egfr is capable of rapid endocytosis, high-affinity binding, and mitogenic signaling.

2. Extensive work has shown that the duration of the Ras signal is a key determinant of cell fate in the sense that growth factor activation of Ras must be transient to promote a proper proliferative response (Henning 2016).

Ras substrates are able to recognise the conformational changes in Ras induced by this GDP/GTP exchange and to interact with Ras-GTP with higher affinity than with Ras-GDP. Ras does not chemically modify its substrates but regulates their activity by recruiting them to the activator localized to the plasma membrane near it. A primary target of activated Ras during growth-factor stimulation is Raf, a serine/threonine kinase localized in the cytoplasm. Three Raf proteins are found in mammalian cells: Raf-1, A-Raf, and B-Raf. The first is the most widely expressed, with significant protein levels detected in all cell type. In addition, mouse knockout studies have revealed that Raf-1 is required for viability, acting in the transmission of normal growth and developmental cues (Dougherty 2005). In a quiescent cell, Raf-1 exists in an inactive state in the cytosol. The inactive conformation is ensured by interaction between the N-terminal regulatory and the C-terminal catalytic domains, and by the binding of a 14-3-3 dimer that contact two phosphorylation sites, S259 and S621. Ras-GTP binding promotes the 14-3-3 displacement and the conformational changes that remove the Raf-1 autoinhibition and facilitate the phosphorylation of activating sites (Hibino 2011). Several mechanisms have been identified that induces the Raf inactivation. The most important involve the Protein phosphatase 5 (PP5) enzyme, that has been reported to dephosphorylate the critical S338 N-region site of C-Raf. Another important mechanism that attenuates the activity of all Raf family members involves a negative feedback loop in which active Erk phosphorylates the RAFs on multiple S/TP sites. Phosphorylation of the S/TP sites disrupts the interaction with Ras as well as Raf dimerization, and results in Raf proteins that are signaling incompetent (Terrell 2019).

Thereafter, Raf kinases transmit the signal to the MAPKKs, Mek1 and Mek2 (Mek1/2), by phosphorylation of two Ser residues in their activation

loop. MEKs are localized in the cytoplasm both resting and stimulated cells due to due the interaction with scaffold/anchoring proteins, such as KSR1 close to the plasma membrane. Beside their monomeric cytoplasmatic localization, MEKs can stay in that location also as heterodimers and this is important for the regulation of Mek2 activity. Upon stimulation the free MEKs, and some Mek molecules that are detached from their anchors, rapidly translocate to the nucleus. However, they usually remain there only for a short time due a rapid export signal by the NES/exportin system, giving rise to their apparent cytoplasmatic localization (Zehorai 2010).

In turn, Mek1/2 activate their only known substrates, native Erk1/2, which function as their sole downstream targets, suggesting that the Mek1/2 serve as the specificity-determining components of the Erk1/2 cascade. The Mek1/2 are the only kinases that can phosphorylate both regulatory Thr and Tyr residues of Erk1/2. Inactivation of ERKs requires the removal of either one or both phosphorylation. A large family of dual specify phosphatases, the DUSPs, can inactivate ERKs. A coordinated action of these phosphatases is important to shape the temporal activity for proper mammalian development and growth (Wortzel 2011).

In resting cells, Erk is localized primarily in the cytoplasm. This localization is mediated by a set of anchoring and scaffold proteins that bind the Erk common docking (CD) motifs and retain it diffused in the cytoplasm or in cytoskeletal elements, surface of organelles. Erk phosphorylation causes not only kinase activation, but also conformational changes. In this way ERKs, free from their anchoring proteins, can shuttling to other parts of the cells, including not only nucleus, but also mitochondria, endosomes/lysosomes. Interestingly, about 30% of Erk molecules are not affected by stimulation and remain attached to their anchoring proteins upon stimulation. These molecules are not anchored via their CD nut rather via loop6, important for anchoring to cytoskeletal elements. This binding is important to regulate the intensity and targets of Erk signal upon stimulation (Wainstein 2016).

Erk nuclear shuttling involves multiple steps. First, Erk is phosphorylated on two Ser residues (Ser244 and 246 in ERK2) within the kinase insert domain by autophosphorylation and mainly by the kinase CK2. The phosphorylated residues, together with seven other adjacent residues, constitute the nuclear traslocation signal (NTS), which mediate the binding of Erk to importin7, which escorts Erk to the nucleus via nuclear pores. Here a Ran GTPase dissociates the complex, making Erk free to phosphorylate its targets (Plotnikov 2011).

*1.2.2 Erk1/2 role in cell proliferation and tumour*

Erk activation is necessary for the cell cycle progression induced by growth factors like EGF, especially in the progression from G1 phase to S phase. The first clue of this came from experiments using synthetic Mek inhibitors, like PD98059, which inhibits both Mek1/2 and Mek5, and was better confirmed by using more specific inhibitors like PD184352, capable of discriminating between Mek1/2 and Mek5: at concentrations below 1 micromolar it is able to completely inhibit Mek1/2, but it does not affect Mek5 function.

Furthermore, serum-induced cyclin D1 expression and DNA synthesis were inhibited by low concentrations of PD184352. The key step for quiescent cells to undergo cell cycle entry is the formation of an active cyclin D-CDK4/6 complex, formed when newly synthesized cyclin D associates with existing CDKs. CDK4/6 kinase activity releases E2F family of transcription factors from Rb repression inducing expression of a second class of G1 cyclins, cyclin E required for S phase entry. It is worth noting that the stimulatory effects of cyclins can be counteracted by CDK inhibitors (CDKIs). Repression of CDKIs is also required for G1/S phase transition, while elevation of the level of cyclin D1 is not sufficient to induce cell cycle entry. Erk activation acts at several levels in order to promote the activity of CDKs in late G1, and in the inhibition of CDKIs. First, Erk leads to c-Fos and Fra-1 expression, which act indirectly to induce cyclin D1 transcription. Second, activation of Erk1/2 enhances c-Myc protein stability as a result of direct phosphorylation of Ser 62. c-Myc is a member of the Myc family of transcription factors that plays a central role in regulating cell growth, cell cycle progression and apoptosis. c-Myc heterodimerizes with its obligate partner Max to activate or repress transcription of a wide set of genes. c-Myc participates directly in the transcriptional induction of the genes codifying for cyclin D1, Cdk4, p21 and Cdc25. It is noteworthy that Erk1/2 signaling and induction of c-Myc are both necessary to drive cells from G0 to late G1 phase (Meloche 2007).

Erk plays a role in multiple pathways concerning cell proliferation (Chambard 2007):

- Erk activation instructs the cell to synthetize more pyrimidine nucleotides, by increasing the activity of CAD by phosphorylating Thr 456 of carbamoyl phosphate synthetase (CPS II), part of the multifunctional enzyme CAD, which is involved in a rate-limiting step of the de novo synthesis of pyrimidine nucleotides pathway;
- Upon Erk stimulation, transcription of the ribosomal RNA genes by RNA polymerase I is rapidly activated.
- The MNK1 protein kinase, direct substrate of Erk, is responsible for inducing phosphorylation of the eukaryotic translation initiation factor 4E

(eIF4E) on Ser 209 following cell stimulation. This phosphorylation increases translation by incrementing the affinity of eIF4E for capped mRNA.

As expected, Erk/MAPK signaling pathway plays an important role in cancer, since it are composed of immortal, constantly proliferating cell. Cancer-associated alterations of MAPK signaling derive from its effector's mutations, in both form of constitutive activation and continuous deregulated signal transduction. This is especially important in solid tumours, which can be characterized by the mutations affecting the genes coding for Ras, Raf, Erk and Mek. They are listed in order of decreasing probability of being involved. For example, Ras-small G-protein is mutated in about 34% of colorectal cancers, whilst mutations of Mek or Erk are found in about 3%. Ras mutations are also relevant as a predictor of tumour aggressiveness and poor prognosis (Burotto 2014).

*1.2.3 Erk1/2 role in tissue invasion and metastasis*

A mutated Ras is found in approximately 50% of metastatic tumours. Cancer metastasis is a phenomenon which occurs in sequential steps (a) detachment from the primary tumor mass; (b) digestion the surrounding extracellular matrix (ECM) and migration through it; (c) penetration into local blood or lymphatic vessels (intravasation), and transportation by the blood or lymph throughout the organism; (d) arrest in the narrow lumen of small vessels, breach the vessel wall, and transmigrate into the extravascular space (extravasation); (e) adaptation to the new anatomic site and outgrowth (Welch 2019). Tumour cells break away from the primary tumour, adhere to the basement membrane and become invasive. These cells infiltrate and grow in the surrounding stroma and enter the circulatory system, where most cells are spotted and killed by the immune system. A small number of tumour cells with strong survival ability reach the target organ and continue proliferating, forming new metastases in the same manner as the primary tumour (Guo 2020). With respect to focal adhesions, Erk activation is associated both in time and space with integrin clustering during cell spreading, which may play a role in regulating the focal adhesion assembly process. Active Erk may be targeted to newly forming focal adhesions after integrin engagement. Both active Erk targeting and cell spreading may be impaired by UO126, a Mek1 inhibitor (Klemke 1997).

Components of the integrin to focal adhesions pathway include several Erk substrates that might be responsible for Erk-induced peripheral effects on cell migration, including myosin light chain kinase (MLCK). The mitogenic signals generated by integrins or surface receptors, impact the actin-myosin cytoskeleton and temporal-spatial organization of cell adhesion contacts with

the extracellular matrix. Previous studies indicate that Erk activates the cell's motility machinery by enhancing MLCK activity leading to increased MLC phosphorylation and enhanced cell migration. MLCK represent a key regulator of cell motility and contraction. It has also been shown that this MAPK-induced cell migration can be blocked by selectively interfering with MLCK activity, suggesting that MLCK is downstream of MAPK and is required for cell migration. Active Erk phosphorylates MLCK that in turn induces phosphorylation of the myosin light chain, promoting myosin ATPase activity.
Thus, MLCK activation might be involved in turnover of focal adhesions and extension of membrane protrusions at the front of polarized cells, which are important for cell migration. Inhibiting the Erk pathway negatively affects MLCK, MLC phosphorylation and disrupts cell migration. On the other hand, expression of active Mek1 promotes phosphorylation of MLCK and MLC and enhances cell migration in COS-7, MCF-7 human breast cancer and HT1080 fibrosarcoma cells (Reddy 2003). Egfr overexpression can circumvent the dependence on Erk for maintaining sustained cell motility. Two distinct pathways regulate cell migration in Egfr overexpressing cells such as MDA-468 breast cancer cells: MAP kinase (MAPK or Erk 1/2) and myosin light chain kinase (MLCK) pathways play a more significant role in regulating early stages of cell migration, and protein kinase-delta isoforms (PKC-d) and ROCK (Rho-kinases) pathways play the most significant role in regulating the sustained cell migration. Inhibition of Erk activity with Mek inhibitor PD 98059 or MLCK activity with ML-7 blocks early stages of cell migration (Sepe 2013, Barillari 2020).

During cancer cell invasion and metastasis, tumor cells digest the ECM molecules by synthesizing and employing a large variety of enzymes; among them, the matrix metalloproteinases (MMPs) are the main contributors to ECM degradation by tumor cells (Gialeli 2011).
Disruption of the integrity of the basement membrane is a key histological marker of tumor's transition to invasive carcinoma. The basement membrane forms a cellular support for tumors and is composed of a mix of extracellular matrix (ECM) proteins, including laminins, collagens and proteoglycans. In response to extracellular stimuli, phosphorylated Erk translocates from cytoplasm to the nucleus, where Erk activates various transcriptional factors like activating protein-1 (AP-1), as a result of the induction of the c-fos gene. AP-1 is involved in expression of matrix metalloproteinases (MMPs), which are proteolytic enzymes that hydrolyze the extracellular matrix (ECM), one of the most important processes in the invasion and metastasis of cancer cells (Maeda-Yamamoto 2003). MMPs derive their name from the finding that they are metallo-enzymes able to degrade ECM, and therefore have been implicated in progression and invasion of cancer in addition to their involvement in

normal tissue remodeling, wound healing and angiogenesis. In tumors, sustained MAPK activation could lead to enhanced induction of proteolytic enzymes in the surrounding environment, leading to destruction of ECM. In addition data showing that direct association of MMPs with specific ECM receptors provides spatial control of MMP activity and directional signals to the invading tumor cells are available. The activation of the Erk/MAPK signaling pathway can increase tumour invasion and metastasis by upregulating MMP expression, while inhibition of this signaling pathway can reduce tumour invasion and metastasis. Maeda-Yamamoto (2003) showed that inhibition of Erk phosphorylation in fibrosarcoma HT1080 cells resulted in inhibition of MMP−2 and MMP-9 expression in these cells. In the same way, Simon et al (1999) found that in oral cancer cells, Erk1/2 activation downregulates MMP-9, thus reducing invasiveness.

Cell morphological alterations and migration, related to expression of cytoskeletal and microfilament-related proteins, occur during tumour metastasis. Several studies showed that protein phosphorylation is associated with regulation of the microfilament cytoskeleton. The human colon cancer cell line SW620 showed a larger number of intracellular microfilaments and a longer migration distance upon treatment with hepatocyte growth factor (HGF) compared with a control treatment. The study showed that HGF enhanced cell migration by activating the Erk/MAPK signaling pathway, thus promoting the invasion and metastasis of tumour cells. Bray et al (2001) demonstrated that Erk/MAPK signaling pathways transduce extracellular signals and regulate the expression of transcription factors that cause cytoskeleton deformation and enhance tumour invasion and metastasis. Blocking the Erk/MAPK signaling pathway may inhibit the role of extracellular signals that promote cell movement, which inhibits tumour invasion and metastasis.

Recently, a novel Erk substrate was characterized: epithelial protein lost in neoplasm (EPLIN). It was originally identified as the product of a gene transcriptionally down-regulated or lost in a number of human epithelial tumor cells, including oral, prostate, and breast cancer cell lines. When present, EPLIN cross-links and bundles actin filaments, thereby stabilizing actin stress fibers. In addition, EPLIN inhibits Arp 2/3 complex-dependent nucleation of new actin filaments. These events result in overall reduction of cell motility. Erk is able to regulate EPLIN activity by phosphorylation of Ser 360, Ser 602, and Ser 692 at the C-terminal region. These phosphorylations reduce EPLIN affinity for actin filaments, and cause destabilization of stress fibers and reorganization of the actin cytoskeleton in protrusive structures, eventually enhancing cell migration. A similar effect may be observed by RNAi-mediated silencing of EPLIN, which also results in enhanced cell motility. On the other hand, expression of an EPLIN mutant, not phosphorylatable by Erk, prevents

stress fiber disassembly and membrane ruffling and inhibits cell migration (Han 2007).

## 1.3 State of the art on Erk 1/2 models

The Erk cascade is one of the most important cellular signaling pathways and has been the subject of intensive study in the laboratory and, more recently, through systems biology approaches.

### *1.3.1 The first simple models*

The first mathematical models developed focused on the study of the properties and behavior of the cascade. The first model was published in 1996 by Huang and Ferrell (1996). Their model starts from the activation of Raf by Ras up to Erk activation, where the dual phosphorylation of Mapkk1/2 and Erk occur by two-step distributive mechanism. They assumed the initial species concentrations by experimental measurement done in Mos and Xenopus oocytes, and in some cases, they have estimated them. The kinetic parameters were set in this way: they arbitrary took all the Km values to be 300 nM, based on the value measured for the phosphorylation of p42 MAPK/Erk2 by active MKK-1. They subsequently varied all of these Km values and concentrations over a 25-fold range. The input stimulus to the cascade was taken to be the concentration of Ras. The model shows how the Erk cascade exhibit ultrasensitivity, i.e. a non-linear sigmoid activation curve, with the degree of ultrasensitivity increasing as one moves down the cascade.

Models now routinely incorporate growth-factor receptors that subsequently activate the Erk cascade. Egfr represents the receptor that most commonly activates the Erk cascade and is the major growth factor receptor incorporated in the Erk cascade models. This is because the Egfr system has been well studied, is expressed in a sensible manner in several cell lines, and good antibodies and molecular reagents are widely available, which allow for a variety of quantitative studies. Kholodenko (1999) studied the short-term pattern of cellular responses to epidermal growth factor (EGF) in isolated hepatocytes combining experimental kinetic analysis and computational modeling. They developed an ODE-based mathematical model of the Egfr signalling with 25 reactions involving 23 different species, that describe for the first time the binding with the receptor with EGF, the dimerization and the receptor auto-trans-phosphorylation. The model includes three adaptor proteins that can directly interact with phosphotyrosine residues on Egfr (Shc, Grb2 and PLCγ). The kinetic parameters in the model were extrapolated from the scientific literature and/or derived from basic physical-chemical quantities. In order to validate the model before analysis, a number of ‘wet’ laboratory experiments were performed, such as time courses of Egfr phosphorylation

induced by several EGF concentrations. The simulation was then compared with these data to demonstrate the validity of the model created. Analysis of the model showed a rapid, short-lived, pattern of Egfr phosphorylation. Sensitivity analysis of the model showed to be robust to significant changes in many of the rate constants of the protein interactions involved.

*1.3.2 Complete cascade models*

Kholodenko's model does not include the core Erk cascade in its description, but it has been used as a basis for many other models of the Egfr system which do include the core Erk cascade. By integrating the recruitment of Sos to Egfr at the plasma membrane where Ras is located, this model can be used to predict the transient activation of Ras and the Erk cascade, as expected for an EGF response. Schoeberl (2002) developed a model describing the behaviour of the EGF signal-transduction pathway to investigate the effects of receptor internalization on the Erk cascade, and also the signal–response relationship between the Egfr activation by EGF at the cell surface and the activation of downstream proteins in the signalling cascade. The model is one of the most comprehensive available (125 reactions involving 94 species), including a large range of dynamic processes. The model includes two pathways of Ras activation (Shc-dependent and Shc-independent), and the Egfr internalization by endocytosis. The majority of kinetic parameters were based on values published in the literature (some from the Kolodenko's receptor model), and the species' concentrations were based on the literature or based on laboratory experiments measurements. The model describes how the initial rate of change of receptor activation determines the cellular response to EGF. The initial velocities of Egfr activation, rather than the peak maxima, are important for signal propagation. As the EGF concentration varies, the initial speed of Egfr activation, rather than the maximum peak reached, influences the behaviour of the signaling. The presence of receptor internalization reactions allows the model to predict how they can attenuate the activation of the pathway at very high doses of EGF.

The researchers Arkun and Yasemi tried to understand how the pathway dynamics can be modulated to generate a diverse array of Erk dynamics that will selectively affect the cellular functions. They focused their attention on the how the negative and positive feedback loops affect the dynamic characteristics of the cascade. To generate their model, they subdivided the cascade in three subsystems, combining three different available models: the EGF-Sos subsystem from Kolodenko model, the Ras subsystem from Das (2009), and the MAPK subsystem from Qiao (2007) . The resulting model was integrating with several reactions that reproduce several positive and negative feedback loops. All the parameters and the species' concentrations involved were

converted to a single set of units, and some parameters were fixed at their literature values. In this way, they show combination of positive and negative feedback loops modulates the Erk behavior (in terms of duration, amplitude, stability and oscillations) selectively to obtain the specific biological effect (Arkun 2018).

Brown (2004). studied the actions of nerve growth factor (NGF) and the EGF in rat pheochromocytoma (PC12) cells. Each of these growth factors stimulates Erk phosphorylation with distinct dynamical profiles. Considering how many of the parameters required remain unknown or at best represent estimates, they chose a different approach. They applied the statistical mechanics to extract predictions of the parameters for the model. An important feature of the approach involves the use of Monte Carlo methods in Bayesian sampling of model spaces. They showed that this approach can make useful biological predictions even in the face of indeterminacy of parameters and of network topology. The generated model consists of 13 different protein species involved in 16 biochemical reactions, which primarily utilize Michaelis-Menten kinetics. The model considers the Sos-Ras-RAF1 pathway leading to Erk activation as well as the Sos negative feedback and Akt negative feed-forward loops.

*1.3.3 Tumour cascade models*

Brown's model was used and expanded by other researchers later. In 2009 Orton proposed this model to investigate what effects various cancerous alterations (such as Ras or an Egfr mutation) had on signalling through the EGF activated Erk pathway. Within the original Brown model, they included the activation of Rap1 by EGF and its capability to selectively activate B-Raf, adding additional complexity to the pathway. The new model consists of 17 proteins involved in 31 reactions, which utilize primarily Michaelis-Menten but also mass action kinetics. The model predicts how different cancerous situations resulted in different signalling patterns through the Erk pathway, especially when compared to the normal EGF signal pattern; cancerous Egfr mutation and overexpression signals via the Rap1 pathway are the best target for drugs.

About 50% of melanomas have B-Raf activating mutations. It is an aggressive tumor of the skin with a poor prognosis for patients with advanced disease and it seems to be resistant to current therapeutic approaches. In this way it represents a good tumor model to investigate the activation of the MAPK cascade. Pappalardo (2016) considerably expanded the Brown model (48 species and 48 biochemical reactions) to study the biochemical and metabolic interactions in the PI3K/AKT and MAPK pathways potentially involved in melanoma development. Moreover, they introduced the features to

reproduce the effect of Dabrafenib inhibitor in the complex dynamics of the PI3K/AKT pathway and modeled two specific reaction: the normal drug degradation and the main effect of Dabrafenib in the inhibition of the B-Raf Mutated species.

Several studies have shown that the Egfr overexpression is very common in their the non-small-cell lung cancer (NSCLC), one of the deadliest and most difficult to diagnose forms of cancer. Bidkhori (2012) developed two mathematical models to relate to the different Egfr signaling in NSCLC and normal cells in the presence or absence of Egfr mutations. The model's reactions were designed in according to previous experimental observations and model (most of them described above). For the first time they simultaneously analyze several mutations in both Egfr and PTEN and over-expression of PI3K, Egfr, Akt, STAT3 and Ras in NSCLC Egfr signaling in one study. Their model shows how mutations in Egfr can increase the levels of pERK, pSTAT and pAkt. Over activation of Erk, Akt and STAT3 which are the main cell proliferation and survival factors is a promoting factor for tumor progression in NSCLCs.

## 2. Aim of the work

The aim of the project was to develop models which describe the behaviour of Erk1/2 cascade and to interface them with a simulation system which reproduces motility and proliferative behaviour of eukaryotic cell cultures, with the intent of studying cellular migration in pathological contexts like tissue invasion and metastasis.

# 3. Materials and methods

## 3.1 Motocell

Motocell (Cantarella 2009) is a bioinformatic tool which was first created as an online application for the evaluation of the motility of cell populations maintained in various experimental conditions. It can simultaneously perform time-lapse experiments, derive motility parameters from the comparative analysis of multiple experiments, and test the models' predictions of cell proliferation and movement. The system also includes additional tools which allow the acquisition and analysis of microscopy images and the graphical representation of datasets describing cell paths obtained by tracking procedures. Statistical analysis is also included: it consists of the evaluation of descriptive parameters (such as average speed and angle, directional persistence, path vector length) calculated for each step of the migration for the entire population as well as for each single cell; in this way the behavior of a cell population may be assessed as a whole or as a summation of individual entities.

Over the years, it was gradually expanded, until it became a development environment containing different integrated analytical modules, where to perform custom routines to extract and manipulate data, access the R environment, plot the results in tables and graphs. Among them, at the moment tools of systems biology are available such as models that predict the behavior of biochemical pathways.

## 3.2 Cultured cell movement simulation

The tool, developed in hosting laboratory, reproduces the behaviour of eukaryotic cultured cells observed in time-lapse microscopy experiments, reproducing migration paths of cultured cells under different experimental conditions. The tool is based on a single cell approach where individual cells are followed through the cell cycle, to generate a synthetic cell population with features comparable to the experimental data used as a reference. Different features of the cell are taken into account in the system, such as:

- The chemical gradient that influence the cell migration as generated by a reagent added to the culture, such as nutrients or growth factors. Combination of different gradients is also supported;
- Interaction between cells is taken into account, both in terms of attraction and repulsion;

All these features can be combined to reproduce a cell population behaviour in different experimental conditions. The tool is able to quantitatively tune the contribution of each movement feature in complex situations.

Originally, the cell proliferation simulation is based on the probability of undergoing mitosis calculated from a cumulative Weibull distribution model adapted to experimental informations produced by observing cell cultures in time-lapse experiments. Given a set of parameters, the use of the Weibull model in combination with random number generation, produces simulated populations that, whilst showing some variability, have a proliferation trend which closely resembles the one of the original experimental cell population.

### 3.3 Database of estimated protein concentrations

In hosting laboratory was developed a database of estimated protein concentrations representative of different cell types and experimental conditions. It is based on a standardization method that calculates intracellular protein concentrations starting from both quantitative data derived either from mass spectrometry-based proteomics experiments, and RNA-seq-based experiments. An automated system retrieve data from proteomics and transcriptomics online databases, and use these data to estimate the concentrations. based on an algorithm that estimates intracellular protein concentrations from literature data and converts them to values directly usable for systems biology approaches. First a method was developed able to normalize quantitative data derived from different experimental sources in a reliable way; the resulting protein concentrations were stored in a database specifically built that can be easily accessed programmatically by a model runner tool. Those data were then used to set up mathematical equations, which describe the biochemical processes under consideration, to create a cellular model that is then simulated to forecast changes over time and under different conditions, in concentrations and phosphorylation degrees in involved molecules.

| Cell line | Cell Type | Molecular source |
|---|---|---|
| NIH3T3 | Mouse fibroblasts | Protein |
| GAMG | Glioblastoma | Protein |
| HEK293 | Human embryonic kidney | Protein |
| LnCAP | Androgen-sensitive prostate adenocarcinoma | Protein |
| RKO | Colon carcinoma | Protein |
| U2OS | Bone osteosarcoma | Protein |
| HeLa | Cervix adenocarcinoma | Protein, transcriptomics |
| HepG2 | Hepatocellular carcinoma | Protein, transcriptomics |
| A549 | Lung adenocarcinoma | Protein, transcriptomics |
| MCF | Breast adenocarcinoma ER+ | Protein, transcriptomics |
| ECC-1 | Endometrium adenocarcinoma | transcriptomics |
| T-47D | Braest ductal adenocarcinoma | transcriptomics |
| HT-29 | Colorectal carcinoma | transcriptomics |
| MCF10A | Fibrocystic disease from breast | transcriptomics |
| TMR | Breast adenocarcinoma Tamoxifen-resistant | transcriptomics |
| HelaS3 | Subclone of HeLa | transcriptomics |
| 786-O RCC | Renal adenocarcinoma | transcriptomics |
| HCT116 | Colorectal carcinoma | transcriptomics |
| OVCAR5 | High grade ovarian adenocarcinoma | transcriptomics |
| A1847 | Ovarian adenocarcinoma | transcriptomics |
| HCC70 | Primary ductal breast carcinoma | transcriptomics |
| CR4 | Metastasis of colon carcinoma | transcriptomics |
| SK-BR-3 | Metastasis of breast adenocarcinoma | transcriptomics |
| C4-2B | Metastasis of prostate carcinoma | transcriptomics |

**Table I - Database of estimated protein concentration**. Cell lines available at the moment.

## 3.5 Software development

### *3.5.1 Use of objects*

The simulator development is based on object-oriented programming (OOP), a style of programming which includes the flexibility to create modules that do not need changes when a new type of object is implemented. New objects are created by extending the already present objects and inherit all of their characteristics. In OOP, every object is an occurrence of a class, which defines the properties and methods of an object. The OOP approach is diffusely used to compute modular softwares where different tools share the same objects and acquire specific characteristics and behavior.

### *3.5.2 PHP*

PHP is the main programming language subtending the development of the simulator. PHP is a scripting language which was born for the design of dynamic web pages and is strictly integrated with the Apache server. The relationship between PHP and the simulator software is that the former was occasionally used as a simple scripting language for web pages, but mostly as a standard language for developing multifile programs. The version currently in use is PHP 5.6.30.

### *3.5.3 R*

Simulations of biochemical pathways were performed by using R (web site: R), through Motocell, where a good software integration with the R environment was already made available. R is an open source environment for statistical computing and graphics, which runs on a wide array of platforms. R is a combination of different statistics packages and programming languages. In the Biochemical simulations were executed with the function *sim()* from the SBMLR package provided by Bioconductor. It is an interface to Systems Biology Markup Language (SBML), a standard language used for the definition of systems biology models and based on a text format (XML), which allows the sharing of models between different softwares. The software version used was R 3.3.2.

# 4. Results

## 4.1 Development of the biochemical model

Simulation of a cell that is followed for a number of hours during which it undergoes changes in cell cycle phases and, quite likely, changes in the surrounding environment, required a quantitative model, that goes beyond the scope of currently available models. To this aim, a new model was created, which, by using kinetic parameters and concentrations derived from literature or locally determined data, can simulate the Erk1/2 cascade and their effects on cell cycle, starting from an extracellular stimulus. The model was generated by decomposing the pathway into four linked functional subsystems:

1. The first subsystem describes the initial part of the pathway from the extracellular growth factor-receptor binding to the formation of the active receptor complex;
2. The second subsystem is necessary to sustain longer execution times and includes the mechanisms involved in internalization and degradation of the receptor complex;
3. The third describes the steps leading to the activation of Ras and Raf proteins;
4. The fourth subsystem describes the phosphorylations cascade of Mek1/2 and Erk1/2 and transfer of phosphorylated Erk to the nucleus.

### *4.1.1 The receptor binding and activation subsystem*

For the purposes of this study, epidermal growth factor (EGF), which is present in the growth medium of many eukaryotic cells, was chosen as the

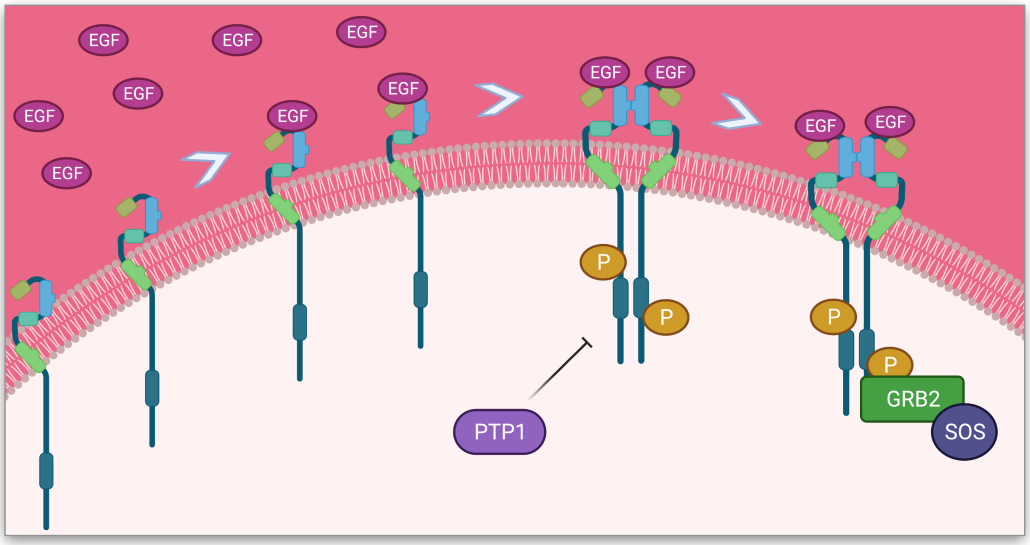



**Figure 1 - Receptor binding and activation.** EGF exerts its stimulating effects through its binding with Egfr, followed by the dimerization between two bound monomers of the receptor, and the dimer auto-trans-phosphorylation. The activated dimer then binds the adaptor protein Grb2 and the GEF protein Sos; it can also be dephosphorylated by the Ptp1 phosphatase. Created with **BioRender.com**

activating stimulus of the cascade. It binds to its receptor, the epidermal growth factor receptor (Egfr), whose activation starts the intracellular response (Fig. 1). In order to simulate these events, a set of reactions was defined (Fig. 2) and an in-depth study of the literature directed to identify acceptable kinetic parameters was conducted. The complete list of reactions and the relative kinetic parameters are listed in the Appendix.

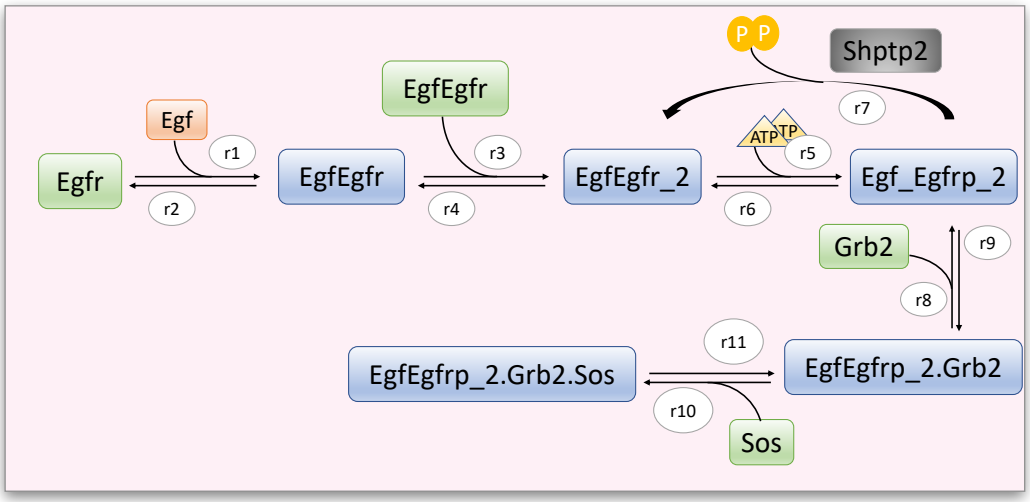



**Figure 2 - Reactions flowchart.**

To simulate the binding between EGF and Egfr, a basic landmark was the study from Berkers (1991), because it was conducted on HeLa cells, and biochemical assays directly measured binding to membrane receptors within living cells, also differentiating them in subtypes with different affinity for the ligand. The binding parameters calculated on the basis of this study were tested by simulating binding and release reactions between EGF and Egfr at different concentrations (Fig. 3). In order to confirm the effectiveness of the simulator software, Berkers' experiment was replicated, reproducing his experimental conditions (cell line, receptor concentrations, EGF concentrations). The results obtained with the software match closely with the results reported in the article, with higher EGF concentrations resulting in more receptors bound and shorter time needed to reach equilibrium. These results also confirmed the good performance of the simulator software.

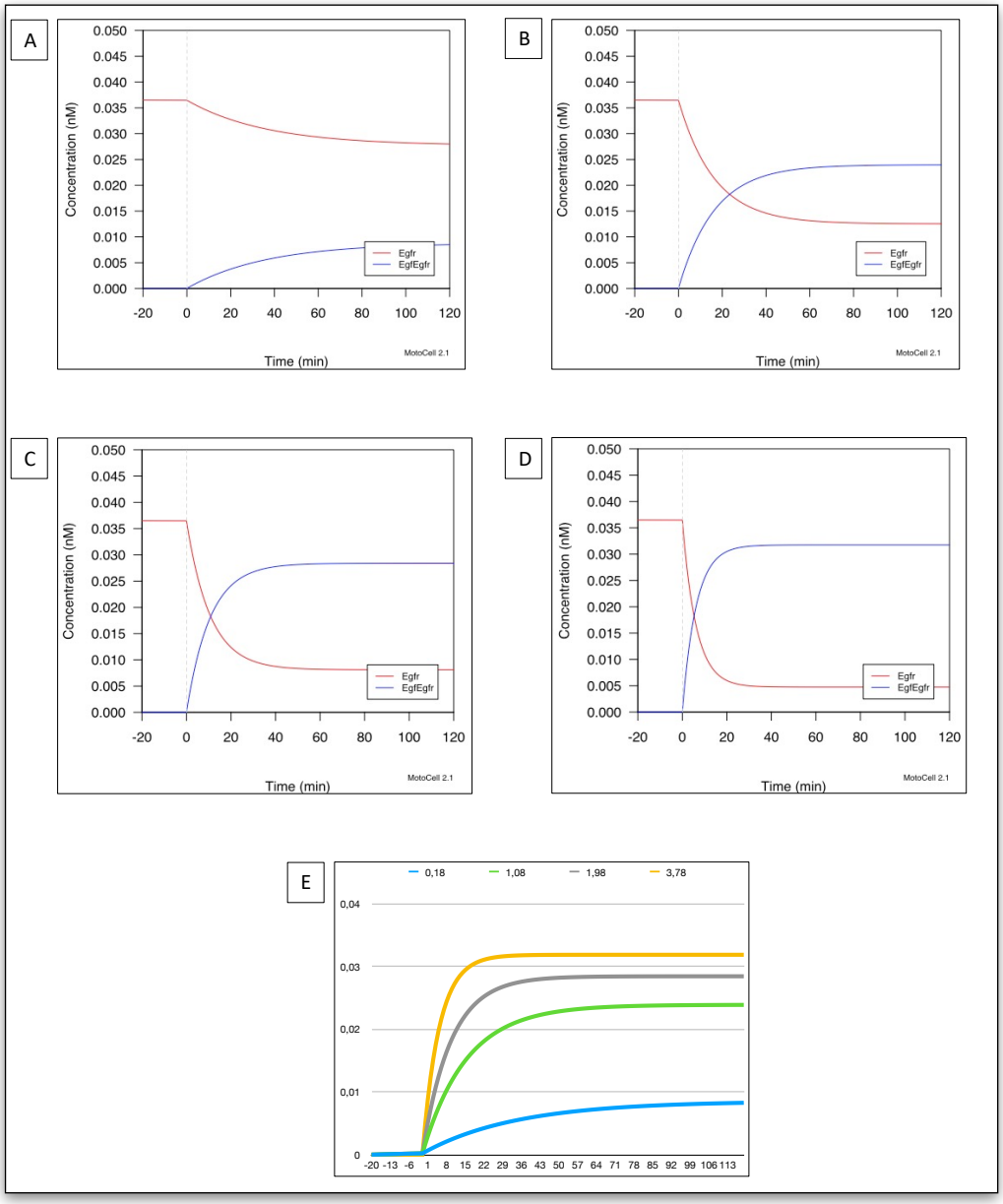



**Figure 3 - EGF-Egfr binding simulation**: The graphs show the association and dissociation between EGF and Egfr, which were evaluated at different EGF concentrations (0.18, 1.08, 1.98, 3.78 nM in Fig. 2 A-D). In each plot, the level of monomeric Egfr (Egfr, red lines) is compared to the concentration of the receptor bound (EgfEgfr, blu lines). The results for EgfEgfr were plotted together (Fig. 3E).

The following steps of Egfr activation after EGF binding were produced through extensive research of scientific evidence from diverse papers, which contain parameter calculations and models; the model by Kholodenko (1999), is one of the first models ever created and still represents a reference point for this kind of research. Figure 4B shows how the model produced in this way predicts a rapid phosphorylation pattern of Egfr:

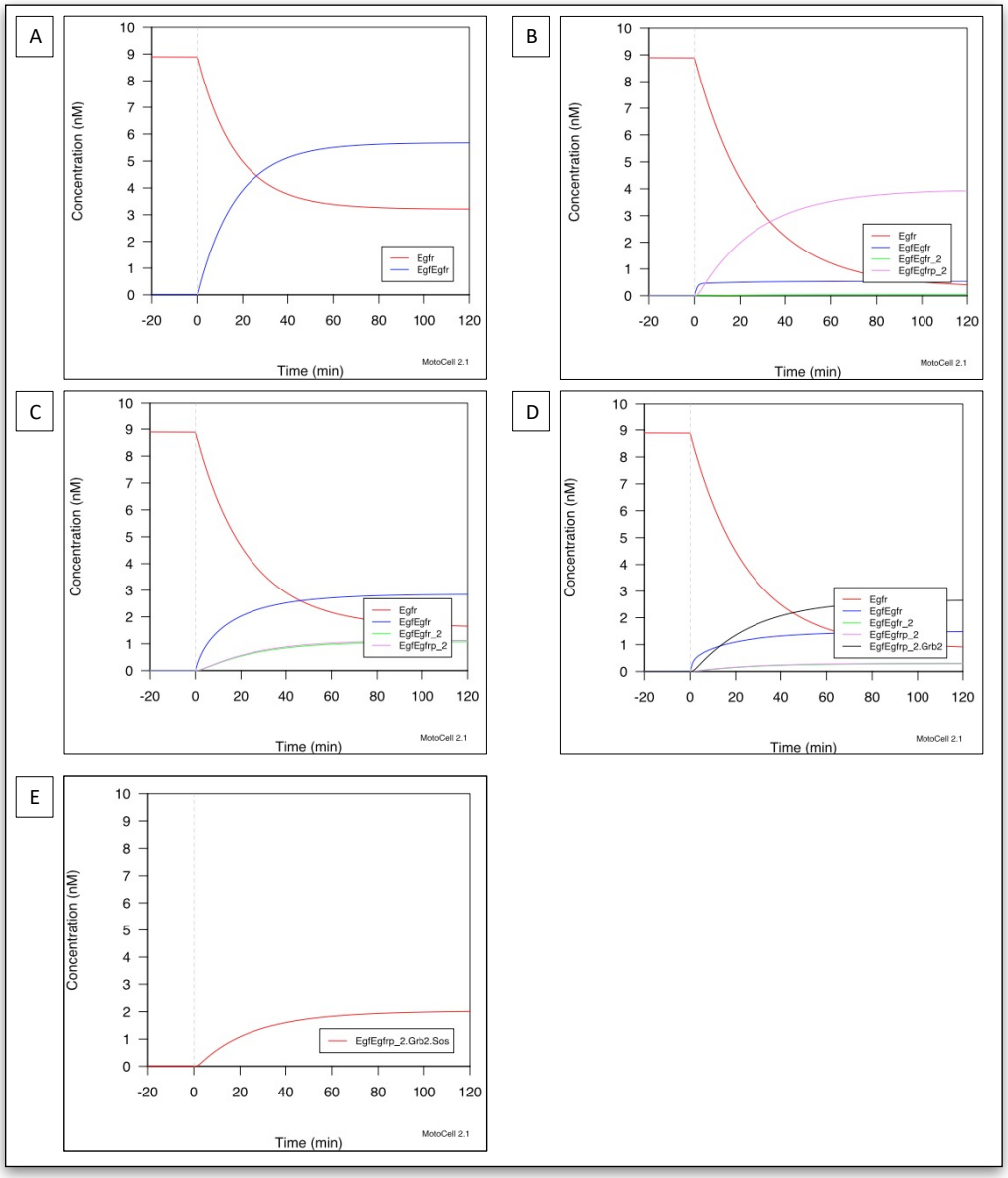



**Figure 4 - Receptor binding and activation subsystem in HeLa cells:** the plots show the predicted behaviour of the cascade's early steps, regarding the Egfr. Its binding to EGF (Fig. 4A) is followed by the dimerization of two bound monomers (EgfEgfr_2, green line) and the phosphorylation of the receptor (EgfEgfrp_2, purple line) showed in Fig. 4B. The effect of the deophosphorylation by phosphatase PTP1 on the complex is depicted in Fig. 4C, partially countered by the binding of Grb2 to the complex (EgfEgfrp_2.Grb2, black line, Fig. 4D). Next, the nucleotide exchange factor Sos binds the phosphorylated complex (EgfEgfrp_2.Grb2.Sos, red line, Fig. 4E).

implementing the dimerization and the phosphorylation steps shifts the chemical balance towards the products, which increases the receptor's saturation. The dephosphorylation step decreases the phosphorylated receptor level (Fig. 4C); on the other hand, binding with Grb2 prevents receptor

phosphorylation, but overall receptor activation still results toned down, as clearly seen from the comparison between Fig. 4D and Fig. 4B. This way, the introduction in the model of the PTP1 phosphatase and the adaptor proteins allows to reproduce not a fast dose-response effect, but an attenuated signal which can be measured over a longer time. The adaptor protein binding to the receptor is followed by the binding of the complex to the nucleotide exchange factor Sos (Fig. 4E).

*4.1.2 Receptor degradation subsystem*

A path for receptor internalization and degradation, based on the process described in Fig. 5, was implemented, in order to confer the ability to describe the cascade over longer times, typical of a eukaryotic cell culture. The

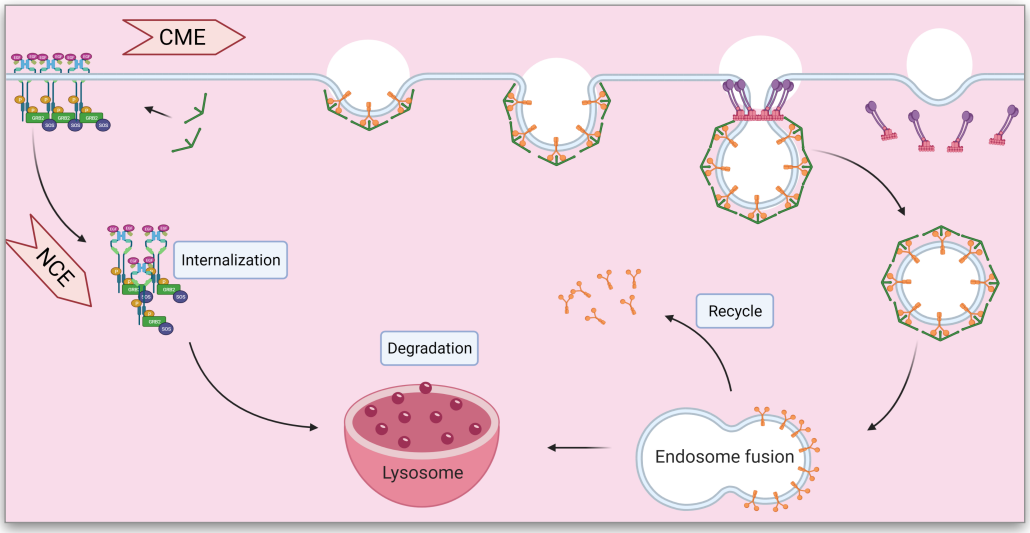



**Figure 5 - Receptor degradation processes.** The subsystem simulates the two internalization processes acting on the receptor complex (NCE and CME), followed by the receptor degradation, and its recycle. Created with **BioRender.com**

path includes two distinct processes acting on the activated receptor complex, the constitutive non-clathrin endocytosis (NCE) mechanism, simulated with an irreversible constant flux law, and a clathrin-mediated endocytosis (CME) one, requiring a more complex simulation including additional molecular species. The reaction scheme describes the trafficking processes of phosphorylated receptor complex, and follows the approach described in Fig. 6, where the two mechanisms act on the various phosphorylated Egfr species present on the plasma membrane. The main outcome of the NCE process is lysosomal degradation, with only a small fraction (10%) of the receptor recycled, in agreement with data reported in literature (Sigismund 2008). CME was instead

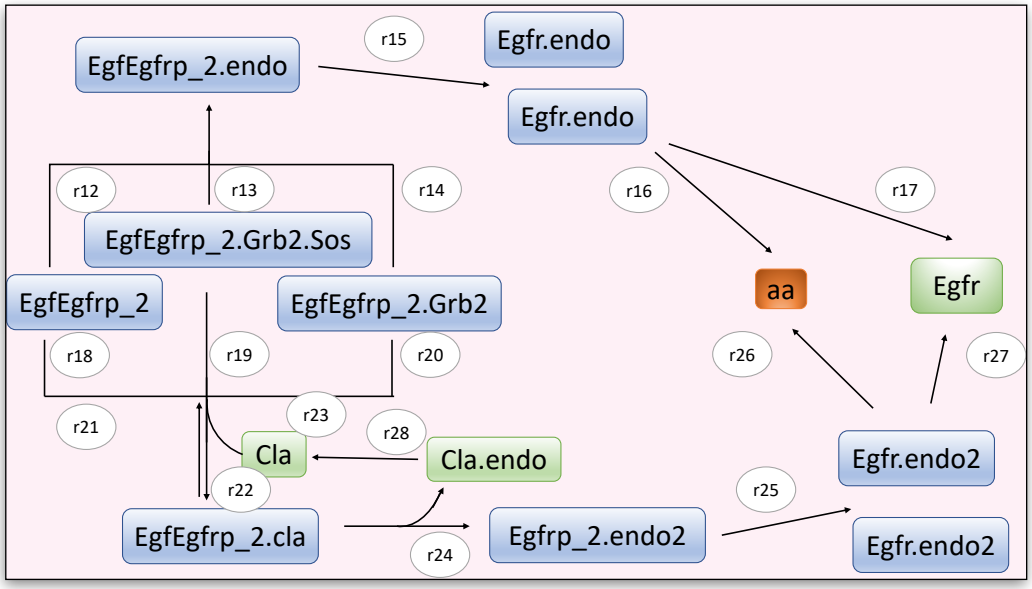



**Figure 6 - Receptor trafficking**. Here is represented the reaction scheme describing the two internalization mechanisms of the phosphorylated Egfr receptor. NCE, on the top, was modelled as an irreversible constant flux. The internalized receptorial dimer (Egfrp_2.endo2) undergoes dedimerization (r15); the monomers are then degraded at lysosomal level (r16). This scheme takes into account also the small percentage of receptor which is recycled to the cell surface (r17). On the other hand, CME starts with the main reversible reaction between clathrin and receptor (r22/r23). The Egfr-clathrin complex is internalized; then, after the clathrin detachment (r24), the receptor undergoes dedimerization (r25), and the monomers proceed to recycle (r26) or degradation (r27). The clathrin molecule goes back to the cell membrane to start another reaction (r28).

modelled through a main reversible clathrin-binding reaction, which leads to membrane recycling and receptor degradation.

The effectiveness of the model in reproducing the different kinetics of these processes, was tested using three different Egfr activation levels (Fig. 7), corresponding to 0.5 nM, a level typical of 10% FBS cultured cells (left), 40 nM, for cultured cells stimulated with 1 nM EGF (center), and 200 nM, representative of cell lines characterized by Egfr over expression (right). The results show that the model correctly predicts receptor decay via the NCE path, with the same halftime at various Egfr concentrations (Fig. 7A). The simulation of CME, instead, shows a faster but saturable behaviour (Fig. 7B). Figure 7C shows the combined effect of the two mechanisms: at low EGF concentrations receptor decay via CME is more effective than NCE, but at higher concentrations it becomes less relevant, being limited by clathrin concentration.

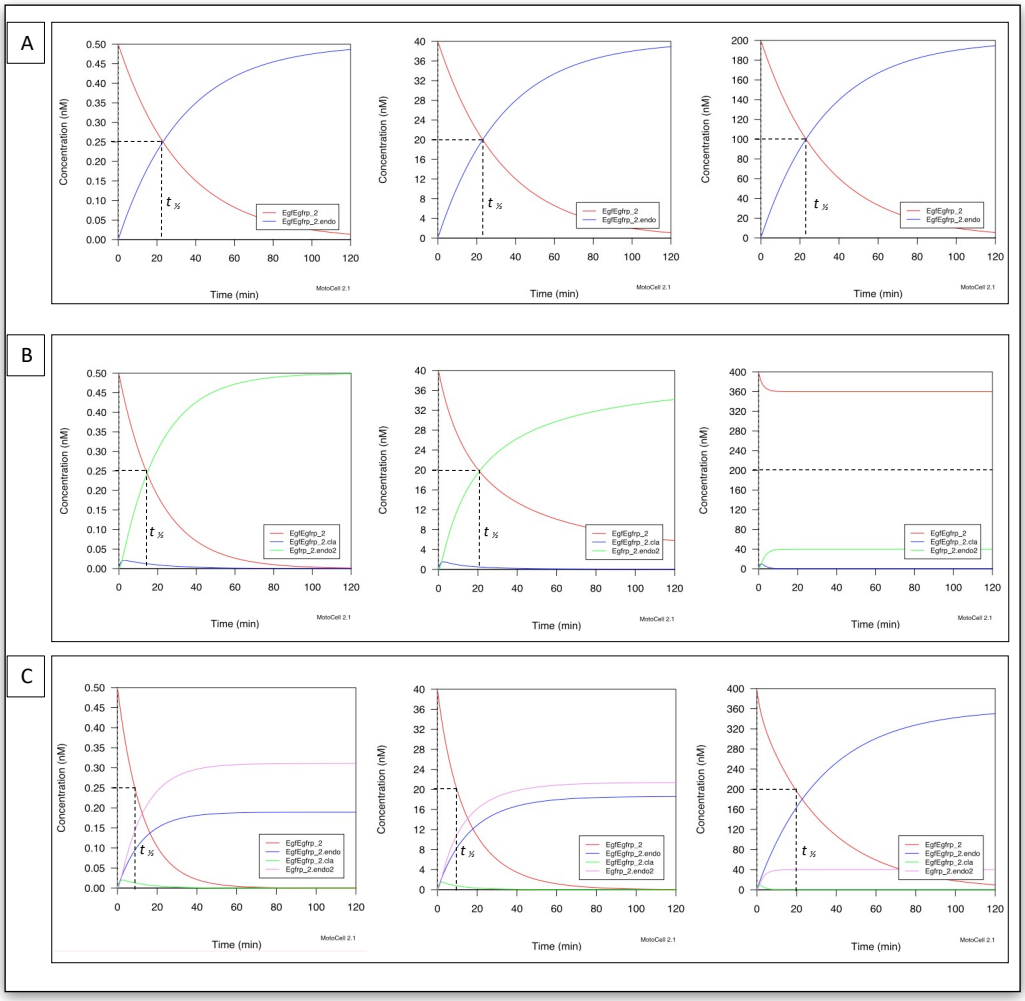



**Figure 7 - The receptor degradation and recycle subsystem**: different Egfr levels (0.5 nM on the left, 40 nM on the center, 200 nM on the right) were used to evaluate the impact of the two internalisation processes on the receptor behaviour. Fig. 7A show the impact of the NCE (expressed by the internalisation receptor EgfEgfrp_2.endo in blue), while the Fig. 7B show the effect of the CME (expressed by the internalizated complex EgfEgfrp_2.cla in blue and the internalised receptor Egfrp_2.endo2 in green). Fig. 3C represent the simulation merge between the two processes.

Figure 8A and B shows, for HeLa and A549 cells respectively, the results of the execution of the complete subsystem including degradation of internalized receptor as well as recycling to the plasma membrane. The plots show that at low EGF concentration, the receptor recycling process has a stronger effect, in agreement with the available experimental evidence (Carpenter, 1976). At an EGF concentration of 1 nM, the activated receptor concentration produces a peak after around 20 minutes, and then slowly decays within about six hours, both in HeLa and A549 cell simulations. The results obtained are consistent with the experimental data by Carpenter and Cohen (1976) who show that, by treating human fibroblasts with iodine tagged EGF,

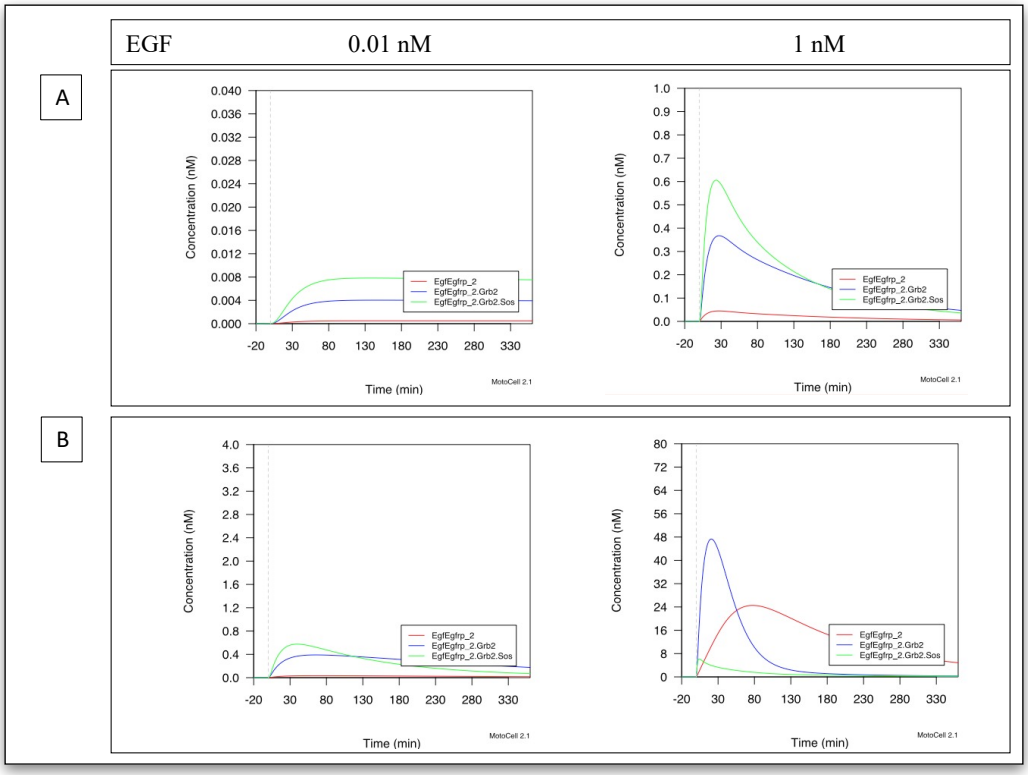



**Figure 8 - Receptor degradation and recycle.** The effects of degradation and recycle of the activated receptor are shown at different concentrations of EGF (0,01nM on the left and 1nM on the right) in HeLa cells (Fig. 8A) and A549 cells (Fig. 8B). Three different forms of the receptor are plotted: the phosphorylated receptor (EgfEgfrp_2, red line), the Egfr-Grb2 complex (EgfEgfrp_2.Grb2, blue line) and the Egfr-Grb2-Sos complex (EgfEgfrp_2.Grb2.Sos, green line).

the radioactive signal from the cell surface peaked after 30 minutes and then rapidly went down (Carpenter, 1976).

In order to evaluate the post-activation behaviour of Egfr receptor as predicted by the simulator under different conditions, several simulations were carried out, adopting experimental conditions corresponding to various experiments taken from literature, in which Egfr phosphorylation levels were measured over time in various cell lines, after treatment with different concentrations of EGF. The simulation reported in Fig. 9A replicates an experiment by Hennig (2016), where Egfr phosphorylation was determined using phospho-selective antibodies in serum-starved HeLa cells challenged with 1.8 nM EGF. The simulated experiment (plot) resulted in predicted results comparable to those obtained through immuno-blotting (western blot). In Fig. 9B and 9C similar validation tests are reported (1.8 nM EGF, 60-360 minutes), based on Egfr activation experiments performed in A549 and LNCaP cells by Wang (2014) and Martin-Orozco (2007), respectively. The simulation performed on the same cell lines predicted a rapid peak of the activated

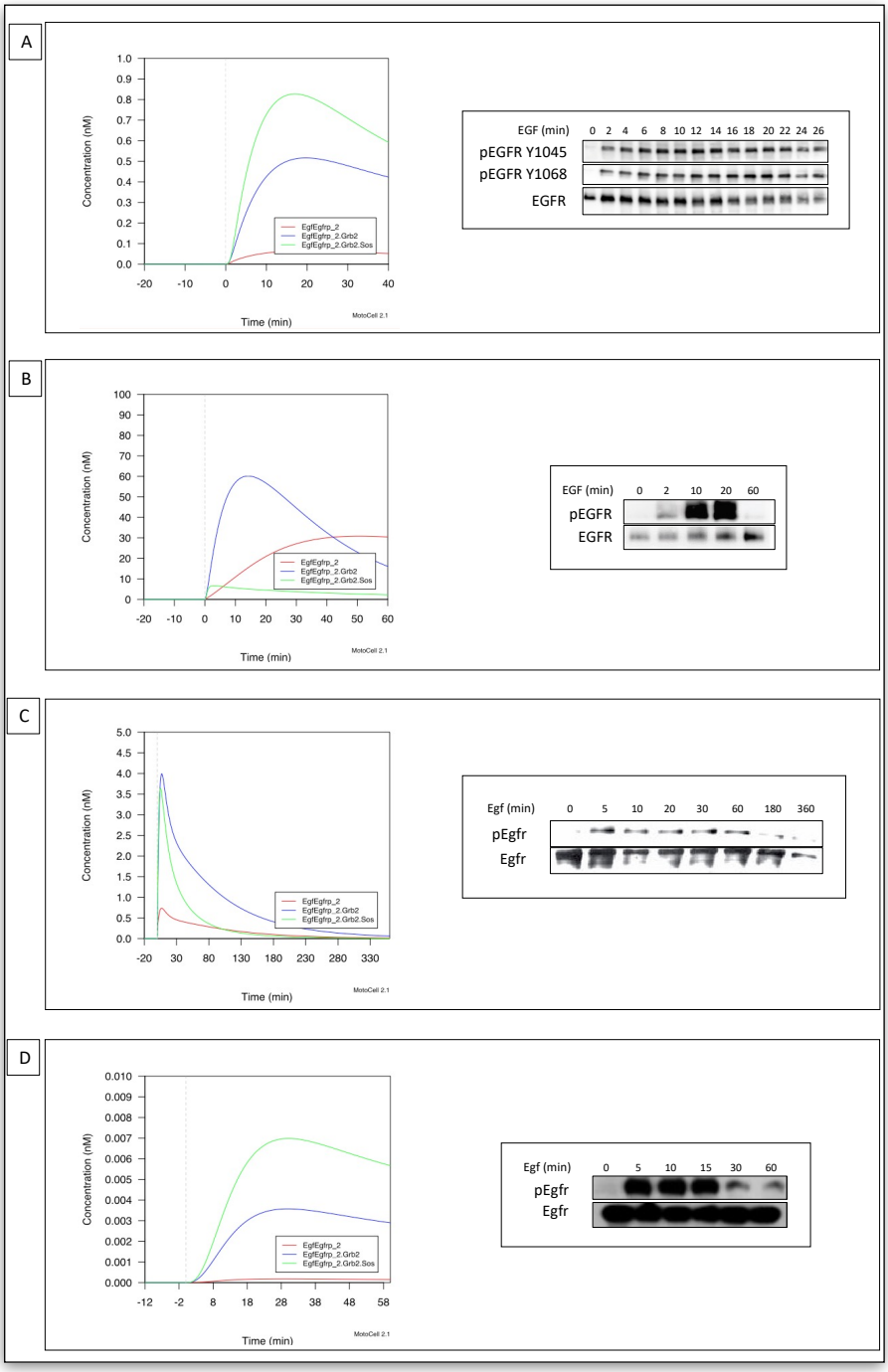



**Figure 9 - Validation of the predicted behavior of activated Egfr.** Stimulations of Egfr are simulated at different EGF concentrations and in different cell lines. In these experiments the Egfr phosphorylation level is measured at the indicated time after treatment with EGF in different cell lines. Three different forms of the receptor are plotted: the phosphorylated receptor (EgfEgfrp_2, red line), the Egfr-Grb2 complex (EgfEgfrp_2.Grb2, blue line) and the Egfr-Grb2-Sos complex (EgfEgfrp_2.Grb2.Sos, green line). In Fig. 9A HeLa cells are stimulated with 1.8 nM EGF for 30 minutes (western blot from Hennig 2016). In Fig. 9B, A549 cells are stimulated with 1.8 nM EGF for 60 minutes (Wang 2017). In Fig. 9C LNCaP cells are stimulated with 10 nM EGF for 360 minutes (Martin-Orozco 2007). In Fig. 9D HEK293 cells are stimulated with 1.8 nM EGF for 60 minutes (Yang, 2018).

receptor at about 10-20 minutes followed by the same rapid decay showed by western blotting. Lastly, in Fig. 9D is reported an experiment performed by Yang (2018), where HEK293 cells were serum starved for 12h, then incubated with 1.8 nM EGF for the indicated durations: also in this case, the simulator predicted an activated receptor behaviour close to the results obtained by western-blotting. The results support the idea that the biochemical model is robust and reliable, as it is applicable to different cell lines, with results very close to those derived by means of biochemical assays.

*4.1.3 The Ras-Raf subsystem*

The cascade subsystem describes the Ras activation processes by the activated receptor, and the start of the cascade activation, from Raf activation (Fig.10).

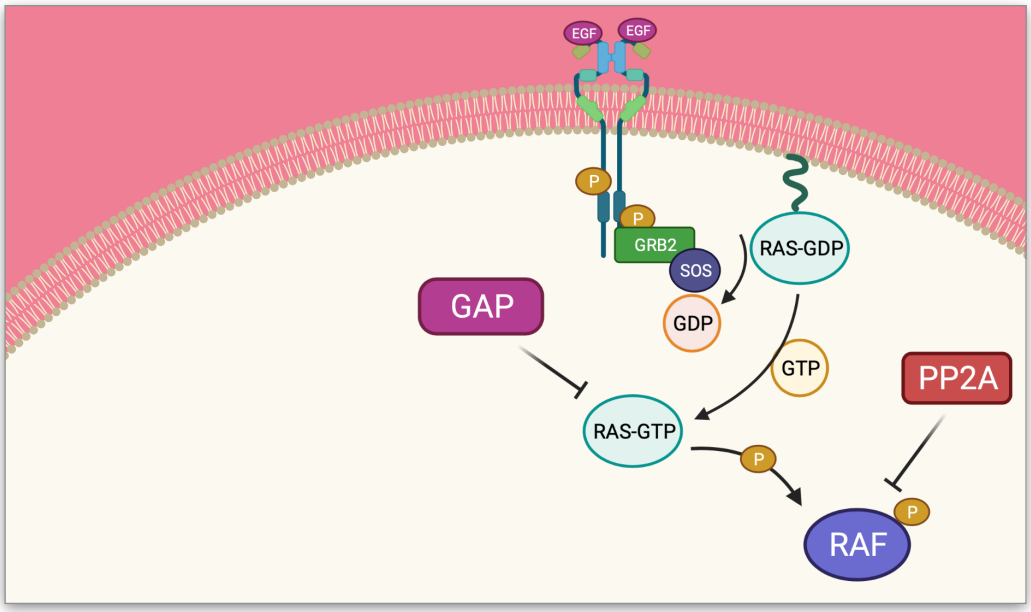



**Figure 10 - Ras and Raf activation.** The third subsystem is composed of 14 reactions. First, the Ras activation is modeled with 3 essential reactions: the activated receptor exchanges the GDP nucleotide on Ras though Sos; Ras bounds the GTP, concentrated in the cell at high levels; GAP protein enhances the hydrolysis of GTP in GDP on Ras**.** The regulation of the activation level of these proteins depends on GTPase. Following, Ras-GTP starts the cascade, first by activating Raf. Phosphatases Pp2a provides the Raf dephosphorylation. Created with **BioRender.com**

Particular attention was paid in modelling the reactions leading to the activation of the Ras protein, also considering the concentrations of GDP and GTP in the process (Fig.11). In this way the model does not treat the receptor as a simple catalyst of a Ras activation reaction, but takes into account its

function as a nucleotide exchanger. It was modelled through three main reactions:

1. The activated receptor complex, through Sos, facilitates the release of GDP nucleotides from Ras;
2. Ras in turn mostly binds GTP, which is more concentrated in the cytosol;
3. GAP protein, acting on Ras, enhances the following GTP hydrolysis which is then converted into GDP.

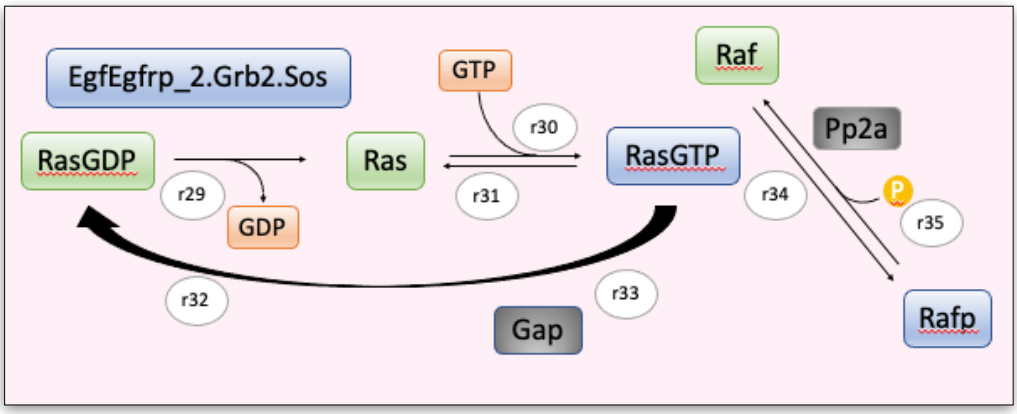



**Figure 11. Reactions flowchart of Ras-Raf activation.**

Ras-GTP is now able to start the Mek1/2 cascade by phosphorylating Raf.

This subsystem was simulated at two different EGF concentrations: 0.01 nM and 1 nM. The results are shown in figure 12. The predicted pattern of Ras-GTP activation is clearly concentration-dependent, since higher EGF concentrations produce higher and earlier peaks (Fig. 12A). A similar pattern was predicted for the first kinase of the cascade, Raf, activated by Ras (Fig. 12B).

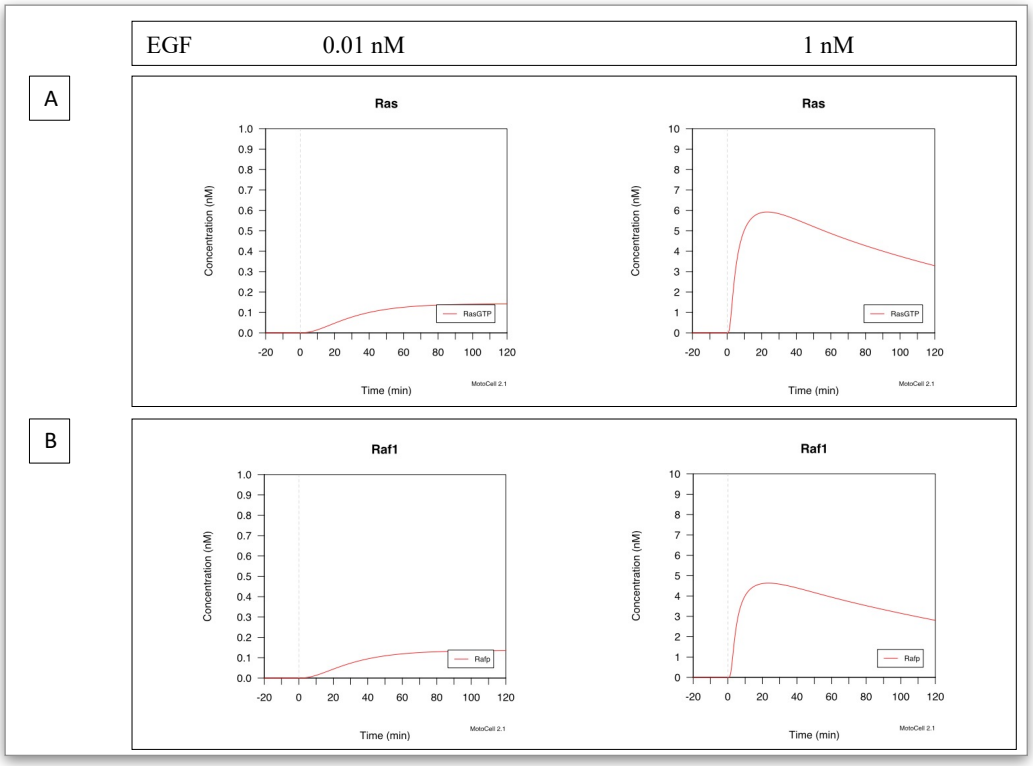



**Figure 12 - Simulation of Ras and Raf activation**. Two different EGF concentration (0.01 nM on the left, 1 nM on the right) were used to simulate Ras (RasGTP) and Raf (Rafp) activation.

Also in this case, simulation results were validated by replicating experiments taken from literature and matching their results with the software predictions. In Fig. 13A the simulation replicated an experiment from Hennig (2016), where serum-starved HeLa cells were challenged with 1.8 nM EGF for in and Ras activation was determined via Ras-GTP affinity pulldowns. The simulation correctly predicted a transient Ras activation, according with the experiment reported. The same experiment was performed by Kiyatkin (2006) in HEK293 cells, stimulated with 20 nM of EGF for 60 minutes (Fig. 13B). The Raf activation was compared with experimental data from Tian-Rui Xu (2010), which describe the Raf-1 and BRAF kinetic activation in HEK293 cells stimulated with 1.8 nM EGF (Fig.13C).

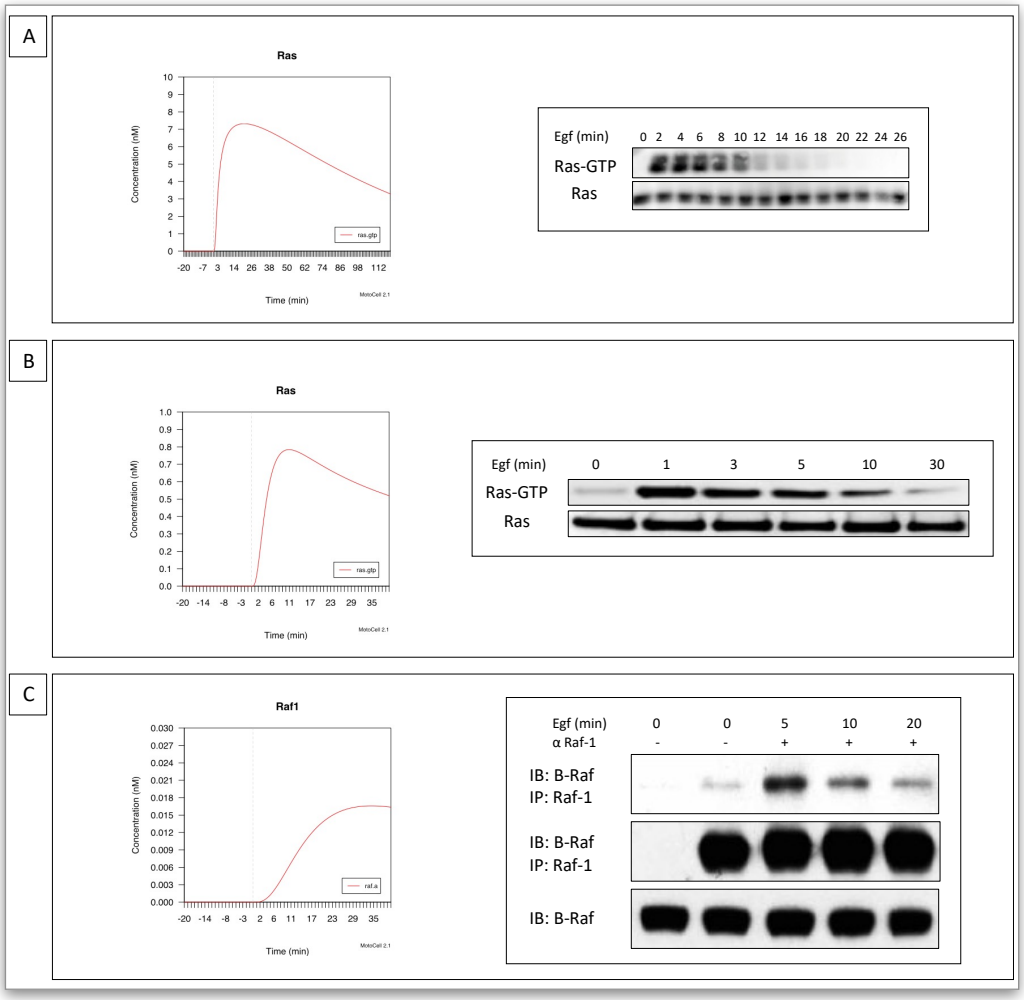



**Figure 13 - Validation of Ras-Raf activation**. In figure 13A, HeLa cells was stimulated with 1.8 nM EGF for 40 minutes (western blot from Hennig 2016). In Fig. 13B was reproduced the experiment of Kiyatkin (2006), which measured RasGTP was by affinity precipitation with Ras-Raf binding in HEK293 cells are stimulated with 20 nM EGF for 30 minutes. In Fig. 13C HEK293 cells were stimulated with 1.8 nM EGF for the time indicated. Endogenous Raf-1 was immunoprecipitated (IP) and blotted (IB) for the presence of B-Raf (experimental data from Tian-Rui Xu 2010).

*4.1.4 The Mek1/2-Erk1/2 subsystem*

Following the Ras and Raf activation, the cascade starts, where Mapkk1/2 and Erk1/2 are activated by sequentially phosphorylating each other. Phosphatases Pp2A and Dusp3 regulate this process by dephosphorylating these proteins. Lastly, Erk translocates to the nucleus (Fig. 14).

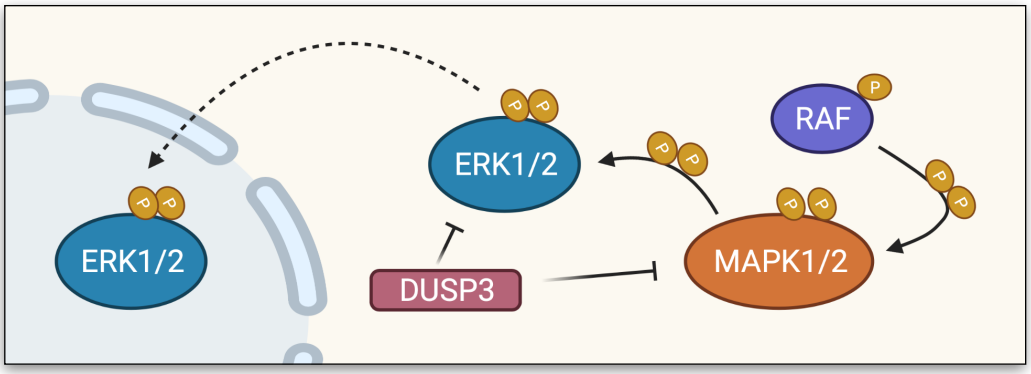



**Figure 14 - Mek1/2 and Erk activation and regulation.** Created with **BioRender.com**

In selecting the reactions needed to model this pathway, particular attention was paid to the description of the nuclear events involving Erk, in particular the trafficking and dephosphorylation of the activated protein (Fig. 15). This was especially important, since the next step would connect the level of Erk activation to its effects on proliferation and movement.

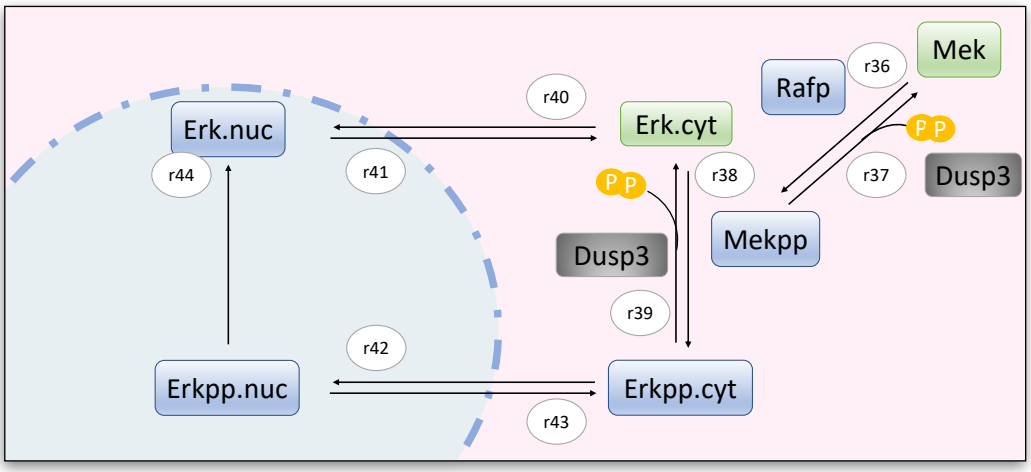



**Figure 15 - Reactions flowchart.** Here are represented the reactions of Mapkk1/2 (Mek) activation to Mekpp, which then activates Erk to Erkpp.cyt. The activated Erk crosses the nuclear membrane (r42) to exert its effects. The model also contains a reaction of nuclear dephosphorylation of the activated Erk (r44).

In order to test the subsystem, two simulations were carried out (Fig. 16) at different EGF concentrations (0.01 nM and 1 nM). The results show that the simulator reproduces, in agreement with other experimental data, an

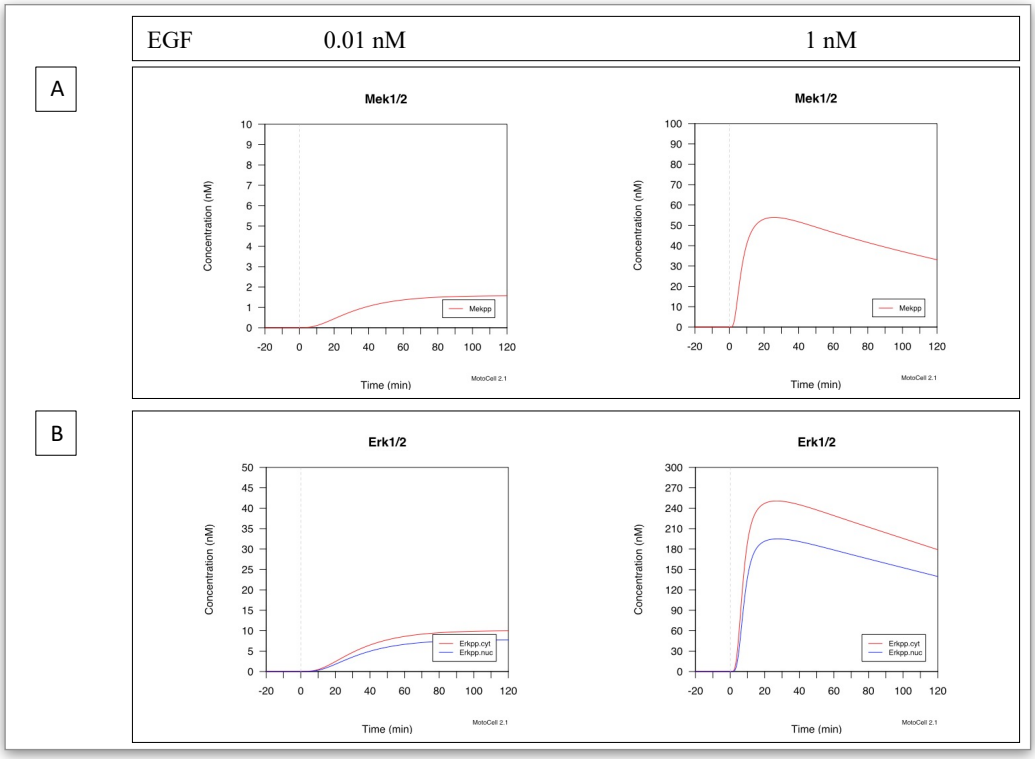



**Figure 16 - Simulation of Mapkk1/2 and Erk1/2 activation.** Two different EGF concentration (0.01 nM on the left, 1 nM on the right) were used to simulate the Mapkk1/2 and Erk activation. The plots show separately nuclear activated Erk (Erkpp.nuc, blue line) and cytosolic activated Erk (Erkpp.cyt, red line).

ultrasensitive behavior of the two kinases: in presence of small stimuli (0.01 nM), the enzyme levels are low, suggesting a basic steady-state activation level, while, when cells are stimulated with high EGF concentrations, they show an upstroke of the response curve. This behavior is compatible with the physiological function of the kinase cascade of producing a signal amplification, in the sense that a small number of activated proteins cause the activation of greater numbers of target molecules.

Simulation results were then validated by replicating other experiments taken from literature, and comparing their results with software predictions. Pinilla-Macua (2016) measured the levels of phosphorylated Mek1/2 and Erk1/2 in HeLa cells, after incubation with 0,72 nM EGF. Likewise, Henning (2016) determined, through immunoblotting, Erk1/2 phosphorylation at 1,8 nM EGF. Lastly, Schoeberl (2002) tested 9nM EGF. In all cases, the predictions

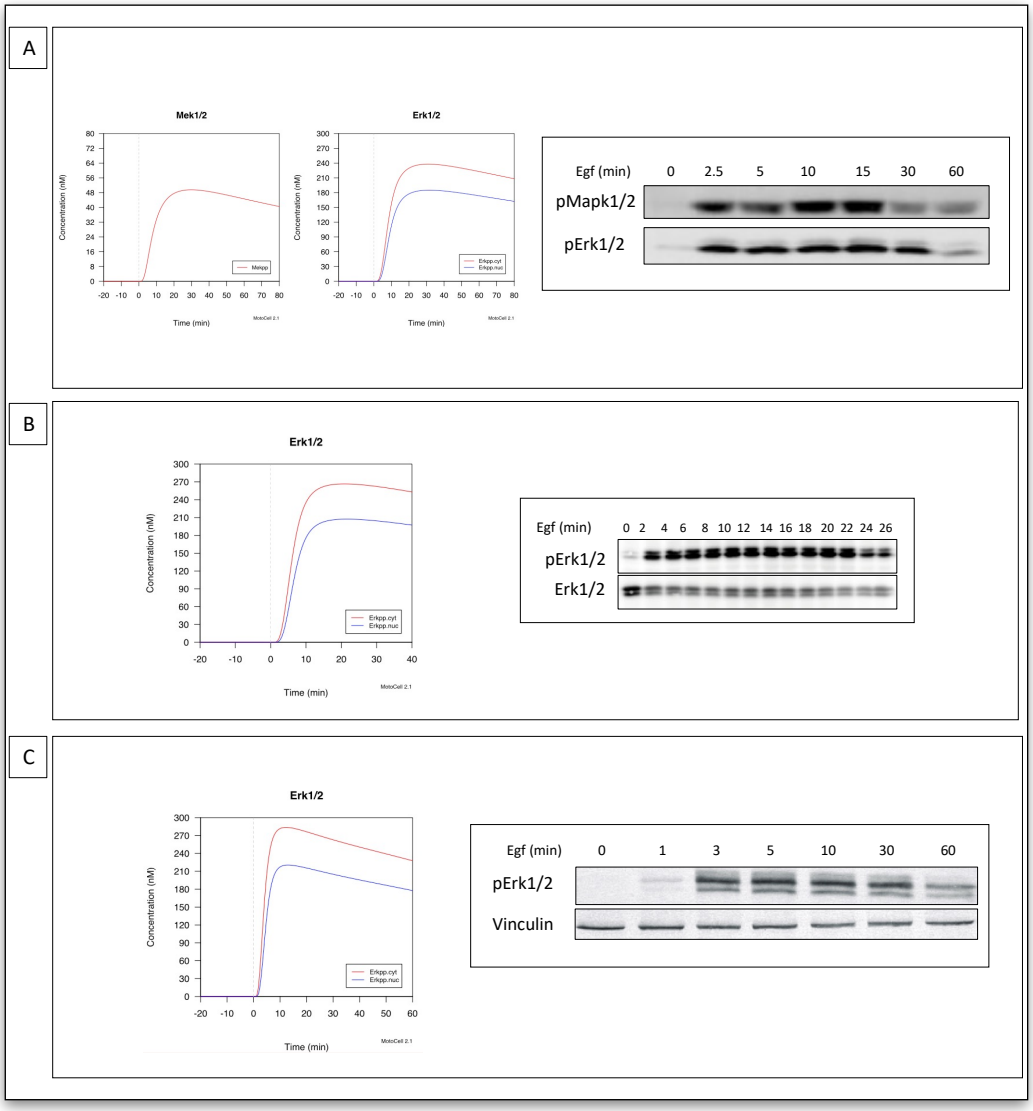



**Figure 17 - Validation of cascade activation.** In figure 17A, the replication of the experiment from Pinilla-Macua (2016), where HeLa cells were incubated with 0,72 nM EGF at 37°C for the indicated times. The phosphorylated species were measured through Western-blotting. In Fig. 17B HeLa cells are stimulated with 1.8 nM EGF for 30 minutes (western blot from Hennig 2016). In Fig. 17C the results from Schoeberl (2002) are reported.

obtained by software match quite well (Fig. 17A,B,C), with all of these reported examples.

Experiments that show the effect of PD184352 and PD98059 on Erk1/2 activation were compared with simulation results were the two molecules were used as inhibitors of Erk phosphorylation. In Figure 18A an experiment from Dokladda (2005) is reported, where HeLa Cells were pre-incubated with the indicated inhibitor concentrations and then stimulated with EGF. The software predicted results very similar to those obtained by Dokladda.

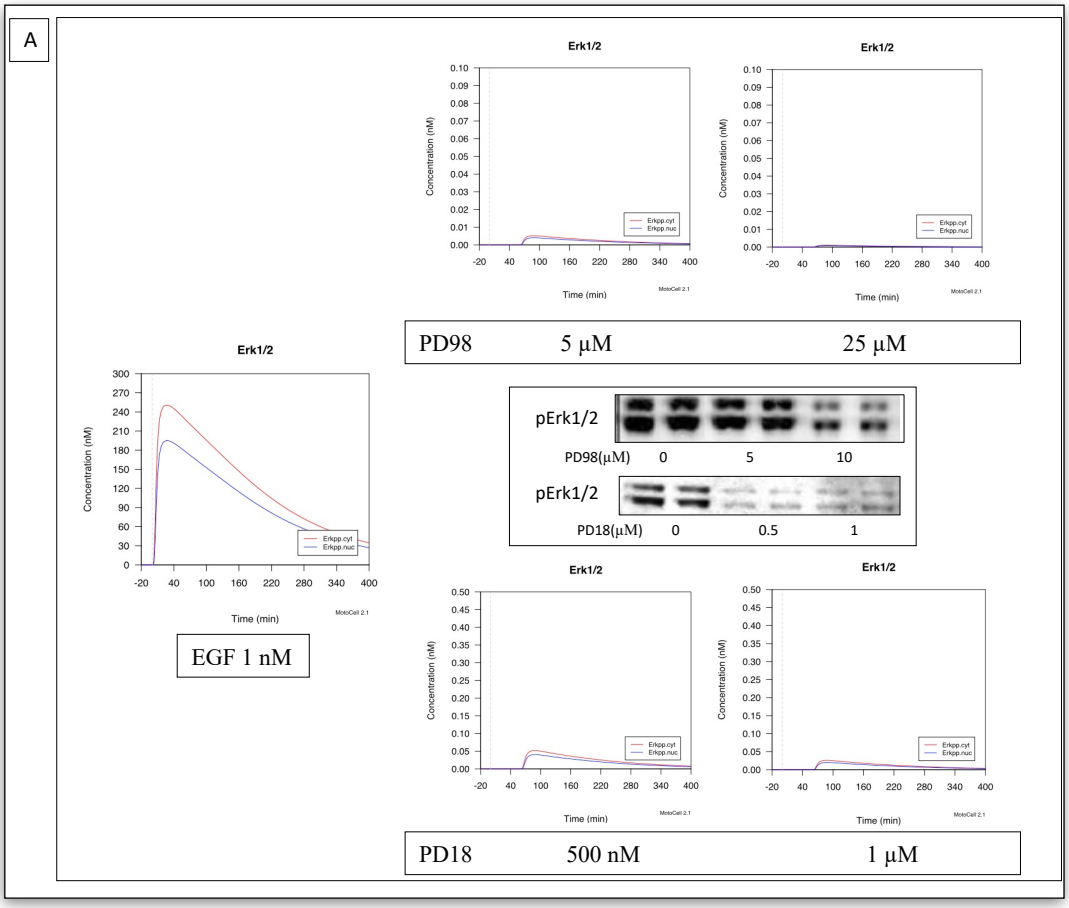



**Figure 18 - Simulation of Erk1/2 inhibitors.** HeLa cells were pre-incubated for 60 minutes with the indicated concentrations of inhibitors, and then stimulated with 1nM EGF. The plot shows the levels of cytosolic activated Erk (Erkpp.cyt, red line) and of nuclear activate Erk (Erkpp.nuc, blue line). The experiment was taken from Dokladda (2005).

4.2 Proliferation and movement analysis

The biochemical model was used to modulate cell proliferation and movement within a cell simulator, in order to create a system able to predict the behaviour of cultured cell populations. To this aim, two mayor steps were identified as necessary:

1. setting up a relationship between the biochemical model of the cascade and a representation of the cell cycle organised as a set of species and a number of reactions describing their succession;
2. creating a path through which the biochemical model is put in communication with the movement simulator, in order to perform analyses of proliferation and motility of an eukaryotic cell culture.

*4.2.1 Cell cycle modulation by activated Erk1/2*

Cell cycle is the series of events that take place in a cell and cause it to divide into two daughter cells. Erk1/2 is known to be an important factor able to modulate cell cycle progression, as it plays a fundamental role in the transition from G0/G1 to S phase, via induction of Cyclin D1 transcription (Meloche 2007). In order to simulate, within a tool for biochemical path simulation, the cell cycle of an eukaryotic cell, a reaction scheme was developed which reproduces the cell cycle phases as different species interconverting into each

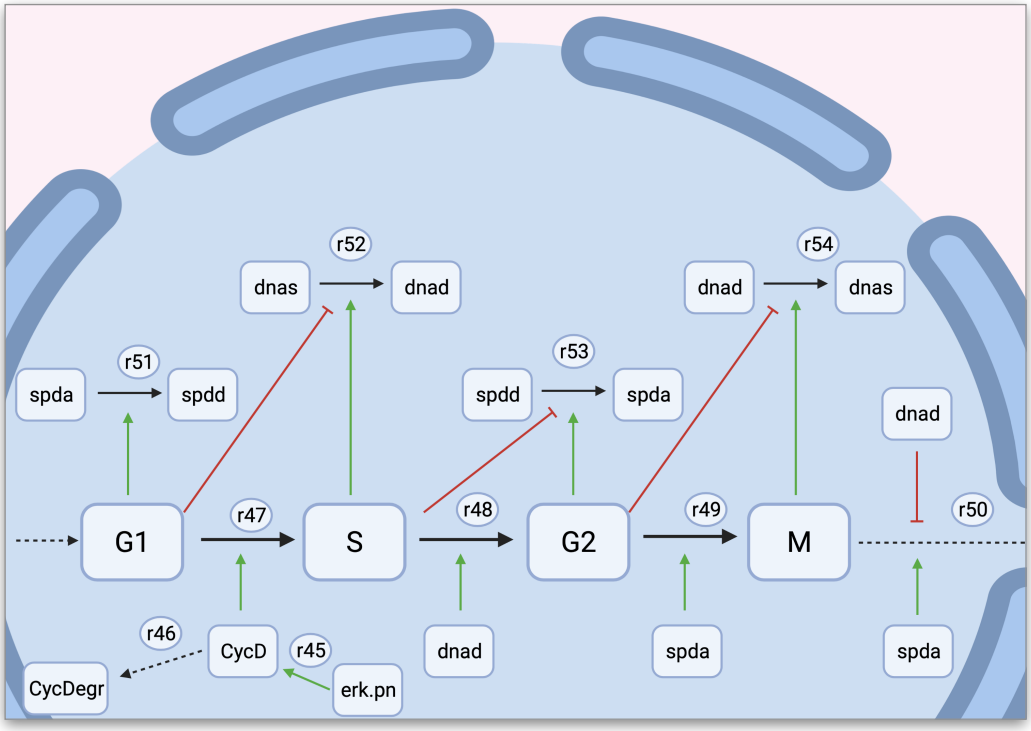



**Figure 19 - Cell Cycle reactions scheme**. Created with **BioRender.com**

other and ranging from 0 to 1 in concentration (Fig.19). This artificial system can relate activated nuclear Erk levels predicted by the biochemical model with G1/S phase switch through an intermediate species named CycD after cyclin D.

Three different culture conditions were used to simulate activation of the Erk1/2 cascade in HeLa cells (Fig.20): 0.001, 0.01 and 0.1 nM EGF

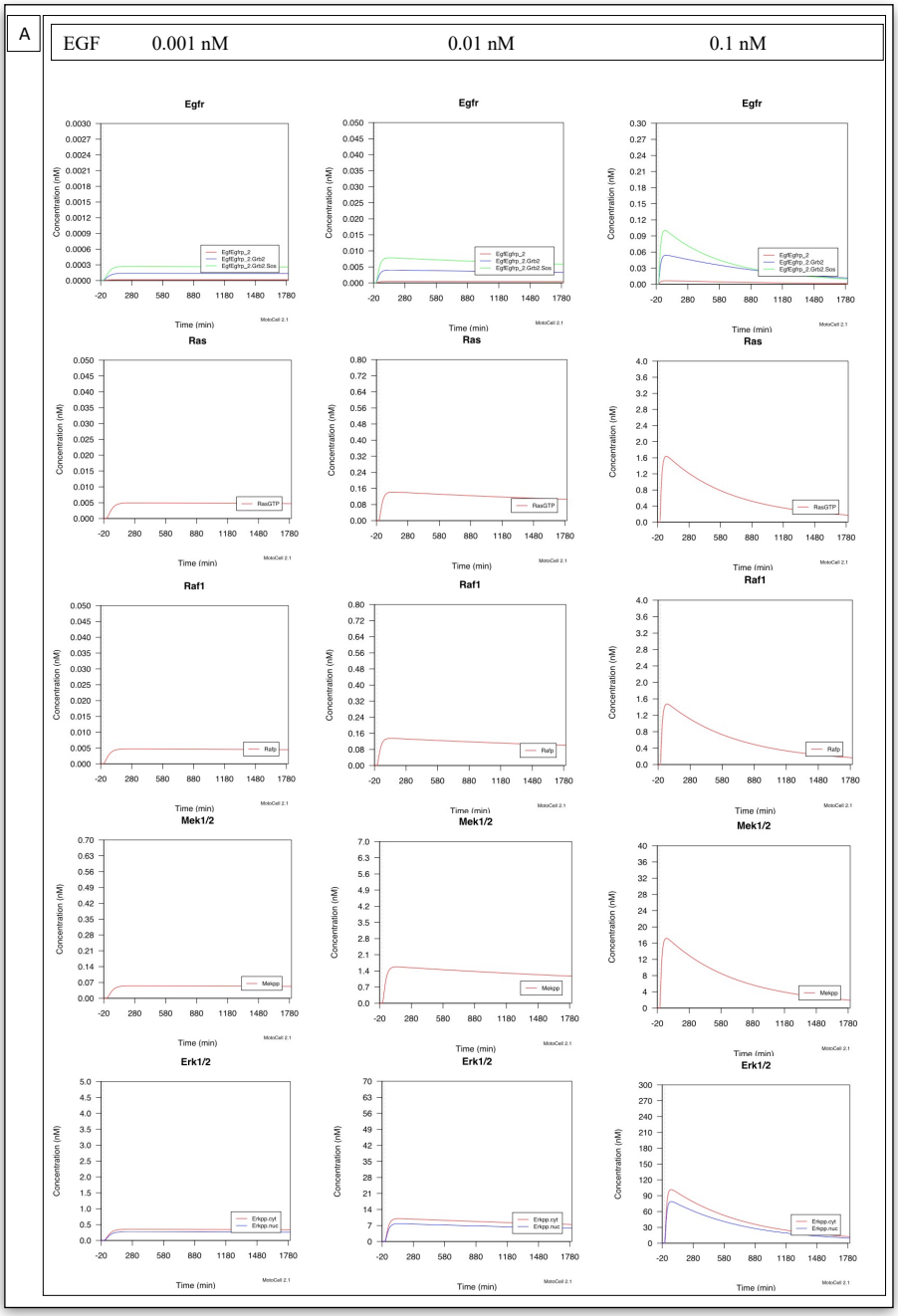



**Figure 20 - Erk1/2 cascade activation.** Three different cell culture condition were used to simulate the cascade activation: starvation (left), 10%FBS (center), stimulated with 0.1 nM EGF (right).

respectively corresponding to growth arrest by serum starvation, 10%FBS and EGF stimulation. Using starved and serum growth conditions, the cascade predicts a relatively stable level of activation, which is maintained for a few hours, under stimulation it shows an early upstroke followed by a decay. The predicted cell phases generate the cell cycles shown in Fig. 21. While in

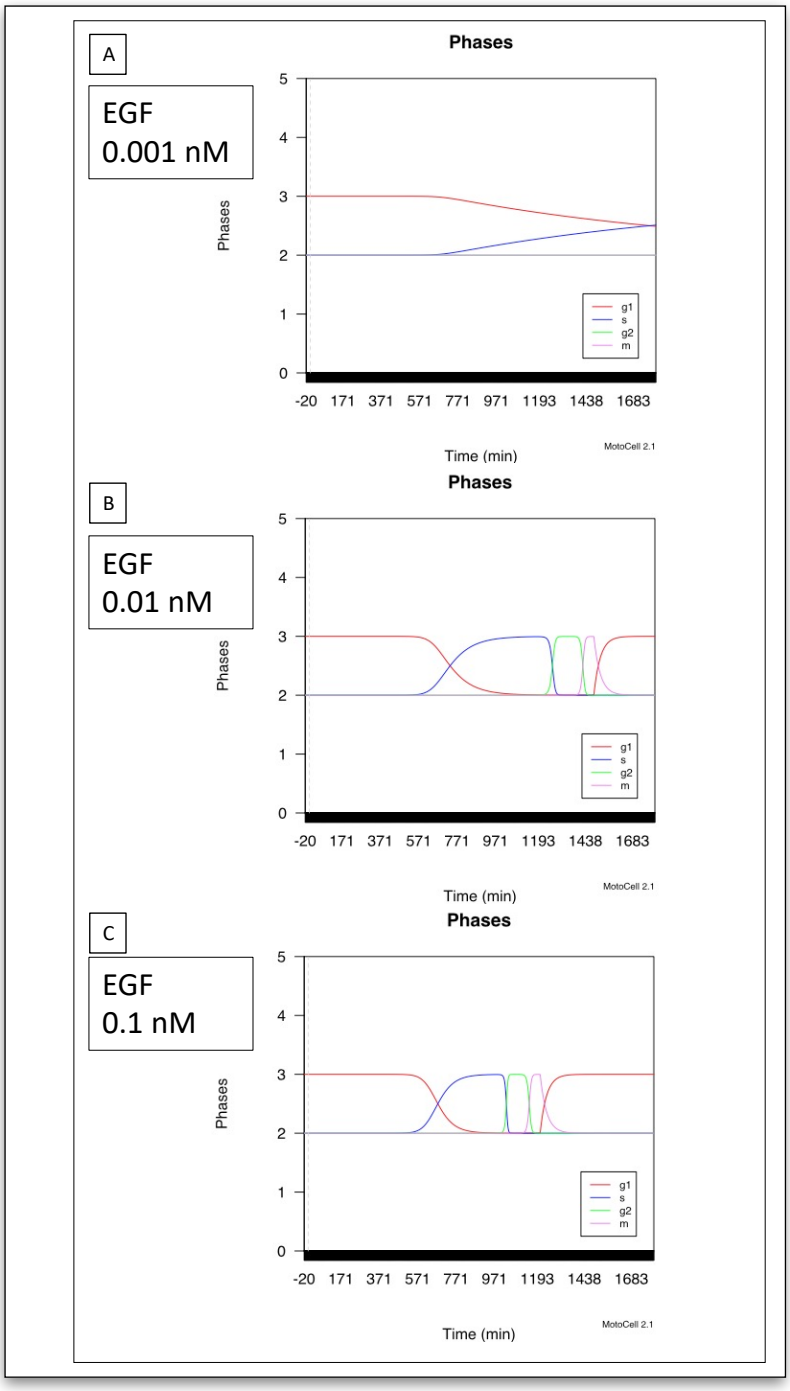



**Figure 21 - Cell cycle regulation by Erk1/2**. HeLa cell cycle was predicted from Erk1/2 activation cascade under three different cell cutter condition: starvation (Fig. 21A), serum 10% FBS (Fig. 21B), stimulation with 0.1 nM EGF (Fig.21C).

starvation conditions the simulator does not predict a switch G1/S within the observation time (Fig. 21A), in serum-like conditions the predictsed phases take about 24 hours to complete a whole cycle. Under stimulated conditions, an accelerated cell cycle is generated which takes about 3 hours less to cycle, compared to the serum condition. These results appear to be consistent with experimental observations on cultured HeLa cells in the same conditions, and may be used to drive the cell simulation.

*4.2.2 Proliferation analysis*

In order to test the ability of the biochemical model the modulate proliferation of synthetic cell cultures produced by the cell simulator (see methods), an interface between the two simulators was setup. The chosen solution was to make the biochemical simulator accessible through a web service, able to support the different features available within the MotoCell analyzer: query to the database of protein concentrations according to the cell type, building of the biochemical model, run execution and result retrieval. Through the web service, the movement simulator sends requests to the biochemical simulator and works on the basis of the information received in return. First, the movement simulator asks the web service to execute a pre-run segment by specifying biochemical model, cell line, run time, and the concentration of extracellular molecular species. For each cell, the web service executes the pre-run and returns a specific ID, which will be used, in the following query, to run subsequent segments, possibly introducing variations, such as the addition of a species or changes of the species concentrations. The results include cell cycle progression status that the simulator uses to predict the behaviour of each individual cell.

In figure 22 three cell populations are presented, which were simulated with this system under three different experimental conditions. On the left the starved condition is represented, for which it is possible to see that the cells do not divide in 24h, since the initial number of cells remains essentially unchanged in subsequent frames. At the center there is a population simulated in 10% FBS serum conditions, which shows a division time of 24h: at the end of the experiment the number of cells is about double than the initial one. Finally, the condition on the right is run at a higher EGF concentration and produces a higher number of cells in 24 hours, corresponding to a division a faster division rate (about 21 hours).

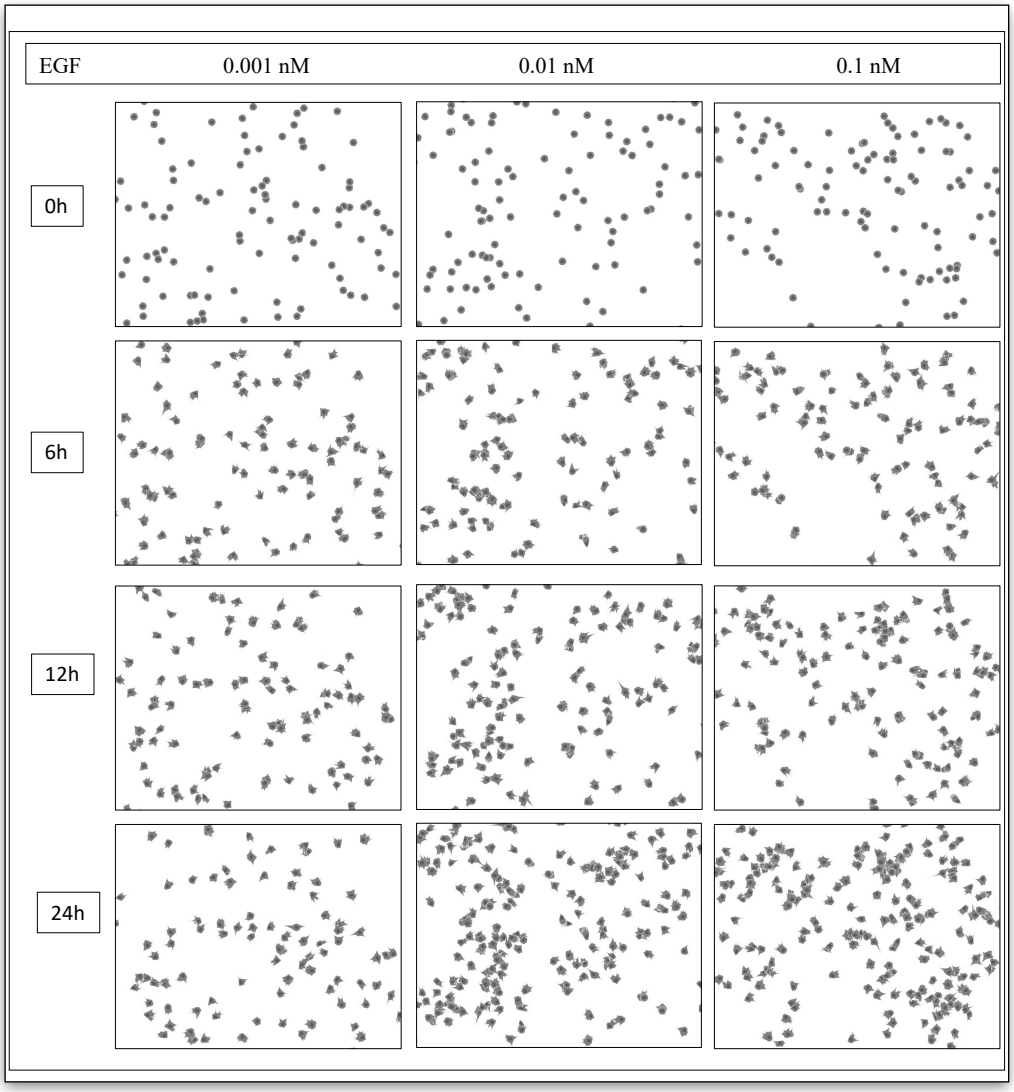



**Figure 22 - Cell proliferation simulation -** Three cell populations were simulated, in three different experimental conditions for the indicated times: starved (left), 10%FBS (center), stimulated with 0.1 nM EGF (right).

To quantitatively analyze this difference in cell division speed, the simulation results were analysed in Motocell with the "*proliferation*" module. The results, reported in figure 23, show that the starved cells do not divide at all within 24 h. Under serum or EGF-stimulated conditions, the cell division events show a typical sigmoid trend similar to that of naturally occurring cell populations, which may be effectively described by Weibull distribution curves corresponding to a $t_{1/2}$ of 670 and 692 min, respectively.

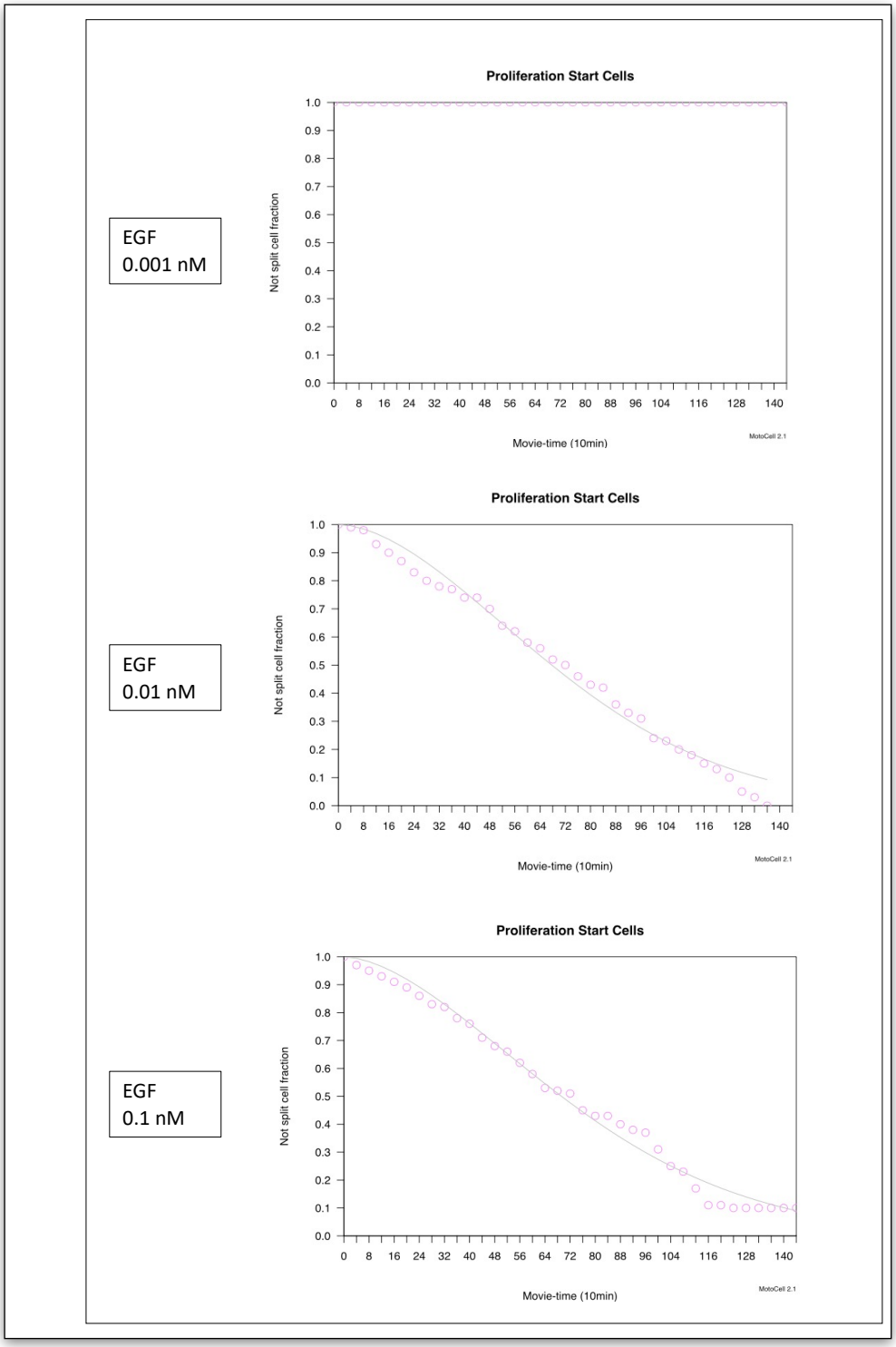



**Figure 23 - Analysis of the proliferation.** Quantitative analysis of the proliferation speed by Motocell's module '*prolifAnalyzer*' of three different cell population: starved (top), 10% FBS (center), stimulated with 0.1 nM EGF (bottom).

Figure 24 shows the growth curve obtain from cell populations simulated under the same conditions and followed for three days. Under EGF-stimulated conditions (blue curve), the cell population shows a faster growth rate than in serum (red curve) and in starving (black curve) condition.

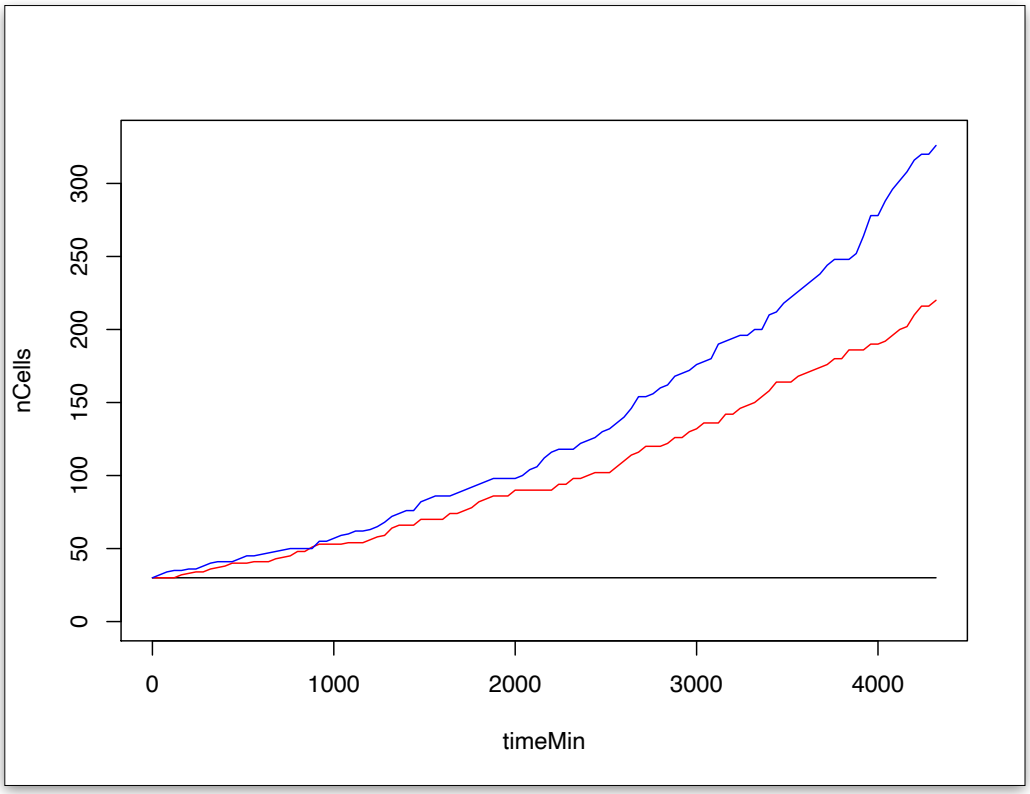



**Figure 24 - Growth curve analysis.** Growth curves show the number of live cells plotted as a function of time for three different cell populations: starved (black), 10% FBS (red), 0.1 nM EGF (blue).

Figure 25 shows another experiment where a rescue from starvation was simulated. Cells were first simulated as living under starving conditions for 10 hrs, after that cells where transferred under standard 10% FBS conditions (arrow): about 6 hrs later cells start to increase in number as cell duplication restarts.

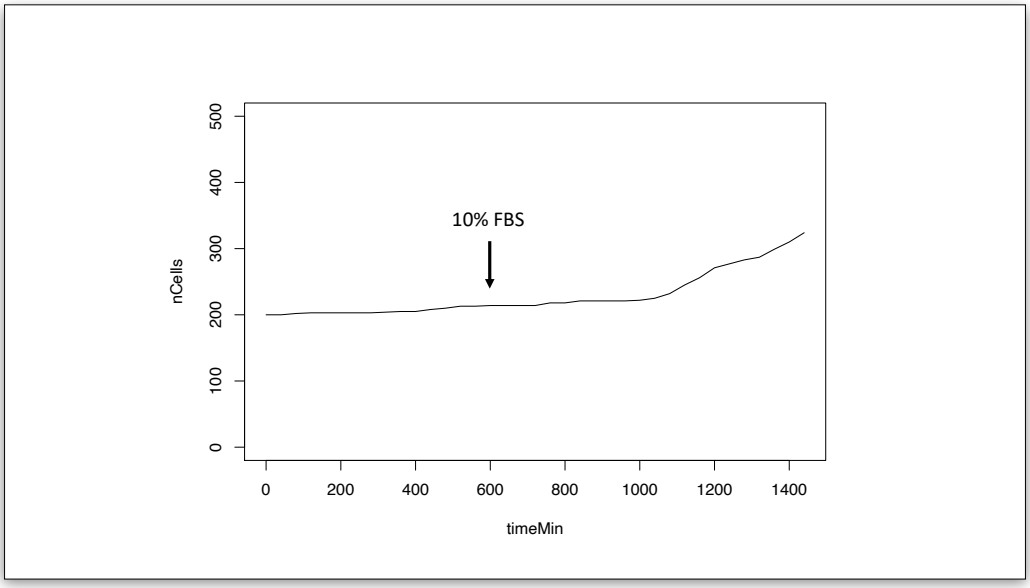



**Figure 25 - Rescue experiment.** The curve shows the number of cells plotted as a function of time. The arrow corresponds to the time when cells were transferred under standard 10% FBS conditions.

4.3 The simulator

The simulator was developed as a plugin hosted within MotoCell (web site: MotoCell), a web application developed in the hosting laboratory, to study cultured cell movement and proliferation. Motocell behaves as a good development system, based an object-oriented programming logic and where good software integration with the R environment is readily available (see Materials and Methods). Within MotoCell, plugins are data analysers, which support many basic functions directed to the analysis of cel culture data.

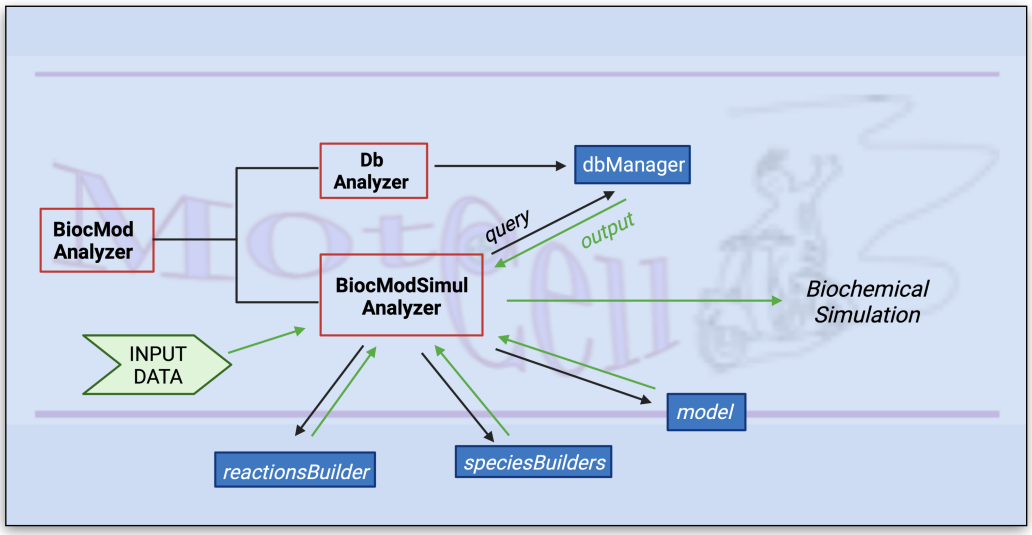



**Figure 24 - Scheme of the developed program.** Red highlighted boxes indicates analyzers, while blue ones are referred to specialized objects. Created with **BioRender.com**

The main object, *BiocModAnalyzer*, extends a common analyser object and contains most of the logic that enables the simulator to receive data and to connect to a database of protein concentrations in various cell lines, described under materials and methods.

*BiocModAnalyzer* is extended by two specialized analyzers:

1. *DbAnalyzer*, responsible for database maintenance and for reporting intermediate analyses during data processing;
2. *BiocModSimulAnalyzer*, which obtains concentration data from the collection and implements model simulations for a given cell type and condition.

The biochemical model is organised in three sheets:

1. In a first sheet we have a list of all the molecules involved. For each entry are indicated the compartment and a key, that the system will use to obtain the initial concentration from the database.
2. In a second sheet we have a list of all the possible compartments. Each compartment is associated with its own size value.

3. In a third sheet is represented the model to be simulated. For any reaction are specified a reaction id, the reactants, the products, eventually a reaction modifier, and the kinetic parameters.

At run time is stored and managed within a *model* object, which is generated starting from the three csv files, two of which are accessed via specialised objects: *speciesBuilder*, *reactionBuilder*. The first creates the species requested by the model and assigns to each its starting concentration; the second uses the content of the csv file, which contains the relations defining the model and converts them to the corresponding equations. All this information is sent back to the analyzer object which, in turn, passes them to the model object that constructs the backbone of the model according to the syntax expected by the SBMLR model simulator, runs it within the R environment and gives the results back to the analyzer, which eventually will show them within the necessary tables and graphs.

From the Motocell web interface it is possible to access the simulator. A dialog window allows to manage several parameters in order to set up the various simulation parameters. For example, it is possible:

1. To choose the model to be used in the simulation;
2. To run the simulation using concentrations or quantities for species;
3. To choose the experimental data (proteomic or transcriptomics) to use as the database of species concentrations construction;
4. To set the time run, time unit and the time step;
5. To set the pathway activator (EGF) concentration;
6. To set a pre-run time

*4.3.2 The Compartment model*

The simulation tool allows the model to define compartments such as extracellular environment, cell, membrane, cytosol or nucleus. It is assumed that the cell volume is 1 and other compartments are expressed relative to it, for example 75% cytosol and 25% nucleus. The extracellular environment is also expressed relative to the cell: considering that in our experiments it is usual to seed 250,000 cells in 1 mL of culture medium, and that HeLa cell volume has been experimentally determined to be about $10^{-12}$ L, it can be assumed that the ratio between the extracellular volume available to a single cell is of the order of 4000:1.

The system is able to solve the model operating either with concentrations or with absolute quantities of the reacting species. In the first case, the system directly uses the concentrations obtained via the database considering a single compartment: cell. In the other case, the system first converts concentrations from the database into quantities by multiplying them for the cell size. The second order reactions and Michaelis-Menten reactions

are then solved relating the kinetic parameters to the compartment they take place into. For example, in a second order reaction where the receptor dimerizes with itself, the reaction is represented by the following equation:

$$v = \frac{k}{vol} * [EGF.EGFR] * [EGF.EGFR]$$

Where k, expressed in $nM^{-1}$ $s^{-1}$, is expressed in nanomolar by dividing for the compartment size where the reaction take place (vol); the reactants are express in nanomolar. Instead, in the reaction where Erk1/2 is phosphorylated by pMEK1/2:

$$v = k_{cat} \frac{[ERK1/2] * [pMEK1/2]}{(K_M * vol) + [ERK1/2]}$$

where Erk1/2 and pMEK1/2 are expressed in nanomolar, and vol is the compartment's size where the reaction takes place.

Lastly, the system converts the resulting protein quantities into the concentration in the compartment where they are located, and graphically plots the results.

*4.3.3 Segments*

To simulate long processes, where conditions may vary in terms of individual molecule presence and concentrations, the concept of simulation segments was introduced.

This feature allows to simulate longer experiments, implementing all of the experimental condition most frequently used in a wet-lab setting, such as starvation, simulation with a determined concentration of EGF, addition of an inhibitor. In order to implement this feature, the software managing the model execution operates as follows:

- the analyzer sends the input data to the *model* object, which constructs the model and sends it back to the analyzer;
- the analyzer sets the model with the initial concentrations taken from the database and starts the first run;
- after the first run is over, the analyzer updates the model with the resulting concentration, which are used as the initial conditions for the subsequent segment
- for each segment the model is not reconstructed but just updated based on the previous run.

From the Motocell interface, in the section dedicated to the biochemical model, a second window may be accessed which allows to split the simulation in up to

10 segments, in each of which it is possible to change activator concentration, or to add a new molecule, such as an inhibitor, setting its concentration.

The software development specifically foresees the execution of a pre-run before the execution of the simulation. In the sheet containing the different species involved in the model, there are several species that at time zero do not have a defined concentration, as they are species subject to synthesis and or modification (for example, phosphorylation or cellular localization). The pre-run has the function of defining a concentration for these species as well. The software allows to define both the duration and the concentration of EGF with which to perform the pre-run.

## 5. Discussion

Several biochemical models built with a systems biology approach are available, and the majority of them studies the MAPK cascade. The search for reliable kinetic parameters describing how the reactions proceed is an essential step in constructing quantitative (biological) models. In the years, different technologies for studying *in vitro* enzymatic activities and reactions' kinetics have been developed so far. For the purpose of this study it was chosen to extract all of the necessary information for the development and validation of the model from literature. The Erk1/2 cascade has been extensively studied, considering its central role in essential cellular events. Searching for "Erk1/2" and "Erk1/2 cancer" on PubMed gives more than 29k and 9k results respectively. Furthermore, different *in silico* studies concerning biochemical models of the Erk1/2 cascade are available, some of which were described in the Background section. Some of them studied Erk1/2 involvement in proliferation and cancer. These represented a point of reference in the construction of a model which has the aim to describe proliferation and motility of eukaryotic cell lines and tumor cells.

Anyway, modelling biochemical pathways presents some ineliminable difficulties. Models often contain a large number of kinetic parameters whose values have not been determined experimentally. *In vitro* measurements of parameters such as a binding constant can give values different from the real ones, because only a few species are considered and the experimental conditions differ from the *in vivo* events. This is especially true for extremely complex and extensive models like the one here described. In addition, kinetic models suffer from a tendential incompleteness, because many protein interactions cannot be included. They usually oversimplify the biological phenomenon they aim to describe, summarizing in a few reactions what in living cells depends on a far more rich network (Golikeri 1974). Additional challenges arise from the continue discover of novel interacting species, or new interactions between known constituents of the pathway (Vojtek 1998); therefore, in order to be effective, a model must also be flexible and capable of incorporating the new informations as soon as they emerge. Looking at this the other way around, the most important role of modeling is to identify missing pieces of the puzzle, since the inconsistencies with the experimental data will point at the critical points to unravel. The purpose of a biochemical model is not to substitute laboratory experiments, but to clarify them and overcome their limitations. It is also true that biochemical models share the same experimental errors with other methods, but the point is the combination of different techniques to achieve a more complete knowledge of the described events. Biochemical models are not supposed to return more precise data, but to allow a more comprehensive integration of what's currently known about cell

biology. For all of these reasons, this approach was chosen to understand how the Erk1/2 behaviour influences cell proliferation and migration over time.

Since experiments are conducted in different culture conditions with different cell lines and various techniques, a desirable feature of a model is the possibility to easily modify the input data. The software running the model was developed with this underlying idea: to make extremely easy the simulation of whichever pathway. It is sufficient to modify the three CSVs to simulate every kind of biochemical pathways, by giving informations about the compartments, the reactions and the species involved. This structure allows to simulate also other mathematical models of different pathways, and furnish an easy mode to bring improvements and changes to the model, without modifying the simulator code. This will allow in the future to describe other pathways that contribute to the regulation of the considered events (i.e. ERK5, AKT/PI3K).

In the model development, the different cellular compartments were taken into account. Generally speaking, a compartmental model is composed of a set of interconnected chambers; in each of them every component of the system is considered as homogeneous and at uniform concentration (Brauer 2008). In this kind of models, biochemical reactions of transmembrane transport and binding processes are treated with first and second order kinetics respectively. This simplicity makes them easy to compute, because the set of ordinary differential equations describing them can be readily solved with numerical and sometimes analytical methods. However, the analysis of the compartmental models available in literature so far revealed that the compartments are merely descriptive, without concrete effects on the model. In contrast, to develop a strong and powerful model, it is of the utmost importance that both the protein concentrations and the reactions are related to the compartment where they are located.

In every mathematical model the concentration at time zero of many cellular species is set as equal to zero. This is especially true for species that are producted in a reaction, like the phosphorylated Erk or the nuclear Erk. However, a cell is not a test tube where chemical balances or all/nothing phenomena are established, but a living organism with dynamic homeostases. In order to have as realistic simulations as possible, it is necessary for the software to perform a preliminary run with very low levels of the stimulus before the simulation, which allows the system to reach a cell-like condition. Then, the simulation will proceed returning a more reliable approximation of the true biochemical events. All of the experiments performed in this study come after a 20 minutes pre-run at 0.001 nM EGF.

The first step in the construction of a biochemical model is the selection of the stimulating input. For the aims of this study it is essential to associate the kind of stimulus with a feature of a typical culture of eukaryotic cells, i.e.

species that are already present or can be added to the culture medium like glutamine, inhibitors or growth factors. The simulator of movement and proliferation is indeed able to take into account these elements, but also the gradient determined when one of these components is inserted. The chosen feature is the epidermal growth factor (EGF), a normal component of FBS. Its receptor, Egfr, is the main tyrosine kinase receptor activating the MAPK cascade. This is why it is the main growth factor receptor incorporated into Erk1/2 models. The Egfr system has been well-studied and is present at substantial levels in various cell types.

Another important point is that various models of Erk1/2 don't consider any degradation reaction, so the receptors remain constitutively activated. The models which take into account the receptor degradation have the pitfall of modelling it only as an irreversible constant flux, with the aim of creating a dose/response system. This is incompatible with the purpose of this study, which is the recreation of cellular events measured in the frame time of 24-48 h, which the cell needs to replicate and move. As reported in literature, the receptor of epidermal growth factor undergoes two degradation mechanisms, one of which (CME) actually promotes the signal sustainment. Drawing from all of the resources available in the published literature, this is the first model capable of simulating both this mechanisms to better describe the biochemical substratum of cell proliferation and motility. A future perspective is the possibility of simulating over even longer durations, when the model will include also the reactions of synthesis and recycle of the main components of the pathway.

The other models generally treat Ras activation through mass action or Michaelis-Menten laws, where the receptor participates as a simple catalyst. The model here discussed instead proposes a different view, where the receptor fills its role as a guanine nucleotide exchange factor, taking into consideration GDP and GTP concentrations as well. This makes the model more realistic and gives space to further expand it by simulating different contexts like cellular quiescence.

The biochemical events describing the cascade activation, namely reactions of phosphorylation and dephosphorylation, are extremely simple and shared by all of the available models. This study focused also on another less investigated aspect, which is the array of events leading to the translocation of Erk to the nucleus to exert its biological effects. Only few *in vitro* measurements of the kinetic parameters of this reactions have been performed; to model it, the work of Fujioka (2005), in which the author collected them via fluorescent probes, was utterly useful. In addition to the determination of translocation rates, in that study a nuclear dephosphorylation rate of Erk was measured, which was implemented in this model.

The connection between the biochemical model and the simulator of movement and proliferation allows to study, both quantitatively and qualitatively, how cells proliferate in different growth conditions. Results are in according to other experimental observations. Future expansions of this simulation system will allow to apply it to pathological context, such as tumour cells, in order to study its proliferation and migration behavior.

# 6. Conclusions

This study presents a computational model which simulates the Erk1/2 cascade in response to the stimulation of Egfr by EGF interfaced with a simulator system that reproduces the behaviour of eukaryotic cell cultures. The system provides results that agree with the available experimental data in different cell lines, and can be used to run simulations over prolonged time frames that reproduce the classic eukaryotic cell cultures. Several novelties were introduced in the Erk1/2 biochemical modelling such as the implementation of a detailed system of internalization of the receptor, the introduction of compartments and the description of the trafficking of Erk between cytosol and nucleus. The model can easily be modified to simulate various contexts, like different concentrations of species or the presence of inhibitors.

This required the development of software to manage the simulator. It gives different opportunities such as the possibility to split the experiments in more segments or to set different experimental conditions. It presents a flexible and simple structure which can easily be adapted to other biochemical pathways for future studies.

Taking all of these points in consideration, this model opens up future perspectives for its use in investigations about the behaviour of various cancers in which the Erk1/2 or other pathways are altered.

## 7. Acknowledgements

Il coronamento di quello che per me è sempre stato il sogno della vita ha richiesto il superamento di una dura prova, ben al di sopra delle mie capacità, che non ha risparmiato momenti di sofferenza, abbandono, delusione. Ho dato tutto me stesso. Solo grazie all'incrollabile ottimismo delle persone a me care sono riuscito a raggiungere questo ambizioso traguardo. Grazie per avermi dato la forza di non mollare, di essere me stesso, di non buttare tutto via.

A mio padre, a mia madre, e a tutta la mia famiglia. A Luca. Grazie per il vostro amore incondizionato.

8. References

Achard F, Vaysseix G, Barillot E. XML, bioinformatics and data integration. Bioinformatics. 2001 Feb;17(2):115-25

Anderson AR, Chaplain M. Continuous and discrete mathematical models of tumor-induced angiogenesis. Bull. Math. Biol. 1998; 60: 857–899

Arkun Y, Yasemi M. Dynamics and control of the Erk signaling pathway: Sensitivity, bistability, and oscillations. PLoS One. 2018 Apr 9;13(4):e0195513

Barillari G. The Impact of Matrix Metalloproteinase-9 on the Sequential Steps of the Metastatic Process. Int J Mol Sci. 2020 Jun 25;21(12):4526

Batchelor E, Loewer A, Lahav G. The ups and downs of p53: understanding protein dynamics in single cells. Nat. Rev. Cancer. 2009; 9, 371,

Berkers JA, van Bergen en Henegouwen PM, Boonstra J. Three classes of epidermal growth factor receptors on HeLa cells. J Biol Chem. 1991 Jan 15;266(2):922-7

Bhalla US, Ram PT, Iyengar R. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. Science. 2002 Aug 9;297(5583):1018-23

Bidkhori G, Moeini A, Masoudi-Nejad A. Modeling of tumor progression in NSCLC and intrinsic resistance to TKI in loss of PTEN expression. PLoS One. 2012;7(10):e48004.

Brauer F. (2008) Compartmental Models in Epidemiology. In: Brauer F., van den Driessche P., Wu J. (eds) Mathematical Epidemiology. Lecture Notes in Mathematics, vol 1945. Springer, Berlin, Heidelberg.

Bray D: Cell movements, 2nd edition. Garland Publishing, New York, pp79, 2001

Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA. The statistical mechanics of complex signaling networks: nerve growth factor signaling. Phys Biol. 2004 Dec;1(3-4):184-95

Burotto M, Chiou VL, Lee JM, Kohn EC. The MAPK pathway across different malignancies: a new perspective. Cancer. 2014 Nov 15;120(22):3446-56

Cancer systems biology: a network modeling perspective. Carcinogenesis. 2010;31(1):2-8

Cantarella C, Sepe L, Fioretti F, Ferrari MC, Paolella G. Analysis and modelling of motility of cell populations with MotoCell. BMC Bioinformatics. 2009;10 Suppl 12(Suppl 12):S12. Published 2009 Oct 15

Capuani F, Conte A, Argenzio E, Marchetti L, Priami C, Polo S, Di Fiore PP, Sigismund S, Ciliberto A. Quantitative analysis reveals how Egfr activation and downregulation are coupled in normal but not in cancer cells. Nat Commun. 2015 Aug 12;6:7999

Cargnello M, Roux PP. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases [published correction appears in Microbiol Mol Biol Rev. 2012 Jun;76(2):496]. Microbiol Mol Biol Rev. 2011;75(1):50-83

Chambard JC, Lefloch R, Pouysségur J, Lenormand P. Erk implication in cell cycle regulation. Biochim Biophys Acta. 2007 Aug;1773(8):1299-310

Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, Hucka M, Jalowicki G, Keating S, Knight-Schrijver V, Lloret-Villas A, Natarajan KN, Pettit JB, Rodriguez N, Schubert M, Wimalaratne SM, Zhao Y, Hermjakob H, Le Novère N, Laibe C. BioModels: ten-year anniversary. Nucleic Acids Res. 2015 Jan;43(Database issue):D542-8

Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, Mellman I. Spatial control of EGF receptor activation by reversible dimerization on living cells. Nature. 2010 Apr 1;464(7289):783-7

Citri A, Yarden Y. EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol. 2006 Jul;7(7):505-16

Cleland WW. Enzyme kinetics. Annu Rev Biochem. 1967;(36):77-112.

Das J, Ho M, Zikherman J, Govern C, Yang M, Weiss A, Chakraborty AK, Roose JP. Digital signaling and hysteresis characterize ras activation in lymphoid cells. Cell. 2009 Jan 23;136(2):337-51

Dokladda K, Green KA, Pan DA, Hardie DG. PD98059 and U0126 activate AMP-activated protein kinase by increasing the cellular AMP:ATP ratio and not via inhibition of the MAP kinase pathway. FEBS Lett. 2005 Jan 3;579(1):236-40

Dougherty MK, Müller J, Ritt DA, Zhou M, Zhou XZ, Copeland TD, Conrads TP, Veenstra TD, Lu KP, Morrison DK. Regulation of Raf-1 by direct feedback phosphorylation. Mol Cell. 2005 Jan 21;17(2):215-24

Erickson KE, Rukhlenko OS, Posner RG, Hlavacek WS, Kholodenko BN. New insights into Ras biology reinvigorate interest in mathematical modeling of Ras signaling. Semin Cancer Biol. 2019 Feb;54:162-173

Erickson KE, Rukhlenko OS, Posner RG, Hlavacek WS, Kholodenko BN. New insights into Ras biology reinvigorate interest in mathematical modeling of Ras signaling. Semin Cancer Biol. 2019 Feb;54:162-173

Ferguson KM, Berger MB, Mendrola JM, Cho HS, Leahy DJ, Lemmon MA. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. Mol Cell. 2003 Feb;11(2):507-17

Fincham VJ, James M, Frame MC, Winder SJ. Active Erk/MAP kinase is targeted to newly forming cell-matrix adhesions by integrin engagement and v-Src. EMBO J. 2000 Jun 15;19(12):2911-23

Fredrickson AG, Megee RD, Tsuchiya HM. Mathematical models for fermentation processes. Adv. Appl. Microbiol. 1970; 13:419-65

Fujioka A, Terai K, Itoh RE, Aoki K, Nakamura T, Kuroda S, Nishida E, Matsuda M. Dynamics of the Ras/Erk MAPK cascade as monitored by fluorescent probes. J Biol Chem. 2006 Mar 31;281(13):8917-26

Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol Cell Proteomics. 2012 Mar;11(3):M111.014050

Gialeli C, Theocharis AD and Karamanos NK: Roles of matrix metalloproteinases in cancer progression and their pharmaco- logical targeting. FEBS J. 2011; 278: 16-27

Gideon P, John J, Frech M, Lautwein A, Clark R, Scheffler JE, Wittinghofer A. Mutational and kinetic analyses of the GTPase-activating protein (GAP)-p21 interaction: the C-terminal domain of GAP is not sufficient for full activity. Mol Cell Biol. 1992 May;12(5):2050-6

Golikeri S and Luss D, Aggregation of many coupled consecutive first-order reactions Chem. Eng. Sci. 1974 (29): 845–855

Guo YJ, Pan WW, Liu SB, Shen ZF, Xu Y, Hu LL. Erk/MAPK signalling pathway and tumorigenesis. Exp Ther Med. 2020 Mar;19(3):1997-2007

Hancock JF. Ras proteins: different signals from different locations. Nat Rev Mol Cell Biol. 2003 May;4(5):373-84

Heinrich R, Rapoport TA. A linear steady state treatment of enzymatic chains. General properties, control and effector strength. Eur. J. Biochem. 1974; 42:89-95

Hennig A, Markwart R, Wolff K, Schubert K, Cui Y, Prior IA, Esparza-Franco MA, Ladds G, Rubio I. Feedback activation of neurofibromin terminates growth factor-induced Ras activation. Cell Commun Signal. 2016 Feb 9;14:5

Hibino K, Shibata T, Yanagida T, Sako Y. Activation kinetics of Raf protein in the ternary complex of Raf, Ras-GTP, and kinase on the plasma membrane of living cells: single-molecule imaging analysis. J Biol Chem. 2011 Oct 21;286(42):36460-8

Huang CY, Ferrell JE Jr. Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci U S A. 1996;93(19):10078-10083

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J; SBML Forum. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003 Mar 1;19(4):524-31

Hunter T Signaling--2000 and beyond. Cell. 2000;100(1), 113-127

Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu. Rev. Genomics Hum. Genet. 2001; 2, 343–372

Ideker T, Thorsson V, Ranish JA, Christmas R,BuhlerJ, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001; 292:929–34

Jens Nielsen, Systems Biology of Metabolism, Annu Rev Biochem 2017;86:245-275.
Jewett MC, Nielsen J. Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. FEMS Yeast Res. 2008; 8:122-31

Kacser H, Burns JA. The control of flux. Symp. Soc. Exp. Biol. 1973; 27:65-104

Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. BioEssays 2004; 26:99-105

Kell DB, van Dam K, Westerhoff HV. Control analysis of microbial growth and productivity. Symp. Soc. Gen. Microbiol. 1989; 44:61-93

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004; 32: 277–80.

Kholodenko BN, Demin OV, Moehren G, Hoek JB. Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem. 1999 Oct 15;274(42):30169-81

Kirschner MW. The meaning of systems biology. Cell. 2005; 121, 503–504

Kitano H. Computational systems biology. Nature 2002; 420:206–10

Kitano H. Systems biology: a brief overview. Science 2002; 295:1662–64

Kiyatkin A, Aksamitiene E, Markevich NI, Borisov NM, Hoek JB, Kholodenko BN. Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops. J Biol Chem. 2006 Jul 21;281(29):19925-38

Klemke RL, Cai S, Giannini AL, Gallagher PJ, de Lanerolle P, Cheresh DA. Regulation of cell motility by mitogen-activated protein kinase. J Cell Biol. 1997 Apr 21;137(2):481-92

Kholodenko BN, Demin OV, Moehren G, Hoek JB. Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem. 1999 Oct 15;274(42):30169-81

Lenzen C, Cool RH, Prinz H, Kuhlmann J, Wittinghofer A. Kinetic analysis by fluorescence of the interaction between Ras and the catalytic domain of the guanine nucleotide exchange factor Cdc25Mm. Biochemistry. 1998 May 19;37(20):7420-30

Liang SI, van Lengerich B, Eichel K, Cha M, Patterson DM, Yoon TY, von Zastrow M, Jura N, Gartner ZJ. Phosphorylated Egfr Dimers Are Not Sufficient to Activate Ras. Cell Rep. 2018 Mar 6;22(10):2593-2600

Maeda-Yamamoto M, Suzuki N, Sawai Y, Miyase T, Sano M, Hashimoto-Ohta A and Isemura M: Association of suppression of extracellular signal-regulated kinase phosphorylation by epigallocatechin gallate with the reduction of matrix metal- loproteinase activities in human fibrosarcoma HT1080 cells. J Agric Food Chem. 2003; 51: 1858-1863.

Martín-Orozco RM, Almaraz-Pro C, Rodríguez-Ubreva FJ, Cortés MA, Ropero S, Colomer R, López-Ruiz P, Colás B. EGF prevents the neuroendocrine differentiation of LNCaP cells induced by serum deprivation: the modulator role of PI3K/Akt. Neoplasia. 2007 Aug;9(8):614-24

Meloche S, Pouysségur J. The Erk1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition. Oncogene. 2007 May 14;26(22):3227-39

Mendes, P. and Kell, D. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 1998; 14, 869–883

Moritz A, Li Y, Guo A, Villen J, Wang Y, MacNeill J, Kornhauser J, Sprott K, Zhou J, Possemato A, Ren JM, Hornbeck P, Cantley LC, Gygi SP, Rush J, Comb MJ. Akt-RSK-S6 kinase signaling networks activated by oncogenic receptor tyrosine kinases, Sci. Signal. 3 (2010) (ra64-ra64).

Noble D. Cardiac action and pacemaker potentials based on the Hodkin-Huxley equations. Nature 1960; 188:495-97

Orton RJ, Adriaens ME, Gormand A, Sturm OE, Kolch W, Gilbert DR. Computational modelling of cancerous mutations in the Egfr/Erk signalling pathway. BMC Syst Biol. 2009 Oct 5;3:100

Orton RJ, Sturm OE, Vyshemirsky V, Calder M, Gilbert DR, Kolch W. Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway. Biochem J. 2005;392(Pt 2):249-261

Pappalardo F, Russo G, Candido S, Pennisi M, Cavalieri S, Motta S, McCubrey JA, Nicoletti F, Libra M. Computational Modeling of PI3K/AKT and MAPK Signaling Pathways in Melanoma Cancer. PLoS One. 2016 Mar 25;11(3):e0152104

Pinilla-Macua I, Watkins SC, Sorkin A. Endocytosis separates EGF receptors from endogenous fluorescently labeled HRas and diminishes receptor signaling to MAP kinases in endosomes. Proc Natl Acad Sci U S A. 2016 Feb 23;113(8):2122-7. doi: 10.1073/pnas.1520301113

Plotnikov A, Zehorai E, Procaccia S, Seger R. The MAPK cascades: signaling components, nuclear roles and mechanisms of nuclear translocation. Biochim Biophys Acta. 2011 Sep;1813(9):1619-33

Purvis JE, Lahav G. Encoding and decoding cellular information through signaling dynamics. Cell. 2013; 152, 945-956,

Qiao L, Nachbar RB, Kevrekidis IG, Shvartsman SY. Bistability and oscillations in the Huang-Ferrell model of MAPK signaling. PLoS Comput Biol. 2007 Sep;3(9):1819-26.

Reddy KB, Nabha SM, Atanaskova N. Role of MAP kinase in tumor progression and invasion. Cancer Metastasis Rev. 2003 Dec;22(4):395-403

Schlessinger J, Cell signaling by receptor tyrosine kinases. Cell, 103 (2000), 211-225

Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol. 2002 Apr;20(4):370-5

Sebolt-Leopold JS, Herrera R. Targeting the mitogen-activated protein kinase cascade to treat cancer. Nat Rev Cancer. 2004 Dec;4(12):937-47

Segel IH in Encyclopaedia of Biological Chemistry , 2nd ed., Elsevier, 2013, pp. 216–220

Sepe L, Ferrari MC, Cantarella C, Fioretti F, Paolella G. Ras activated Erk and PI3K pathways differentially affect directional movement of cultured fibroblasts. Cell Physiol Biochem. 2013;31(1):123-42

Shi T, Niepel M, McDermott JE, Gao Y, Nicora CD, Chrisler WB, Markillie LM, Petyuk VA, Smith RD, Rodland KD, Sorger PK, Qian WJ, Wiley HS. Conservation of protein abundance patterns reveals the regulatory architecture of the Egfr-MAPK pathway. Sci Signal. 2016 Jul 12;9(436):rs6

Sigismund S, Algisi V, Nappo G, Conte A, Pascolutti R, Cuomo A, Bonaldi T, Argenzio E, Verhoef LG, Maspero E, Bianchi F, Capuani F, Ciliberto A, Polo S, Di Fiore PP. Threshold-controlled ubiquitination of the Egfr directs receptor fate. EMBO J. 2013 Jul 31;32(15):2140-57

Sigismund S, Argenzio E, Tosoni D, Cavallaro E, Polo S, Di Fiore PP. Clathrin-mediated internalization is essential for sustained Egfr signaling but dispensable for degradation. Dev Cell. 2008 Aug;15(2):209-19. doi: 10.1016/j.devcel.2008.06.012.

Simon C, Hicks MJ, Nemechek AJ, Mehta R, O'Malley BW Jr, Goepfert H, Flaitz CM and Boyd D: PD 098059, an inhibitor of ERK1 activation, attenuates the in vivo invasiveness of head and neck squamous cell carcinoma. Br J Cancer. 1999; 80: 1412-1419.

Srinivasan B. Explicit Treatment of Non-Michaelis-Menten and Atypical Kinetics in Early Drug Discovery*. ChemMedChem. 2021 Mar 18;16(6):899-918

Starbuck C, Lauffenburger DA. Mathematical model for the effects of epidermal growth factor receptor trafficking dynamics on fibroblast proliferation responses. Biotechnol Prog. 1992 Mar-Apr;8(2):132-43

Tanimura S, Takeda K. Erk signalling as a regulator of cell motility. J Biochem. 2017 Sep 1;162(3):145-154

Tavassoly I., Goldfarb J. and Iyengar R., Systems biology primer: the basic methods and approaches, Essays in Biochemistry. 2018;62(4):487-500

Terrell EM, Morrison DK. Ras-Mediated Activation of the Raf Family Kinases. Cold Spring Harb Perspect Med. 2019 Jan 2;9(1):a033746

Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA. E-CELL: software environment for whole-cell simulation. Bioinformatics 1999; 15, 72–84

Xu TR, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie GS, Ketley D, Dunlop AJ, Milligan G, Houslay MD, Kolch W. Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. Sci Signal. 2010;3(134):ra20

Vidal M. A unifying view of 21st century systems biology. FEBS Lett. 2009; 583:3891–94

Vojtek AB, Der CJ. Increasing complexity of the Ras signaling pathway. J Biol Chem. 1998 Aug 7;273(32):19925-8

Wainstein E, Seger R. The dynamic subcellular localization of Erk: mechanisms of translocation and role in various organelles. Curr Opin Cell Biol. 2016 Apr;39:15-20

Walsh R. Are improper kinetic models hampering drug development? PeerJ. 2014 Oct 28;2:e649

Wang X, Chen K, Yu Y, et al. Metformin sensitizes lung cancer cells to treatment by the tyrosine kinase inhibitor erlotinib. Oncotarget. 2017;8(65):109068-109078.

Welch D.R., Hurst D.R. Defining the Hallmarks of Metastasis. Cancer Res. 2019; 79:3011–3027

Wiśniewski JR, Hein MY, Cox J, Mann M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol Cell Proteomics. 2014;13(12):3497-3506.

Wortzel I, Seger R. The Erk Cascade: Distinct Functions within Various Subcellular Organelles. Genes Cancer. 2011 Mar;2(3):195-209

Yamaguchi H, Minopoli G, Demidov ON, Chatterjee DK, Anderson CW, Durell SR, Appella E. Substrate specificity of the human protein phosphatase 2Cdelta, Wip1. Biochemistry. 2005 Apr 12;44(14):5285-94

Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol. 2001 Feb;2(2):127-37

Young MW, Kim Y. Dephosphorylation of Epidermal Growth Factor Receptor by Protein Tyrosine Phosphatase 1B. The FASEB Journal, 2019 (33): 645.5-645.5.

Zehorai E, Yao Z, Plotnikov A, Seger R. The subcellular localization of Mek and Erk--a novel nuclear translocation signal (NTS) paves a way to the nucleus. Mol Cell Endocrinol. 2010 Jan 27;314(2):213-20

Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell. 2006 Jun 16;125(6):1137-49

Zhang ZY, Walsh AB, Wu L, McNamara DJ, Dobrusin EM, Miller WT. Determinants of substrate recognition in the protein-tyrosine phosphatase, PTP1. J Biol Chem. 1996 Mar 8;271(10):5386-92

Appendix

1. Reactions

All of the biochemical reactions used in this model can be therefore divided into four main classes:

1. Enzymatic activation/deactivation reactions modelled with a modified Michaelis-Menten law (for example the phosphorylation activating Erk1/2 operated by Mek1/2);
2. Binding/unbinding reactions, modeled with mass action law (for example the binding reaction between Egfr and EGF);
3. Protein degradation, modelled with mass action law (mostly concerning the phosphorylated Egfr);

Several published models describe enzymatically catalized reactions through the Micahelis-Menten kinetics. When describing activation or deactivation of proteins through kinases and phosphatases, in order to consider both the modifier and its substrate, we introduced light changes of the Michaelis-Menten equation. The modified law is:

$$v = k_{cat}\frac{[S][modifier]}{K_M + [S]}$$

where $k_{cat}$ is the number of enzymatic reactions occurring per second. In this way, this version is more adequate because the modifier's affinity for its substrate is taken into account.

Inhibitors bind to enzyme following the same laws that govern ligand-receptor interactions. This means enzyme inhibitors are subject to the same mass action kinetic principles from which Michaelis-Menten equations are derived (Walsh 2014). Our model is able to consider the effect of activators or noncompetitive inhibitors which can be added to the culture medium for experimental purposes. We already mentioned PD18 and PD98 in their capability to inhibit Mek1/2 activation. In these cases, an extra term was introduced:

$$v = k_{cat}\frac{[S]}{K_M + [S]}\frac{K_i}{K_i + [I]}$$

where $k_i$ is the inhibition constant and I is the inhibitor species. In the case of an activator, this was the equation implemented:

$$v = k_{cat}\frac{[S]}{K_M + [S]}\frac{[A]}{K_a + [A]}$$

where $k_a$ is the activation constant, and A is the activator species.

2. The model

To include all the entities and their respective interactions useful to the target of this study, all the needed information were retrieved from KEGG (Kyoto Encyclopedia of Genes and Genomes) PATHWAY Database (Kanehisa 2004), on Kegg reference ko05200 (Fig. A1).

A biochemical model that describes the behavior of the Erk1/2 cascade in response to extracellular stimuli was developed. The model connects its outcome with a simplified model that reproduces the behavior of the cellular cycle. Globally, the model consist of 42 species and 54 biochemical reactions.

| N | Reactants | kon | koff | kcat | KM | Ref. |
|---|---|---|---|---|---|---|
| **r1** | Egf + Egfr -> EgfEgfr | 6.2E-04 | | | | Berkers (1991) |
| **r2** | EgfEgfr -> Egf + Egfr | | 3.5E-04 | | | Berkers (1991) |
| **r3** | EgfEgfr + EgfEgfr -> EgfEgfr_2 | 0.01 | | | | Kholodenko (1999) |
| **r4** | EgfEgfr_2 -> EgfEgfr + EgfEgfr | | 0.1 | | | Kholodenko (1999) |
| **r5** | EgfEgfr_2 -> EgfEgfrp_2 | 1 | | | | Kholodenko (1999) |
| **r6** | EgfEgfrp_2 -> EgfEgfr_2 | | 0.01 | | | Kholodenko (1999) |
| **r7** | EgfEgfrp_2 + Shptp2 <-> EgfEgfr_2 | | | 69.3 | 3300 | Zhong-Yin Zhang (1996) |
| **r8** | EgfEgfrp_2 + Grb2 -> EgfEgfrp_2.Grb2 | 3 | | | | Kholodenko (1999) |
| **r9** | EgfEgfrp_2.Grb2 -> EgfEgfrp_2 + Grb2 | | 0.05 | | | Kholodenko (1999) |
| **r10** | EgfEgfrp_2.Grb2 + Sos <-> EgfEgfrp_2.Grb2.Sos | 0.01 | | | | Kholodenko (1999) |
| **r11** | EgfEgfrp_2.Grb2.Sos -> EgfEgfrp_2.Grb2 + Sos | | 0.06 | | | Kholodenko (1999) |
| **r12** | EgfEgfrp_2.Grb2.Sos -> EgfEgfrp_2.endo | 5E-04 | | | | Starbuck (1992) |
| **r13** | EgfEgfrp_2.Grb2 -> EgfEgfrp_2.endo | 5E-04 | | | | Starbuck (1992) |
| **r14** | EgfEgfrp_2 -> EgfEgfrp_2.endo | 5E-04 | | | | Starbuck (1992) |
| **r15** | EgfEgfrp_2.endo - > Egfr.endo, Egfr.endo | 0.5 | | | | |
| **r16** | Egfr.endo -> aa | 0.9 | | | | Estimated on the base of Sigismund (2008) |

| N | Reactants | kon | koff | kcat | KM | Ref. |
|---|---|---|---|---|---|---|
| **r17** | Egfr.endo -> Egfr | 0.1 | | | | Estimated on the base of Sigismund (2008) |
| **r18** | EgfEgfrp_2 + Cla <-> EgfEgfrp_2.cla | 1.7E-05 | | | | Starbuck (1992) |
| **r19** | EgfEgfrp_2.Grb2 + Cla -> EgfEgfrp_2.cla | 1.7E-05 | | | | Starbuck (1992) |
| **r20** | EgfEgfrp_2.Grb2.Sos + Cla -> EgfEgfrp_2.cla | 1.7E-05 | | | | Starbuck (1992) |
| **r21** | EgfEgfrp_2.cla -> Cla + EgfEgfrp_2 | | 0.0017 | | | Starbuck (1992) |
| **r22** | EgfEgfrp_2.cla -> Cla + EgfEgfrp_2.Grb2 | | 0.0017 | | | Starbuck (1992) |
| **r23** | EgfEgfrp_2.cla -> Cla + EgfEgfrp_2.Grb2.Sos | | 0.0017 | | | Starbuck (1992) |
| **r24** | EgfEgfrp_2.cla -> Egfrp_2.endo2, Cla.endo | 0.017 | | | | Starbuck (1992) |
| **r25** | Egfrp_2.endo2 -> Egfr.endo2, Egfr.endo2 | 0.5 | | | | |
| **r26** | Egfr.endo2 -> aa | 9.7E-04 | | | | Starbuck (1992) |
| **r27** | Egfr.endo2 -> Egfr | 3.67E-05 | | | | Starbuck (1992) |
| **r28** | Cla.endo -> Cla | 0.0017 | | | | Starbuck (1992) |
| **r29** | RasGDP + EgfEgfrp_2.Grb2.Sos -> Ras | | | 0.5 | 500 | Bhalla (2002) |
| **r30** | Ras + GTP -> RasGTP | 2.2E-03 | | | | Lenzen (1999) |
| **r31** | RasGTP -> Ras + GTP | | 23.6 | | | Lenzen (1999) |
| **r32** | RasGTP -> RasGDP | 2 | | | | Bhalla (2002) |
| **r33** | RasGTP + GAP -> RasGDP | | | 0.1 | 172.41 | Gideon (1992) |
| **r34** | Ras + RasGTP -> Rafp | | | 1 | 500 | Bhalla (2002) |
| **r35** | Rafp + Pp2a -> Raf | | | 5 | 15700 | Bhalla (2002) |
| **r36** | Mek + Rafp-> Mekpp | | | 0.105 | 159.1 | Bhalla (2002) |
| **r37** | Mekpp + Dusp -> Mek | | | 2.4 | 36000 | Yamaguchi (2002) from BRENDA |
| **r38** | Erk.cyt + Mekpp -> Erop.cyt | | | 0.105 | 159.1 | Bhalla (2002) |
| **r39** | Erkpp.cyt + Dusp -> Erk.cyt | | | 2.4 | 36000 | Yamaguchi (2002) from BRENDA |
| **r40** | Erk.cyt -> Erk.nuc | 0.0086 | | | | Fujioka (2005) |
| **r41** | Erk.nuc-> Erk.cyt | | 0,018 | | | Fujioka (2005) |
| **r42** | Erkpp.cyt -> Erkpp.nuc | 0.0070 | | | | Fujioka (2005) |
| **r43** | Erkpp.nuc -> Erkpp.cyt | | 0.013 | | | Fujioka (2005) |

| N | Reactants | kon | koff | kcat | KM | Ref. |
|---|---|---|---|---|---|---|
| **r44** | Erkpp.nuc -> Erk.nuc | 0.013 | | | | Fujioka (2005) |

**Table A1 - Reactions and Kinetic Parameters**. In order to be consistent with the database concentrations, all the concentration was converted to nano-molar (nM). Michaelis Menten constants are given in [nM], first order rate constants in 1/s and second order rate constants in [$nM^{-1}$ $s^{-1}$].

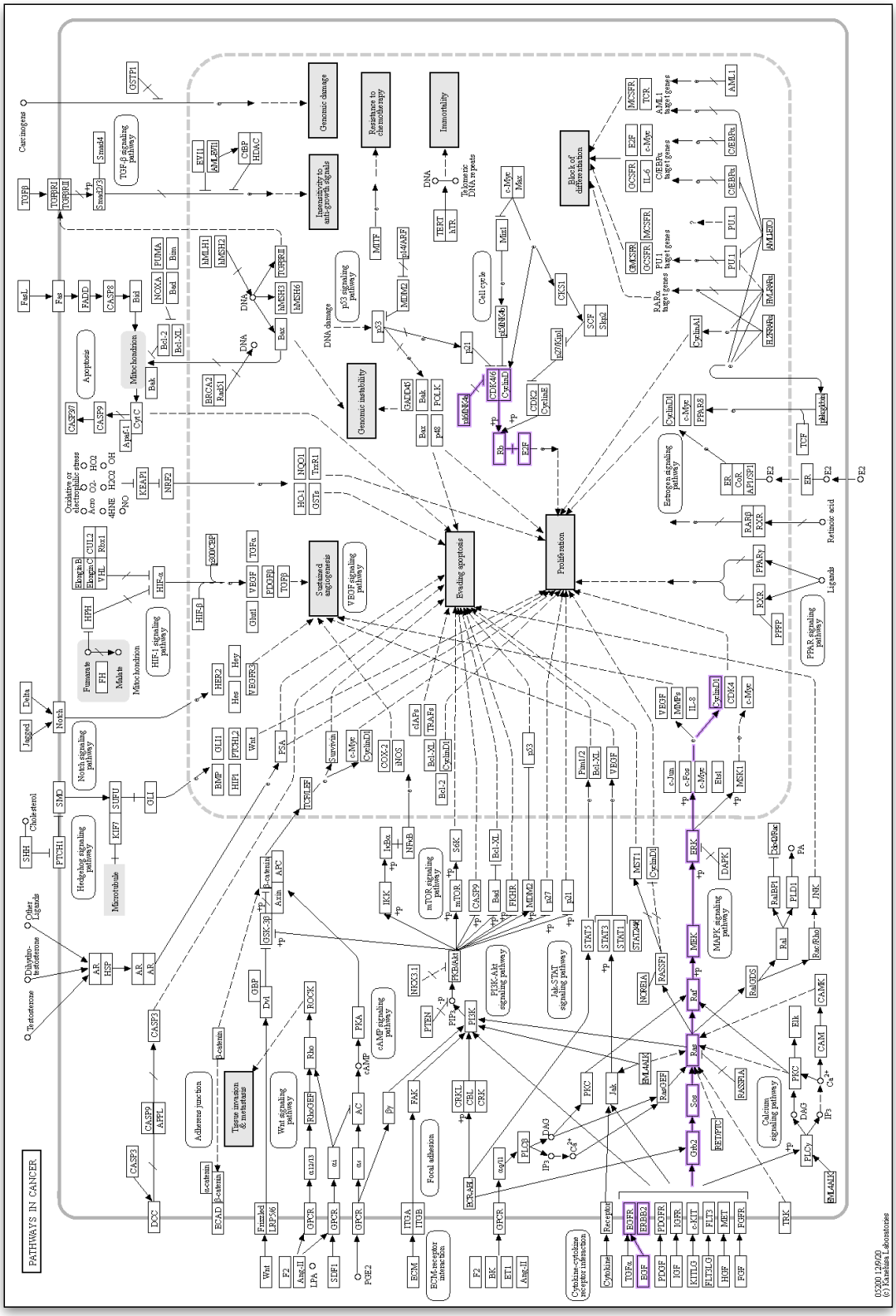



**Figure A1 - KEGG reference for the model development**

La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”