MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **10**, Number **1**, February **2013**

doi:10.3934/mbe.2013.10.103

pp. **103–120**

PARAMETER SPACE EXPLORATION WITHIN DYNAMIC SIMULATIONS OF SIGNALING NETWORKS

Cristina De Ambrosi, Annalisa Barla
Lorenzo Tortolina and Nicoletta Castagnino

DIBRIS Department of Informatics, Bioengineering, Robotics and Systems Engineering
Università degli Studi di Genova - Via Balbi, 5 - 16126 Genova, Italy

Raffaele Pesenti

DiMa - Department of Management
Università Ca' Foscari - Dorsoduro 3246 - 30123 Venezia, Italy

Alessandro Verri

DIBRIS Department of Informatics, Bioengineering, Robotics and Systems Engineering
Università degli Studi di Genova - Via Balbi, 5 - 16126 Genova, Italy

Alberto Ballestrero, Franco Patrone and Silvio Parodi

Di.M.I - Department of Internal Medicine
A.O.U. IRCCS San Martino IST, Italy
Università degli Studi di Genova - Via Balbi, 5 - 16126 Genova, Italy

Abstract. We started offering an introduction to very basic aspects of molecular biology, for the reader coming from computer sciences, information technology, mathematics. Similarly we offered a minimum of information about pathways and networks in graph theory, for a reader coming from the biomedical sector. At the crossover about the two different types of expertise, we offered some definition about Systems Biology. The core of the article deals with a Molecular Interaction Map (MIM), a network of biochemical interactions involved in a small signaling-network sub-region relevant in breast cancer. We explored robustness/sensitivity to random perturbations. It turns out that our MIM is a non-isomorphic directed graph. For non physiological directions of propagation of the signal the network is quite resistant to perturbations. The opposite happens for biologically significant directions of signal propagation. In these cases we can have no signal attenuation, and even signal amplification. Signal propagation along a given pathway is highly unidirectional, with the exception of signal-feedbacks, that again have a specific biological role and significance. In conclusion, even a relatively small network like our present MIM reveals the preponderance of specific biological functions over unspecific isomorphic behaviors. This is perhaps the consequence of hundreds of millions of years of biological evolution.

1. **Introduction.** We start with a general biological introduction, certainly redundant for the bio-medical component of people working in systems biology and medicine, but perhaps useful for people coming from the mathematic/information

2010 *Mathematics Subject Classification.* Primary: 92C42; Secondary: 92C40.

Key words and phrases. Systems biology, Molecular Interaction Maps, signaling networks, dynamic simulations, parameter space exploration.

sector. Conversely we will give some basic didactic information at the level of systems biology and information technology, including graph theory, to some extent redundant for people coming from the mathematic/information sector, but probably useful for people coming from the bio-medical sector.

On our earth, both living and non-living entities are composed of molecules made of atoms, such as Carbon, Hydrogen, Oxygen, and Nitrogen. The organization of these atoms into organic molecules, including macromolecules, of organic molecules in biochemical-interaction networks, and finally of multiple signaling-networks and structures into cells capable of reproduction, is one feature that distinguishes living entities from all other matter. The cell is the smallest unit that can carry on all the processes of life.

Most biomolecules contain Carbon and many contain Hydrogen, Oxygen, Nitrogen, Chlorine, Sodium, Potassium, Calcium, Magnesium, Iron, less common atoms, are also essential constituents of living cells. During more than three billion years of evolution cells have 'learned' to use atoms whose unique chemistry is compatible with carrying on the reactions necessary for life. Atoms can be arranged into a series of small molecules known as building blocks. Building blocks include compounds such as amino acids, nucleotides, sugars and fatty acids. The building blocks are organized into larger compounds, known as macromolecules. Macromolecules are components of different signaling-networks and structures that are found in the cells.

Four major different types of macromolecules are present in a cell:

- nucleic acids
- proteins
- lipids
- carbohydrates

Each type of macromolecule is used for a specific purpose in the cell. There are many examples of macromolecules being combined in different configurations to form larger cell structures.

The nucleic acids can be subdivided into DNA and RNA. **DNA** is composed of two kinds of building blocks, the bases (adenine, guanine, cytosine, and thymine) and a sugar-phosphate backbone. DNA is used by the cell as a repository of replicable information (from parental to daughter cell), necessary for the programs that direct synthesis of the macromolecules, production of energy for this synthesis, and control of specific cell differentiations. RNA has a very similar composition to DNA. The two major differences between DNA and RNA are in the sugar used in the sugar-phosphate backbone (ribose for RNA and deoxyribose for DNA) and in one of the bases (uracil for RNA and thymine for DNA). DNA usually exists as a double-stranded molecule, often in a diploid form: some redundancy of the information was evolutionary useful. **RNA** in the cell has at least four different functions.

1. Messenger RNA (mRNA) is used to direct the synthesis of specific proteins.
2. Transfer RNA (tRNA) is used as an adapter molecule between the mRNA and the amino acids in the process of making the proteins.
3. Ribosomal RNA (rRNA) is a structural component of a large complex of proteins and RNA known as the ribosome. The ribosome is responsible for binding a mRNA and tRNAs, and directing the synthesis of proteins.

4. The fourth class of RNA is a catch-all class. There are small and less small relatively stable RNAs whose functions are progressively discovered. Some of these RNAs have been shown to be involved in regulating expression of specific regions of the DNA.

Proteins are composed of amino acids. Most proteins are made from a unique combination of 20 different amino acids. The order in which amino acids appear in a protein are specified by the mRNA used to direct synthesis of the protein. All amino acids have a common core of repeating amino-carbon-carboxyl groups, with varying side chains on the central carbon. Proteins, therefore, have a repeating backbone with an amino terminus and carboxyl terminus. The amino acids can be grouped together and described by physical properties such as charge (acid or basic), size, interactions with water (hydrophobic-water "hating", or hydrophilic-water "loving"), a specific element (sulfur containing) or an organic structure they contain (aromatic rings). The primary sequence and the types of amino acids used to make up a protein, specify what the protein is capable of doing. Proteins perform many duties in the cell, including functioning as structural and motor components, enzymes, signaling molecules, and more in general regulatory molecules. Some proteins perform only one function while others are multifunctional. Each cell is an integrated device made of several thousand types of interacting proteins, plus other organic molecules [1, 2] We hope to have offered to our mathematical/modeling people a smattering of the biological side of this integrated game.

1.1. **How to define systems biology?** It is difficult to come up with a concise and comprehensive definition of systems biology. In fact Systems Biology is already a vast and multifaceted discipline touching many different types of analysis (mostly through mathematical modeling and other informatics approaches) toward a deeper and more integrated understanding of a variety of biological phenomena, seen in many different perspectives.

A few examples of definitions that have been proposed in the last ten years:

- *Systems biology studies biological systems by systematically perturbing them (biologically, genetically, or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately, formulating mathematical models that describe the structure of the system and its response to individual perturbations* [3]
- *To understand complex biological systems requires the integration of experimental and computational research - in other words a systems biology approach* [4]
- *Systems Biology - studying the interrelationships of all of the elements in a system rather than studying them one at a time* [5]
- *The objective of systems biology [can be] defined as the understanding of network behavior, and in particular their dynamic aspects, which requires the utilization of mathematical modeling tightly linked to experiment* [6]
- *By discovering how function arises in dynamic interactions, systems biology addresses the missing links between molecules and physiology. Top-down systems biology identifies molecular interaction networks on the basis of correlated molecular behavior observed in genome-wide* omics *studies. Bottom-up systems biology examines the mechanisms through which functional properties arise in the interactions of known components* [7]

Systems biology is also directly associated with bioinformatics and computational biology.

1.2. **What is a Molecular Interaction Map?** Since the 1990s the use of mathematical and computational models has become a valuable tool to deal with the rapid growth of information concerning biological networks. Dynamic simulations of networks of biochemical interactions among signaling-proteins / molecules have been applied to different biological systems, among others: concerning chemotaxis in bacteria, a seminal work of [8]. Small signaling circuits made of few molecules, involving specific types of feedback interactions have been described; discontinuous bistable dynamics or oscillations can be generated by some of these circuits [9]. In recent years the size of biochemical networks increased progressively from less than a dozen molecules to about 30-60 molecules [10, 11]. The number of different mammalian signaling-network regions to which a computational approach was applied is limited, probably in the order of the fingers of our two hands.

Larger simulations have been implemented in the field of bacterial metabolic models, to analyze metabolic control and flux balance [12, 13]. Different methods have been used to implement dynamic simulations of signaling networks, we mention only some of them: discrete approaches as Petri nets, logic-based descriptions like Boolean networks, rule-based methods, systems of ordinary differential equations and/or partial differential equations, stochastic methods involving the interaction of finite numbers of molecules [14, 15, 16]. It is important to be able to move easily a given signaling-network model from a given software tool to a different one. The Systems Biology Markup Language (SBML) is a machine-readable exchange format for computational models of biological processes [See http://sbml.org/Documents/. For collections of SBML models: http://www.ebi.ac.uk/biomodels-main/]. Its strength is in representing phenomena at the scale of biochemical reactions, but it is not limited to that. By supporting SBML as an input and output format, different software tools can operate on the same representation of a model, removing chances for errors in translation and assuring a common starting point for analyses and simulations. At the biochemical level, we may consider a normally differentiated cell as a very complex network of pathways, and we can interpret recent progress in molecular oncology as a description of a cancer cell bearing in the order of two dozen mutated pathways [17, 18]. Potential mutations belonging to the same pathway are hypothesized as being mutually exclusive [19] especially if very close along the pathway. Each pathway might contain a dozen signaling-molecules. In principle, one of them could be mutated/altered through gain or loss of function. The conclusion of these considerations is in agreement with the Vogelstein group's observation that about 20-40 different alterations that are present in an individual tumor are selected out of a pool containing about 200-400 potential oncogenes [18]. It makes sense to hypothesize that the most frequent mutations in a given tumor type, in some way, give a more important contribution to malignant transformation. Signaling-network analysis and modeling could contribute to find answers to this relevant question. In our work, signaling-network molecular pathologies move to the front stage. In this work we have primarily considered Breast Cancer (BC), and a fraction of the G0-G1 cell cycle transition. We adopted the approach of reconstructing the molecular anatomy of our network through a Molecular Interaction Map (MIM). We simulated the attainment of a stationary state in our biochemical network through hundreds of ordinary differential equations (ODEs). This approach belongs to a sub-field of Systems Biology. The opportunity, even the necessity, for

this dynamic simulation approach is a consequence of the fact that the behavior of a 30-40 molecules signaling-network is not intuitive a priori for a "naked-mind", even of an expert molecular oncologist. With some effort the mind of a cancer investigator is however capable of understanding "a posteriori" the suggestions coming from a computational approach.

A Molecular Interaction Map (MIM) is a diagram convention that is capable of unambiguous representation of networks containing multi-protein complexes, protein modifications, and enzymes that are sometimes substrates of other enzymes. This graphical representation makes it possible to view all of the many interactions in which a given molecule may be involved, and it can portray competing interactions, which are common in bio-regulatory networks. To avoid an overcrowded map, each molecular species is represented only once in a diagram. A formal description of the MIM notation can be found in [20, 21, 22]. An updated formal specification for software implementation can be found in [23]. In a Molecular Interaction Map (MIM), a variety of defined connecting lines serve to describe the interactions between the molecules (shown only once). A summary of the conventions used to depict binary interactions between molecules is shown below. Multi-molecular complexes or modified forms are depicted by "nodes" placed on the lines. A line may originate either at a named molecular species or at a node, and may terminate at a molecular species, a node, or at another line (contingency symbols, which modulate another reaction). Lines that cross do not imply an interaction.

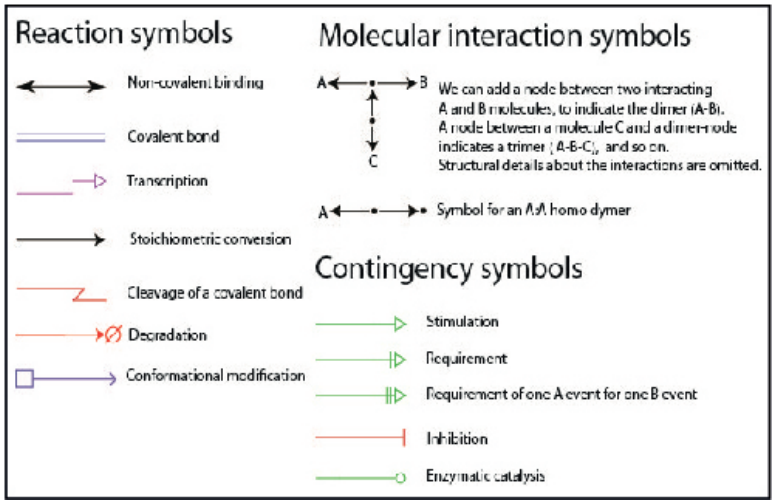

FIGURE 1. Syntactic rules for drawing a Molecular Interaction Map (MIM). According to [21]

1.3. **Networks and pathways.** We can analyze a MIM at two different levels: pathways and networks. Pathways imply 'paths', simple sequences of objects (molecules) that transmit information: pathways are the basic multi-molecular structure of a MIM, through which knowledge of biochemical interactions among proteins and other molecules is organized. Pathways are often not completely linear, but have ramifications: they are sub-components of larger networks. Networks usually represent a broader structure with a more complex connectivity, involving several pathways. A larger network suggests at least a portion of a function

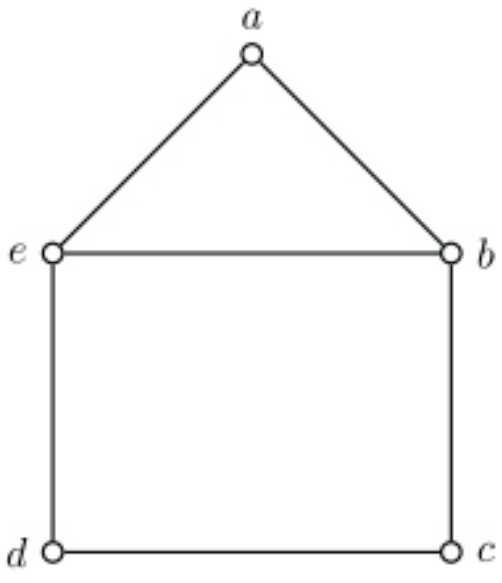

FIGURE 2. The graph G is undirected, meaning that we do not impose direction on any edges. Without any direction on the edges, the edge *ab* is the same as the edge *ba* [24]

displayed by a cell (for instance, in our case, a portion of the G0 - G1 cell cycle transition). A pathway may exhibit the name of some characterizing molecular component. Pathways and biochemical networks are both represented by graphs.

Definition 1.1. Graphs. A graph G=(V,E) is an ordered pair of sets. Elements of V are called vertices or nodes, and elements of E⊆V×V are called edges or arcs. We refer to V as the vertex set of G, with E being the edge set. The cardinality of V is called the order of G, and |E| is called the size of G.

Definition 1.2. Directed graphs. A directed edge is an edge such that one vertex incident with it is designated as the head vertex and the other incident vertex is designated as the tail vertex. A directed edge uv is said to be directed from its tail u to its head v. A directed graph or digraph G is a graph such that each of whose edges is directed. The in-degree of a vertex v∈V(G) counts the number of edges such that v is the head of those edges. The out-degree of a vertex u∈V(G) is the number of edges such that u is the tail of those edges.

2. Methods.

2.1. **Case study: Breast cancer.** Breast Cancer (BC) is a heterogeneous disease whose progression depends on its specific biological / molecular characteristics; this has consequences on the prognosis as well as on the response of individual patients to treatments. Considering women worldwide, BC is the most common form of cancer, both in terms of incidence 22.9% and in terms of mortality 13.7% [25]. Like other forms of cancer, BC can be considered as a genetic disease, since it is linked to sequential accumulation of mutations/alterations in genes (oncogenes) that control

growth, differentiation, correct position of a cell in the tissue architecture. These alterations, both genetic and epigenetic (like gene expression silencing) are in principle inherited at a somatic level, from parental cell to daughter cells. BC incidence can also be modulated by estrogens and progestins levels during life, because these hormones modulate breast cells differentiation. For instance, the use of estrogens and progestins increases moderately but significantly the risk of breast cancer in postmenopausal women [26]. Newly introduced DNA sequencing technologies allow us to draw and analyze new and broader landscapes of genes involved in the process of tumor onset and development. We observe a few genes mutated at high frequency together with a large number of less frequently mutated genes. It has been recently suggested that for each individual colorectal cancer (CRC) tumor the average number of the so-called driver mutations (including low frequency driver mutations) is about twenty [18], which is significantly higher than the number estimated up to few years ago. Probably these estimations are complicated by the fact that we have stronger and progressively weaker driving mutations. The gene products, primarily proteins, use molecular interactions in order to build up articulated communication systems (pathways) closely interconnected (molecular signaling-networks) which encode, process and transmit the information necessary to regulate all cellular functions. Cancer should be considered as a disease of cross-talks between normal and mutated-protein signals, rather than a simple sum of altered genes, a pathology of altered pathways and altered network regions.

The signaling-network immediately downstream of the ErbB-family is crucial in BC and other tumors, especially in the perspective of treatment strategies focused on signaling-protein inhibitors within an altered pathway. Within a pathway a protein could be mutated/altered through gain or loss of function. Exome sequencing works (Vogelstein group and other groups) tend to suggest the existence in CRC of 20-40 semi-autonomous pathways. Mutations in these distinct pathways are not mutually exclusive, but rather complementary, concurring to the completion of an overall cellular malignant transformation.

We make reference to the mutations database of COSMIC v58 Release, 15th March 2012) [27]. In BC (all BC types), among the signaling-proteins present in our MIM, in the pathways downstream of ErbB2, ErbB2 is amplified in 20-30% of cases [28], PI3KCA is mutated in $\approx$ 26% of cases [27], PTEN is hypo-expressed or inactivated in $\approx$ 40% of cases [29, 30, 27], CDH1 (E-Cadherin) is mutated in $\approx$ 17% of cases , APC is mutated in $\approx$ 4% of cases, AKT in $\approx$ 4% of cases , BRAF in $\approx$ 3% of cases, and β-catenin is mutated in $\approx$ 2% of cases [27]. A smaller pathway (one that we explored less extensively) downstream of the ErbB-family receptors is represented by an activated, mutated, amplified EGFR receptor which can phosphorylate β-catenin in Y-654 and make it independent from E-Cadherin, thus making β-catenin able to migrate to the nucleus and co-operate with the transcription factor TCF/LEF. E-Cadherin (CDH1) is frequently mutated in BC and is then incapable of binding β-catenin, even in the absence of any EGFR stimulation. Nuclear β-catenin is a co-transcription factor for the transcription factor TCF/LEF (TCF-4). Cyclin D1 and c-myc (both transcribed by TCF/LEF + other transcription factors) are among the genes that are important to open the way to the G1-S transition.

As we have repeatedly discussed, close mutations along a given pathway tend to be mutually exclusive [19]. During cancer progression, not much will happen by adding two adjacent or close mutations within the same pathway, in the same cell.

At the same time, and as a consequence of the above considerations, the addition of all the mutually exclusive alterations along a given pathway represents the overall pathway alteration frequency for BC. It must be pointed out that along a given pathway even the loss of function of the gene product of a recessive oncogene (for instance PTEN) can contribute to the excess of function of the overall pathway.

The study of molecular-network alterations in cancer, in the presence of onco-protein mutations and onco-protein inhibitors, is a quite modern strategy of crucial importance, and in order to implement this type of research a computational approach is essential. Even for intensively explored network regions, parameter knowledge is often incomplete. To study the degree of tolerance of a network to parameter uncertainty, becomes a very important task, for a critical evaluation of the dynamic modeling of a given network region, after careful training of the model with direct and indirect literature inputs. This was the intent of the present study, obviously it is a work in progress.

2.2. **Molecular Interaction Map.** Our MIM, see Figure 3 , has been created using the symbol table originally proposed by [20, 21], slightly modified/adapted to fit to some new semantic requirement of our MIM [31].

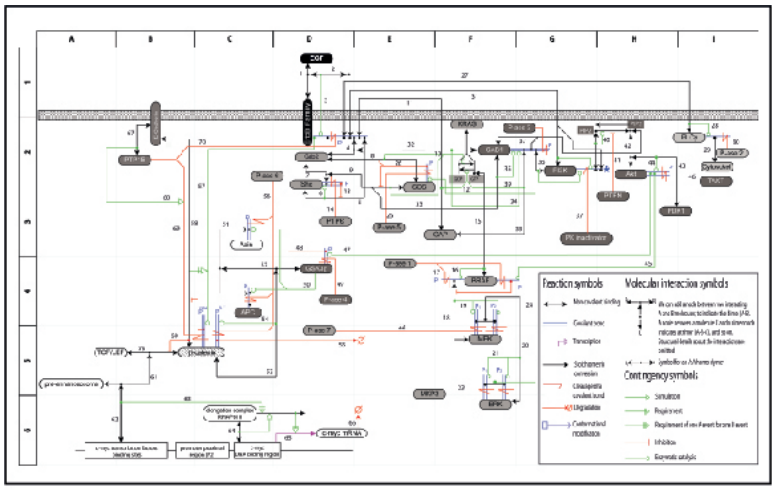

FIGURE 3. MIM of the signaling-network downstream of ErbB-family receptors.

A List of Abbreviations (a very synthetic definition of each molecule) is reported in a Glossary, Table 2SM in [31]. Our MIM describes a network downstream of the ErbB-family receptors that is relevant for BC. Similar networks are also operative in colon cancer, in Non Small Cell Lung Cancer (NSCLC) and perhaps in most tumors. In our mathematical modeling, a stationary, temporary equilibrium is assured by a growth factor (EGF), 10 kinases (ErbB1, ErbB2, ErbB3 counted separately), 14 phosphatases (including GAP),10 signaling / adaptor proteins, and 4 small signaling-molecules, for a total of 39 basic molecular species. Following the suggestion of [32, 33] we introduced in our simulation a phenomenon of 'piggyback' binding of SOS and GAP to an activated ErbB receptor in the sub-membrane region where KRAS is also anchored. This was equivalent to local association rate

increases of about 250 times [31], Table 3SM. Table 3SM shows a list of 279 reversible reactions and 110 catalytic reactions, rate-constants included, which represent the complete set of our dynamic simulations. Table 3SM also reports the concentrations of the 39 basic species. It is accompanied by the references source of the data. Some numerical values have been interpolated by taking into account the constraints imposed by: existing values, molecular anatomy of the network, indirect evidence at the molecular, cellular and clinical level. Narrow ranges of the interpolated values were practically imposed by the rest of the network system. The GDP and GTP species, as well as the cytoskeleton-protein, were considered in large excess (non-consumable).

2.3. **Ordinary differential equations.** The dynamics of the signaling network is numerically simulated by solving a system of ordinary differential equations (ODEs), with the help of dedicated software, such as the SimBiology toolbox of Matlab (http://www.mathworks.com/products/simbiology/?BB=1).

This kind of numerical approach has been pursued by different authors, among them [34, 35, 36, 37, 33, 38, 39], and all the other authors whose models are available in the BioModels Database [BioModels Database http://www.ebi.ac.uk/biomodels-main/]. The numerical solver adopted in our mathematical modeling was ode23tb, a solver for stiff differential equations that is an implementation of TR-BDF2, an implicit Runge-Kutta formula [40]. In our simulations, we started from a situation out of equilibrium. The total concentration, relative to a given basic protein/small-molecule species and all its complexes and post-translational modifications, was initially entirely attributed to the corresponding unbound factors. We brought the reactions to a quasi-stationary equilibrium, causing a redistribution of each factor among all its forms/complexes with binding partners. We verified (in a significant number of explored cases) that given constant molar concentrations of each component and constant virtual reaction rates inputs, regardless of the out of equilibrium starting state (and resultant variable transitional products), within a virtual time of 7-10 hours, end products consistently converged toward the same stationary equilibrium. For the limited complexity of the present MIM, few seconds of pc computation were in general sufficient for coming very close to a stationary equilibrium. We make reference to our pc model (Dell Optiplex 960, Intel Core 2 Duo processors @3.00 GHz, 4.00GB of RAM) and the SimBiology toolbox of Matlab software.

2.4. **Signaling network simulation.** In this paper we do not provide a detailed analysis, but when the system is already in a quasi-stationary state and we vary only the EGF concentration, for instance from a physiological EGF concentration (.1 nM) to a pharmacological EGF concentration (10 nM), a new quasi-stationary state is reached within few seconds of our pc computation. To simulate the signaling-network we considered in this paper, we mathematically formalized the reaction scheme of Table 3SM of [31], in terms of the reactions' kinetic laws [41]. The kinetic laws of a reaction describe the velocity at which the reactants are transformed into the products of the reaction. We assumed that all reactions followed a mass action kinetic law (a consequence of the IInd law of thermodynamics). According to this kinetic law, the velocity of the reaction is directly proportional to the concentration of the reactants multiplied by the reaction rate. As an example, given the reversible reaction:

$$[\mathrm{A}] + [\mathrm{B}] \leftrightarrow [\mathrm{A\text{-}B}]$$

the velocity of the [A-B] formation reaction is:

$$k_1[\text{A}][\text{B}] \text{ - } k_{-1}[\text{A-B}];$$

where each [X] indicates the concentration of a given reactant, k_1 and k_{-1} are forward (association) and backward (dissociation) rates, respectively, of the reversible reaction. At equilibrium

- $k_1 \cdot [\text{A}][\text{B}] = k_{-1} \cdot [\text{A-B}]$
- $([\text{A}][\text{B}])/[\text{A-B}] = k_{-1}/k_1 = K_d$ (equilibrium constant K).

We can also have an irreversible catalytic reaction of the type:

$$[\text{XP-Phosphatase}] \rightarrow [\text{X}] + [\text{Phosphatase}] + \text{P (P goes into the phosphates pool)}$$

$$\text{v} = k_{cat}[\text{XP-Phosphatase}]$$

where k_{cat} is a catalytic rate (a turnover number). Knowledge of the kinetic laws of the reactions has allowed us to describe the rate of change of each complex concentration by means of an ordinary differential equation in which the velocities of the reactions that produce or consume the reactant are algebraically summed. The collection of this type of differential equations for all 242 complexes plus 39 basic species included in the signaling-network fully describes the dynamic behavior of our biologic system. Unfortunately, the non linear nature of the above differential equations has prevented us from determining the analytical expressions for the system evolution over time. We could only obtain numerical solutions for a quasi-stationary equilibrium stage.

The simulation of the entire signaling network represented in the MIM is performed by SimBiology. Within this toolbox, we explicitly write all the reactions occurring in the signaling network of interest. SimBiology takes care of associating the ODEs to their corresponding reaction. For the simulations presented in this paper, we infer the rates from current literature and additional interpolations in agreement with numerous preclinical and clinical experimental papers [31]. We keep them constant throughout in silico experiments.

In order to explore sensitiveness/robustness toward perturbations of the introduced parameters, in this report we varied the concentrations, using (an arbitrary but reasonable choice): 10x and 10/ perturbations. The problem of exploring all possible combinations of parameters is computationally unfeasible; therefore we adopted a random strategy to explore the parameter space. We considered 10,000 (randomly sorted) 5-tuples combinations of perturbing species applying both 10x or 10/ perturbation factors (changes of the concentrations). The numerical simulation provides as output the final concentration at quasi-steady state of all the species belonging to the signaling network. Perturbed species were always not coincident with the perturbing species. Only perturbations on 34 basic species are reported herein (we excluded the constant not-consumable basic species). Perturbation effects on modified species and complexes are not reported.

2.5. **Exploration of random and non-random perturbations.** An aim of this paper was to explore the preliminary hypothesis that the effect of perturbing species that are far from the perturbed species could be negligible in a large majority of cases. We ignored however how far the effects of random perturbations could be propagated in our graph. We discovered that there are dramatically important exceptions along pathways that convey a signaling message of biological significance. These biologically significant exceptions (preferential paths) are unidirectional in our MIM. As mentioned before, a biochemical signaling-network can be

represented as a graph where edges represent the interaction between two adjacent species (nodes). The considered MIMs were first associated to a matrix of distances between nodes reporting the total number of edges between node pairs. We implemented perturbations of the concentrations concerning 34 consumable basic molecular species. EGF, GDP, GTP, PIP3 and cytoskeleton-protein were not considered. All PLCγP can convert to a [cytoskeleton-protein:PLCγP] complex (cytoskeleton-protein was implicitly considered in large excess). We introduced combinations of 10x and 10/ perturbations in one, or two, or three or four or five of the $n = 34$ consumable total molecular concentrations. Perturbed species were always not coincident with the perturbing species. Only perturbations on basic species are reported herein, while perturbation effects on modified species and complexes are not reported. For the time being we have not reported perturbations of rates either. Being n the number of perturbing species, we examined concentration levels of the 34-n perturbed basic molecular species, individually, always in the presence of the physiologic .1 nM EGF concentration. In our computational approach each of the 34 simulations (for just one perturbing species) gives information about all the remaining 33 perturbed species. The corresponding computing time on our desktop PC (Dell Optiplex 960, Intel Core 2 Duo processors @3.00 GHz, 4.00GB RAM) was 3x34 seconds (1.7 min approximately). In the case of five perturbing species and two perturbing factors (10x or 10/ concentration changes, see also the Methods section), the number of perturbing combinations is: The computing time on our PC would have been 3x8.9x106 seconds. This computing time is obviously too long for just one PC and we resorted to a random sampling strategy, sorting out randomly 10,000 combinations. This implies a computer time of about 3x10,000 seconds, approximately equivalent to 8.3 hours, a still acceptable length of computing time. Notice that sorting randomly 10,000 (perturbing) combinations we generate a subset of 10,000x29 (perturbing x perturbed) species. 29 = (all the 34 species considered) - (5 perturbing species). Note that each simulation is characterized by the following actors:

perturbing species: $\{S_i\}$ *s.t.* i=1:5
perturbed species: $\{S_j\}$ *s.t.* $\forall$j$\neq$i
perturbing factors: $\{v_i\}$ *s.t.* $\{v_i\}\in\{10X; 10/\}$
d_{ij} *s.t.* i=1:5

Distance was measured as a summation of edges for each of the n perturbing species, toward the perturbed species.

3. **Results.** Following our hypothesis that adjacent or close nodes are in general stronger perturbing species (or perturbators) than distant nodes, toward a perturbed species, we investigated a system of five perturbing species (randomly selected) toward a perturbed species, in a randomly sampled set of 10,000 cases.

In Figure 4 we measured the ratio between a given perturbed value (10x or 10/) and its corresponding unperturbed value (values expressed in a semi-log scale). We compared the original sampling defined above with subsets in which the (up to 3) closest nodes (edges) had been excluded. The perturbation became progressively weaker.

Figure 5 reports the results in terms of "relative variation" , measured according to the following formula:

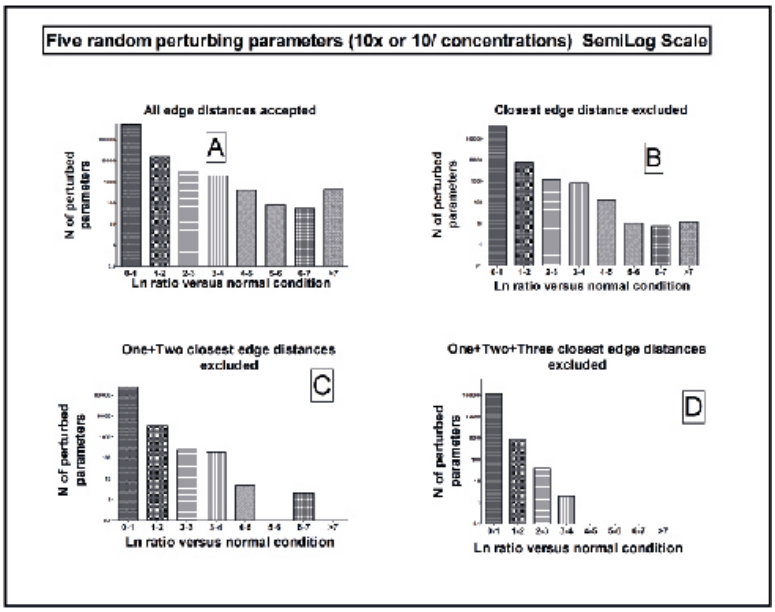

FIGURE 4. Attenuation of perturbation, excluding progressively a larger number of closer edges.

$$Relative\,Variation = \frac{\Delta Ratio}{Ratio(F)_{[A,B,C,D,E]}} = \\ = \frac{abs(Ratio(F)_{[A,C,E]} - Ratio(F)_{[A,B,C,D,E]})}{Ratio(F)_{[A,B,C,D,E]}}$$

where the term Ratio stands for a ratio between $[F]_{perturbed}$ and $[F]_{physiological}$. The previous formula will be generalized as:

$$Relative\,Variation = \frac{\Delta Ratio}{Ratio(F)_{[A_n]}} = \frac{abs(Ratio(F)_{[A_{n-x}]} - Ratio(F)_{[A_n]})}{Ratio(F)_{[A_n]}}$$

where n are all the perturbing species and x are the most distant perturbing species in terms of nodes/edges.

In Figure 5 we moved to progressively closer edges: out of five randomly sorted perturbing species, we chose the four closest ones to the perturbed species, the three closest one, the two closest ones, the closest of all perturbing species. Figure 5 (arithmetic scale), shows that average perturbations become stronger as we restrict our analysis to closer and closer perturbing species, toward the perturbed one. Going from histogram A to histogram D, the relative proportion of more perturbed species increases progressively. The extreme left column of each histogram (unperturbed species) decreases from about 90% (histogram A) to about 60% (histogram D).

Figure 4 and Figure 5 tend to confirm our hypothesis. When we select five random perturbator molecules (perturbing species), the largest part of the untargeted (outside of biologically significant pathways) effects induced on the concentration of the perturbed molecules depends prevalently on edge distance (the closest perturbator molecules). We could say that our network has a low level of random noise. In

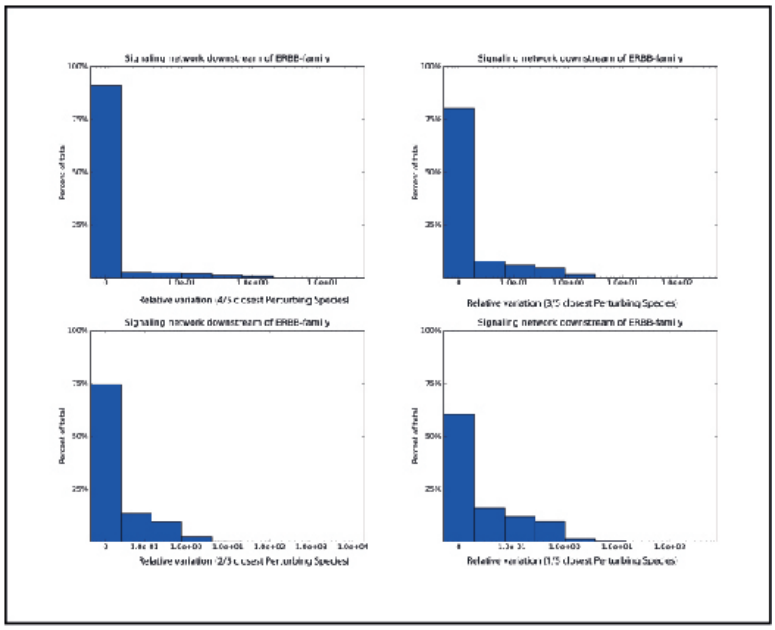

FIGURE 5. Perturbation effects including only progressively closer perturbing species.

other words, a non-specific propagation of the information in the graph of our MIM is quasi-inexistent. This makes sense: biochemical-interactions networks acquired specific-non-random functions during hundreds of millions of years of evolution.

3.1. **Our pathways are directed graphs.** The peculiar feature of the graph represented by our MIM is to be not isomorphic: there are privileged directions of propagation of the information. These privileged directions are recognized at the biological level as pathways. The entire MIM is essentially a limited network-region composed of few pathways and fragments of pathways. Along a pathway the propagation of the information is mostly unidirectional, with the exception of cases in which a downstream molecule sends a feedback signal of regulation to a molecule upstream in the pathway. This is another example of biologically-specific behavior. We come from the experience of mathematical modeling on our MIM, with the purpose of studying normal and altered pathways, and the role of virtual drugs inhibitors of signaling-proteins affected by excess of function [31]. We had already noticed that the transmission of information is directed and preferential along pathways which acquire biological relevance and significance for a molecular oncologist studying networks of biochemical interactions. Along these relevant pathways we notice the presence of very frequent and important mutations for the process of BC neoplastic transformation.

We present in the four tables (see Figure 6, 7, 8, 9) listing some of the major asymmetries in the propagation of the information present in our MIM. Notice that the general rule is that the propagation of information along a given pathway is unidirectional, except for the specific presence of backward feedbacks.

Minimum distance (number of edges)	1 edge downstream	2 edges downstream	2 edges upstream
Pase7 (phosphatase perturbing species)	MEKPP perturbed	ERKPP perturbed	BRAFP perturbed
Pase7 5 nM	102.48 nM	105.49 nM	0.54 nM
Pase7 50 nM	5.46 nM	1.17 nM	2.13 nM
Pase7 500 nM	0.012 nM	3.38E-06 nM	2.21 nM
|Fold change| (going from 5 to 500 nM)	8540	31228316	4.06

FIGURE 6. Along the pathway from KRAS to ERK (see MIM), this is one of the most dramatic known examples of specific and amplified propagation of signal. The activations of both MEK and ERK (ERK is immediately below MEK in the pathway) require a double parallel phosphorylation. Concentration dependence is therefore quadratic for two consecutive times, rather than linear.

Minimum distance (number of edges)	1 edge downstream	2 edges upstream	2 edges downstream via the negative feedback ERKPP - SOS
MKP3 (phosphatase perturbing species)	ERKPP perturbed	MEKPP perturbed	SOSP perturbed
MKP3 5 nM	20.10 nM	1.14 nM	0.0079 nM
MKP3 50 nM	1.17 nM	5.46 nM	0.00088 nM
MKP3 500 nM	0,0082 nM	6.30 nM	6.56E-6 nM
|Fold change| (going from 5 to 500 nM)	2441.83	5.53	1203.78

FIGURE 7. SOS would be apparently 5 edges upstream of MKP3, but it is directly connected through the negative feedback ERKPP - SOS. MEKPP is more weakly perturbed because it is 3 edges downstream of an inactivated SOSP (negative feedback SOSP-KRAS-BRAF-MEK).

Minimum distance (number of edges)	2 edges downstream	2 edges upstream
AKT (perturbing species)	GSK3β_P perturbed	PI3K_P perturbed
AKT 10 nM	0.099 nM	0.228 nM
AKT 100 nM	0.402 nM	0.229 nM
AKT 1000 nM	0.789 nM	0.235 nM
|Fold change| (going from 10 to 1000 nM)	7.94	1.03

FIGURE 8. Along the pathway from PI3K to GSK3β the propagation of signal is attenuated. Some experimental reports [42, 43] have suggested a weak or controversial connection between the PI3K-AKT pathway and GSK3β(engaged in canonical Wnt signaling). In our MIM signal propagation is relatively weak but not negligible in the downstream direction; it is practically absent in the reverse biologically non-significant direction.

4. **Conclusions.** For the time being, we have looked only at the free molecules as perturbing and perturbed objects of our observations, concentrations of complexes

Minimum distance (number of edges)	2 edges downstream		Minimum distance (number of edges)	2 edges upstream
EGFR (perturbing species)	β-catenin_P_Y654 perturbed		β-catenin (perturbing species)	EGFR_P perturbed
EGFR 10 nM	0.056 nM		β-catenin 5 nM	0.0137 nM
EGFR 100 nM	1.323 nM		β-catenin 50 nM	0.0137 nM
EGFR 1000 nM	7.09 nM		β-catenin 500 nM	0.0136 nM
|Fold change| (going from 10 to 1000 nM)	126.6		|Fold change| (going from 5 to 500 nM)	1.005

FIGURE 9. The physiological direction of signaling is from EGFR to β-catenin, the propagation of signal is evident. In the reverse direction there is practically no signal propagation.

and alterations of rates will be introduced in subsequent studies. The investigations reported in this work suggest that, as an average behavior, a 10x or a 10/ change in concentration in a given molecular species, tends to display rapidly decaying perturbing effects on the neighboring species, according to their distance measured in edges, when we deal with a-specific directions. We have however very important exceptions. These exceptions can be associated to asymmetries of the biochemical network and have a crucial biological significance. A perturbation significant at longer distances makes more relevant a corresponding mutation during malignant transformation. Notice that, from the perspective of signal transmission, mutations tend to be "all or none" changes. Among the asymmetries we have observed in our network (because of the existence of positive and negative feedbacks and local peculiar network architectures) we have noticed that downstream effects along biologically significant pathways tend to be amplified, or at least not attenuated, at variable degrees; attenuation of signaling suggests a weak connection between pathways (Figure 8). In conclusion, these biochemical interactions networks are basically non isomorphic, functional developments during evolution. Without performing a systematic analysis, we noticed that rare significant perturbations observed with the random perturbation approach tended to correspond to biologically significant directions of signal propagation. Signaling-proteins displaying high-frequency mutations/alterations, present in an altered network and involved in a specific malignant transformation, could perhaps represent biological hubs, linked not to the crude number of connecting edges, but rather to specific asymmetries of the networks, important in terms of signal propagation and its relative function.

Acknowledgments. This paper was partially supported by: Liguria Region Project (N. 280 - 2010-2011): "Study and simulation of molecular interaction networks relevant in malignant transformation; search and study of inhibitors of the oncoproteins c-Myc and Bcl-XL". CARIGE Foundation Project (N. 2010/228-16): "Analysis of molecular alterations in signal transduction networks downstream of EGFR-family receptors, in HER2 positive breast cancers and triple negative cancers. Rationalization at the clinical level of personalized antineoplastic therapies with onco-protein inhibitors". Compagnia di San Paolo Project (1471 SD/CC N. 2009.1822): "Models and computational methods in the study of physiology and

pathology of signaling networks of oncologic interest". Istituto Superiore di Oncologia (ISO): Grant 2006 from Istituto Superiore di Sanità: "Development of new drugs capable of modifying the cancer micro-environment". Grant RF-CAM-2006-353005 Regione Campania, from Italian Ministry of Health: "Molecular Diagnostic and Prognostic Markers of Thyroid Neoplasias". S. P. is grateful to Kurt W. Kohn for his pioneering inspirations toward the study of Molecular Interaction Maps of relevance in cancer.

REFERENCES

[1] J. Kendrew, "The Encyclopedia of Molecular Biology," Blackwell Science Ltd. Reprinted, 1995

[2] N. Trun and T. Trempy, "Fundamental Bacterial Genetics," Blackwell Publishing Company, 2004.

[3] T. Ideker, T. Galitski and L. Hood, *A new approach to decoding life: Systems biology*, Annu Rev Genomics Hum Genet., **2** (2001), 343–372.

[4] H. Kitano, *Computational systems biology*, Nature, **420** (2002), 206–210.

[5] L. Hood, *Systems biology: Integrating technology, biology, and computation*, Mech Ageing Dev., **124** (2003), 9–16.

[6] M. Cassman, *Systems biology: International research and development*, World Technology Evaluation Center. SpringerLink, Chapter I, (2007), 1–13.

[7] F. J. Bruggeman and H. V. Westerhoff, *The nature of systems biology*, Trends Microbiol., **15** (2007), 45–50.

[8] N. Barkai and S. Leibler, *Robustness in simple biochemical networks*, Nature, **387** (1997), 913–917.

[9] B. N. Kholodenko, *Cell-signalling dynamics in time and space*, Nat. Rev. Mol. Cell Biol., **7** (2006), 165–176.

[10] N. Borisov, E. Aksamitiene, A. Kiyatkin, S. Legewie, J. Berkhout, T. Maiwald, N. P. Kaimachnikov, J. Timmer, J. B. Hoek and B. N. Kholodenko, *Systems-level interactions between insulin-EGF networks amplify mitogenic signaling*, Mol Syst Biol., **5** (2009), 256.

[11] L. Tortolina, N. Castagnino, C. De Ambrosi, E. Moran, F. Patrone, A. Ballestrero and S. Parodi, *A multi-scale approach to colorectal cancer: From a biochemical-interaction signaling network level, to multi-cellular dynamics of malignant transformation. Interplay with mutations and onco-protein inhibitor drugs*, Current Cancer Drug Target (CCDT), **12** (2012), 339–355.

[12] D. Segré, D. Vitkup and G. M. Church, *Analysis of optimality in natural and perturbed metabolic networks*, Proc Natl Acad Sci U S A., **99** (2002), 15112–15117.

[13] D. Segré, A. Deluna, G. M. Church and R. Kishony, *Modular epistasis in yeast metabolism*, Nat Genet., **37** (2005), 77–83.

[14] W. Materi and D. S. Wishart, *Computational systems biology in drug discovery and development: methods and applications*, Drug Discov Today, **12** (2007), 295–303.

[15] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys.Chem., **81** (1977), 2340–2361.

[16] D. J. Wilkinson, *Stochastic modelling for quantitative description of heterogeneous biological systems*, Nat Rev Genet., **10** (2009), 122–33.

[17] T. Sjöblom, S. Jones, L. D. Wood, D. W. Parsons, J. Lin, T. D. Barber, D. Mandelker, R. J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. D. Markowitz, J. Willis, D. Dawson, J. K. Willson, A. F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. H. Park, K. E. Bachman, N. Papadopoulos, B. Vogelstein, K. W. Kinzler and V. E. Velculescu, *The consensus coding sequences of human breast and colorectal cancers*, Science, **314** (2006), 268–274.

[18] L. D. Wood, D. W. Parsons, S. Jones, J. Lin, T. Sjöblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezso, V. Ustyanksky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P. A. Wilson, J. S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J. K Willson, S. Sukumar, K. Polyak, B. H. Park, C. L. Pethiyagoda, P. V. Pant, D. G. Ballinger, A. B. Sparks, J. Hartigan, D. R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S. D. Markowitz, G. Parmigiani, K. W. Kinzler, V. E. Velculescu and B.

Vogelstein, *The genomic landscapes of human breast and colorectal cancers*, Science, **318** (2007), 1108–1113.

[19] C. H. Yeang, F. McCormick and A. Levine, *Combinatorial patterns of somatic gene mutations in cancer*, FASEB J., **22** (2008), 2605–2622.

[20] M. I. Aladjem, S. Pasa, S. Parodi, J. N. Weinstein, Y. Pommier and K. W.Kohn, *Molecular interaction maps–a diagrammatic graphical language for bioregulatory networks*, Sci STKE., **222** (2004), pe8.

[21] K. W. Kohn, M. I. Aladjem, J. N. Weinstein and Y. Pommier, *Molecular interaction maps of bioregulatory networks: A general rubric for systems biology*, Mol. Biol. Cell, **17** (2006), 1–13.

[22] K. W. Kohn, M. I. Aladjem, S. Kim, J. N. Weinstein and Y. Pommier, *Depicting combinatorial complexity with the molecular interaction map notation*, Mol Syst Biol., **2** (2006), 51.

[23] A. Luna, E. I. Karac, M. Sunshine, L. Chang, R. Nussinov, M. I. Aladjem and K. W. Kohn, *A formal MIM specification and tools for the common exchange of MIM diagrams: An XML-Based format, an API, and a validation method*, BMC Bioinformatics, **12** (2011), 167.

[24] D. Joyner, M. Van Nguyen and N. Cohen, "Algorithmic Graph Theory," Version 0.5 2010 November 30.

[25] a *GLOBOCAN project* http://globocan.iarc.fr/.

[26] G. A. Colditz, S. E. Hankinson, D. J. Hunter, W. C. Willett, J. E. Manson, M. J. Stampfer, C. Hennekens, B. Rosner and F. E. Speizer, *The use of estrogens and progestins and the risk of breast cancer in postmenopausal women*, N Engl J Med., **332** (1995), 1589–1593.

[27] a *COSMIC 2012: Catalogue of somatic mutations in cancer*, http://www.sanger.ac.uk/genetics/CGP/cosmic/.

[28] M. Mukherji, L. M. Brill, S. B. Ficarro, G. M. Hampton and P. G. Schultz, *A phosphoproteomic analysis of the ErbB2 receptor tyrosine kinase signaling pathways*, Biochemistry, **45** (2006), 15529–15540.

[29] N. R. Leslie and C. P. Downes, *PTEN function: how normal cells control it and tumour cells lose it*, Biochem. J., **382** (2004), 1–11.

[30] E. Tokunaga, E. Oki, Y. Kimura, T. Yamanaka, A. Egashira, K. Nishida, T. Koga, M. Morita, Y. Kakeji and Y. Maehara, *Coexistence of the loss of heterozygosity at the PTEN locus and HER2 overexpression enhances the Akt activity thus leading to a negative progesterone receptor expression in breast carcinoma*, Breast Cancer Res. Treat., **101** (2007), 249–257.

[31] N. Castagnino, L. Tortolina, A. Balbi, R. Pesenti, R. Montagna, A. Ballestrero, D. Soncini, A. Nencioni and S. Parodi, *Dynamic simulations of pathways downstream of ERBB-family, including mutations and treatments: Concordance with experimental results*, Current Cancer Drug Targets (CCDT), **10** (2010), 737–757.

[32] B. N. Kholodenko, J. B. Hoek and H. V. Westerhoff, *Why cytoplasmic signalling proteins should be recruited to cell membranes*, Trends Cell Biol., **10** (2000), 173–178.

[33] J. Wolf, S. Dronov, F. Tobin and I. Goryanin, *The impact of the regulatory design on the response of epidermal growth factor receptor-mediated signal transduction towards oncogenic mutations*, FEBS J., **274** (2007), 5505–5517.

[34] B. N. Kholodenko, O. V. Demin, G. Moehren and J. B. Hoek, *Quantification of short term signaling by the epidermal growth factor receptor*, J Biol. Chem., **274** (1999), 30169–30181.

[35] N. I. Markevich, G. Moehren, O. V. Demin, A. Kiyatkin, J. B. Hoek and B. N. Kholodenko, *Signal processing at the Ras circuit: what shapes Ras activation patterns?*, Syst Biol (Stevenage), **1** 2004, 104–113.

[36] A. Kiyatkin, E. Aksamitiene, N. I. Markevich, N. M. Borisov, J. B. Hoek and B. N. Kholodenko, *Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops*, J. Biol. Chem., **281** (2006), 19925–19938.

[37] M. R. Birtwistle, M. Hatakeyama, N. Yumoto, B. A. Ogunnaike, J. B. Hoek and B. N. Kholodenko, *Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses*, Mol. Syst. Biol., **3** (2007), e144.

[38] W. W. Chen, B. Schoeberl, P. J. Jasper, M. Niepel, U. B. Nielsen, D. A. Lauffenburger and P. K. Sorger, *Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data*, Mol. Syst. Biol., **5** (2009), e239.

[39] T. Nakakuki, M. R. Birtwistle, Y. Saeki, N. Yumoto, K. Ide, T. Nagashima, L. Brusch, B. A. Ogunnaike, M. Okada-Hatakeyama and B. N. Kholodenko, *Ligand-specific c-Fos expression*

emerges from the spatiotemporal control of ErbB network dynamics, Cell., **141** (2010), 884–896.

[40] G. Ernst and G. Wanner, "Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems," Springer-Verlag, 1996.

[41] J. J. Tyson, B. Novak, G. G.M. Odell, K. Chen and C. D. Thron, *Chemical kinetic theory: understanding cell-cycle regulation*, Trends Biochem. Sci., **21** (1996), 89–96.

[42] S. S. Ng, T. Mahmoudi, E. Danenberg, I. Bejaoui, W. de Lau, H. C. Korswagen, M. Schutte and H. Clevers, *Phosphatidylinositol 3-kinase signaling does not activate the Wnt cascade*, J Biol Chem., **284** (2009), 35308–35313.

[43] D. Voskas, L. S. Ling and J. R. Woodgett, *Does GSK-3 provide a shortcut for PI3K activation of Wnt signalling?*, F1000 Biol Rep., **2** (2010), 82.

Received April 19, 2012; Accepted July 14, 2012.

E-mail address: cristina.deambrosi@unige.it
E-mail address: annalisa.barla@unige.it
E-mail address: lorenzo.tortolina@unige.it
E-mail address: nicoletta.castagnino@unige.it
E-mail address: raffaele.pesenti@unive.it
E-mail address: alessandro.verri@unige.it
E-mail address: aballestrero@unige.it
E-mail address: fpatrone@unige.it
E-mail address: silvio.parodi@unige.it