
# *Network Based Simulation on HPC for Translational Medicine: an Application to Anticoagulation*

A dissertation

by

CASTALDI DAVIDE FABIO

University of Milano-Bicocca - Italy

in partial fulfillment of the requirements for the degree of

DOCTOR of PHILOSOPHY

July 2014

Department of Informatics, Systems and Communication -DISCo

Advisors: Prof. Francesco ARCHETTI

*To my well-beloved family, my unique Erika, who knows the scientific research challenges and my irresistible child Leonardo, who gives new meaning to my life. Thanks to them for their patience, understanding, and moral support during these years of work.*

*I hope that the time taken away from them for my interdisciplinary research can be useful and inspiring to the scientific community.*

# Contents

# List of Figures

*A theory has only the alternative of being right or wrong.*
*A model has a third possibility: it may be right, but irrelevant.*

Manfred Eigen
Nobel Prize for chemistry in 1967

# Acknowledgements

Through these three years of intensive research, several people have contributed with their knowledge and support to my research, to my academic career and professional training.

All my old and new colleagues at DISCo University department and at Consorzio Milano Ricerche have contributed to my growth in the field of computer science. Each of them with their own experience helped me to overcome small obstacles along this path, as well as with them I could have had fruitful exchanges of views. For all of this I have to thank Antonio, Dante, Divna, Elisabetta, Gaia, Paolo and my PhD colleague Raul.

In particular, I want to emphasize the support given me by my colleague Ilaria, who has been able to guide me in the doctoral program and provided me with relevant advices on how to move in the academic world; and my colleagues Daniele and Davide, which, by means of their computer science skills, have been able to give me a valuable help in complex situations.

A special Thanks to my Supervisor Prof. Archetti F. He believed in me from the beginning, appreciating my bioclinical background. His ways of guiding me in academic matters have been motivating, encouraging, and supportive. He has made me constantly strive to perform my very best. I am indebted to him for giving me another way of understanding science.

A continuous gratitude to Professor Mari, Ospedale Maggiore Policlinico of Milan, as well as having contributed and validated my thesis by her expertise in hematology, she never left me aside despite my change in the scientific field.

In conclusion, I cannot forget how essential was the support of my parents who allowed me to continue to study for many years, believing in me and giving me financial security

# 1 Summary

A very brief description of the thesis project can be summarized as:
A machine learning investigation on Oral Anticoagulation Therapy (OAT) in elderly patients to improve the dynamic network modelization of the bioclinical process of coagulation cascade.
In particularly, after the construction of a database, including the features of therapeutic interest and new index originated by computational approach, we compared and tested techniques of supervised Machine Learning for classification and regression analysis, in order to develop an innovative model, compared to those currently available in the literature. The results of this investigation can provide adequate support for dual purpose: give an advantage in managing the therapy in patients with an increased risk of adverse drug events (ADRs), as elderly patients are; define pharmacodynamics characteristics of the drug employed in the oral anticoagulant therapy to finely configure a mechanistic model of involved pathway, providing further insights in therapy management.
Therefore, simultaneously, we investigated and tested techniques of modeling and simulation in order to analyze the system in terms of both structure and dynamics to specify the qualitative and quantitative properties of the system. Such an approach would make it possible to profile a manageable model of the coagulation cascade which is able to adapt to the need for integration of recent pharmacological and genetic data, distinctive of modern pharmacogenomic characterization of anticoagulant therapy.
Among the different mathematical modeling frameworks that have evaluated, the Temporal Extension of Petri Nets (Timed Petri Net), in particular the Stochastic Petri Net (SPN), seem to offer the most suitable characteristics. They permit an approach to modelization, combining an intuitive graphical representation to a valid solid mathematical, which give rise to a model numerically solving, able to represent the relational components and the quantitative time-dependent behaviors, including randomness, of a complex biological system such as coagulation cascade. The employment of such a framework, compared to the most

popular models to differential equations provides a tool that can greatly enhance the interaction with biologists and clinicians enabling them to both evaluate the entire structure of system and cooperate for the development of any new modules.
The biological system was mapped onto Petri Net by means of Snoopy editor, and the simulations were executed employing as stochastic solver the Tau- Leaping Stochastic Simulation Algorithm on Stochkit2 software. In order to simulate a high number of trajectories of the system, representative of several bioclinical scenarios, we employed HPC technology.

The stochastic approach to simulate clinical events and interventions, due to its ability to represent the variability of the real system, could provide an interesting innovation for the integration of simulation techniques in clinical decision support according to recent translational medicine perspective.

# 2 Introduction

The sudden increase of molecular information has resounded throughout all areas of biological study. Scientists are increasingly able to suppose dense processes in terms of the interaction of multiple pathways, and this knowledge of life at different molecular levels has illuminated relationships that were previously unknown. The development of these complex networks of heterogeneous data has required, more than ever, the support of computer science.

Information Technology has taken part in this field of applications, complementing an existing branch of computer science dedicated to the biological issues, the bioinformatics. Thus, it was necessary to found a new discipline, with a broad multidisciplinary approach and the attempting purpose to understand the mechanisms that rule the biochemistry of life [Ideker et al., 2001]. Using this cutting-edge perspective, we more strongly connect the different areas of Biology, Biotechnology, and Computer Science to be increasingly able to describe complicated processes in terms of the interaction of specific molecules, entire pathways or multiple biochemical systems to get inside in the innate hierarchical organization of biological systems [Raven et al., 2014][Ravez et al, 2002].

# 3 The Bioclinical Pathway

A biological pathway which has recently attracted strong interest in supporting bioclinical knowledge with computational modeling is the coagulation cascade [Chatterjee et al., 2010] [Wajima et al., Isbister and Duffull, 2009). This model was biochemically deeply investigated, resulting in complex networks where two sub-pathways, Intrinsic and Extrinsic interacting to produce active fibrin which affects the haemostasis process [MacFarlane, 1964] (Butenas and Mann, 2002).

Both pathways are required for normal haemostasis and there are different feedback loops between the two pathways that amplify reactions to produce enough fibrin to be able to crosslink each other to form a fibrin clot.

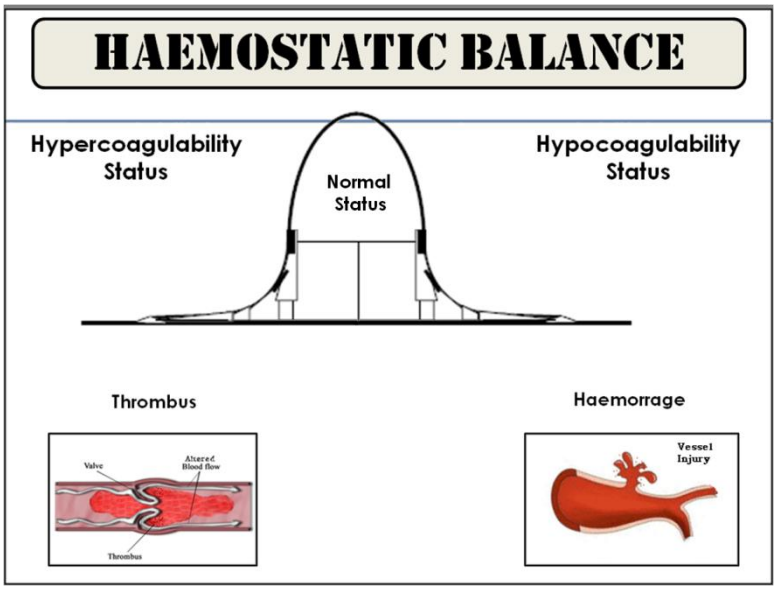



Figure 3.1 - Haemostatic Balance: The diagram shows as the haemostatic process is characterized by a dynamic equilibrium to assure simultaneously blood fluidity and responsiveness in carrying out anti-hemorrhagic function, action taken slower in the event that it was in a quiescent equilibrium

Modulation of extrinsic (Tissue Factor Pathway) and intrinsic (Contact Pathway) pathways is acted either by several positive and negative feedback mechanisms, or by a set of natural proteins that restrict coagulation progression. In healthy physiological conditions constant generation of small amounts of coagulation active factors (active proteases) affect the constitutive haemostasis, while another biological mechanism, the fibrinolytic system, counterweigh the process. All together, coagulation, anticoagulation and fibrinolysis define a delicate physiological balance, named “Haemostatic Balance” [Gaffney et al., 1999]; significant deviations from this equilibrium, hypercoagulability or hypocoagulability unbalance, can evolve in cardiovascular adverse events.

Many different biochemical techniques are employed to evaluate the coagulation performances from clot-based, direct chemical measurements, ELISAs test, to chromogenic. One of the most used, is the Prothrombin Time test (PT Test) which, belonging to the clot-based test, provide a global assessment of coagulation function. The PT test is performed by adding a thromboplastin reagent that contains Tissue Factor (which can be recombinant) and calcium to plasma to measure the clotting time. Typically, PT reagents contain excess phospholipid, so platelet activity does not influence the performances of the test. The PT is prolonged with deficiencies of factors VII, X, and V, prothrombin, or fibrinogen and by antibodies directed against these factors.

The classical *cascade model* of the coagulation mechanism is reevaluated by the new, *Cell-based model* of *coagulation* which emphasizes the interaction of coagulation proteins with the cell surfaces of *platelets, subendothelial cells* and the *endothelium.* According to this model coagulation is initiated (*The Initiation Phase)* by the formation of a complex between tissue factor (TF) exposed on the surface of fibroblasts as a result of vessel wall injury, and activated factor VII (FVIIa), normally present, in tiny quantities, in the circulating blood. The TF-FVIIa complexes convert FX to FXa on the TF bearing fibroblasts. FXa then activates prothrombin (FII) to thrombin (FIIa). The next phase is the *Amplification Phase* in which this limited amount of thrombin activates FVIII, FV, FXI and platelets, on the surface of blood platelets. Thrombin-activated platelets change shape and as a result will expose negatively charged membrane phospholipids, which form the perfect template for the assembly of various clotting factors and full thrombin generation, involving FVIIIa and FIXa (*The Propagation Phase*)[ Smith, 2009].

## 3.1 Coagulation Process Significance

Studies to investigate the molecular mechanisms of coagulation have maintained a great interest among researchers because of the direct

and indirect implications of coagulation disorders in the development of various diseases. Impairments in haemostatic balance directly cause diseases cardiocirculatory highly disabling (thrombosis, myocardial infarction), but also have an important role in the one set of complications in diseases not haemostatic (anticoagulation cirrhotic coagulative paraneoplastic syndromes, immunological syndromes) [Byrne et al., 2014] [Lisman and Porte, 2010].

In addition, functional impairments of this pathway contribute both as a causal factor and as a worsening factor in other diseases with high clinical impact and increasing social impact as the frailty syndrome, neurological disorders (dementia and neurodegeneration) and systemic inflammation as a risk factor for chronicization and susceptibility to infections [Levi et al.,2003].. In latest investigations on coagulation trigger mechanism of extrinsic pathway, were identified three overlapping procoagulant stages] [Hoffman and Monroe,. 2007].that have changed the way we consider the coagulation cascade, which more than a series of sequential events (waterfall of course) proves to be characterized by an intricate mechanism of modulation of events. The revised coagulation cell-based model further reinforces the concept that some proteases (active factor) involved in enzymatic coagulation signal diffusion, may function a notable roles outside of haemostasis, such as inflammation, vessel wall functionality, and cell proliferation.

## 3.2 New Insight of Coagulation

Ongoing research has elucidated other components of the coagulation process. These components include microparticles

Microparticles are irregularly shaped vesicles that arise from the plasma membrane of blood-borne cells during cell activation, programmed cell death or exposure to shear stress [Davizon and Lopez, 2009]

Microparticles seems play a role in promoting thrombosis during the amplification phase of coagulation. The molecular thrombus formation

process, include platelets accumulate at the site of vascular damage, which turn in activated status and express P-selectin adept to TF-bearing microparticles binding. TF from the microparticles then binds to and activates Factor VII activating the tissue factor process leading to clot formation [Furie and Furie , 2008].
TF-bearing microparticles are also implicated in thrombosis associated with diabetes, metabolic syndrome, specific malignancies (e.g. cancer of the colon, pancreas, breast, ovary and lung) and inflammatory and haematological disorders. [Davizon and Lopez, 2009]

## 3.3 The Anticoagulation Therapy

Treatments are currently available to protect AF patients from stroke or related complications due to blood clots. The most common and effective treatment is the use of the oral anticoagulant medication Warfarin. Despite its proven efficacy, Warfarin therapy is not fitting with all patients, has a very narrow therapeutic range, and entails risks for 2010]bleeding complications. Additionally, the effectiveness of this blood thinner varies because of interactions with certain foods and medications, and surely due to genetic implications [Siguret et al.,2008] Therefore, to be safe and effective, it requires frequent blood tests and dose adjustments, which can be troublesome. Even with dose adjustments, patients are usually outside the therapeutic range about half the time.
When prescribing Warfarin therapy in patients with atrial fibrillation (AF), physicians must weigh the risks of significant bleeding complications and non-compliance issues against the risk of ischemic stroke without Warfarin. Fewer than 50% of the patients eligible for long-term Warfarin are currently being treated due to either noncompliance or tolerance issues [Gouin-Thibault et al.2010].
With the known disutility of Warfarin, and the questionable effectiveness of aspirin, a computer assisted-based solution may provide

added protection against migration of blood clots in certain patients with AF.

### 3.3.1 Oral anticoagulation Therapy in Elderly

The problems associated with Warfarin therapy are particularly significant in the elderly population, which has a greater risk of falling, thereby increasing the chances of serious outcomes in the case of bone fracture. With ageing they acquire a procoagulant status [Mari et al., 2008][Forest et al., 2010] which can mislead a therapist not used to dealing with elderly patients , who therefore may underestimate the therapy in the induction phase to subsequently risking to overestimate it trying to reach the maintenance dose. Moreover, elderly suffer of vascular fragility [Mayer and Rincon, 2005] which can contribute to the more dangerous drug adverse effect [, the intracranial hemorrhages. All this issues make the anticoagulation therapy in elderly particularly challenging greater challenges with maintaining the therapeutic range. This is of importance since individuals over the age of 75 constitute approximately half of AF-associated stroke patients.
Issue of therapy for elderly can be streamlined by an accurate personalized treatment managed with information technologies assistance, in particular a coupling pattern discovery process and profile modeling [Schera et al., 2014]

### 3.3.2 Computer Aid to Warfarin Therapy

Computer-assistance and self-monitoring lower the cost and may improve the quality of anticoagulation therapy. The main purpose of this clinical investigation was to use computer-assisted oral anticoagulant therapy to improve the time to reach and the time spent within the therapeutic target range compared to traditional oral anticoagulant therapy by physicians. This approach must be contemplated especially in older patients due to the fragility status distinctive of ageing process [Gouin-Thibault et al., 2010].

Recently, particularly with the aid of advanced technologies capable of reducing the simulations computational cost, is developing the idea to use a combination of traditional decision-making algorithms, with mathematical models to simulate clinical an therapeutics aspects Computational simulations have been employed to gather valuable insights into the molecular mechanisms of cancer [Gallasch et al., 2013], assisting oncologists to improve diagnosis and prognosis of the disease, and as an attractive decisive goal, to identify novel therapeutic targets. To this emerging trend of *Computational Clinics*, does not fail to join also the fund cardiovascular clinical division, which dealing with wall geometry, structure, and properties and the hemodynamic features to manage anorthic narrowing [Coogan et al., 2013].

# 4 Network Models in Life Sciences

Robert Rosen and Nicholas Rashevsky were probably the firsts (in 1950s) activists of *Relational Biology*. The two theoretical biologists had foresighted the need to understand the importance of relationships in biological systems, which is the valuable perspective considered in the present by Network Biology research. Their idea of biological systems consisted in the definition of *Relational Biology* itself: *it keeps the organization and throws away the matter; function dictates structure, whence material aspects are entailed*. Hence, the organization that all systems buid-up, is quasi-independent from the material particles that constitute the system and that is based on the nature of the interactions among those system's elements, as reported in a leading treatise on the origins of life firstly published in 1991 [Rosen, 2005]. This point of view is not far from the interpretation that is given by the modern biological networks theory, which aims to reconstruct the complex biological systems in complex networks can be analyzed in their whole, or in their fundamental constituents organization [Feng , Jürgenj and Qian, 2007][Kaneko, 2006].

The basic unit of living systems organization is its smallest vital unit, the cell. As we will explain during this section, interpret the functional fundamental architecture at the base of such autonomous element, since the biochemical functional structures reappears at different scale levels and in different organisms [Zinman et al., 2011], allows us to extend this general model both to systems at higher complexity scale (tissues, organisms , ecosystems) and lower complexity (organelles and molecules). This imply that analyzing generic properties that arise from the topology of the network motif, we argue that all instance of a particular motif display characteristics that can be studied comprehensively in biological network organization, such the robustness, which seems characterized by feedback loops or motif repetitions [Kim et al., 2008] [Kitano, 2004].

In this context of "shelf-organization", make an abstraction of biological processes through network representation is almost "*natural*".

The main functions that regulate cellular activity consist in environmental communication, functional and structural maintenance, and metabolic activity. These processes can be described with three, apparently disjointed, networks, reported in the sequence corresponding to stimulus, response, and effect: *Signaling cascades Networks; Gene Regulatory Networks;* and *Metabolic Networks*.

Signaling cascades mediate external stimuli from environment to the inner of cellular compartment (mainly the nucleus), or spread an internal signal to coordinate cellular activities or phenomena. Gene regulatory networks react to stimuli modulating the transcription level of genes, which consist in induction of protein production. These proteins are involved in all cellular functions, playing roles as enzyme, structural proteins, carriers and receptors (back to the signaling functions). Metabolic networks, essentially directed by gene products, use external sources transform them into energy and organic compound that are together employed in protein and nucleic acids synthesis, signaling energy consumption, homeostatic balance supports and products secretion, essentially in cell life.

Regarding to the extensively modeled *Protein-Protein Interaction Networks* (PPIs), initially effectively exploited to reveal the general organization and modules of the directly linked proteins involved in a biochemical process, were gradually adapted to graphically characterize the wet-lab measured protein-coexpression clusters and all kind of biochemical pathway, assuming only the mutual aspecific interactions.

A very simplified depiction of the main cellular processes organized in a cooperation network that as a whole guides the cell life, is shown in figure 4.1 below.

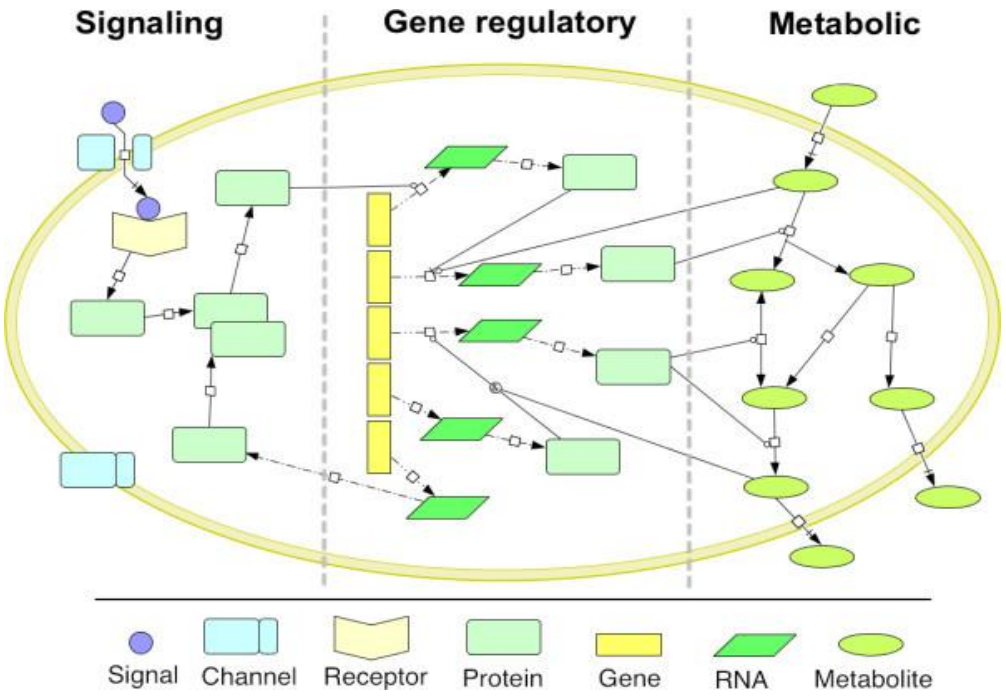



Figure 4.1- The life cellular processes: Schematic representation of the main interacting cellular processes that guide the cell life. Signaling cascades, Gene regulatory networks and Metabolic networks. mediate external stimuli from environment to the inner. Gene regulatory networks react to stimuli modulating the transcription level of genes, which consist in protein production induction. These proteins are involved in all cellular functions, playing roles as enzyme, structural proteins, carriers and receptors (back to the signaling functions). Metabolic networks use external sources transform them into energy and organic compound that are together employed in protein and nucleic acids synthesis, signaling energy consumption, homeostatic balance supports and products secretion, essentially in cell life.

## 4.1 Modeling Biological System

The widespread use of computer techniques in application domains related to biology hail from the late eighties, even if the first seventies, although in the early seventies, sporadic pioneers of computational biology have produced relevant results with regard to available technology [Garfinkel et al., 1970] [Newman and Rice, 1971] when was coined the term *Bioinformatics* [Lim, 2010]. Essentially, this approach concerns the storage, transformation and the examination of the data generated by biologists. Thus, the skillfulness most employed by computer scientists were the data mining algorithms, databases construction, and some artificial intelligence techniques, to try to extract meaningful data and make predictions from raw sets of data produced by *in vitro* experiments.
A more practical application of computer science in the field of biological sciences is summarized with the most recent definition of bioinformatics provided by Hwa A. Lim, who strengthened the definition of bioinformatics provided by who coined the term is

> "*the study of information content and information flow in biological systems and processes* "[Lim, 2010]

In the first decade, this new discipline has looked almost exclusively to the information content applying static techniques and ignoring the flow of information. Modern biology, and its application to contemporary clinics, highlights the clear need to integrate data from all disciplines called "*-omics* "(*genomics, proteomics, metabolomics*, mainly). With the aid of high-resolution technologies (high throughput technologies), these disciplines have generated in less time a larger amount of data that, merged each other, defined models of complex biochemical systems that can be analyzed within an acceptable time-span only by means of appropriate information technologies (IT) infrastructures and scalable approaches.

Though currently there is no doubt about the need to integrate biological investigations whit computer science, to improve the insights, is appropriate to point out how this collaboration has provided benefits to both discipline. If the support of information technology in biology applications is easily to get, it is not so renown that, thanks to the possibility to engage in this area, IT bumped into new exciting tasks to solve. Such stimulation, induced by a field in which the mathematical foundations are often overlooked, motivated computer scientists to create new mathematical theories and computational methods often specifically targeted for biological question, or, sometimes, directly inspired from biology, as genetic programming done. While this joint venture was already recognized to be effective 40 years ago, when we talked mostly about biomathematics [Bender, 1978], today it is at least undeniable perspective [Marvalee, 2008].

Hence, new dedicated methodologies and algorithms joined, with advances in computing power, extended the boundaries of bioinformatics at a higher level. Its goal was, so no longer limited to organization and extraction of knowledge from data, but rather to engage in the *in silico* reproduction of biological systems to generate new hypotheses about the complex regulatory mechanisms which direct their behavior.

The validation of the generated potential hypothesis has required the contribution of biological domain's experts not only supporting the computational results, moreover testing them in the real context, by *in vivo* examinations and in vitro experiments.

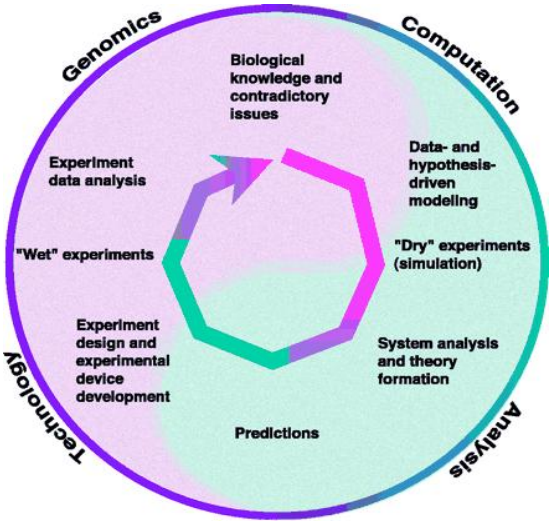



Figure 4.2 - The iterative workflow of System Biology approach

Hence, the required approach to comprehend complex biological systems, gathering both static and dynamic properties, recommends the mutual continuous integration of the experimental data and the computational results, which consist in the notion of the Systems Biology perspective.

The *in silico* component of this discipline is the Computational System Biology, which

> "*through pragmatic modeling and theoretical exploration, provides a powerful foundation from which to address critical scientific questions head-on"* [Kitano, 2002].

Computational techniques have contributed in three main ways, making up three main distinct computer science approaches in Computational Systems Biology [Kitano, 2002].

One is *knowledge discovery* or data mining, where hidden patterns are extracted from experimental data with the aid of machine learning techniques. The second, more recent regard to the traditional bioinformatics proposal is the branch of *Network Biology* originated from joined cooperation of graph theory and the biological pathway

discovery, which give rise to the field. The last one, often the most demanding method in terms of *a priori* knowledge is the *modeling and simulation-based analysis*, which tests hypotheses and validates models with *in silico* experiments, as shown in Figure 3.3:

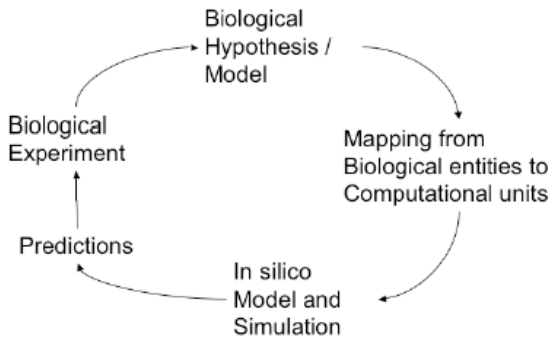



Figure 4.3 – The iterative workflow of Computational System Biology approach

During the last decade the Network Biology approach, has acquired high relevance in the representation of biological systems as resource to analyze and interpret relevant features of complex system [Emmert-Streib and Glazko, 2011][Kitano et al., 2005].

Not all biochemical pathway representations are necessarily able to reveal the biological insights of the process due to loss of information largely related to a static modeling. However, each of different level of graphical representation abstraction can reveal some peculiar structural features or define the general behavior of phenomena.
Currently, two main formalisms are used to graphically represent biological systems: informal graphical representation (wiring diagrams or pathway diagrams) typically employed in public databases such KEGG, Reactome, TRANSPATH, MetaCyc, or as a corporation services, like Quiagen Gene & Pathway and Ingenuity; and formal graph representation based on the network theory.

The informal graphical representation, even known as *Wiring Diagram*, often has the disadvantage to have not a standard notation for type of relations, misplacing the possibility to discriminate the type of biological process, thus their interpretation can be often ambiguous. Moreover, even the elements that compose the graph (nodes and edges features) have no reference standards to be clearly defined. As a result, these models differ in graphic symbols and their semantics.

An illustrative example of informal depiction of a biochemical pathway is reported below (Figure 3.4)

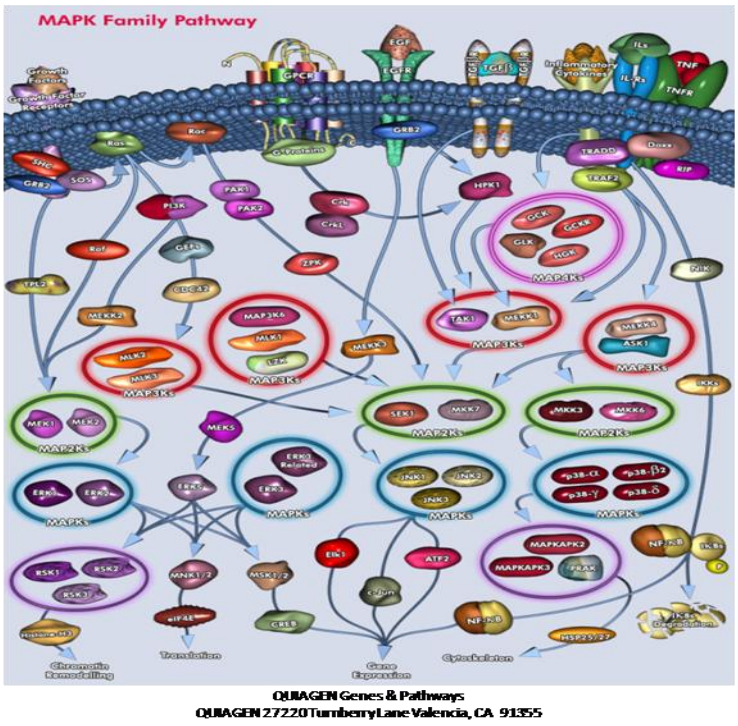



Figure 4.4 – Wiring diagram depiction of biochemical signaling network of MAPK family. MAPKs deliver extracellular signals from activated receptors to various cellular compartments, notably the nucleus, where they direct the execution of appropriate genetic programs, including activation of gene transcription, protein synthesis, cell cycle machinery, cell death, and differentiation. The colored circles identify the active subfamily or complex formation.

The network representation is more formal and even if exist different data format (GML, KGML etc.) which are often interconvertible. Network modeling can includes quantitative information on relations and allows to perform topological analyses, which points out the general properties

and structure of the system, generating new information about relations, signal flow an system resilience.
Networks can be inferred directly from data, easily [Cline et al., 2007] or sophisticatedly [Nacher et al., 2009], and modularity of this approach allows compositional modeling for complex networks that analyzed afford to captures general insight to describe the promising dynamical features of real complex systems.

Network models recently have become ubiquitous in *natural sciences* at different complexity scales resulting from size, time, and interaction features. One of the key features of natural networks is their ability to adapt to changing environments, maintaining an appropriate pattern of behavior. Network biology perspective captures the logical step beyond both single gene and pathway analysis, attempting to identify and reproduce a model of the complex multi-dimensional interactions of cells, organs and organisms. They in fact includes, considering to go through from macroscopic to microscopic scale, Phylogenetic trees, Ecosystem Networks, Food webs, Physiological networks, Cellular networks and Molecular networks (Figure.4.5)
Really, each of this broad networks group can be considered as an autonomous entity, which, as a fractal, can be further organized at different scales.

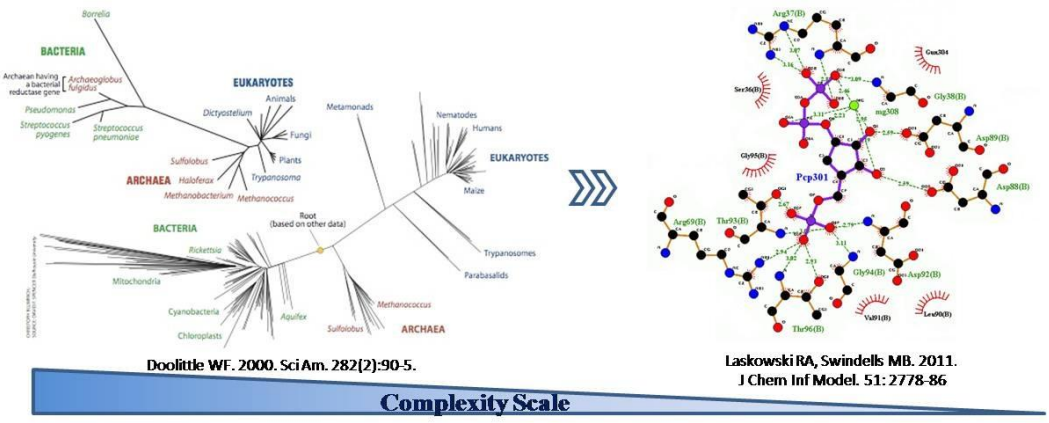



Figure 4.5 - The ability of Network Biology to investigate organizeing principles of biological systems at different levels of complexity.

Networks, therefore, offer a new means to characterize systems of very different origin in a single framework. Network analysis has revealed unexpected structural similarities between various complex systems, indicating that the graph representation is able to capture the general laws that govern them, in terms of organization and behavior. Defining the fundamental theories of different networks organizations with similar topological features, enables systems biology to reasonably combine the sparse details of submodules into a single framework, to achieve an overview of the process that is gainful both for the bottom-up and top-down modeling methods [Kestler et al., 2008].

A simplification of Biological systems consists of a large number of functionally assorted components, which interact both selectively and generally, often in non-linear relation, to produce highly well-structured behaviors. These components may be individual molecules (such as in signaling or metabolic networks); assemblies of interacting complexes; sets of physical factors that guide the development of an organism (genes, mRNA, associated proteins and protein complexes); cells in tissues, whole organs or physiology; and even entire organisms.
Due to the intrinsic organization of Biological system in network structure, the emerging field of Network Biology, based on the sound mathematical graph theory, seems to be the most appropriate methodology to represent, analyze, and interpret these complex systems.
Biological networks have naturally several concomitant features as recurring feedback loops, several bifurcations, parallel paths and oscillations, which, all together, can lead to interesting dense dynamical behavior.
Although there are several similarity among networks properties, two aspects, in particular, highlight biological networks character:
Firstly, in biological systems, copy numbers of many species are very low, which can give rise to significant fluctuations in behavioral patterns.

Secondly, biological systems have innate well-organized structural hierarchies, in accordance with the “building-blocks of life” concept [Raven et al., 2014].

The rising interest of researchers in this discipline is adequately represented by the trend of publication in this field during the last years; particularly, if we consider that it is generated only taking into account the composed word “network biology”.

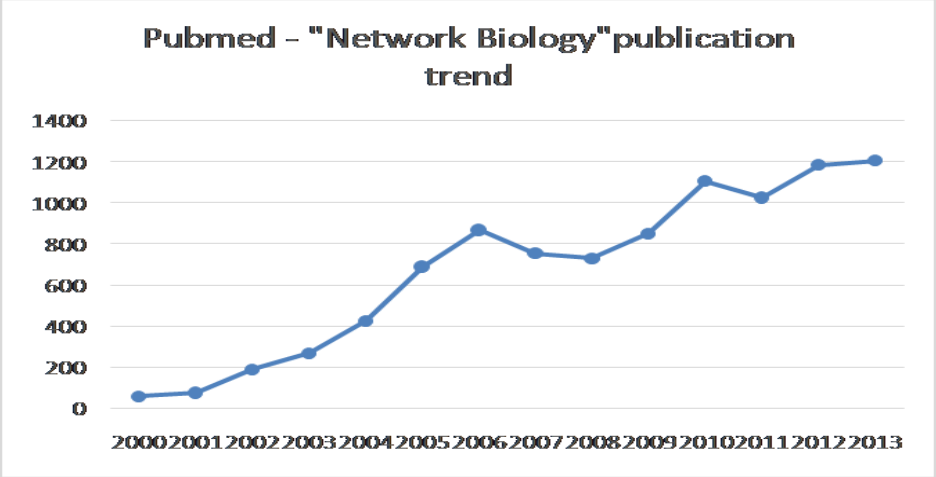



Figure 4.6 – Pubmed publications trend relative to scientific papers that contain in the titles, abstracts, and in the key words "Network Biology". The search was made by Pubmed advanced search engine.

Among the biological networks, those attracting the most attention of computational biologists in last decade have been *Biochemical networks*. These networks represent the molecular level patterns of interaction and mechanisms of control in the biological cell, or in entire organism. As we said, the development of these extremely intricate networks has been possible thanks to the combination of *"-omics"* sciences that has strongly contributed to establish the current Big Data scenario. The fusion, or better, the merging of these information into a single network give rises to what is generally defined as *interactomes* [Sanchez, 1999]. Indeed, the fundamental question of what determines the function of a cell is gradually shifting from the standardized focus on genes patterns

to the awareness that each cellular function is characterized by a cooperative network modules composed of genes, transcription factors, coding and non coding RNAs, enzymes, and metabolites differently interconnected each others.

## 4.2 Biochemical Networks

The types of networks widely investigated in this area are metabolic networks, signal transduction networks, genetic regulatory networks and protein-protein interaction networks, according to their general organization, usually formalized as Bipartite Networks, Bipartite or Directed Networks, Directed Networks and Undirected network, respectively.
This association between biochemical process and type of graph used to represent it, is not exclusive, but according to the descriptive requirements concerning to the goal of the investigation, as we shall see, can be modified.

### 4.2.1 Biochemical Network properties

The structural features of interaction graphs can be quantified by network measures that characterize both properties across the entire network, as extent of a graph or the density, and local features, as single nodes or edges information.

A general description of network features can ease the comparison among different networks, which results particularly useful in Network Biology. In this domain often happen to have to compare the characteristics of networks modified by changing conditions with background network (e.g. pre- and post-perturbation), or to have to compare similar pathways across species. Therefore, even a sketched measurement of graph properties can be valuable.
The minimal descriptive properties for a biological network can be summarized as follow:

The *node degree*, a local node property, consists in the number of edges incident to the node $i$, $k_i$

Denoted an undirected graph *G (n,m),* the degree $k_i$ is:

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{4.1}$$

In directed graphs, we can consider two values of the node degree: $k_i^{in}$ (for ingoing arches number); and $k_i^{out}$ (for outgoing arches number).

Then, *average degree* for the two kind of graph, a global property, can be defined as:

$$\langle k \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i \quad \langle k_i^{in} \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i^{in} \quad \langle k_i^{out} \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i^{out} \tag{4.2}$$

The *degree distribution* of a network is defined as:

Let *P(k)* be the percentage of nodes of degree $k$ in the network. The degree distribution is the distribution of *P(k)* over all $k$. *P(k)* can be understood as the probability that a node has degree $k$.

In a characteristic biological network, the degree distribution follows the *power-law distribution* [Clauset et al., 2009][Barabasi and Oltvai, 2004] given by the probability density function (pdf)

$$P(k) = Ak^{-\gamma}, \;\; 2 \leq \gamma < 3, k \geq k_{min} \tag{4.3}$$

where *A* and γ are constants. γ is often called the power law exponent, which in biochemical context usually assume values $2 \leq \gamma < 3$ ; in natural life networks (web, social, food, etc.) it is rare that it assumes values $\gamma < 1$. This distribution is known as, *heavy-tailed distribution* due to its shape in log-log scale representation (Figure 4.7).
Networks with power-law degree distributions are often called *scale-free* networks.

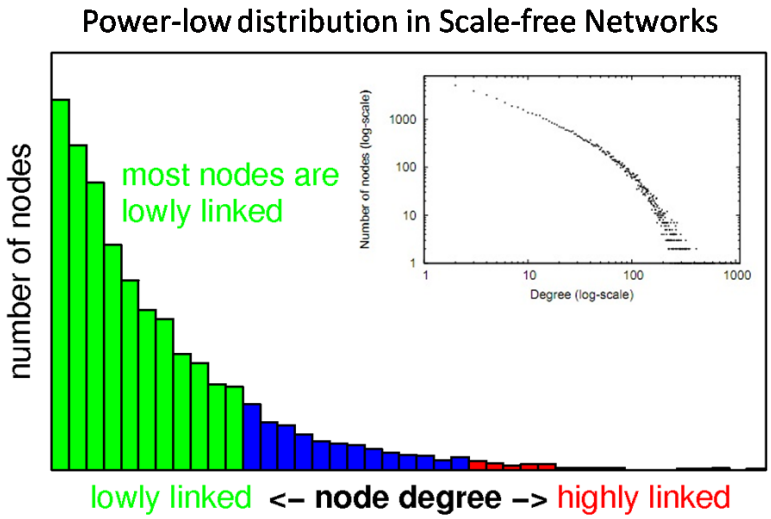



Figure 4.7 – Power-law Distribution diagrams. In the frame is reported the log-log scale of the colored bar diagram

As we said, Biological networks are scale-free (degree distribution approximates a power law, $P(k) \sim k^{-\gamma}$ ), that means that the majority of their nodes have low degree, but there is a small number of nodes with high degree (highly-linked nodes), called *hubs*. The value of γ determines many properties of the system. The main characteristic is that, smaller is the value of γ, more important is the role of the hubs is in the network. In general, this property of scale-free networks are valid only for γ<3, when the dispersion of the *P(k)* distribution, defined as $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, increases with the number of nodes, σ diverging, resulting in a series of unexpected features, such as a high degree of robustness against accidental node failures [Albert et al., 2002].

Distance in networks is measured with the *path length*, which define how many links are needed to move across the path that connect two vertices. Usually, there is not a unique path between two nodes, although the *shortest path*, a local property that consists of the smallest number of links between the two nodes, absolves a distinctive role in network connectivity.

In directed networks, the distance from node A to node B ($\ell_{AB}$) is often different from the distance $\ell_{BA}$ , which connect B to A.

The *average shortest path length*, $<\ell>$ is

$$<\ell> = \frac{\sum_{i,j}^{N} len\_sp\ (v_i, v_j)}{\sum_{i,j}^{N} len\_G\ (v_i, v_j)} \qquad (4.4)$$

where $\sum_{i,j}^{N} len\_sp\ (v_i, v_j)$ is the value of all-pairs shortest-path length in the graph, $\sum_{i,j}^{N} len\_G\ (v_i, v_j)$ is the sum of the all paths in the graph *G*, and *N* is the number of nodes of the network.

For connected undirected networks, $\sum_{i,j}^{N} len\_G\ (v_i, v_j)$ assumes the value $N \times (N-1)$, because exist at least an edge connecting every node pairs of the network.

The *average shortest path length* measure, provide an indication about networkwide navigability. Generally, the "life networks" show a relatively short path between any two nodes, and its length is assumable corresponding to the logarithm of the network size, defining the *Small World Properties* [Albert and Barabási, 2002].

Formally, networks are said to show the small-world effect if the value $\ell_{i,j}$ scales logarithmically or slower with network size for fixed mean degree.

Strictly linked to shortest path is the global measure of the *average network diameter* is the average of shortest path lengths over all pairs of nodes in a network.

Scale-free networks are inclined to form clusters, groups of nodes with a high number of reciprocal connections among them. It may entail that if vertex A is connected to vertex B, and vertex B is connected to vertex C, and then there is a heightened probability that vertex A is also connected to vertex C.
In terms of network topology, transitivity implies the presence of an amplified number of triangles in the network. These triangles belong to those structures generically defined *network motifs*, significant patterns of interconnections that identify over-represented partial subgraph in networks. The *clustering coefficient*, as local vertex feature, can be specified as follows [Duncan and Strogatz, 1998]:

$$C_i = \frac{3 \times number\ \ of\ network\ \ triangles\ \ connected\ \ to\ v_i}{number\ \ of\ connected\ \ triples\ of\ vertices\ \ i}$$

In particular, the *connected triples* are represented by a single vertex with edges running to an unordered pair of other vertices.
Mathematically it is expressed as

$$C_i = \frac{1}{k_i(k_i - 1)/2} \sum_{jk} v_{ij}\ v_{ik} v_{jk} \tag{4.5}$$

that characterizes the overall tendency of nodes to form clusters or groups
Therefore, the average of the local clustering coefficients of all the vertices can be denoted as

$$\langle C_i \rangle = \frac{1}{k_i(k_i - 1)/2} \sum_{jk} C_i \qquad (4.6)$$

Another measure derived from cluster coefficient is the *clustering spectrum, C(k)*, which is the distribution of the average clustering coefficients of all nodes of degree *k* in the network, over all *k*.

In directed graphs, the subset of nodes connected by paths in both forward and reverse directions shapes the so-called *strongly connected component,* or strongly connected *cluster* (i.e. strongly connected direct graph if defined by the existence of a path in each direction between each pair of vertices of the network). Due to the directed character of arches, it is appropriate to define the *in*-cluster (nodes that can reach the strongly connected cluster but that cannot be reached from it) and *out*-cluster (vice versa).
Nodes belonging to these different subsets, in biological networks, seem to perform different shared task. Regarding to signal transduction networks, for example, the nodes of the in-cluster tend to be involved in trigger signal binding (e.g ligand-receptor); the nodes of the strongly connected cluster compose the central signaling subnetwork assigned to signal amplification and diffusion ; and the nodes of the out-cluster finalize the trigger signal affecting phenotypic changes [Ma'ayan et al., 2005].

Networks' global connectivity originated by a succession of adjacent edges through nodes is characterized by the merge of these local and global properties.

### 4.2.2 Metabolic Networks

*Metabolism* is the chemical process by which sets of successive chemical reactions, mainly involving a sequence of enzymes reactions that convert initial inputs into useful end-products by a series of biochemical steps. The complete set of all reactions in all pathways forms a *Metabolic Network.*

Generically, an individual chemical reaction involves the transformation of one or more reactants that are cleaved, spliced, combined, or energetically modified (catabolism or anabolism) to produce one or more others molecules, often involved in the subsequent reaction. The metabolites modified during the reaction are called the *substrates*, while those produced are called the *products,* and the elements that operate the conversions are the *enzymes*.

Therefore, the most proper representation of a metabolic network is as a *directed bipartite network* (or direct *bigraph*) which connects the metabolites to the abstraction of transforming enzymes, the enzymatic reaction. It is even possible to represent direct links between metabolites, considering them as the two different kinds on nodes, implying the reaction. This assumption can be useful to trace a metabolite path more easily [Isermann et al., 2004].

Bipartite Networks, hence, consist of two distinct types of vertex, with edges running only between vertices of unlike sort.

The bigraph representation is not so common in Computational System Biology modelization; however, it allows performing, useful topological analysis.

In usual metabolic networks interpretations, vertices represent *metabolites,* usually circles, and *metabolic reactions*, often rectangles with edges joining each metabolite to the reactions in which it participates. (Another case of typical bipartite networks is the affiliation networks, which have two types of vertex corresponding to people and the groups they belonged to).

Moreover, a metabolic network is really a *directed bipartite network* [Jeong et al., 2000], since some metabolites go into the reaction (the substrates) and some come out of it (the products). By placing terminal arrows on the edges, transforming it in an arch we are able to discriminate between the ingoing and outgoing metabolites. An example is sketched in Figure 4.8 that shows a network representation of the up regulated metabolic submodule response of *S. cerevisiae* to glucose depletion stimulus. [Bryant et al., 2013].

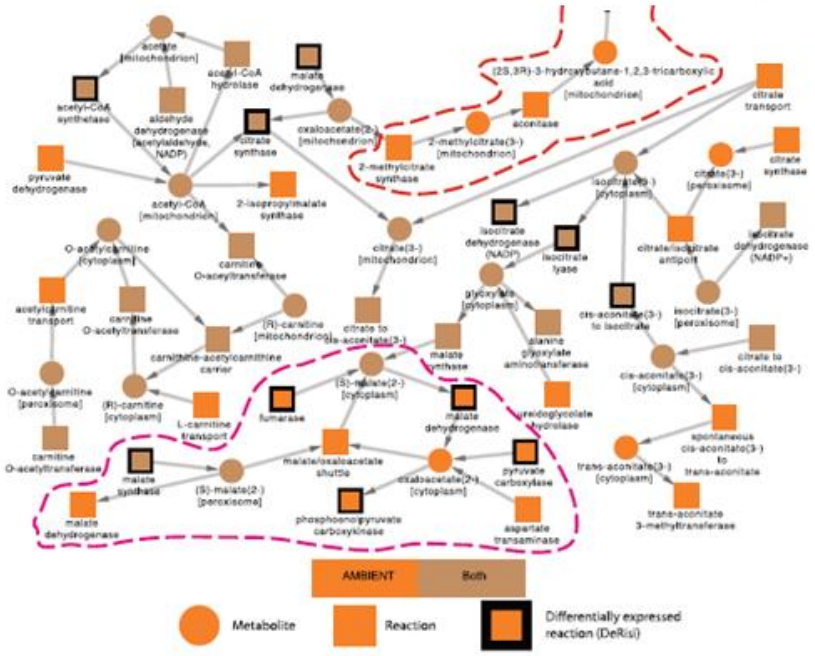


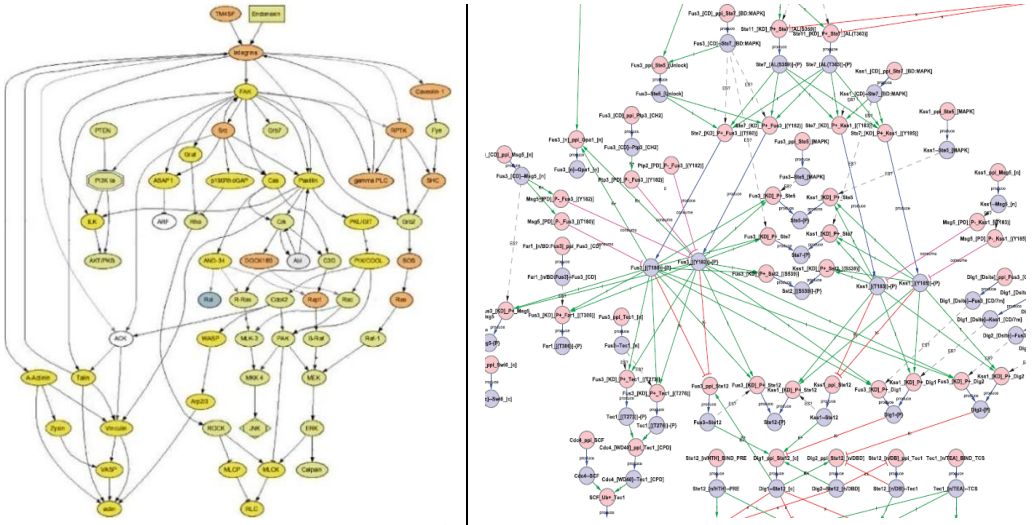

Figure 4.8 – Directed Bipartite Metabolic Network: section of up-regulated modules of diauxic shift network in Saccharomyces cerevisiae. The different colors arise from the overlay of two different methods applied in submodule identification.

### 4.2.3 Signaling networks

Life is a chemical phenomenon, and metabolism is the bio-chemical process that takes place in an organism to safeguard its vitality. If biological scientists almost universally accept this statement, it is appropriate to highlight the equally important function that covers the transduction of signals [Hlavacek et al., 2006], which are often blended

in metabolic networks [Frauwirth et al., 2002] or are able to modulate their activity [Rathmell et al., 2000].

Biochemical networks, in particular those containing signaling pathways, display a wide range of regulatory properties. These include the ability to spread information across different time scales and to function as switches and oscillators. Depending on their active signaling molecules, they can alter metabolism (metabolic signaling), change gene expression (transcription signaling), or induce cellular movements or structural changes [Gomperts, Kramer and Tatham, 1999].
The mechanisms underlying these multifaceted behaviors engage many interacting components or entire sub-pathways and, so, it is quite impossible to get an overview of the whole pattern even by actual *in vitro* advanced array alone.
Signal transduction networks involve both protein interactions and biochemical reactions for signal modulation, so, their abstraction should be accomplished by means of bipartite graphs, or at least, given the effort in managing bigraphs mentioned in previous section, their edges should always be directed, mimicking the direction of signal propagation.
An example of the two possible abstractions is viewable in the directed bigraph describing MAP kinase signaling pathway the inferred by Tiger and colleagues [Tiger et al., 2012]; the directed graph display the integrine signal network in [Martin et al., 2002]. In both networks is recognizable a central hierarchical structure, prevalent for cascade signals description as depicted in Figure 4.9.



a) b)

Figure 4.9 – Directed Signaling Networks: In a): Directed information flow graph of the Integrin Signaling Pathway collected across different species. In b): Portion of the regulatory graph that highlights the causality between reactions revealing the regulatory structure of the network. This bipartite graph illustrates the relationships between the reactions (red nodes) and states (blue nodes) within the network. Edges from reactions to states define how states are produced (green) or consumed (red), and each such edge corresponds to a single essential reaction.

These types of complex behavior from relatively simple networks highlight the necessity of using theoretical approaches in understanding higher order biological functions

The first approach for revealing the dynamics of signal transduction pathways is the derivation of models able to trace the principal path of information flow, starting from the putative input or output signals, across all networks. Particularly relevant is the identification of a main flow channel, often represented by the strongly connected cluster [Ma'ayan et al., 2005].

However, achieve a valuable abstraction of signal transduction network is quite difficult due to several cross-talks among different patterns,

which can originate feedback loops, and parallel developments of the signal flow paths, which can vary in their relevance according to cell state. This modulating or alternative paths are fundamental to provide the system robustness [Kitano, 2004][ Barkai et al., 2001][Gomperts, Kramer and Tatham, 1999]. Thus, data of signal transduction network tend to be heterogeneous and noisy such that the system will have a large degree of uncertainty with respect to both its structure and parameter values. Consequently, validate and refine signal transduction pathway models is crucial.

According to the ascertained large-scale organization of biological networks [Jeong et al., 2000] the concept of flow of information thought network has been intuitively extended from network science to the macroscopic level involving entire physiological systems [Csermely et al., 2013]. Among them, the circulatory system [Blinder et al., 2013] and the most complex natural system, the nervous system, about to which it is facing the challenge of shaping the neural signal transduction [Craddock et al., 2013][Nakamura et al., 2006], are actually widely investigated (https://www.humanbrainproject.eu/).

### 4.2.4 Bipartite Network

The bipartite network, also called a *two-mode network* in the sociology literature, an *bigraph* in graph theory, are characterized by two distinct type node (vertices) , one representing the original vertices and the other representing the groups to which they belong; and connecting edges running only among vertices of distinct nature.
A small example of a bipartite network is shown below in Figure 4.10

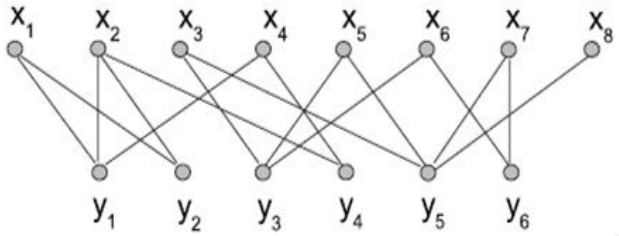



Figure 4.10 – Bipartite Graph: A bipartite network composed of two different type of vertices, the $x_{\#}$ and $y_{\#}$ nodes

Staring from formal definition for a generic graph (undirected):

G is an ordered triple *G=(V, E, f)*, where

- *V* is a set of vertices or nodes or points
- *E* is a set, containing elements known as edges or arches
- *f* is a function that maps each element of *E* to an unordered pair of vertices in *V*

A Bipartite graph can be defined as *G*

Whit the set of vertices *V* partitioned into 2 sets $V_1$ and $V_2$
such that $(u,v) \in E$ which implies either $u \in V_1$ and $v \in V_2$ OR $v \in V_1$ and $u \in V_2$
Let $e = (u,v)$ be an arc that is directed from $u$ to $v$, such that $e$ is *incident from* $u$ and is incident to $v$ ; $v$ is incident from $e$ and $u$ is incident to $e$.
A graph can also be represented by its *adjacency matrix*, or vertex matrix, which describes a graph by representing which vertices are adjacent to which other vertices. For unweighted graphs, it can assume 1 or 0 values in position$(v_i, v_j)$; to describe weighted graphs the values in the matrix assume the weight values in position$(v_i, v_j)$.
Let *G* be a graph with $n$ vertices numbered $0, \ldots . n-1$.

The Adjacency matrix of $G$ is the $n \times n$ matrix

$$A = (a_{ij}) 0 \leq i, j \leq n-1$$

with $a_{ij}$ specifies the number of edges from vertex $i$ to vertex $j$:

$$a_{ij} = \begin{cases} 1 \; if \; (v_i, v_j) \in E \\ 0 \quad \text{otherwise} \end{cases}$$

The equivalent of an adjacency matrix for a bipartite network is a rectangular matrix called an *incidence matrix.* If $n$ is the number of elements in the network and $g$ is the number of groups of elements then the incidence matrix $B$ is a $g \times n$ matrix

$$B = (b_{ij}) 0 \leq i \leq n-1; 0 \leq j \leq g-1$$

having elements $b_{ij}$ such that

$$b_{ij} = \begin{cases} 1 \;\; if \; v_i \in g_j \\ 0 \;\; \text{otherwise} \end{cases}$$

for instance the 8×6 incident matrix of the simple network in Figure 4.10

$$B = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

We can use the bipartite network to directly showing the relations among a particular set of nodes, producing the compression of the two-mode bipartite by the *one-mode projection network* [Latapya et al., 2008].

The one-mode projection onto *X* consists of a network containing only *X* nodes, in which two *X* nodes having at least one common neighboring *Y* node are defined connected. The Figure 4.11 shows the resulting subnetworks of *X* and *Y* one-mode projections, respectively.

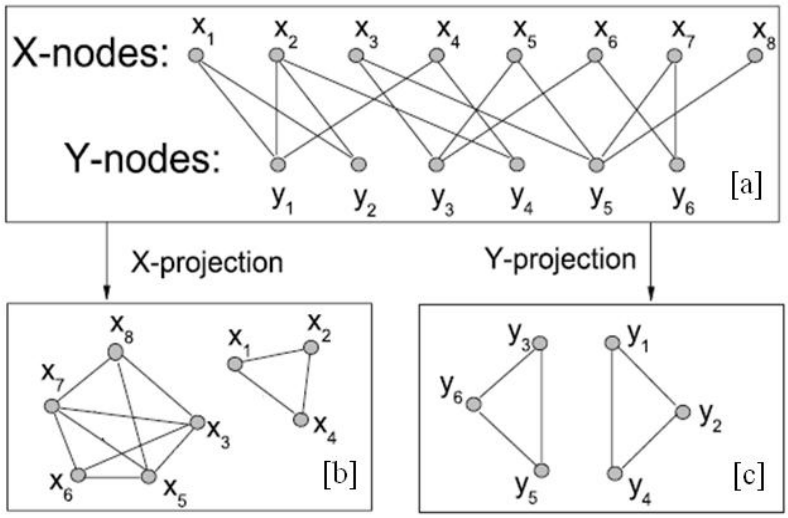



Figure 4.11 – Bipartite and Projected networks: Illustration of a bipartite network [a], as well as its the two possible one-mode projection, the X projection in [b] and the Y projection in [c].

When we form a projection, each group in the bipartite network corresponds to a network module composed of a subset of the vertices such that every vertices of the set is connected by an edge to with all the other, defining a *clique* [Roland, 1968][Cicerone and Di Stefano, 1999]. For instance, if a cluster contains five elements in the bipartite network, then each of those five vertices is connected to each of the others in the one-mode projection due to common membership in that group. The mentioned clique of five vertices is visible in the left side of section [b] in Figure 4.11.

Thus the projection is, generically, the union of a number of cliques, one for each group in the original bipartite network.

The one-mode projection is often useful and is widely employed, but its construction removes much information present in the fundamental bipartite network structure. Indeed, this complexity reduction method conduces to power lost regarding to data representation; for example the projection loses any information about how many groups composed of two vertices share in common; information that could be useful to evaluate a potential strong connection between the same two vertices.
We can capture information of this kind in our projection by making the *projection weighted* (Figure 4.12) giving each edge between two vertices in the projected network a weight equal to the number of common groups the vertices share.

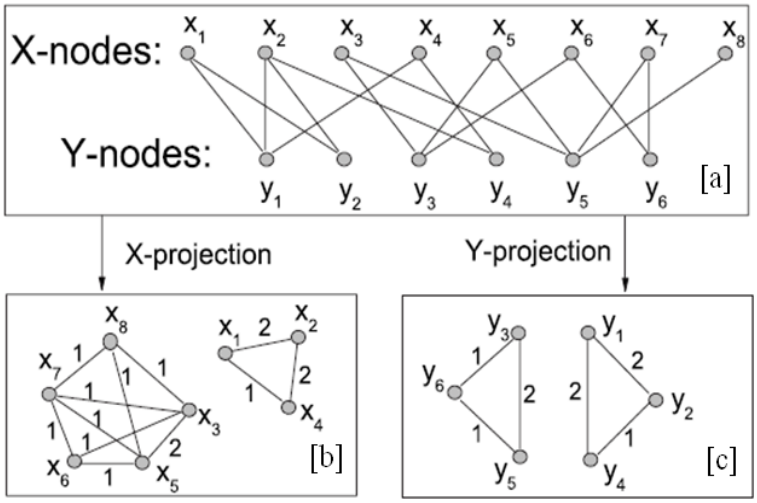



Figure 4.12 – Bipartite and weighted networks: Illustration of the weighted version [b] and [c] of the two possible one-mode projection originated from network in [a].

Unfortunately, even the weighted one-mode network does not capture all the information in the original digraph (it is not able to record the number of groups or the exact membership of each group for instance) but it is an improvement on the unweighted version and is quite widely used.

Mathematically the projection can be written in terms of the incidence matrix $B$ as follows.

The product $B_k \cdot B_{kj}$ will be 1 if and only if $i$ and $j$ both belong to the same group $k$ in the bipartite network. Thus, the total number $P_{ij}$ of groups to which both $i$ and $j$ belong is:

$$P_{ij} = \sum_{k=1}^{g} B_{ki} B_{kj} = \sum_{k=1}^{g} B_{ik}^{T} B_{kj} \tag{4.7}$$

in which $B_{ik}^{T}$ is an element of the transpose $B^T$ of $B$.

The $n \times n$ matrix $P = B^T B$ is similar to an adjacency matrix for the weighted one-mode projection onto the $n$ vertices. Its off-diagonal elements are equal to the weights in that network, the number of common groups shared by each vertex pair, $P$ is not quite an adjacency matrix, however, since its diagonal elements are non-zero, even though the network itself, by definition, has no self-edges. The diagonal elements have values

$$P_{ii} = \sum_{k=1}^{g} B_{ki}^{2} = \sum_{k=1}^{g} B_{ki} \tag{4.8}$$

By this equation, considering that $B_{ki}$ only takes the values 0 or 1, $P_{ii}$ results equal to the number of groups to which vertex $i$ belongs.

Thus to derive the adjacency matrix of the weighted one-mode projection, we would calculate the matrix $P = B^T B$ and set the diagonal elements equal to zero. Differently, to obtain the adjacency matrix of the unweighted projection, we have to take the adjacency matrix of the weighted version and replace every non-zero matrix element with value 1.

The other one-mode projection, onto the groups, can be represented by a $g \times g$ matrix $P' = BB^T$, whose the off-diagonal element $P_{ij}$ returns the number of common members of groups $i$ and $j$, as the diagonal element $P_{ii}$ provides the number of members of group $i$.

In natural life, as we said, is not rare to observe native bipartite networks that are directed; an example of such a graph is the metabolic networks showed in previous section 4.2.2

The bipartite representation of a metabolic network does not include often the physical modeling of the enzymes, which, though they are not both metabolites and the reactions that they perform, they are a relevant part of the metabolism. Even if can be demanding in large pathway, it is possible include the enzymes in the graph by introducing a third class of vertex connecting them to the reactions that they catalyze. The beta-alanine metabolism is an example of small metabolic network composed of three vertices.
Since enzymes are not consumed in usual reactions, these edges can be drawn as undirected, linking neither into nor out of the reactions that they conduct. An example of such a network is presented in Figure 4.13.
Technically this is indeed a *tripartite network,* partially directed and in part undirected.

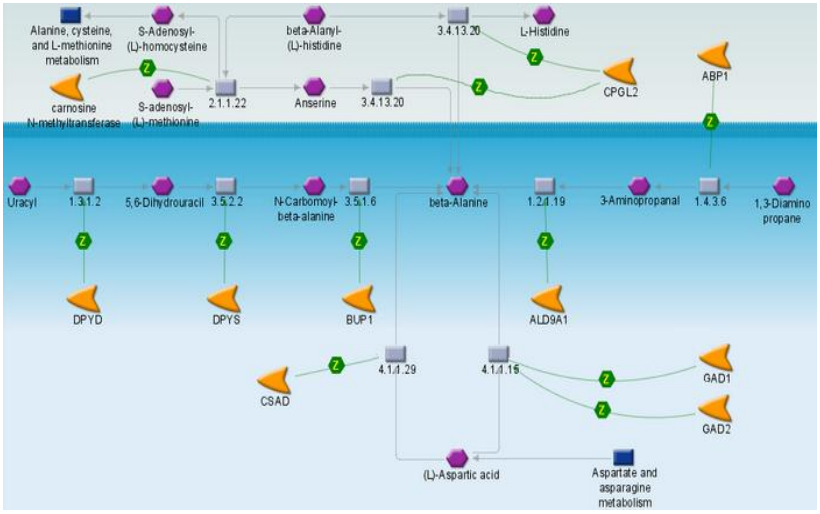



Figure 4.13 - Beta-alanine metabolic tripartite graph: metabolites are depicted as hexagon, reaction as gray squares and enzymes are similar to arrow end. In particular colored edges, due to the specific biological application, are labeled with the type of enzymatic reaction (often non standardized, but repository-specific: Z is catalysis) and green colored for the positive effect. [Merck Millipore Pathway tools]

Large bigraph representation is typically too complex to be designed or interpreted and biochemists are more inclined to deal with wiring diagrams, where only metabolites are represented as vertices with arcs denoting participation in the same reaction.

In network theoretic terms, this representation is equivalent to the projection onto the set of nodes. The most common representation of metabolic network projects the network onto the set of nodes, the metabolites.

Clearly, this projection loses much of the information contained in the full bipartite network, but, as we have said, it is nonetheless widely used. This most widespread alternative approach, contains more of the information from the full network, but is still somewhat unsatisfactory since a reaction with many substrates or many products appears as many edges, with no easy way to tell that these edges represent aspects of the same reaction.

The popularity of this representation arises from the fact that for many metabolic reactions only one product and one substrate are known or are considered relevant , and therefore the reaction can be represented

by only a single directed edge with no confusion arising. An example is displayed in figure below

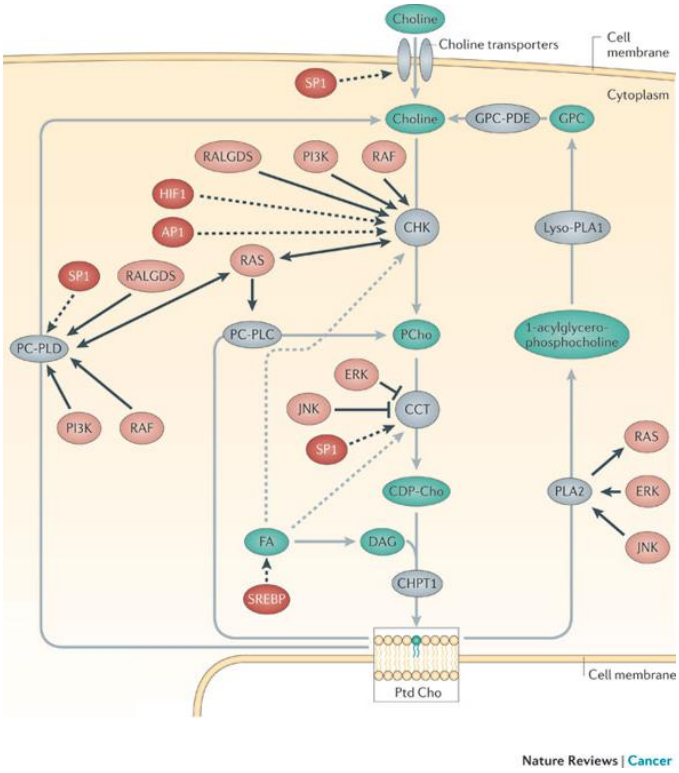



Figure 4.14 – Malignant Choline Metabolism: Grey arrows represent the choline metabolism pathway, proteins in grey catalyze the reaction that is depicted by the corresponding grey arrow on the metabolites in green Dashed grey arrows indicate the regulation of enzyme activity in the choline metabolism pathway. Black arrows indicate connections Choline metabolism in malignant transformation [Glunde et al., 2011].

The inhibition of a reaction is usually attained by substrate competition or removing an enzyme necessary for the reaction.

First method use *substrates competitive*-proteins, which are biochemically-engineered proteins that bind to an enzyme and saturate the active sites for the "natural" substrate, make it unable performing its normal function as a catalyst.

The other one can genetically alter the organism under study to silent or delete the DNA sequence dedicated to enzyme production (a so-called gene *knockout* experiment). The same techniques can also be used to

determine which reactions are catalyzed by which enzymes and to discover the structure of the third, enzymatic part of the tripartite metabolic network reported in Figure 4.13. The construction of a complete or partial metabolic network implies the combination of data from different pathways, certainly reached by different biochemical assay, and using many different techniques

Actually, several database deals with metabolic pathway data giving a useful starting point to build up the corresponding network diagram. Networks inference from this informal data is a non-trivial task because they often derive from many sources differently curated. The majority of these databases have missing step in their pathways. This is ascribable to different reasons: due to the challenging *in vitro* evaluation of the omitted part of the pattern (this deficiency is compensated with the support of computational techniques, mainly, amino acid affinity sequence and molecular docking); for the lack of attested evidence; or, sometimes, due to inadequate efficiency in updating or revising pathway data.

Often, so, we have to infer the missing data imputed by deduction based on biochemistry and genetics knowledge. Several computer tools or software plug-in have been developed to merge several pathway diagrams or to reconstruct networks from raw metabolic data. Among them, the well-designed also allow you to perform enrichment analysis on node or on edges [Cline et al., 2007]. Generally, the networks created by the support of domain's experts are more intuitive and accurate, although the computationally inferred are an evaluable initial frame to network building process.

Even though the projected network, which is the Wiring Diagram itself, is suitable for the biochemist whose aim is to reconstruct the network from experimental data, it is not appropriate for a system biology perspective. The wiring diagram representation also causes a loss of information concerning the dynamics features of the systems, consisting of the kinetic rate of enzyme, strictly related to reaction rate, concentrations of the reactants or metabolites. Despite these models

being useful to represent integrating information from differently curated pathway repository (sbml.org), they are not adequate for computational modeling. A vast list of biological pathway related and molecular interaction related resource is available in the most detailed portal for bioinformatics applications [Bader et al., 2006].

To gain an insight into biological phenomena is essential to integrate the biology whit computational techniques to generate plausible hypothesis about the functional roles of network elements. This considerably complex methodology needs the establishment of a comprehensive workbench able to range from "static" models of network topology to dynamical and stochastic simulations [Schaff et al., 1997] Ideker et al., 2001][Kitano, 2002][ Hucka et al.,2002].
This aspiring goal could be reached considering an accurate graph representation ascribable to the bipartite networks supported to a sound mathematical theory necessary to reproduce as accurately as possible the dynamics and behavior of the a motley system.

These characteristics are coupled efficiently in Petri Net mathematical modeling languages. In particular, Timed Petri Nets (TPN) have been proposed again as a building-block of the computational system biology workbench to realize the challenging biological modeling goal [Machado et al., 20011][Koch, Reisig and Schreiber, 2010] [Chaouiya, 2008, 2007][ Heiner et al., 2008][ Bos, 2008][Peleg et al., 2005][Hardy and Robillard, 2004] after the farsighted publications of Reddy and coworkers [Reddy et al., 1993] more than twenty years ago.
Among the considerable literary production produced by the supporters of the so-called BioPN, not missing,of course, a good portion dealing witht the mapping of signaling pathways onto Petri Net [Miwa et al., 2010][ Breitling et al., 2008][ Sackmann et al., 2006]][Li C. et al., 2005 - 2007].
The figure below (figure 4.15) provides a fine example of how the Petri Nets, in particular employing their numerically solving extensions (Timed and Functional PNs), are one of the most proper tools to the

simultaneous modeling of static and dynamic properties of a biological system.

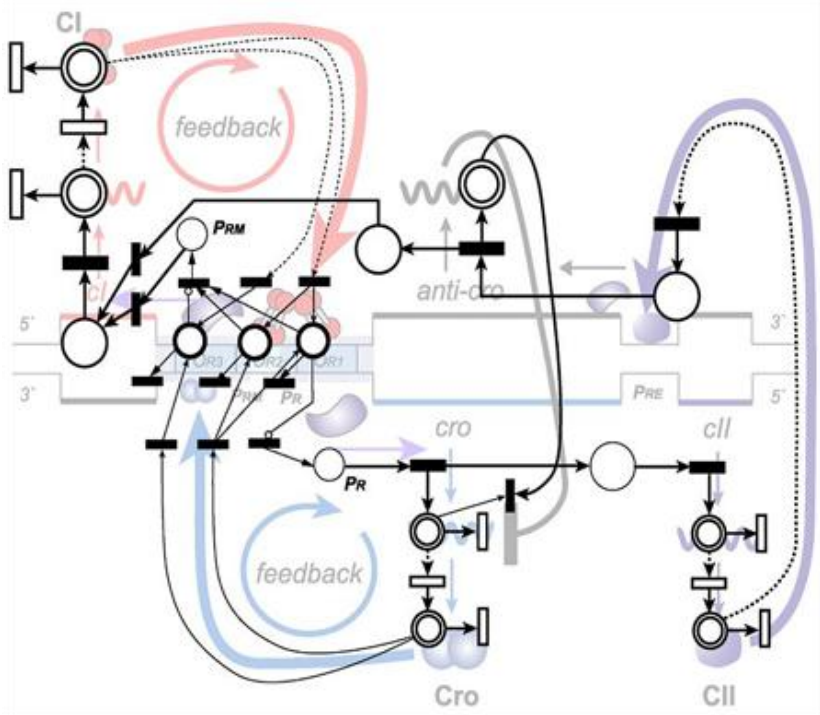



Figure 4.15 – A Bio Petri Net:The section of the switching mechanism of lambda phage and its hybrid functional Petri net (HFPN) reflection. Readapted from [Matsuno et al., 2000]

### 4.2.5 Protein-Protein Interaction (PPIs) Networks

The wide scenario of protein–protein associations have been considered with several research perspectives. *Protein-Protein Interaction Networks* (PPIs) were initially exploited to reveal the general organization of the directly linked proteins involved in a biochemical process (by means of affinity binding experiments) often integrated whit inferred interaction data resulted from computational analyses based on protein sequences and structures affinity. Their major goal consisted in pointing out the network submodules structures included within them, without particular regard to their functional properties.

Usually these submodules disclosed protein interaction motifs, characteristic of protein complexes, which performs their function after assembling.
Afterward PPIs networks were employed to graphically characterize the huge protein-coexpression clusters obtained from emerging high-throughput wet-lab experiments (transcription assay), which make possible the measurement of expression levels (mRNA quantification) for thousands of genes. The resulting *Coexpression Networks*, even without directly interacting proteins, can be represented as PPI networks, with edges underlying the relation revealed by biochemical experiment [Lee and Tzou, 2009] .
Moreover, a PPI network can mapped onto all kind of biochemical pathway, assuming only the mutual aspecific interactions among nodes, by the one-node projection network described in sec. 4.2.4

Due to the challenging identification of such a wide interacting panorama, many different computational techniques have been applied to investigate protein–protein functional linkages and interactions. These methods range from the identification of a single pair of interacting proteins, to the recognition and analysis of a multimeric complex with a wide-ranging function (e.g. the transcriptome), to go through the evaluation of large network of thousands of proteins, until as large as an entire organism, defining a *proteome* (the all interacting protein in the organism). A portion of the protein-protein interactome network of *C. elegans* is shown in Figure 4.16.
A computational method to be valuable in Network Biology should have as scalability as possible to manage such a large gamma of interacting data.

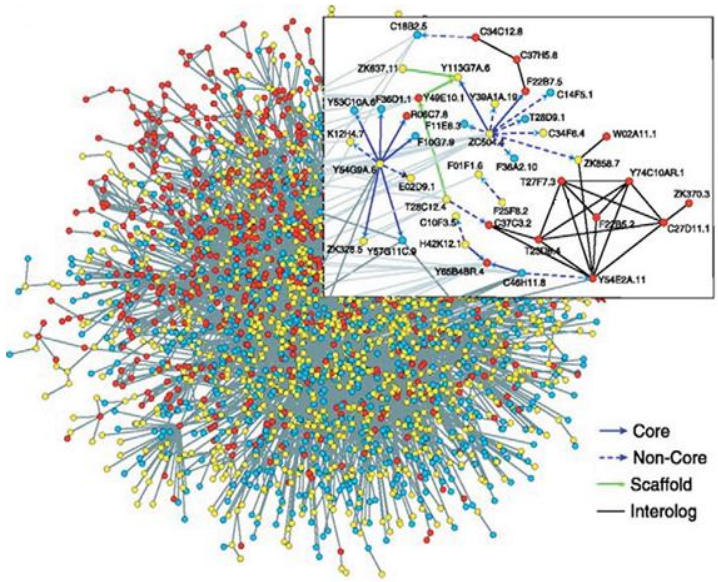


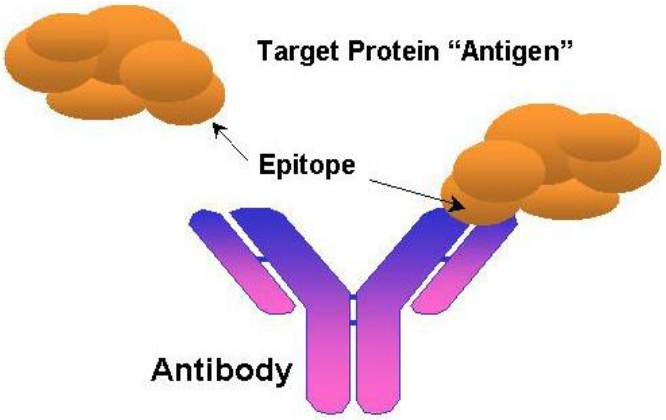

Figure 4.16- C. elegans protein interaction network. The nodes are colored according to their phylogenic class: ancient, red; multicellular, yellow; and worm, blue. The upper framemake a zoom of a small part of the network. Figure modifyed from [Li et al., 2004].

The set of all interacting protein are the vertices of the network connected by an undirected edge. Although this representation of the network is the one commonly used, probably due to the current relative ease of obtaining the data to build (actually exist more than 100 data base dealing with protein-protein interaction, see Pathguide.org for redirecting),this graph formalization omits much useful information about the real interactions. Interactions that involve three or more proteins, for instance, are represented by multiple edges, and there is no way to infer from the network which edges represent aspects of the same interaction. Nonetheless, they, appropriately analyzed, are able to provide a good degree of understanding of the system, in particular about very large system networks. [Behrends at al., 2010][Uetz et al., 2006][Vazquez et al., 2003].

This problem can be escaped by bipartite graph modelization, as we have ascertained to be effective in the design of metabolic and signaling networks. We therefore can define, even in this case, two kinds of

vertices standing for protein and interaction that are mutually linked by undirected edge. Such representations, however, are rarely employed due to the complexity of the design process that can distort the nature of these graphs that aim at a direct representation of the experimental data obtained from biochemical explorations.
As we mentioned Coexpression Networks are mainly represented as PPIs, almost exclusively by means of directed graphs with one kind of vertex that is the traditional representation assigned to PPI. The only gain of information expressiveness of such networks is the presence of a weight on the arch, generally assigned by using the statistical values obtained from the assay of coexpression micro-array [Jeffery et al. 2006.], which permit a more in-depth analysis of relation among network elements. It must be remembered, however, that some topological analysis typical of PPI, such as the search for high-interaction motifs, macro-complexes, should be interpreted with due care.

Another peculiar type of PPI networks is provided by the immune system, which, similarly to the receptor binding protein, is an example of characteristic physical interaction between proteins, which should be the basic concept of the PPIs modeling.
The simplified summarization of the main Immune System activity, strictly connected with circulatory systems, can be the immune response. It consist of the cells (B cell) or free proteins (Immunoglobulins) of the immune system functionally comparable the through their hypervariable regions (V region), which is able to specifically recognize and bind a restricted set off proteins, which contain the amino acid sequence with high affinity for the V region itself (epitope).




4.17 – Immune antibody: Depiction of the immune binding of an antibody (IgG) or one of its portions (IgM) with the epitope on the target protein usually exposed on exogenous microorganisms surface.

Important properties of this system, including memory, are then properties of the network of cells as a whole, rather than of the individual cells. This was a revolutionary paradigm for immunology, and it is actually challenging the investigation aimed to understand the immune system in these terms [Sun et al., 2014].

An open issue, in particular for clinical pediatrics is the inheritance of the immune system [Levy and Netea, 2014][ Quon et al., 2013], aligned with the wide employment of PPIs which makes the research of developmental biology [Jin et al,. 2013][Sharan et al. 2005]. The understanding of the immune deficiencies phenomenon resulting from hereditary and congenital defects is well-founded in the literature. It act by a inherited genetic alterations that can affect all major aspects of the immune system, often making it unable to act the immune response [Janeway et al., 2001]

### 4.2.6 Genetic Regulatory Networks

Over to protein-protein interaction, the most recurrent event among the biochemical process, the other fundamental example of biochemical physical interaction is provided by gene-protein binding. Conversely to the protein-protein associations, which are exploit thank to specific

binding sites with high degree of specificity (the maximal specificity is showed by enzyme-substrate and immune recognition), the DNA-protein binding and are often not so specific. A transcription factor can bind different genes in conserved regions; unlikely, most of transcription modulators (enhancers, silencers or insulators) can attach themselves to different DNA sequences directly or indirectly, mediated by a cofactor [Maston et al., 2006][West et al.,2002]. Thus, this kind of networks often identifies cause and effect relationships without specific knowledge of whether the interactions are direct.

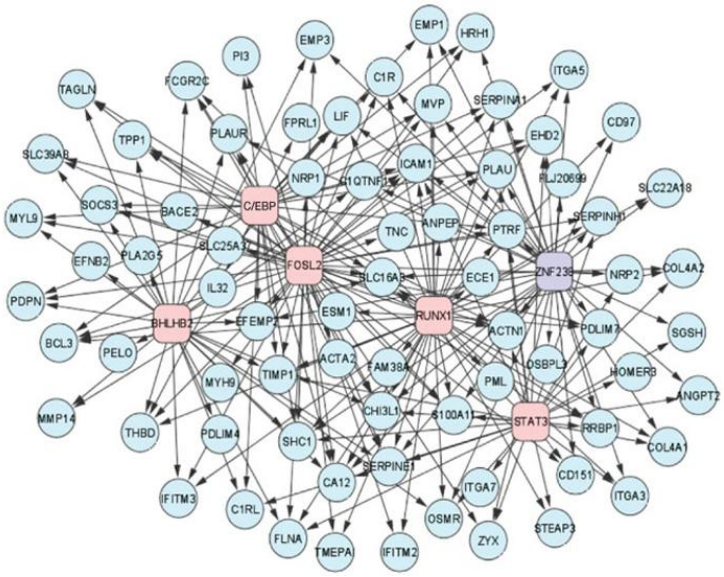
The cell does not, in general, transcripts constitutively every coding regions of its DNA originating all proteins. Individual proteins serve specific purposes, and in particular, it is important for the cell to be able to respond to environmental changing or specific stimuli by turning on or off the production of defined proteins as required. It does this by the use the mentioned transcription factors and modulators.

The interesting point is that transcription factors are themselves produced by transcription from genes.

The complete set of such interactions origins a *genetic regulatory network.* The vertices in this network are proteins or equivalently the genes that code for them and a directed edge from gene A to gene B indicates that A regulates the expression of B. This *directed unipartite* networks are the easiest way of describing the relationships in a genetic translation control system. They involve the assumption of a hypothetical direct gene-gene interaction, which does not occur in biomolecular processes. Many large genetic regulatory network give arise from the overlay of protein-DNA interaction obtained from affinity experiments and coexpression networks obtained from the same *in vitro* culture. Than the ensemble of data concerning the over-translated or low transcribed proteins of the coding gene to the gene-specific binding factor let suppose the regulative function of the DNA binding factor itself [Madhamshettiwar et al, 2012] [Bansal et al.,2007].

An explanatory example of the more plausible graph representation of a genetic regulation pattern is given in figure 4.18, by the *bipartite*

*directed* graph of the transcriptional network for mesenchymal transformation of brain tumors.




4.18 - The mesenchymal signature of high-grade gliomas is controlled by six transcription factors. TFs involved in activation of MGES targets are shown in pink, those involved in repression are in purple. By permission of Columbia University Department of Systems Biology-Irving Cancer Research Center. New York. U.S. 2014

### 4.2.7 Disease Networks

A sound application for assembled knowledge obtained by mapping and analyzing biochemical physiological networks is certainly the application in the clinical domains. The broadly investigated disease network is related to protein networks to point out hidden genes/proteins or new mechanism of action of known genes/proteins involved in diseases onset; or, moreover, to investigate which pattern is more fitting to be targeted by a pharmacological intervention. These networks are mainly generated by overlaying protein interaction networks with networks resulted from transcriptional experiments aimed to identify the over- or under-transcriptional pattern.

The main methods to investigate these graph is based on *topological modules* and *functional modules* identification.

The topological modules are described by locally dense neighborhood in a network,

Such modules can be identified using network clustering algorithms that are blind to the function of individual nodes [Palla et al., 2005][Ravasz et al.,2002]. By contrast, a functional modules consist of the aggregation of nodes of similar or related function in the same network neighbourhood, investigated by different clustering technique as in as in [Ulitsky and Shamir, 2007].

Obviously, unlikely as a pathological condition arises from a single cause (except for some genetic diseases), not only the protein-protein interactions are appropriate to be examined in disease pattern. In particular, there are strong evidences of the interconnection of more complex pathway, mutually interconnected, in the disease onset metabolic disorder with disease predisposition or development [Finkelstein et al., 2012].

The relevance of Signaling networks is well emphasized by the continuous literature production on searching relationships between these pathways and diseases and possible therapeutic approaches suggested by the resulting analysis of these models. Indeed, the identification of pathological-inducing features in the topology of the signal transduction network may enable the improvement of treatment options for diseases arisen from alteration in the related transduction path [Wang et al., 2014] [Lee and Ozcan, 2014] [Laplante and Sabatini, 2012] or in cross-talk between a metabolic and signaling pathway [Masson and Ratcliffe, 2014] [Lee and Ozcan, 2014].

The clinical intervention may consist in restoring altered stretch of the path with molecules (mainly synthetic) that mimic the correct transduction mechanism; or in the modulation of the signal transduction in case of imbalances in its functionality. The coupled relation of metabolic and signaling network imbalance with pathological conditions, is proved even in the macroscopic scale, concerning

cardiovascular disorders and coagulation pathway [Williams et al. 2014] [Samad and Ruf, 2013].

Whereas several recent studies have applied graph theory design to evaluate relationship between the genes products implicated in a selected disorder to gain a better understanding of genes related disease profiling; new approaches on large and comprehensive pathological pattern definition are emerging. These system biologist researchers consider, as the new bioclinical perspective recommends [Rees, 2002][ Gibson, 2009], the disease as a complex multifactorial disorder to be analyzed in a wider concept. Ki Goh and coworkers. [Gho et al, 2007] described firstly the "*The human disease network",* comprehensively exploring the relationships between human genetic disorders and the corresponding disease genes, as well as between disease genes themselves, from a macroscopic organism level of organization. More recently, Pinto and many other colleges from different field of research [Pinto et al., 2010], reconsidered the visionary concept of Gho, adapting the comprehensive approach to complex diseases of lesser scale, to achieve a higher level of network details in a multi-causal disorder currently not completely clear, the autism spectrum disorders (ASD). The latest work on disease bipartite network analysis [Darabos et al., 2014 ] underline as the graph theory is able to give a deeper understanding about genes-phenotypes relations in a complex panorama (Figure 4.19).

Hence, the main Disease networks authors combined the human "disease phenome", the systemic phenotype of genetic disorders linked to a pathological condition, with the "disease genome", which specify the complete set of genes linked to diseases, resulting in a global view of the "*diseasome*", the combined set of all known associations between disorders and disease genes.

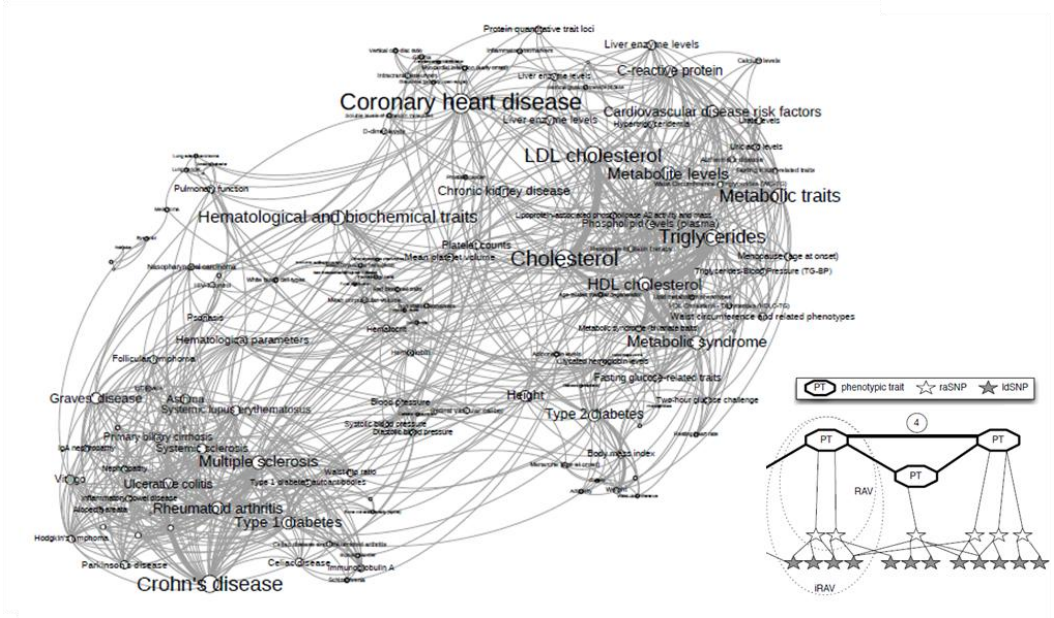



Figure 4.19 – Imputed Genetic Variations Network based Human Phenotype Network (iRAV-HPN). The nodes and labels sizes are proportional to the degree of the phenotype and the edge width is proportional to the number of overlapping iRAVs. In the lower frame, the schematized details of the bipartite network elements [Pac Symp Biocomput. 2014:188-99].

## 4.3 Petri Net Modeling for Stochastic Representation

Apart from the need for a formal graphical representation, which is essential for large networks, it is equally relevant detailed investigation of the quantitative aspect of biological system, usually performed on models of reduced dimensions compared to which mapped on networks. Quantitative models can be used to numerically compute both the biological functions and behavior of networks. Although several modeling languages have attempted to encode quantitative information, visualizing pathways is not always considered fundamental issue in the modeling language. The resulting models and their simulation process, thus, have been considered as a black box, in particular by non-mathematicians.

In accordance with the need following a multidisciplinary perspective suggested by Computational System Biology is evaluable to include

standardize graphical notations in biological pathways quantitative modeling [Dandekar et al., 2014[Kohn et al., 2006], in order to facilitate the compositional modeling, the exploration, the validation and the analysis, and complex pattern.
Typically, quantitative and qualitative models are handled as separate tasks. The Petri Net modeling language fulfills this ensemble requirement.

### 4.3.1 Petri Net Panorama

The most general description of the *Petri Net* (PN) is ascribable to a graphical and mathematical instrument to model discrete event systems. It consists of two types of nodes, places (*circles*) and transitions (*squares*), and direct arcs (*arrows*) that connect them.
Places are passive nodes refer to conditions or states (e.g. molecules, proteins, resources in general) that may contain zero or more tokens (*dots*). Tokens are the variables elements of a Petri Net represented by dots or numbers within a place; they indicate the discrete value of a state of the system (e.g. number of molecules, concentration levels or discrete number of any system components).
Transitions are active nodes that describe state changes, system evolution and, activities in a network (e.g. biochemical reactions, molecular interactions, or any displacement of components); transitions consume tokens from its *pre*-places and produce tokens within their *post*-places according to the arc features.
Directed arcs are inactive elements identifying causal relationships between transitions and places (e.g. identify reactants and products of a reaction). Ingoing arcs connect places (pre-) with transitions, while outgoing arcs arise from transitions to reach places (post-). Arcs may are labeled with a non-negative positive number greater than 1, indicating that they are weighed, otherwise, their weight is implied to be 1. Arcs weight, in biochemical context, represents the stoichiometry of a reaction.

The current discrete state of the Petri Net (*the marking*) is given by the number of tokens in each place, the starting state of the net represent the initial configuration of the model (*the initial marking*), usually indicated as $M_0$.
The development of the Petri Net formalism has been firstly attributed to Carl Adam Petri in 1962 in his dissertation [Petri, 1962], although Petri CA claimed to have originally invented them in 1939, at the age of 13, for the purpose of describing chemical processes [Rölke and Heitmann, 2010]. It was the first time a general theory for discrete parallel systems was formulated. The language is a generalization of automata theory such that the concept of concurrently occurring events can be expressed [Yen, 2006].

*Place/transition nets* (P/T-nets), introduced in the early eighties [Genrich and Lautenbach, 1981] soon became the generally most used version, since they permit multiple tokens per place, making this formalism more suitable for a better representation of real-world systems.
This mathematical formalism, simplified in its general definition given at the beginning, shows notable practical features for modeling:

- Natural expression of causal dependencies, conflicts, and concurrency.
- Plain and powerful synchronization mechanism, simplifying the description of mutual exclusion constraints.
- Local representation of states and actions which permits the hierarchical and building-block development of large net models.

The qualitative description can be therefore enhanced by the abstract representation of discrete quantities of species, achieved in Petri nets by the use of tokens at places. Tokens may represent the discrete number

of elements, or continuous values. A particular arrangement of tokens over a network is called, as we said, a marking.

A Petri Net $N$ is formally defined by the quadruple $N = (P,T,f,m_0)$, where:

$P$ and $T$ are finite, non empty, and disjoint set of nodes.
$P$ is the set of the places and $T$ is the set of the transitions.

$f : ((P\times T) \cup (T\times P)) \rightarrow \mathbb{N}_0$ is the set of directed arcs, weighted by non-negative integer values, which define the flow relation of the net

$M_0 : P \rightarrow \mathbb{N}_0$ is the initial marking, i.e. the non negative number of tokens associated to each place. Tokens in graphical representation consist of black dots inside places and represent the numerable elements that compose the systems.

Moreover, might be appropriate to define some additional notations:
$M(p)$ refers to the number of tokens on place $p$ in the marking $M$, representing a state of the system different from the initial sate $M_0$.
The overall markings reachable by means of a sequence of firing transition from the initial marking forms the *state space* of the model

According to the directed nature of graph, it is also useful to define the pre-set and post-set of a node $x \in P \cup T$ as:

Pre-set: ${}^\bullet x := \{y \in P \cup T \mid f(y,x) \neq 0\}$
Post-set: $x^\bullet := \{y \in P \cup T \mid f(x,y) \neq 0\}$

then also, the bipartite nature of Petri Net let us to define for places and transitions, four types of sets:

${}^\bullet t$ pre-places of transition $t$ (reaction's precursor)
$t^\bullet$post-placesof transition $t$ (reaction's products)

${}^{\bullet}p$ pre-transitions of place $p$ (all producing reactions of a component)
$p^{\bullet}$ post-transitions of place $p$ (all consuming reactions of a component)

The execution of a marked Petri Net (*Transition Firing* ) is able to represent the dynamic behavior of the modeled system by moving the tokens across places (from *pre*-place to *post*-place), which may reflect the occurrence of events or execution of operations, underling the evolution of the system.
A Petri net executes by *firing* transitions according to the enabling rule and firing rule of a transition, which govern the flow of tokens:

*Enabling Rule*: A transition $t \in T$ can occurs, is *enabled*, in marking $M$ if each input place $p$ contains at least the number of tokens equal to the weight of the directed arc connecting $p$ to $t$.

$$\forall p \in {}^{\bullet}t: M(p) \geq f(p,t)$$

else it is disabled

*Firing Rule*: Only enabled transition can fire. The firing of an enabled transition $t$ removes from each input place $p$ the number of tokens equal to the weight of the directed arc connecting $p$ to $t$. It also deposits in each output place $p$ the number of tokens equal to the weight of the directed arc connecting $t$ to $p$.
Mathematically, firing $t$ at $M$ yields a new marking $M'$

$$\forall p \in P: M'(p) = M(p) - f(p,t) + f(t,p)$$

Notice that since only enabled transitions can fire, the number of tokens in each place always remains non-negative when a transition is fired. The firing action does not consume any time.

All the information about of the structure of the net, $N=(P,T,f,m_0)$, can be described alternatively, with the *incidence matrix*, which can be defined as the $|P| \times |T|$ rectangular matrix $C$ in which the entry $(p,t)$ express the effect of the transition $t$ on the place $p$ in marking independent way.

This concept is formally defined as:

$C: P \times T \to \mathbb{Z}$ indexed by $P$ and $T$, where $C(p,t) = f(t,p) - f(p,t)$

This can be rewritten, according to the Pre and Post definition of Petri Net semantic, as $C(p,t) = Post(p,t) - Pre(p,t)$

### 4.3.2 Petri Net Properties

The Petri net resulting from process abstraction allows us to characterize its static system properties, which, so, can be evaluated without modeling any dynamic feature (e.g. kinetics parameters) and, therefore, without considering the time-dependent behavior.

Some topological properties, revealed by network structural analysis, is important for the consistency of the model because prove the observance of fundamental rules for a biological system.

However, the properties of a Petri Net can be used to validate the model and to check relevant system features. Each structural property must be interpreted to be converted in the corresponding significant biological meaning.

Therefore, topological properties depend only on the arrangement of places, transitions and arcs. They characterize the network structure and are independent of the marking and transition rate parameters, which, consequently, mean that they are independent of time-related system dynamics.

The main qualitative properties of a Petri Net is ascribable to:

*Boundedness*: For every holding place, the maximum number of tokens on this place is bounded by a constant amount. This property guarantees that, whatever the initial marking and the evolution of the

net, the number of tokens in each place is bounded, limited. As a concrete biological example in metabolic cycle, boundedness means that a product cannot be overstocked.
As regards to places, place *p* is *k-bounded* if exists a positive integer number *k* consisting in the maximum amount of tokens contained in the place *p* for every reachable markings of the Petri Net. Then A Petri net is k-bounded if all its places are k-bounded.
*Liveness*: For every holding transition, it will always possible to reach a state where this transition is enabled. In a *live* PN, all transitions can occur perpetually, reproducing the net behavior infinite times, without reaching any *dead states* (dead states are state of the net in which the marking permit none transition enabled). In biological systems it corresponds to a reactions cycle that characterize the maintenance of biochemical process.

*Reversibility*: For every holding state, the net will be always able to reach the initial state again. In biology, it means that the initial conditions of a closed system can be reproduce by any possible sequence of reactions belonging to i.
Belonging to structural features are, instead, the *invariants*
In mathematics, an invariant of a system is a predicate, which is not changed by the involved processes in the system. If the predicate is true at the initial state before the start of a sequence of processes, then it must be true during all processes phases, including the end.
In Petri Net notation, invariants indicate states in the net graph that does not change as a result of an event or a sequence of events. We can distinguish two type of invariants, place invariants and transition invariants.

*P-invariant* (place invariant): it is a set of places over which the weighted sum of tokens is constant and independent of any occurring transition. P-invariant is characterized by the conservation of tokens number. Then, each place belonging to a P-invariant is bounded. The biological

meaning of a P-invariant is to assure mass conservation and avoid an infinite increase of molecules in the model.

Mathematically, a vector of places is called P-invariant if it is a non trivial non-negative integer solution of the linear equation system

$x^T \cdot C = 0$ ,

where $C$ is the incidence matrix $m \times n$ whose elements $c_{ij}$ denote the change in the number of tokens in place $j$ due to the firing of transition $i$ of the bipartite graph, and x is a $m \times 1$ vector ≥ 0

*T-invariant*: it is a sequence of transitions able to lead back the system to the initial sate of the occurrence sequence, getting enabled all the transition belonging to the T-invariant itself. The biological meaning of T-invariants is to allow to the system to reinitialize a given initial condition. Firing the transitions of a T-invariant leads to a steady state behavior.

Mathematically, a vector of transition is called T-invariant if it is a non trivial non-negative integer solution of the linear equation system

$C \cdot y = 0, \quad y \geq 0$

with y is a $n \times 1$ vector ≥ 0

As we noticed in network section, the topological analysis of networks the qualitative properties and the structural motifs earlier defined, a qualitative analysis of a Petri net representation of a biological pathway would entail comparison among different Petri net models.

In Petri Net formalism, as we have sketched in sec 4.3, the most abstract representation of a network consists of its topology (the bipartite graph) and its initial state (the initial marking) by using the mentioned place-transition Petri Net (P/T-nets, or qualitative Petri Nets, QPN).

The need for including timing variables in the models of various types of dynamic systems is noticeable since these systems are real time in nature. In the real world, almost every event is time-related.

When a Petri Net contains a non predefined time variable, it becomes a *Timed Petri Net* (TPN)[Wang, 1998].

The topological structure of a timed Petri net generally takes the form that is used in a conventional PN, but differentiates itself in the firing rules that are defined depending on the way the Petri net is labeled with time variables. The firing rules defined for a TPN control the process of token flow.

The time implementation lead to several different types of timed Petri nets. Among them, the two most widely used extended Petri Nets are the Continuous Timed Petri Nets and Stochastic Timed Petri Nets [Wang, 1998][Molloy,. 1982]. [Goss and Peccoud ,1998] in which different type of time functions are associated with transitions.

The two variants of Petri Net included with time information on the firing rule can be summed up as follow

A Continuous Petri Net is a quintuple is the marked untimed PN underlying the SPN *(P, T, f, υ,$m_0$)*, where:

– *P, T* and *f* are defined as before.

– $\upsilon : T \rightarrow H$ is a function which assigns a firing rate function $h_t$ to each transition t; $H := \{h_t | h_t : \mathbb{R}^{|{}^\bullet t|} \rightarrow \mathbb{R}^+, t \in T\}$ is the set of all firing rate functions, and $\upsilon(t) = h_t$ for all transitions $t \in T$

– $m_0 : P \rightarrow \mathbb{R}^+{}_0$ gives the initial marking.

The semantic evolution of a continuous Petri Net is defined by a system of ODEs, whereby one equation describes the continuous change over time on the token value of a given place, influenced by transitions that increase or decrease the token value. Each place has its own equation:

$$\frac{dm(p)}{dt} = \sum_{t \in {}^\bullet p} f(t,p) h_t - \sum_{t \in p^\bullet} f(t,p) h_t$$

Timed information is introduced by the association of specific deterministic rate information with each transition, permitting the

continuous model to be represented as a set of ordinary differential equations (system of ODEs).
A Continuous PN contains all the information needed to generate the system of ODEs. Each place in CPN will determine a ODE based on its pre- and post-transitions: for example, from a simple example of Continuous PN we can generate an equation for each place by looking at the transitions directly connected to the places.

A Stochastic Petri Net is defined as a quintuple ($P, T, f, \upsilon, m_0$), where:
– $P$, $T$ and $f$ are defined as before.
– $\upsilon : T \rightarrow H$ is a function, which assigns a *hazard function* $h_t$ to each transition $tH := \left\{h_t | h_t : \mathbb{N}_0^{|\bullet t|} \rightarrow \mathbb{R}^+, t \in T\right\}$
is defined as the set of all hazard functions, and $\upsilon(t) = ht$ for all transitions $t \in T$.
– $m_0 : P \rightarrow \mathbb{N}_0$ gives the initial marking.

### 4.3.3 Petri Nets in Biological Modeling

A Petri Net is a weighted, directed and bipartite graph, whose representation is widely used to study complex model such as biochemical pathways. Starting with Reddy et al.[Reddy et al., 1993] and the universally cited work of Goss and Peccoud,[Goss and Peccoud, 1998] many other researchers have esteemed this modeling framework in biochemical domain [Simão et al. 2005][Chaouiya et al. 2008][Heiner et al. 2008].
The graphical representation of the Petri net is quite similar to a biochemical network: for physiologic pathways, we can use places to represent compounds (such as metabolites, enzymes, cofactors etc.) participating in a reaction, which will be represented by transitions. The arcs that connect places and transitions will represent the stoichiometry of the reaction, while the tokens associated to a pre-place

can represent the amount of molecules of that particular reactant; tokens in post place the resulting products of reaction (Table 4.1)

| Pathway Elements | Petri Net Elements |
|---|---|
| Molecular Species | Places |
| Biochemical Reactions | Transition |
| Reactants | Input Places |
| Reaction Products | Output Places |
| Amount of Molecules | N° of Tokens on Places |
| Stoichiometric Coefficients | Arc Weights |
| Kinetic Constants | Transition Rates |

Table 4.1 – Pathway mapping: Elements of biological pathway mapped onto Petri Net graph

Above all, places and transition are able to describe every component of a biological system consistent with the scalability prerequisite for a valuable modeling tool.
Onto places can be mapped elements at molecular level (atoms, ions, proteins, nucleic acids, functional complexes), at mesoscale level (organelles, entire cell types, unicellar organisms), as well at multicellular (functional cellular organization, tissues, entire organisms) and ecological level. Transitions can describe all sort of biological occurrence to underline the dynamic of system, ranging from biochemical event (molecular changes, enzymatic reaction, binding, complex association/dissociation, transcription, synthesis/degradation, and so on) to macroscopic dynamics as life activity (movement, eating, reproduction).

Beyond a qualitative analysis, one can uses extensions of the original formalism to model biochemical networks. In particular, stochastic PNs

(SPNs) and continuous PNs (CPNs), the cases of signaling and metabolic networks are dealt.

The standard semantics for qualitative Petri nets does not associate a time with transitions or with the layover of tokens at places, and thus these descriptions are time-free. The qualitative analysis considers however all possible behavior of the time-independent system.

We defined our model by using the timed extension of PN called Stochastic Petri Nets (SPN). The particularity of SPN is that the firing of a transition is not instantaneous. There is a delay following a probabilistic distribution, thus the delay is a random variable.

In SPN biological models, this delay is interpreted as the reaction rate, and it is given by the weight function of the corresponding transition. The delay mean time is obtained by the transition reaction rate, which is a function of a stochastic rate constant and the quantity of each molecular species involved as a reactant or a catalyst.

The SPN description preserves the discrete state description, but in addition associates a probabilistically distributed firing rate (waiting time) with each reaction. All reactions which occur in the QPN can still occur in the SPN, but the probability that a reaction will happen within a period of time depends on the probability distribution of the associated firing rate.

Heiner and colleagues, [Heiner et al., 2008] defined two types of *hazard function* (defining the marking-dependent transition rate for the transition t) for biological systems, based on a molecule number (discrete amount) or concentration (continuous measure). In our model, and in its dynamical investigation, we focus primarily on molecule numbers, therefore, to give explanation of mathematical foundation of our model, we consider only the following hazard function for each transition:

$$h_t := c_t \cdot \prod_{p \in {}^\bullet t} \binom{m(p)}{f(p,t)} \tag{4.9}$$

where $c_t$ is the transition-specific stochastic rate constant (calculated by biochemists from reaction rate constant), and m(p) is the current number of tokens on the pre-place p of transition t.

In order to include time in a SPN, the transitions are associated with a function that defines the waiting time before that transition can fire, i.e. the time before a reaction occurs.

As in QPN, a transition must be enabled before it can fire, but this eventuality is already included in the definition of hazard function

if at least one pre-place has

$m(p) < f(p,t)$

Then $\binom{m(p)}{f(p,t)} = 0$

and the transition would not fire.

The stochastic hazard function $h_t$ defines the marking-dependent transition rate $\lambda_t(m)$ for the transition *t*. The transition waiting time is a negatively exponentially distributed random variable $X_t$ with the following probability density function:

$$f_{X_t}(\tau) = \lambda_t(m) \cdot e^{(-\lambda_t(m)\cdot\tau)}, \qquad \tau \geq 0 \tag{4.10}$$

Petri Nets rapidly became a modeling tool widely used to represent biological systems, thanks to their intuitive graphical representation which helps to model and understand even systems with a complex structure, with many applications also in real life systems. In particularly, in the last twenty years the graphical aspects of the Petri Net attracted the attention of computational biology modelers due to their quite similarity to biochemical network representation, which is able to give superior communication ability to models simplifying their design in particular thanks to the building block approach [Reddy et al. 1993,1996][Chaouiya, 2007].

## 4.4 Stochastic Discrete simulation technique for Biochemical Systems

There are many several equations for modeling systems of biochemical reactions.
We mainly deal with stochastic simulation methods that intend to emulate a random process *X(t)* of system evolution. Having enough sampling pathways of the random process, it is possible to calculate the probability density function *P(x, t)* and other statistical behaviors including the mean trajectory and correlations.

The deterministic approach can only be used to understand the average behavior of the model, but it cannot detect the variability between different individuals, in particular given by the underlying discreteness of the model and the fluctuations in subsystems which contains only few particles.
Instead, the use of a stochastic approach allows to introduce a degree of randomness into the simulation, thus representing the variability between individuals given by internal perturbations in the model.
Also, a stochastic simulation can be used to study the robustness and fragility of the model, because a variation of the initial setting might alter the behavior of the model only in a slight number of cases.
Stochastic approaches are based on the notion of "*stochastic reaction constant*" $c_\mu$ for each reaction $R_\mu$ in the system. In the case of mass action kinetics, $c_\mu$ is derived from the conventional deterministic reaction rate $k_\mu$.
$c_\mu\, dt$ is the average probability that a specific combination of substrate particles in the system will react in the next infinitesimal time step *dt*. Multiplied with the number of possible $R_\mu$ substrate particle combinations $h_\mu$, we get a propensity $a_\mu$ for reaction $R_\mu$ as $a_\mu = h_\mu c_\mu$ with:
$a_\mu dt$ = the average probability that a reaction $R_\mu$ will occur in the system in the next infinitesimal time step *dt*.

The biochemical system can be defined more extensively and formally identified according to a Markov process represented by the *Chemical Master Equation* (CME), which describes the time evolution of the system state probability distribution. [Gillespie, 1992]

### 4.4.1 Chemical Master Equations

Because chemical reactions result fundamentally from random collisions of discrete molecules, a more faithful numerical simulation technique might try to explicitly model these molecular interactions. Rather than tracing the evolution of the concentration of each species in time by approximating it as a differentiable function, such methods would instead tally the number of molecules of each species, and rely on discrete-event simulation techniques to evolve the population vector in time.

The chemical master equation (CME), shows that random fluctuations in species concentration or reaction rates scale with (N)1/2, where N is the number of molecules of a given species in the compartment [Gardiner, 2005].

Similar to a Markov process, therefore, the CME is a description of the stochastic state of a set of reactions and is specified as follows:

$$\frac{\partial p(x,t)}{\partial t} = \sum_{u=1}^{r} [-p(x,t)a_u(x) + p(x - v_u, t)a_u(x - v_u)]$$

where $x$ is the state vector listing the concentrations of each species, $p$ is the probability that the species concentrations will be $x$ at time $t$, $r$ is the number of reactions, $a_u$ is the probability that reaction u will take place and $v$ is the change in $x$ due to reaction $u$.

### 4.4.2 Stochastic Simulation Algorithm SSA

As solving CME (analytically or numerically) can be a challenging task, stochastic simulation methods are used to generate instances of the stochastic process according to the Master Equation. Gillespie, in several works published over the years (for a review se Gillespie, 2007) developed different algorithm to simulate or realize a CME model efficiently.

In this context of stochastic methodological evaluation of biological processes, we emphasize that the SPN thanks to their ability to specify the Markov process (the SPNs are, in effect, isomorphic to continuous time Markov chains due to the memoryless property of the exponential distribution of firing times [Molloy, 1982]) appear to be particularly suitable mathematical tools for this purpose.

#### 4.4.2.1 Direct Method

The most common stochastic simulation algorithm (SSA) is the "Direct Method" proposed by Gillespie [Gillespie, 1976,1977]; it explicitly simulates each reaction event in the system, therefore it correctly accounts for all stochastic fluctuations. Thus, the algorithm has time complexity approximately proportional to the overall number of particles in the system.

Direct Method uses two random numbers per step to compute which will be the next reaction and when it will happen (the time step $\tau$).

The algorithm follows this procedure:

-Set time t = 0;

-While t is less of an end time value $T_{end}$

-Compute the propensity value $a_\mu$ for each reaction $\mu$ in the set of reactions M

-Compute the sum $a_0$ of all propensity values $a_\mu$

-Generate a random number $r_1$ uniformly distributed between 0 and 1.

The stochastic time step τ is calculated as follows:

$$\tau = \frac{1}{a_0} \ln r_1$$

Determine which reaction will take place. Another random number $r_2$ is generated, and the reaction μ is chosen according to the following criteria:

$$\sum_{i=1}^{\mu-1} \frac{a_i}{a_0} \leq r_2 \leq \sum_{i=1}^{\mu} \frac{a_i}{a_0}$$

The chosen reaction $R_\mu$ is resolved (the reagents are decreased and the products increased), and the time is incremented by τ.

This method has the best accuracy among all Stochastic Simulation Algorithms, but it has a prohibitive time complexity for a system with a high number of reactions.
Slight changes to the Direct Method, such as First Reaction Method [Gillespie, 1976,1977] and Next Reaction Method [Gibson and Bruck, 2000], do not reduce the overall time complexity. This is due to the exponential increase in the total number of reactions the algorithm must simulate, because the number of particles increases in time.

#### 4.4.2.2 Tau-leaping

In order to accelerate simulation of the coagulation model, we used a faster SSA called 'Tau-leap method' [Gillespie, 2001]. This is faster than Direct Method, because it avoids the simulation of every single reaction event. Instead, it "leaps" along the time axis in steps of length τ, which contains many single reaction events.

τ must fulfill a property called "Leap Condition", which means it has to be small enough that no significant change in the propensities $a_\mu$ occurs during [t, t+τ].

Note that τ, differently from direct method, is not given by the reaction system, but it is fixed and user defined. Choosing the correct value for τ is a problem much discussed in literature [Gillespie and Petzold, 2003][Cao et al., 2006]

For each reaction, the number of firings $K_\mu$ during time τ (starting at time t from state x) can be approximated by a Poisson distributed random variable:

$$K_\mu(\tau; x, t) = P(a_\mu(x), \tau)$$

where

$$\text{Prob}\{P(a_\mu, \tau) = k\} = \frac{(a_\mu \tau)^k}{k!} e^{-a_\mu \tau}.$$

In each step, a Poisson random number $k_\mu$ is drawn for each reaction $R_\mu$ in the system, and the number of molecules is updated as in the following formula:

$$x(t + \tau) = x(t) + \sum_{\mu=1}^{M} k_\mu s_\mu$$

where $M$ is the total number of reactions, and $s$ is the state change given by reaction $R_\mu$ (the number of molecules produced or consumed by that reaction, for each reactant in the model).

A single step in Tau-Leap usually takes more time than a step in Direct Method, because computing a Poisson random number is expensive. However, Tau-Leap method allowed us to simulate the coagulation model much faster than the Direct Method, because when the particle number increase, many reactions are simulated at once, respecting the strict approximation limits imposed by the method.

# 5 Pattern Discovery

Starting from the visionary prediction of Frawley and colleagues [Frawley et al., 1992]

> "*It has been estimated that the amount of information in the world doubles every 20 months"*

we arrived to date that is called the "era of Big Data". This term that is frequently abused (one can hardly attribute the term Big data to database is not less than dynamic hundreds of terabyte) identifies a new global need.

New data sources, ranging from sparse business transactions to healthcare data through social media, collected by means of high-throughput technologies and the "Internet of Things" [Ashton, 2009], are creating a digital flood of big data that must be captured, processed, integrated, analyzed, and archived by the aid of unconventional Relational Database Management System (RDBMS) software. Big data systems storing and analyzing petabytes of data are becoming increasingly common in many application fields, and seen to be major, long-term investments requiring considerable financial commitments and massive scale software and system deployments (source: McKinsey Quarterly 2011. McKinsey & Company 1996-2014).

Every field of application need to process, integrate, analyze and store data, remains fundamental to improve the understanding of the phenomenon under investigation. To fulfill these recommendations it advantageous to generate an appropriate database during the execution of an accurate knowledge discovery in data base (KDD) process [Maimon and Rokach, 2005]. A schematic workflow inspired to Fayyad, Shapiro and Smith's work [Fayyad et al., 1996] is shown in the Figure 5.1 below.

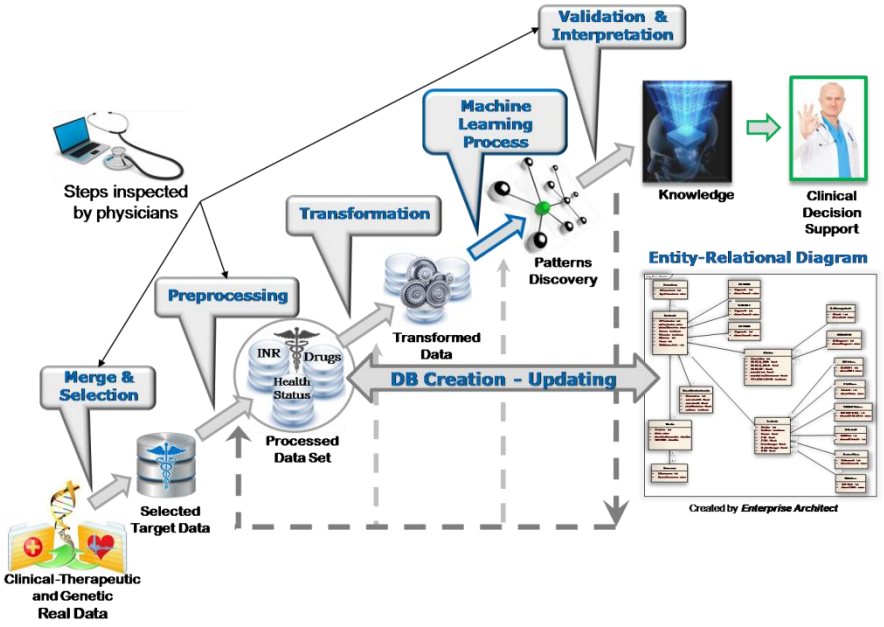



Figure 5.1 - DB Creation and Pattern Discovery in Databases workflow: the two processes are carried out in conjunction to take reciprocal advantages

This convention for addressing to knowledge is particularly conceptualized in Computational System Biology, which, further suggests to couple it with a mathematical modeling language for testing *in silico* the hypothesis gathered from the same KDD framework, as well as by biological experiments [Kitano, 2002][ Ideker et al., 2001].

## 5.1 Database Construction

### 5.1.1 Data Preprocessing

The first step that was required have concerned the integration of data from two different clinical research institutes, each of which had two differently structured databases for clinical and genetic data. This step was performed manually and controlled with simple java control data scripts to avoid loss, duplication, or translation of data during the merging. The starting raw data consisted in extracting and collecting therapy and clinical data on 3949 patients who are prescribed oral anticoagulant therapy in two specialized centers. The first selection of

the data was aimed to isolate elderly patients in an age range between 65 and 99 years (and the resulting average was 78 years), which generate a selected dataset composed of 1013 elderly patients. By means, mainly, of statistical technique, based on data dispersion data, we cleaned the noisy and outliers data. Was applied no policy on missing data both on single missing data, and on tuples ones.
New patient classes were obtained through variable discretization, according to epidemiologic and clinical-based guidance supervised by experts. Continuous features as age, Body Max Index (BMI), Creatinine Clearance (Ccr)
The resulting cleaned data set to carry on the DB design and analysis consisted of 748 patients (instances) and 86 features.

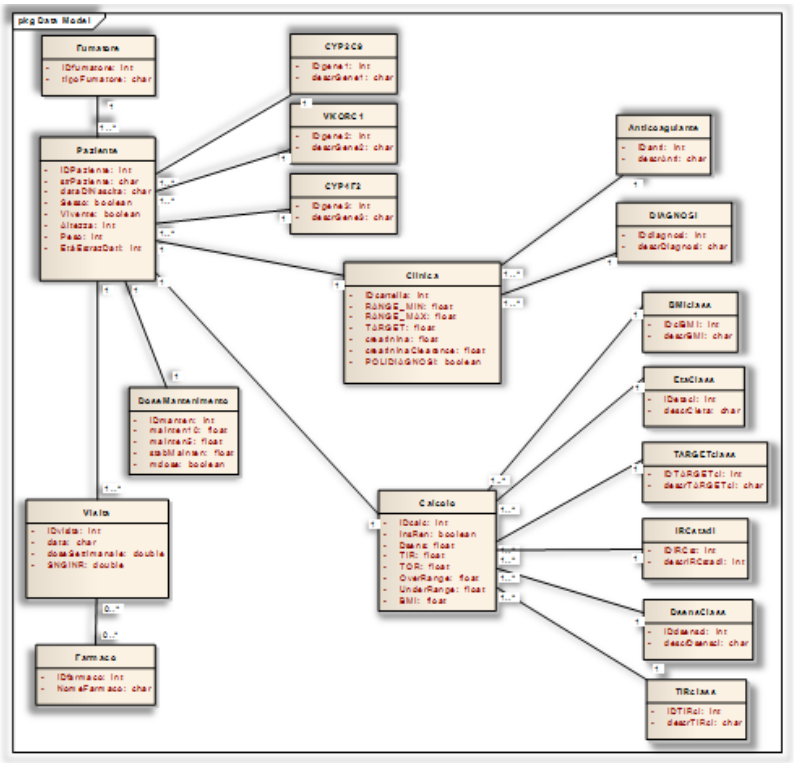
### 5.1.2 Data Set Description

The data set was composed of 748 patients (384 males and 364 females; mean age of 78,4±6,9 years), consecutively recruited at two Anticoagulation Clinics in Milan, Italy. Inclusion criteria were age of 65 years or over, absence of overt liver disease and of alcohol abuse.
Each patient underwent a visit and blood tests. The following data are recorded in electronic patient records: age, gender, indication for OAT, concomitant medications, clinical history, weight, height, smoking habit, biochemical parameter, pharmacogenomics data as well as the at least seven time series of INR monitoring at corresponding oral blood thinner doses (time series of therapy response).
Among elderly patients there is a high prevalence of comorbidity, so 54.8% of patients were in polipharmacotherapy; these patients had a chronic drug therapy administration ranging from 5 to 11 drugs daily.
Two different drugs were administered to patients according to their clinical history end the therapist's know-how: warfarin was prescribed to 453 patients (60,6%) and acenocoumarol to 295 (39,4%) patients.
In particular, the therapy effectiveness was assessed by the criterion of three consecutive in range INR measurements derived by of quasi-

constant drug dose regimen, with the tolerance of ±10% of discard [International Warfarin Pharmacogenetics Consortium, 2009]. This dose, defined as *maintenance dose*,(the effective drog dose) is our *dependent variables* for the predictive regression model.
All variables included in the dataset were assigned to appropriate types (boolean, nominal, categorical and continuous).

### 5.1.3 Conceptual schema design

After identifying the entities, analyzing and enumerating their relations, we have design the conceptual scheme in order to obtain a wide view on the organization of data, essential to valuable query execution.
By means of a graphical software, Enterprise Architect, we realize the Entity-Relational diagram as shown in Figure 5.2




5.2 - ER Diagram of Oral AnticoagulationDataset

### 5.1.4 Database Implementation

We investigated for an appropriate type of DBMS, Database Management System, that would satisfy the needs for logical design according to data representation. We selected a RDBMS, Relational-DBMS, which can fulfill the concept of relationship, as MySQL. Therefore, we have been able to proceed with the implementation of the database. In particular, we have provided the declaration of a logical relational schema using DDL (Data Definition Language). The output of the logical relational schema DDL was then used for the final creation of the tables. All tables were subsequently populated by using the DML (Data Manipulation Language)

## 5.2 Explorative statistics for features relations

Data were statistically analyzed with R statistical software package [R Development Core Team, 2008] and IBM® SPSS® version 18 and 21 (SPSS Inc., a IBM Company**,** Chicago, Illinois, USA). We employed, t-test, Mann-Whitney U test ANOVA test, Kruskal-Wallis for to perform tests of independence; and Pearson correlation, Spearman rho and $x^2$ test to test correlation or association among variables. The appropriateness of test were evaluated according to their data distribution and variable type of features to compare [Peacock and Peacock, 2010]. This explorative analysi allow us to define a statistical relational characterization of patients' features (associations and correlations).

## 5.3 Linear Regression Models

Generally, multiple regression models is used to forecast the variance in an interval dependent, based on linear combinations of interval, dichotomous, or dummy independent variables. Multiple regression can

establish that a set of independent variables (regressors) explains a proportion of the variance in a dependent variable at a significant level (significance test of fitting parameter $R^2$), and can establish the relative predictive importance of the independent variables (comparing beta weights, in statistics, or the coefficient for the other regression approaches) the assigned). Power terms can be added as independent variables to explore curvilinear effects.
Multiple Linear regression model can be mathematically represented as

$$Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj} + \varepsilon_j \quad (5.\ 1)$$

which is equivalent to

$$= \alpha + \sum_{i=1}^{k} \beta_i X_{ij} + \varepsilon_j = E(Y_j \mid X) + \varepsilon_j \quad (5.\ 2)$$

where:

$\beta_i$ is the partial slope coefficient (also called partial regression coefficient, metric coefficient). It represents the change in *E(Y)* associated with a one-unit increase in $X_i$ when all other regressors (in are held constant.

$\alpha$ is the intercept. Geometrically, it represents the value of *E(Y)* where the regression surface (or plane) crosses the Y axis. Substantively, it is the expected value of Y when all the regressors have value equal to 0.

$\varepsilon_j$ is the deviation of the value $Y_j$ from the mean value of the distribution given X. This error index may be interpreted as the effects on *Y* of independent variables not explicitly included in the equation, and a residual random element in the dependent variable, the “dummy” variable.

The parameter estimation process, to assess the metric coefficients, consists in estimating their values from a finite sample across the population. The sample regression model is written as:

$$Y_j = a + b_1 X_{1j} + b_2 X_{2j} + \ldots + b_k X_{kj} + e_j \quad (5.\ 3)$$

which is equivalent to

$$= a + \sum_{i=1}^{k} b_i X_{ij} + e_j = \hat{Y}_j + e_j \quad (5.\ 4)$$

where $a$ is the sample estimate of $\alpha$ and $b_k$ is the sample estimate of $\beta_k$.

The models performances is evaluated by means of goodness-of-fit of linear regress measures $R^2_{Adj}$ (adjusted coefficient of determination) and *F global statistics.*

$\boldsymbol{R^2_{Adj}}$ measures the percentage of the variability in independent variable y that can be explained linearly by regressors $x_1$, $x_2$, …, and $x_k$, adjusted for degrees of freedom related to employed regressors ($k$) and instances ($n$) numbers.
Formally it is defined by the formula

$$R^2_{Adj} = 1 - \frac{n-1}{n-k-1} + \frac{RSS}{TSS} \quad (5.\ 5)$$

where *RSS*, residual sum of squares, is the sum of the squared differences between the observed $Y$ and the predicted $Y$, and *TSS* total sum of squares, corresponds to sum of the squared difference between the observed $Y$ and the mean of $Y$.

***F*** *global statistics* tests about all coefficients computed for all $x_k$, $b_k$, especially if any of the betas $\neq 0$. It is proportionally related to $R^2$ (coefficient of determination) by the formula

$$F_{Glob} = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} \quad (5.\ 6)$$

where $R^2 = 1 - RSS/TSS$

# 6 Model Development

Due to the complexity of biochemical pathway interactions and large number of involved components, it is almost impossible to intuitively understand the behavior of biological networks. Mathematical modeling and computer simulation techniques have proved useful for understanding the topology and behavioral dynamics of such networks. The *In silico* biology, more effectively defined as Computational System Biology, provide a significant advantage to conventional experimental biology in terms of cost, ease and speed, feasibility in general. Therefore, experiments that are infeasible *in vivo* applications or non-measurable *in* by *in vitro* assays can be conducted *in silico*.

Development of predictive *in silico* models present opportunities for gaining exceptional insights about specific and general organization and mechanism of biological systems, and offer the possibility to mimic extreme conditions to assess the behavior and robustness of the real system. In particular, it is impossible to test extreme concentration of drug or of cytotoxic compounds to monitor the modification induced in the *in vivo* system, because the organism or cell exposed to such a treatment would not survive. Moreover, the computational approach can reveal novel experiments for testing hypotheses, based on the modeling experiences.

When is feasible for computational costs, a valuable aspect of modeling biochemical networks is the incidence of stochastic or random events, and the relevance and impact of the inherent variability.
Stochasticity in biological systems is ascribable to the coupled intrinsic and extrinsic phenomena. The former establishes its law on physical basis of randomness; the latter originates from external environmental conditions (e.g. inducing an external stimulus or modification of physic-chemical conditions) [Hilfinger and Paulsson, 2011] [Swain et al., 2002]
To achieve the challenging goal to capture the native stochasticity it is recommended to define the proper workflow about modeling approaches, mathematical formalisms, and simulation algorithms.

## 6.1 Modeling Biochemical Reaction

Biological computational modeling rests on mathematical frameworks to describe the structures and behaviors of real natural systems. A biochemical reactions systems can be generalized as a triple $(S,C,R)$ of three nonempty finite sets, where $S$ is the set of species (ions, molecules and enzymes), $C$ is the set of complexes (association of two or more species), and $R$ is the set of reactions which consist of chemical equation law and kinetic parameters. Therefore, modeling seeks to describe molecular transformations in terms of equations and parameter derived from established biochemical theory. Equations in physicochemical models refer to explicit processes (such as catalysis and assembly) and parameters have physical interpretation (such as concentration, binding affinity, and reaction rate)
The complexity of real biological systems request several assumptions to obtain a manageable model, especially in case it is valuable to simulate its dynamic behavior. The first challenge is then to carry out a plausible model reduction and parameters estimation without changing the fundamentals of the real system.

Models can therefore exhibit different levels of details (model granularity) for the variables and different degrees of precision for reaction assumption and kinetic parameters [Segel, 1988].

Recalling that the aim of the project of the thesis is to establish a pharmacological model in order to mimic the effectiveness in anticoagulant drug therapy, it should be emphasized firstly that is essential to characterize both the physiological and the pathological mechanisms of the system in the absence of the drug administration. This procedure, as well as assures us to satisfy the compliance of the most appropriate criterion of modular approach to model a biological system [Lorenz et al., 2011][Randhawa et al., 2009] [Kitano, 2004] [Ravasz et al., 2002], also permit us to reuse partially or totally the physiological model for other pharmacological applications or interacting processes, by integrating the specific submodule.

## 6.2 Survey on Coagulation Mathematical Modeling

The significance of coagulation pathway, ascribable to its impact in the systemic physiological homeostasis and to its involvement in the onset and acutization of several pathologies, has made it one of the most investigated biological systems in various research fields. The studies aimed at the characterization of the processes that contribute to defining the biochemical properties of this complex system has involved both the biomedical field both the computer science and physics.
The query “mathematical coagulation model” submitted to most common web search engines returns more than 2 millions results concerning journals specialized in biochemistry, clinics, system biology, computer science, applied mathematics and physics that have considered the both rheological and corpuscular nature of coagulation in blood flow, the implication of vascular membranes, as the pure enzymatic reaction events. In their approaches, the authors have considered different levels of abstraction, and reaction schemes ,

representing the coagulation process by disparate mathematical methods.
The complexity of this system lies in the co-occurrence of parallel or cross-talking pathways; in the development of processes in different time phases; several feed backs occurrence; and involvement of membranes. These elements suggest a modular approach to model such a pathway, permitting to describe general behaviors as well as detailed submodules functionality.

The first evidence of the general organization of the coagulation system, dating back to the mid-60s, when MacFarlane published on *Nature* the first enzyme cascade model concerning the blood clotting mechanism [MacFarlane, 1964]. In his work have been defined two path of activation (extrinsic and intrinsic), joining at the end of cascade to commit the same goal, the clot induction.
Subsequently mathematical models representing the coagulation pathway have expanded to investigate assorted aspects of the cascade, taking into account the new sets of conditions and biochemical pathways. Lavine was the first to begin the deterministic modeling approach with a linear system of first order ODEs to describe the sequential enzymatic reactions [Lavine, 1966]. By means of ODEs system have been modeled several different pattern, mostly focused on single kinetic reactions pattern [Nesheim, 1992], according to the continual advances in mechanism understanding. In particular, it has begun to take shape the impact of the extrinsic coagulation pathway as the most responsive pathway in coagulation process [Pohl et al., 1994][Khanin and Semenov, 1989].
Models of greater complexity have been rapidly published. Different works joined the kinetics of various sets of reactions, and firstly described feedback loops [Jesty et al.,1993], the relation of coagulation factors with cellular surfaces [Baldwin and Basmadjian, 1994], and the roles of natural inhibitors to fine modulate the process. Several advanced models describing the Tissue Factor as the trigger molecules

to obtain a quick response to procoagulant stimulus through extrinsic path, have been proposed [Jones and Mann, 1994][Pohl et al.,1994].

The growing knowledge about the biochemical system improved the complexity of relations among the constituting elements, giving rise to the development of models representing extremely large systems of equations, both ODEs and reaction-diffusion equations. The improved mathematical methods have consented the inclusion of several numbers of features in emerging models (membrane binding site densities, availability of phospholipid sites, small molecules cofactors, extent of activating stimulus, and external affections). Examples of these improvements are evident in the model of and Kuharsky and Fogelson [Kuharsky and Fogelson, 2001], who evaluated the contribution of surface membranes in coagulation modulation. Moreover, the development, by Leipold end colleagues [Leipold et al., 1995], concerning in, probably, the first mathematical model of blood coagulation that considers the effects of exogenous inhibitor (the precursor of warfarin) on the main factors that compose the extrinsic pathway (II, V, VII, VIII, IX e X).
The introduction of the spatial features, only as an homogenous space [Baldwin and Basmadjian, 1994] or with reaction an diffusion [Ataullakhanov et al., 2002] in addition to, physical characteristics, as flow rates [Byun et al., 1996], rheological aspects, and membranes [Beltrami and Jesty, 2001], engaged new disciplines in the study of coagulation and haematology in general.
Particularly relevant for the period are the work of Bungay, and and Hockin. The first author described one of the most complete model, consisting of a system of 73 coupled ODEs to fully represent the extrinsic end common pathway including the end part of the intrinsic one to demonstrate the dependence of a threshold response on the availability of phospholipid surfaces [Bungay et al.,2003]. Hockin end colleagues [Hockin et al., 2002], in their paper describe the kinetics of extrinsic and common ways with scrupulous awareness on natural anticoagulant counteracting the TF increase; end validated the results

of simulation by means of biochemical experiments. His name is one of the most cited by coagulation modelers, and the model has been the starting point for further model developments and the basis for testing new analysis methods [Brummel-Ziedins et al., 2005].

At this stage, investigators are also beginning to consider whether these model systems can effectively characterize the clots development and thrombus formation. Within this perspective, researchers groups have focused on certain important questions related to the coagulation response and thresholds.

Khanin and coworkers after the earliest models of thrombin generation in plasma via extrinsic pathway that integrated five zymogen conversion reactions [Kanin et al., 1989], extended the research to model the sensitivity of clotting time to concentrations of zymogens. In this work [Khanin et al. 1998] on the same perspective of [Pohl et al., 1994] modeled an *in vitro* bioclinical test to monitor the performance of coagulation, the Prothrombin Time test (PT test), specifically to measure the time of fibrin formation, end in particular thrombin generation, under different conditions.

Ataullakhanov and coworkers who studied this process from 1985 to 2002 investigated the growth and termination of clot formation in spatially in-homogenous unstirred systems firstly representing the contact activation and then included the role of the extrinsic pathway [Ataullakhanov et al., 2002].

To integrate the modelization of flow rates influence (sometimes overlooking the reaction kinetics of clotting) to pure enzymatic reaction cascade (excluding the blood flood), was introduced the mathematical modeling by means of partial differential equations (PDEs). They were able to overcome the shortcomings of the two approaches, coupling, with various degrees of detailing, the rheological properties of blood to the biochemical essence of the process. Anand and coworkers [Anand et al., 2003] introduced convection-reaction-diffusion equations taking into account platelet activation, the extrinsic coagulation pathway and fibrinolysis; via a two step system biology approach Panteleev and colleagues reproduced the time - series of the diffusion of TF induced

thrombus formation mechanism in a non-uniform space, involving platelet and C protein [Panteleev et al.2006]; extending the previous work Anand [Anand et al., 2008], with a careful reduction of the model (23 reaction-diffusion equations), has examined the entire process of formation, growth, and lysis of clots.

Concerning the relevant issue of pharmacological approach to coagulation modeling, Some models include the cascade of physiological interaction with exogenous molecules, such as anticoagulants. Wajima described a really "comprehensive" model [Wajima et al., 2009] using the differential equations not just to represent the whole process of blood coagulation (considering the delayed aid of activation of the intrinsic to the extrinsic one); but also implementing the model with concentration-time and time-effect dynamics of warfarin, heparins, and vitamin K and the corresponding test to evaluate their effects on different subject conditions.

However, none of these stochastic approaches allows a detailed analysis of the coagulation system and thrombin generation variability, whose results can be representing existing clinical tests, in particular the most widely used test for monitoring oral anticoagulation therapy, the "Prothrombin Time" (PT) test.

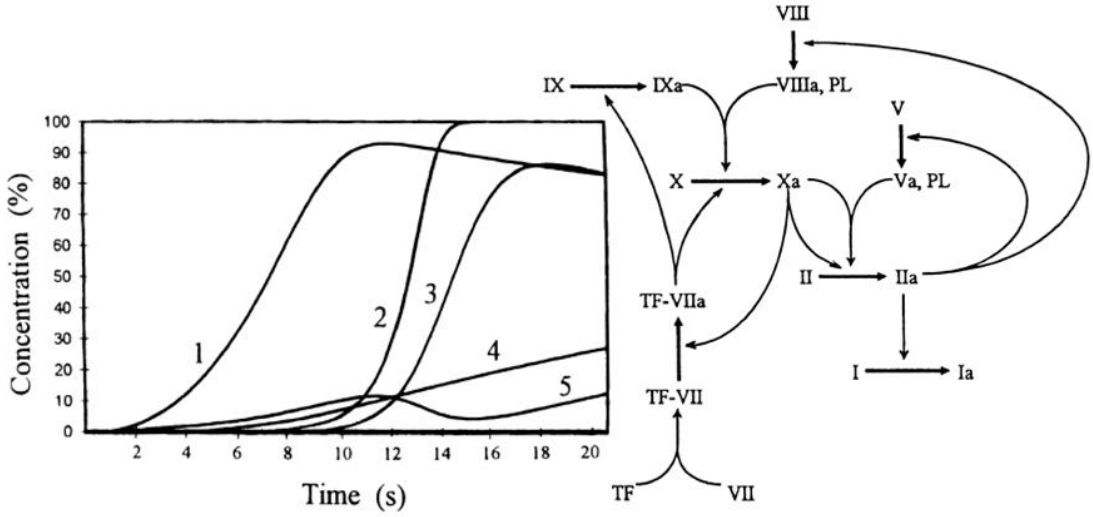



Figure 6.1 - PT test time-course and underlined path in Khanin et al. 1997: Time-courses of enzymes concentrations during the PT test execution (Time series list: 1-factor TF-VIIa; 2-factor Ia; 3-factor IIa; 4-factor Xa and 5-factor IXa).On the right the scheme of the underlined *in vitro* TF pathway of coagulation. Khanin et al. 1998.

## 6.3 Coagulation Modeling Assumptions

### 6.3.1 Environmental modeling assumption

The classical coagulation cascade model was biochemically deeply investigated in the years to yield essential advanced understanding of the enzymatic coagulation pattern of *in vitro* coagulation. The recent inclusion of the active role of cells in coagulation allows for an integrated view of the mechanisms by which coagulation may work in the dynamic of the vascular flow (sec. 3). In our model, we have not contemplated the active coagulatory role of cells, because the mechanisms of action of the considered anticoagulant drug (warfarin), and the related PT monitoring tests, are not conditioned by platelets activity (see sec.nn,). We, instead, merged the two coagulation functional models to obtain a more realistic improvement of the underlined quick coagulation mechanism [Porwit et al., 2011].

We therefore represent a qualitative and quantitative humoral coagulation processes, based on the extrinsic pathway (Tissue Factor pathway) that consist in a well-stirred 100 picoliters ($1\times10^{-10}$ liters) molecular environment, set up with a realistic molecules amount in accordance with the considered plasma volume, which is in a steady state flux-dynamics. This assumption is often quite recommended to gain more information about different tiers of function or modules interactions both for metabolic or signaling networks in intracellular and extracellular environment [Bro et al., 2006][ Schuster et al.1999]
However we incorporated into the model the recent insights on the extrinsic pathway, concerning the microparticles roles and the pre-activated complex VIIa-TF as inflammatory prothrombotic marker [Monroe and Key, 2007][Girard and Nicholson, 2001]. Furthermore, consistent with the contemporary clinical guidelines we followed the recommendation to consider the pharmacogenetics features of the warfarin therapy [Johnson et al., 2011][Camm et al., 2010].

The built mechanistic model represents a hypothetical *in vitro* test (PT test) which is able to *in situ* measures the activity of this pathway in physiological conditions (healthy), in the onset of a thrombotic event (unhealthy) and during therapy. The biochemical *in vitro* PT test is not able to capture the development of somewhere intracorporeal thrombus formation. This test is based on the analysis of peripheral blood and the thrombus trigger factor, fortunately, does not affect haemostasis at the systemic level. Otherwise, during oral anticoagulation therapy, the PT test, by INR transformation, measures the systemic pro-hemorrhagic effect that prolongs the coagulation time of whole blood, measuring the thrombin plasmatic level. This is in accordance whit the goal of anticoagulant treatment based on the prevention of any thromboembolic event in the whole cardiocirculatory system, resulting in a potential serious and possibly fatal event.
Therefore, to describe the effectiveness of blood thinner to counteract the procoagulant stimulus *in situ*, it has been necessary to develop a mathematical model that represents, as we said, a local *in vitro* clinical

test. Furthermore, the PT test is not a kinetic test that records the temporal evolution of thrombin production, but it is an end-point test, differing from the time series evaluation gained whit the model simulation.

### 6.3.2 Model Granularity Refinement

The real Tissue Factor pathway consists of many other molecules compared to those included in our model. Some of these molecules act as cofactors, others mediate the complex formation process, or are negative modulators.

Among the more relevant cofactors that is not considered in the model there is the ion $Ca^{2+}$. It acts in almost all reactions and than it should be handled as logical place, with the resulting increase of the species to be modeled by the equations. Its cofactor effect was then included in the kinetic constants of enzymatic reactions in which it takes part.

von Willebrand factor, the main modulator of pathway's complexes formation, was not modeled due to its multimeric structure that generates extreme functional variability and would be uselessly tangle to represent it, in particular in pharmacogenomics model [Cao at al., 2008].

Finally, the inhibitory effect of Protein C was not considered as additional negative modulators of the process since its mechanism of action involves an entire sub-pathway [Esmon, 2003] considerably increasing the number of modeled molecules.

All this neglected molecules would significantly increase the computational complexity of simulation if they had been included in the model without, necessarily, providing a significant contribution to system understanding [Lebiedz et al., 2008].

About the submodule developed to describe the action of the blood thinner, it was necessary to model the cycle of vitamin K. The complete cycle, which includes vitamin K redox compounds and their related transforming enzymes, has been simplified in the number of composing

elements. Relatively to the states of vitamin K states, we considered only the active form of vitamin K (VKH2) and the oxidized one (VKO), whereas, the enzymes considered to model the reactions were only GGCX and VKORC1. This model reduction simplifies the pathway as well enables us to evaluate the main target reaction of warfarin and the effect on pre-activation of vitamin K-dependent clotting factors (II, VII, IX and X).

### 6.3.3 Biochemical Reaction Law Assumptions

The most common reaction that can be molded in a biological system is driven by the approximation of the Law of Mass Action. According to this physicochemical kinetics, the chemical reaction rate is directly proportional to the molecular concentrations of the reactants (substances concentration is usually indicated by species name or chemical formula enclosed in square brackets), which is a simplification compared to the real laws of thermodynamics that guide the process [Keener and Sneyd 2009].
Therefore, for an association/dissociation reaction schematized below

$$\mathrm{A} + \mathrm{B} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{AB} \tag{6.1}$$

the forward reaction rate, $R_f$, for the production of the AB specie is calculated by the equation $R_f = k_a \cdot [A] \cdot [B]$, a *second order* kinetics with $k_a$ standing for the association constant, expressed in 1/concentration·time units (usually $uM^{-1} \cdot sec^{-1}$), and the capital letter enclosed in square brackets are the concentrations of the species. The reverse reaction rate, $R_r$, is given by the equation $R_r = k_d \cdot [AB]$, a *first order* kinetics with $k_d$ specifying the dissociation constant, which have units of 1/time (usually $sec^{-1}$).

Model several enzymatic reactions can be quite tricky if you consider the real dynamics of the process, which does not follow the law of mass action directly. The biochemical law that directs the enzyme kinetics is, as most of all biochemical process, nonlinear [Qian, 2011]. Its behavior adheres rather to an asymptotic trend, induced by limited catalytic capability. When the concentration of substrate is progressively augmented, the reaction rate increases only to a threshold, reaching a maximal reaction velocity only at saturation of the catalytic sites, at very high substrate concentration (example of the enzymatic saturation limit is shown in Fig 6.2 below).

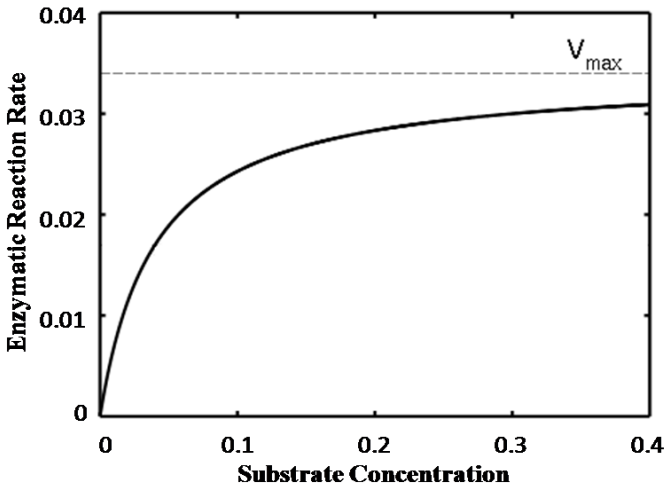



Figure 6.2 – Asymptotic trend of enzyme reaction rate, according to the increasing concentrations of substrate. Vmax is the maximum reachable rate of enzyme, and is expressed in molecules of product formed per fixed time.

About the number of chemical events that are involved in substrate transformation, this reaction consists of two coupled association-dissociation reactions driven by four different four different reaction rates, usually modeled with a set of four different equations [Nelson and

Cox, 2000]. The chemical notation shown below could schematize this reaction:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{E{\cdot}S} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E} + \mathrm{P} \tag{6.2}$$

Where E is the enzyme amount (in continuous ODE models correspond to concentration [E]), S is the substrate, ES is the enzymatic complex and P is the amount of the product of the reaction. $k_1$ and $k_{-1}$ are association and dissociation rate, and $k_2$ and $k_{-2}$ are the production rate of reaction product and the rate of the inverse reaction, respectively .
To model the enzymatic reaction in our system we assumed the *quasi-steady state approximation* kinetics proposed by Briggs and Haldane [Briggs and Haldane, 1925] based on the Michaelis–Menten kinetics law [Johnson and Goody, 2011], It not considers the $k_{-2}$ constant by the hypothesis of highly enzymatic catalytic efficiency, and it is universally represented by the schematic reaction as below:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{E{\cdot}S} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{6.3}$$

Where the enzyme, the reactant, the complex, and the product are the same described in (4.3), $k_1$ is the enzyme association rate (sometimes known as *affinity constant*), $k_{-1}$ is the enzyme complex dissociation rate, and $k_2$ is the enzymatic kinetic constant, known, in Michaelis-Menten terms for unimolecular reaction, as $k_{cat}$, expressed in sec-1. The approximation consist in assuming the concentration of ES remains essentially constant with time (in ODE terms $\frac{d[ES]}{dt} \cong 0$).

This assumption has permitted us to use the formulas derived from the approximation to obtain useful equation to compute the $k_2$ rate when it was not reported:

$$v = V_{max} \frac{[S]}{[S]+K_M} \quad (6.4)$$

$$K_M = \frac{k_{-1}+k_2}{k_1} \quad (6.5)$$

$$V_{max} = k_2 E_0 \quad (6.6)$$

The equations (6.4, 6.5 and 6.6) define:
$v$ is the rate of the production of the product related to the availability of substrate;
$V_{max}$ is the maximum rate at which the enzyme catalyzed a reaction in substrate saturation condition;
$K_M$ is the Michaelis-Menten constant, even if in the steady state approximation it assume a different meaning [Johnson and Goody. 2011]. It is experimentally obtained measuring the substrate concentration that determines the rate of the enzyme reaction is equal $\frac{1}{2} V_{max}$. $E_0$ in (6.6) is the total amount of the enzyme in the initial condition.
As a complete description of enzyme kinetics is not the objective of this thesis, a detailed explanation can be found in [Sauro, 2012].

This assumption, more than allowed us to model readily the enzymatic reactions through the Petri Net notation, was recommended to derive some of the rate parameters during parameter fitting process described in the section on to the model setting.

The *Quasi-steady state approximation* has become a fundamental assumption to deal with enzyme kinetics modeling [Li at al. 2008].

## 6.4 Biochemical Stochasticity Representation

In order for *in vivo* chemical reaction occurs, the reacting species must collide, have sufficient energy, and be well oriented.
The number of collisions is proportional to the concentration of the reacting species. For the reaction (6.1), as we said, the reaction rate is given by the Mass Action Law; but not all collisions are reactive.
The rate constant $k_1$ accounts for the probability that the molecules are well oriented and have sufficient energy to react, and, conversely, the $k_{-1}$ represents the probability that this reaction does not occur.
The average behavior of the process, usually obtained from the solution of a system of ODEs, may be informative in case it has to deal with systems with huge number of molecules that stays almost stable during all the transition state of the system.
This condition is not always easy to know *a priori*, in particular, in highly interconnected concurrent systems, in which it is not unusual that some products, arisen from upstream reactions, could become the reactants quickly used by the next reactions. Such sequential events gives rise to states of the system with low token number in the involved places, condition that is more accurately manageable whit a stochastic simulation approach.
Another reason that consistent with using a probabilistic simulation method was the large intrinsic variability shown by the coagulation cascade, as mentioned during the description of the biological process. The ability to capture this uncertainty in the evolution of the process, which mimics the observed physiological variability, is an exclusive feature of the probabilistic simulation.
The above reported observations suggested us that the more fitting *in silico* description of a biochemical complex system behavior consists in a stochastic modelization/simulation approach.

The Stochastic Simulation Algorithm and its approximations employed, strictly related to SPN models simulation, needs each biochemical transformation is expressed as distinct elementary reaction for forward and reverse transition of discrete molecules. The mathematical representation needs to be based on Law of Mass Action kinetics that is an approximation of a more fundamental, discrete, and stochastic description based on the chemical master equation (CME) that we will expose in next section.

## 6.5 Model Building Process

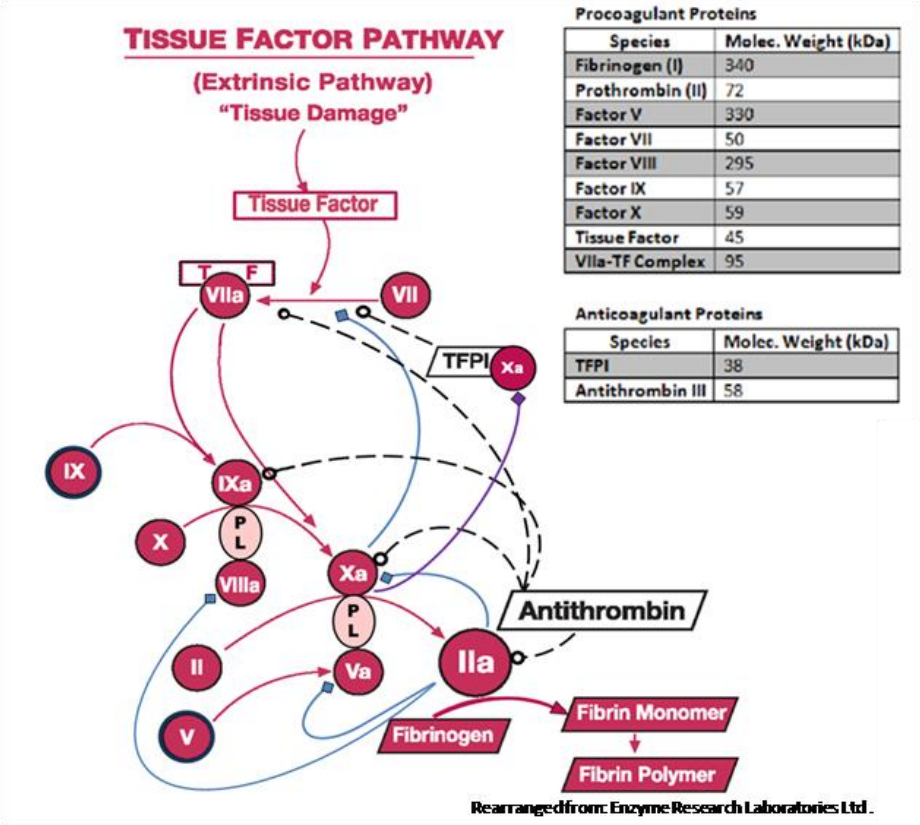
To obtain a usable representation of a biological system is necessary to combine in the same model both static and dynamic aspects of the system with a special emphasis on the semantics to determine dependencies and relationships between the model elements in an unambiguous way [Kitano, 2002]. Therefore, the modeling process can be viewed as translations of intuitive biological pathway maps into a formal graph that can be enhanced whit mathematical equations. This approach becomes easier and more transparent with the adoption of schematic standards for modelization.
Among the several formalisms examined, we chosen the Petri Net formalism, which, not only in our opinion, is the most appropriate modeling formalism to comply with the recommendation for an effective biochemical system modelization. Petri Nets are able to reproduce the structure of the biochemical networks and are founded on sound mathematical theory to investigate the dynamical systems behaviors.

### 6.5.1 From Wiring diagram to Petri Net

The Wiring Diagram, exploit to minimally define a system schema, informally reproduce the network one-mode projection onto vertices needed to identify which variables compose the system and how they interact each others. The directed graph, slightly modified in the arches formalism, shows an easily understandable conceptual representation

of how the system works. It underlines the trigger factor signals flow; emphasize the positive and negative feedbacks; and the direct inhibitions (fig.6.3).

This preliminary graphical representation acquire a fundamental relevance because in the mechanistic model the minimally elements will be modeled with enhancing features. Each node will become a molecular species (places) that thanks to tokens marking identifies a quantifiable status of the system that changes over time, whereas each edge, provided with equations and rates, will compose the set of reactions (transitions) equipped with differently directed arches (input and output) formalizing how the species affect each others. The resulting network properly fine-tuned in its parameter and rate laws, will became able to outline the behaviors of the modeled system.




Procoagulant Proteins

| Species | Molec. Weight (kDa) |
|---|---|
| Fibrinogen (I) | 340 |
| Prothrombin (II) | 72 |
| Factor V | 330 |
| Factor VII | 50 |
| Factor VIII | 295 |
| Factor IX | 57 |
| Factor X | 59 |
| Tissue Factor | 45 |
| VIIa-TF Complex | 95 |

Anticoagulant Proteins

| Species | Molec. Weight (kDa) |
|---|---|
| TFPI | 38 |
| Antithrombin III | 58 |

Figure 6.3 - Tissue Factor Procoagulant Network: red arrows corresponds to the procoagulant signal flow started with trigger factor (Tissue Factor), blue arches indicates the positive feedbacks, purple arch is a indirect negative feedback and dash arches correspond to direct inhibition. The table on the right shows the molecular weight of molecules used to obtain the number of molecules in the fixed volume of 1x10-10 liters of model.

The formal bipartite graph, the Petri Net, was drawn detailing the projection network on nodes with information achieved from databases, repository, literature, wet lab data, and experts' advices.
Reproducing the directed network structure, using the Snoopy Petri Net framework [Rohr et al., 2010] were drawn the formal SPN model, which grew up according to the sub-pathways hierarchy of the coagulation process described before.

Hence, our building process consisted of sequential composition of the active submodules reproducing the overlapping initiation, amplification, and propagation phases differentially modulated by the inhibitory one, also known as termination phases.

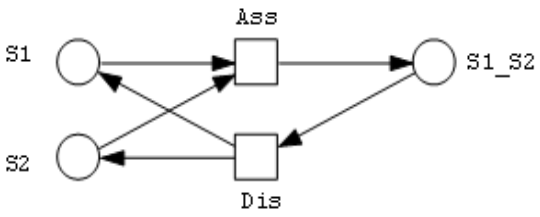
### 6.5.2 Hierarchical Stochastic Petri

Large and intricate Petri Nets development can be quite demanding to draw and risky for the readability of the network, vanishing, consequently, their recognized ability in resembling the represented biochemical pattern.
However, similar to modular programming, the construction of a SPN can be simplified by incorporating more transitions into macro-nodes utilizing the facilities within Snoopy framework to crate hierarchical network using its *coarse transitions* [Heiner et al., 2008].
Conceptually, Petri Nets with "macro-transitions" are nets with multiple layers of detail in the transitions, which are not directly visible in the model. This graphic simplification allows to computational modelers to deal whit the building-block approach, particularly appropriate to reproduce *in silico* biological networks [Breitling et al,. 2008][Fu, 2006].
This modeling method supports both the top - down and bottom - up

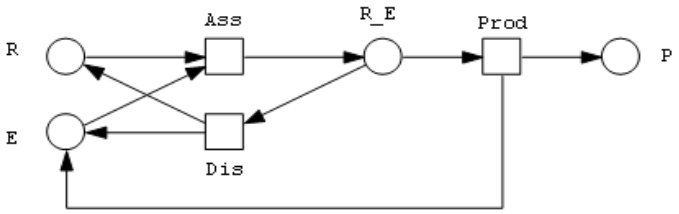
design, underlining the typical high modularity of biological systems [Albert, 2005].

The hierarchization of our model gives rise to a hierarchical high-level SPN (HHSPN) which is in all perfectly equivalent to ordinary flatten SPN model [Buchholz, 1994]. Hence, the hierarchical structure of the model does not change any static and behavioral properties of the flat model [Fehling, 1993] and can facilitate the management of large-scale models preserving the semantics of a complex biological pathway [Breitling et al., 2008].

In the HHSPN were used two different kinds of biochemical reactions in hidden layers, which underline the most prevalent reactions of natural biochemical patterns: the association/dissociation complex formation (AD#) and Michaelis-Menten enzymatic reactions (MM#). We applied this Petri Net enhanced features to describe all AD transition and the MM transition chemically defined in sec 6.3.
These two classes of sub-networks included in the hierarchical net were modeled in their flatten version as shown in the figure below (Fig.6.4)




(a)




(b)

Figure 6.4 - Description of flatten macro-nodes. (a)Representation of association/dissociation of a complex (AD# in hierarchical high-level SPN). (b) Representation of a Michaelis-Menten enzyme reaction (MM# in hierarchical high-level SPN) according to chemical reaction scheme 6.2 and 6.3 respectively.

### 6.5.3 Sub Module Definition

The different subnets were then processed to annotate, extend and complete them with mathematical expressions corresponding to each transition. The system of equations defined by the SPN is enumerated using appropriate rate laws, represented by mass action kinetics, which uses the product of a transition rate and the amount of molecules in pre-places to calculate the firing rates (reaction rates).

The building process considered depiction of all coarse transitions, which were then connected to each other by direct arcs. The arches linked output-places of the previous macro-node to transitions of the following one as schematized in the flatten exemplificative net in Fig 6.4

Each integrated submodules were tested in the resulting network for its effect on previous model version and the effectiveness in Thrombin (IIa) modulation (generation or inhibition) and Fibrinogen Ia production.

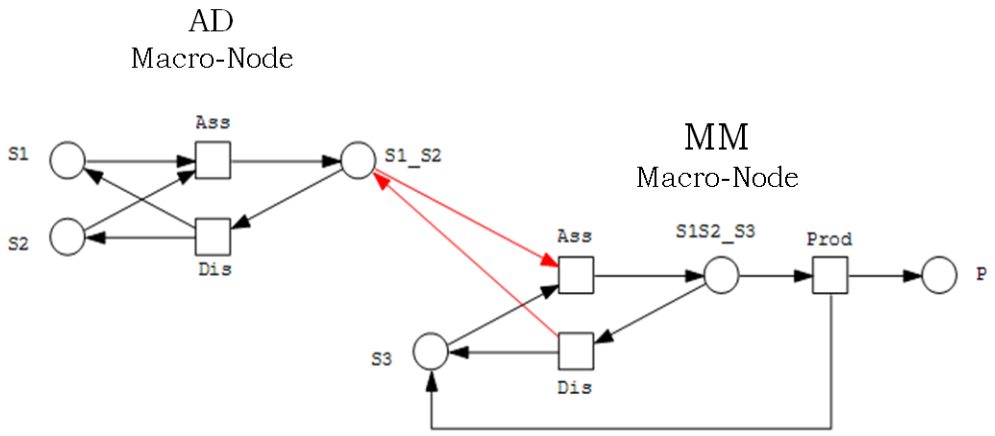


Figure 6.5 – The flatten depiction of macro nodes connection in the SPN models. It shows an example of AD reaction for S1_S2 complex formations that becomes the active enzyme for S3 reactant of the further MM reaction.

#### 6.5.3.1 The Initiation Net

The initial sub-model (*Initiation Net*) describes the initiation phase corresponding to the exposure of the trigger factor. The tissue factor-expressing cells assisted by TF-microparticles expose their TF to plasma as a consequence of mechanic or clinical event. The complex that arise from the contact of factor VII and TF, together with its constitutively pre-activated form (VIIa-TF), activate directly Factor X (pre-place X) which, in its active form (post-place Xa), give rise to the minimal amount of thrombin required to initiate the clotting assembly (I_fibrinogen transformation) and to start the chain of positive feed-back events. The network that origins from these first excerpts of the biological system may be represented by underlying hierarchical stochastic Petri Net showed in figure 6.6 below:

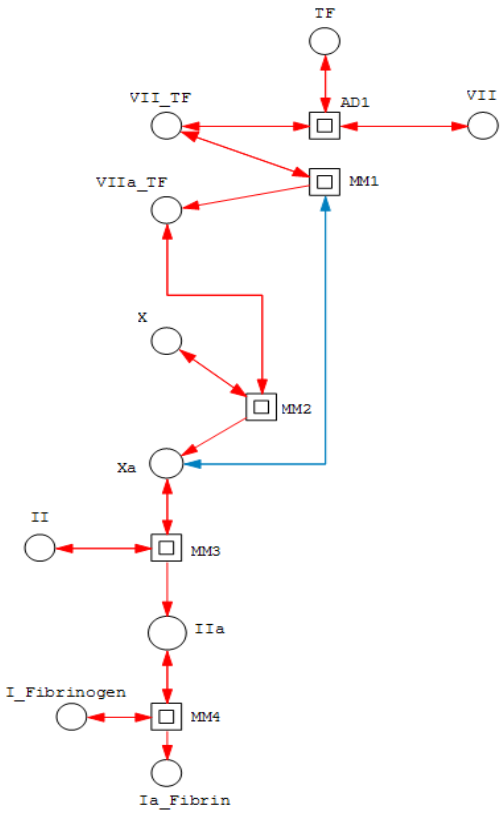



Figure 6.6 – Stochastic Petri Net submodule related to the starting process of the coagulation model. Its hierarchical notation consists of 5 macro-nodes, 10 places and 26 arches (11 double, and 4 single edges).

#### 6.5.3.2 The Amplification Net

The first submodule added to initial model, which define the *Amplification Net*, allows to the initial signal spreading in the system to activate the species that subsequently boost the signal itself, with a resulting increase of the effectiveness of the coagulant event. Part of the previously activated Thrombin (IIa) is used to activate the new upstream added factors (input places VIII and V), as well as to increase the amount of the factor that directly generates it (Xa). To realize this network expansion, we added to the previous sub-model three new MM Macro-node and their directly connected nodes.

This phase, therefore, initially deviates from overall objective of cascade process (Ia_fibrin formation) in order to establish better system condition to the following progress of coagulation stimulus. Indeed, part of the molecules (tokens) in place IIa are captured by coarse transitions MM5, MM6 and MM7 that create a conflict affecting the place itself.

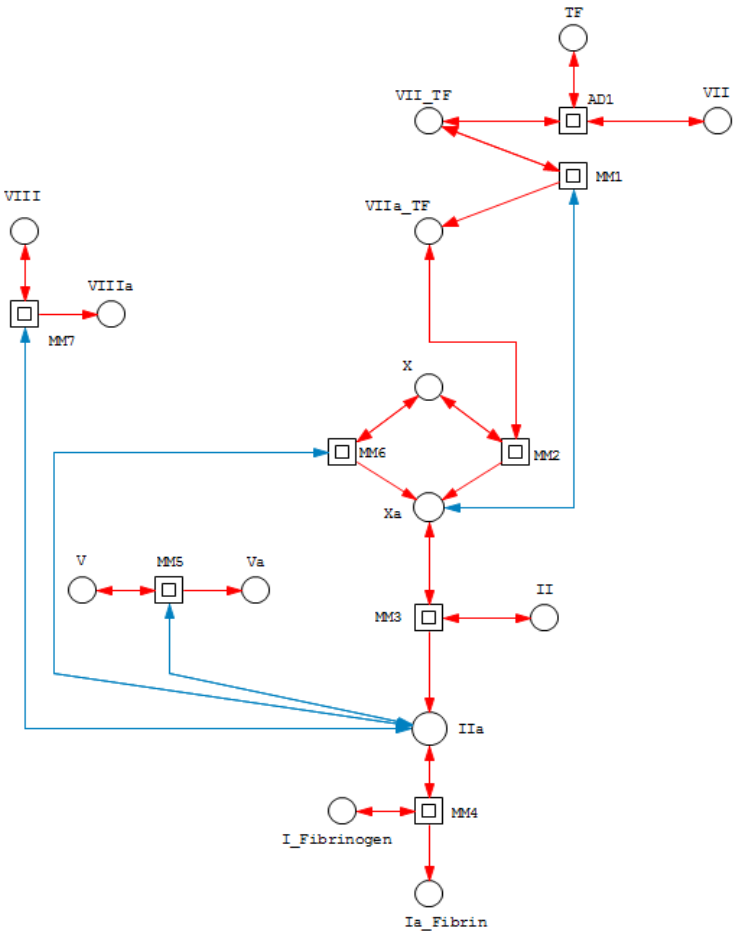



Figure 6.7 - The Amplification Net: The Petri Net originated by the integration of the *Amplification* sub-module that spreads the trigger signal into the system. Blue arches added stand for the biochemical positive feedback process. The resulting bipartite graph consists of 8 macro-nodes, 14 places and 41 arches (17 double, and 7 single edges).

The real effectiveness of the trigger factor is reached in this phase during which the highly functional complexes for thrombin generation are assembled and activated (IXa_PL_VIIIa and Xa_PL_Va). Simultaneously, the general profusion of procoagulant stimulus, in

particular the abundance of molecules in Xa place, pre-activates the counterbalancing modulation system , to avoid itself excessive propagation. This initial negative modulation is obtained drawing the additional AD3 macro-nodes (the purple sub-net in Fig 6.8 below).

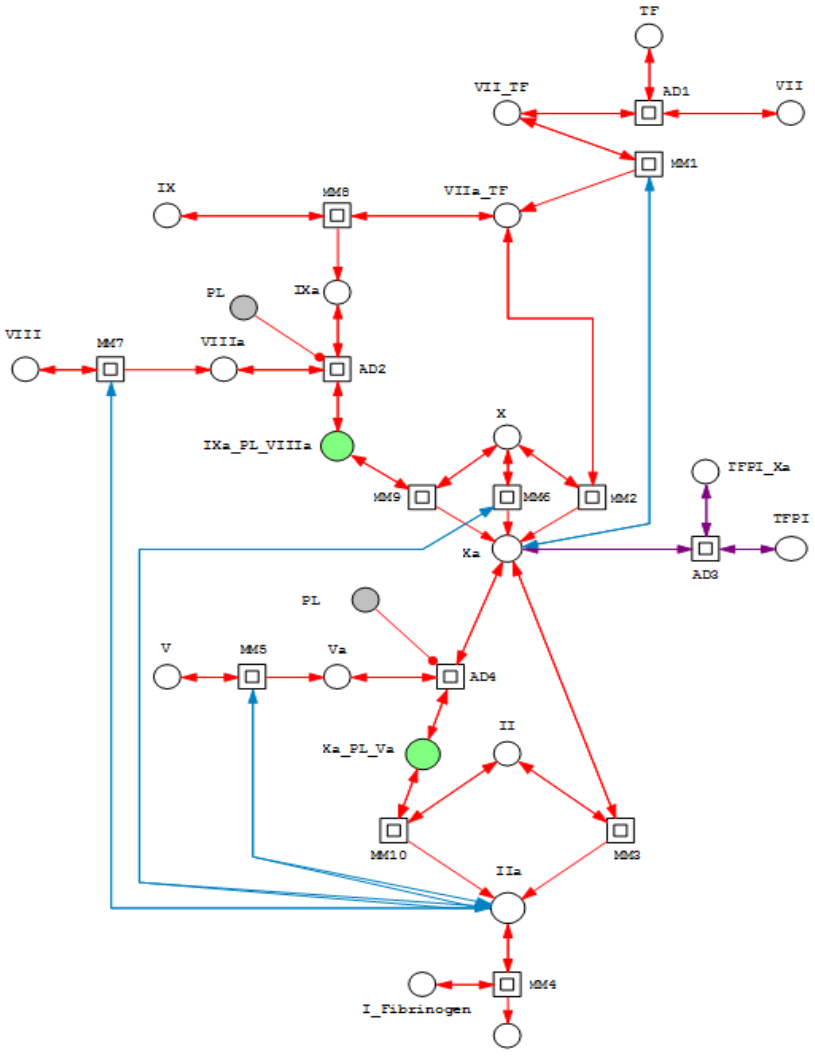



Figure 6.8 - The Propagation Net: The network arisen from the incorporation of enzyme complexes (green places) is able to describe the burst of thrombin production which maximize the coagulative activity of the system and simultaneously pre-activate one of the counterbalancing pattern, as indirect negative feed-back (purple AD3 sub-net) to finely modulate the process. In this submodule we employed 2 red arches to model the qualitative

Notice that in this net we employed an extension of Petri Notation to reinforce its expressiveness, the read edges (or test edges), to model the qualitative occurrence phospholipids (PL place ) in complexes formation. In particular, read edges have the property that upon firing, the marking on the tested place is not changed [Heiner et al., 2008].

#### 6.5.3.3 The Physio(patho)logical Model

The complete physiological model has to consider the natural anticoagulation pattern represented by the inhibitory proteins that act at different level of the cascade to contain the coagulation effect by negative modulation [Malý et al. 2003]. The inhibitory sub-pathway was drawn by means of supplementation of AD macro-nodes (AD5-AD8) which act displacing the pro-coagulant factor out of the process assembling them in an inactive complex and preventing, therefore, to perform their function. The pre-places that hold the inhibitory task are the TFPI and ATIII (Figure 6.9)

This SPN, opportunely set in the $M_0$ configuration (token number in TF and VIIa_TF places) is able to represent even the pathological scenarios, which leads the imbalance of the system. So, this model will be henceforth termed as the *(Un)Healthy model*.

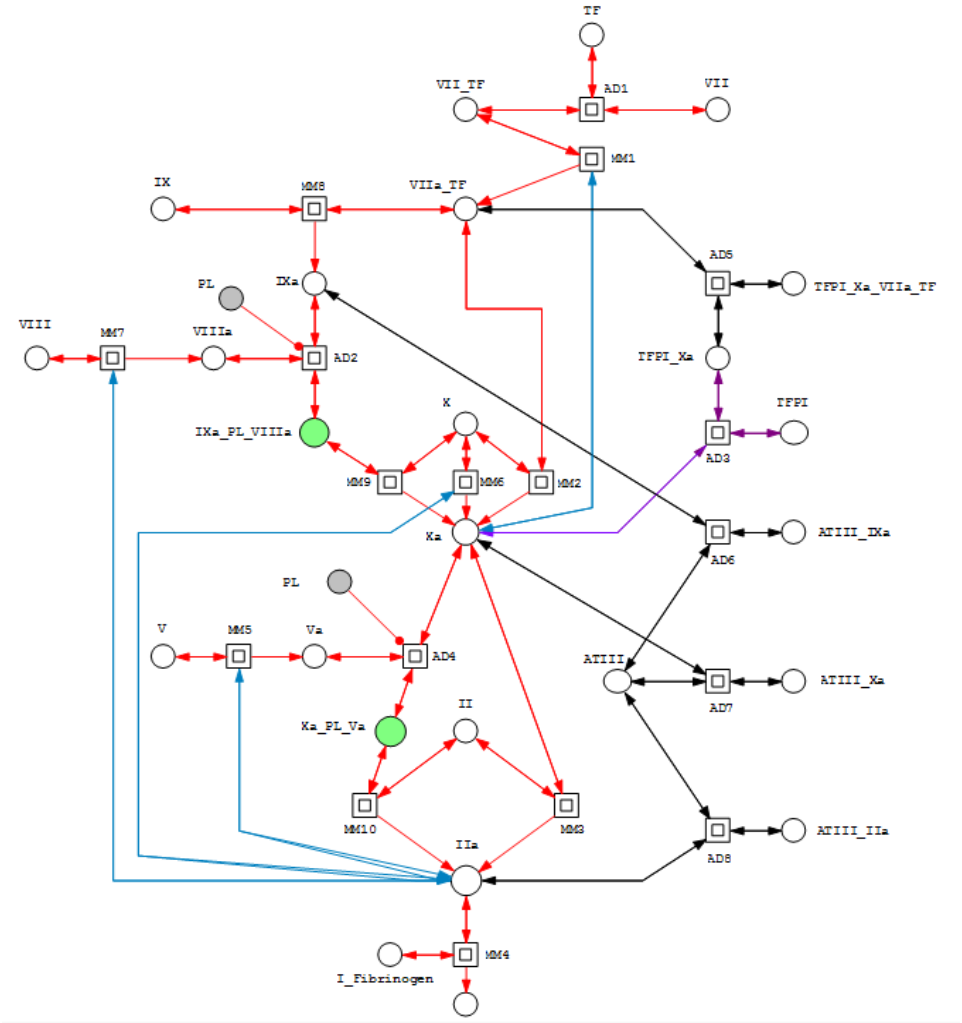



Figure 6.9 – The *(Un)Healthy* HHSPN: resulting by the four compositional phases modeling, it summarizes the entire extrinsic coagulation pathway.

The resulting hierarchical PN is composed of: 8 association/dissociation macro-nodes (AD1-AD8); 10 Michaelis-Menten coarse transitions (MM1-MM10); 44 double edges (44 forward and 44 reverse) and 12 arches connecting, overall, 26 main nodes (the non-visible 10 nodes are included into macro-nodes and are not relevant for Petri Net readability because they are reaction intermediates that are relevant only for simulating purpose). Whereas, the resulting flatten Petri Net is composed of 36 places, 46 transitions, and 142 arcs (two are read arches).

#### 6.5.3.4 The Pharmacogenomic Model

The previously developed (*Un)Healthy* Stochastic Petri Net model considers the biochemical elements (places) and reaction (transitions) to define the time course of coagulation cascade triggered by coagulation factors of the extrinsic path and evolving through the common one, including the natural anticoagulants proteins.

The new model presented in this under section includes, additionally, the components related to the vitamin K cycle, to map the interaction of hepatic coagulation proenzymes, required for modeling the response to blood thinning medication, warfarin. From the biochemical diagram of warfarin pharmacodynamics (Figure 6.9) we extract the noteworthy features to describe the drug reaction required to implement the previous model by pharmacogenomic submodule, starting again from mapping the wiring diagram entities onto the previous SPN.

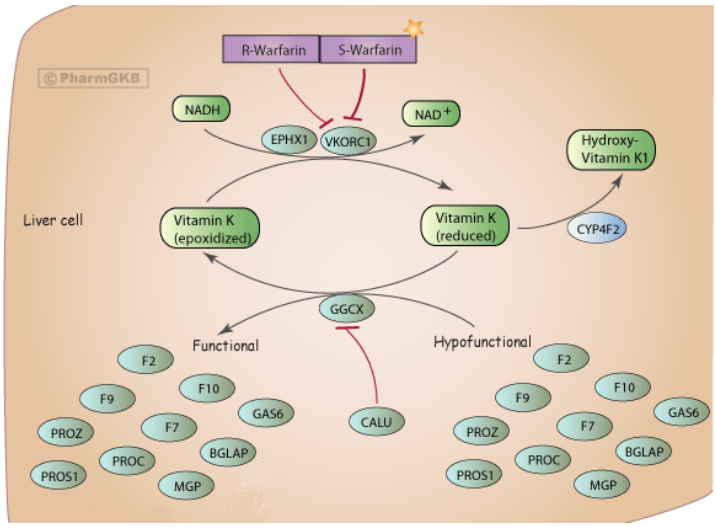



Figure 6.10 - Warfarin pharmacodynamics diagram: simplify wiring diagram of the target of warfarin action and downstream modulatory effects on gene products (from PharmGKB, a registered trademark of U.S. Department of Health & Human Services

To implement this model we added three transitions: VKORC1 is the enzyme affected by genetic variations; the GGCX is the enzyme that catalyze the activation of pro - enzymes to active form (it is replicated

four times as gray square indicating logic transition, a unique transition repeated in the model to give emphasis to the interaction pattern); and the W_Act transition, to model the negative modulation of warfarin inflicted to vitamin k cycle and so to related coagulation factors.
Moreover, we added the drug place (Warf), the two biochemical forms of vitamin k (Vitk and VitK_Ox) and the four pharmaco-dependent coagulation factors (IIp, VIIp, IXp and Xp).

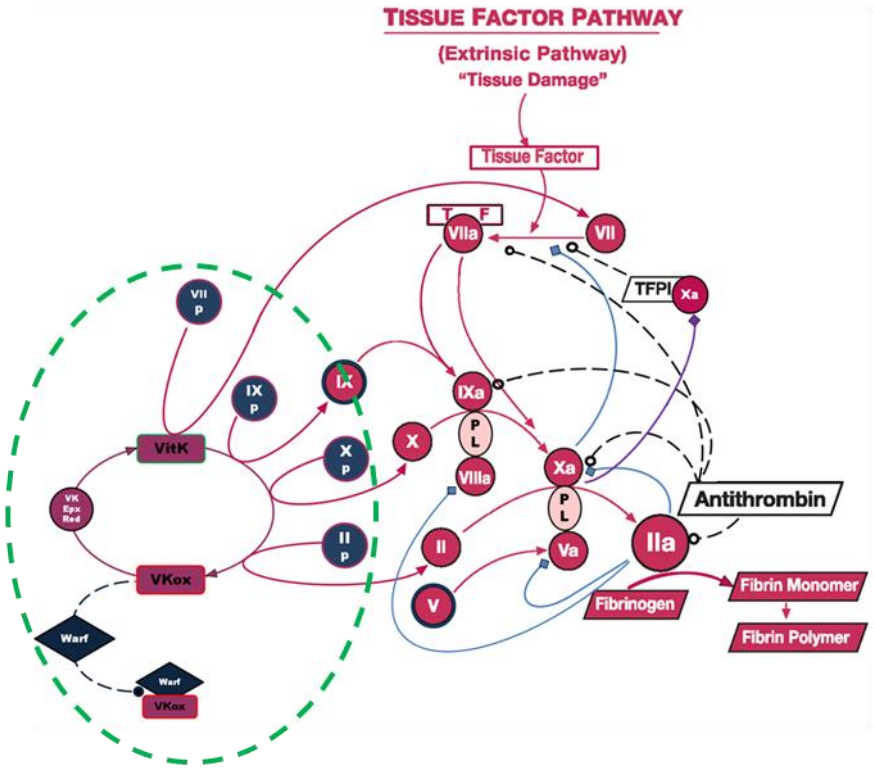



Figure 6.11 - Pharmacogenomic submodule integration: the previous physiological wiring diagram has been integrated with the warfarin related vitamin k-pharmacogenomics submodule (dashed green circle).

The resulting Stochastic Petri Net defines the aspired model for our research project (figure 6.12). In effect, it encompasses the most advanced knowledge about the described biological process, giving a cutting-edge tool to test the safety and effectiveness of a widespread treatment, as oral anticoagulation is. Besides, the model allows taking

into account the recent recommendations about the relevance of pharmacogenomics in a personalized medicine perspective. Above all, the possibility to simulate all these model parameter (marking sets and transitions rates) applying a huge variety of settings, makes such a model useful for real applications, at least in the context of clinical decision support.

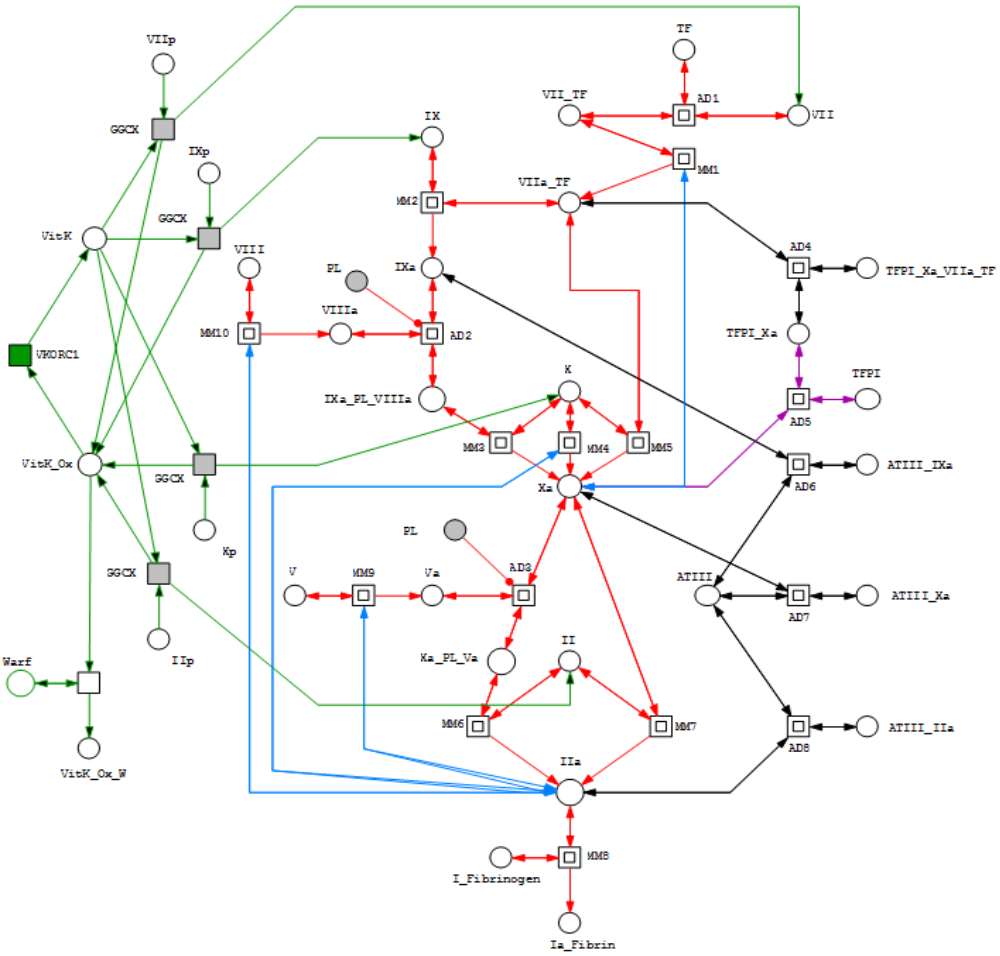



Figure 6.12 - Pharmacogenomic Model: the green subnetwork complete the Petri Net model concerning the pharmacological action of anticoagulation drug modulated by VKORC1 genetic variants (SNP polymorphism)

The resulting anticoagulation flatten Petri Net consists of 44 places, 52 transitions, and 163 connecting arcs. This time, we preferred to avoid any hierarchical notation for the description of the integrated

submodule, to emphasize the interaction of pharmacologic pattern with the directly affected coagulation factors.

### 6.5.4 Models Setting Assessment

Quantitative methods for modeling biological networks require accurate knowledge of the biochemical reactions, their stoichiometric and related kinetic parameters, and in particular, to deal with physiological pathway modeling, their initial concentrations of biochemical species and enzymes involved.

The interdisciplinary collaboration among the computer science institutions, biological labs, pharmaceutics and clinics, have made possible the creation of several public database e repository that enclose the information concerning to of various disciplines of natural sciences. Some of them are particularly structured and curated, and are regularly updated whit the new research discoveries.

The data contained in this DB are essential for the computational modelers, since they can obtain useful parameter and kinetic law data, short descriptions of the belonging pathway and web links to literature that consent them to investigate in different life science's domains without being experts [Li et al., 2010] [Swainston et al., 2010].

#### 6.5.4.1 Initial Marking of the SPNs

**(Un)Healthy Model**

To attribute a realistic amount of tokens that reside in the places assigned to coagulation factors in initial marking, we have referred mainly to Bioclinical labs and Pharmaceutical corporations that deal in products for the biochemical determination of coagulation protein levels:

Thermo Scientific brand of Thermo Fisher Scientific, Inc. USA.

R&D Hematological Division - Bio-Techne Inc. USA

Affinity Biologicals Inc. CA, U.S.

Enzyme Research Laboratories Inc. IN U.S.

Another resource widely employed for literature surveys is Pubmed med line algorithm [Darmoni et al., 2006]. It helped us to find in the huge literature production about haematological and cardiocirculatory disciplines, the most revealing publications. We defined a query consisting of key word list considering: all possible names variable for coagulation factor (e.g. for the Tissue Factor: Tissue Factor, Factor 3, FIII, Thromboplastin, etc.) and the investigated coagulation pathway (extrinsic coagulation cascade, Tissue Factor pathway, blood coagulation cascade, etc.); the most common methods to measure the protein plasma level concentration or activity (coagulometry, ELISA, spectrophotometry, etc.); all possible synonyms for haematological disorder (prothrombotic status, hemorrhagic condition, blood disease, etc); and many other words used singularly or coupled, enclosed in inverted commas ( " " ) or divided by AND. The result of query returned more than 2000 publications that were filtered for journal (e.g. Journal Impact Factor, 5-Year Journal Impact,guideline inclusion, etc) and scientific community relevance (Citation Number, e quando possible mediante il recente CitImpact score at http://www.biowebspin.com/pubadvanced/).

All values extracted from published works were integrated and compared with reference values (mean and range) of coagulation factors obtained from enzyme immunoassays (ELISA) performed in national hematological laboratories specialized on elderly coagulation factors titrations (see acknowledgments).

Most of values reported in literature and repositories comply with the standard for biochemical clinics, expressed in moles or grams per liter (mol/L and g/L, respectively). This continuous unit of measurement, in agreement with discrete stochastic simulation, must be converted in discrete molecules amounts.

This calculation is made possible by the following sequence of conversion formulas:

Let [*M*] the molarity, i.e. the concentration *c* of a solution expressed as the number of moles (*n*) of solute per liter of solution (*V*)

$$[M] = \frac{n}{V} = c \tag{6. 7}$$

It is possible to convey the concentration in terms of the Avogadro constant, $N_A$, which expresses the number of elementary entities per mole of substance and it has the value $6.02214179 \times 10^{23}$ mol-1 [De Bièvre and Peiser, 1992].

$$c = \frac{N}{N_A \cdot V} \tag{6. 8}$$

with $N$, the number of molecules, we can derive the number of molecules in our modeled fixed plasma volume of $1 \times 10^{-10}$ liters.

The average values obtained from these international sources were slightly refined with the advice of domain experts and specialized laboratories (see acknowledgments) to reproduce the average *in vivo* configuration of the coagulation factors in elderly patients.

**Pharmacogenomic Model**

The initial marking set of the *Pharmacogenomic Model* was adapted from the marking of previous model. The corresponding marking value of places affected by new modules (II, VII IX and X, linked by green arch), was replaced in the new analogous places, named with same roman number with a final "p" addition (IIp, VIIp, IXp and Xp) and the "contributor places" was zero marked. This has allowed us to make the marking of these new places subjected to fairing rate of transitions included in the pharmacological module. The molecules amount of places concerning vitamin k redox reactions was extracted from the previously mentioned enzyme data base, and the warfarin place W, was marked according to the real data drug amount (converted in number of molecules for fixed volume) administered during therapy, gained from pattern discovery investigation (Figure 6.13).

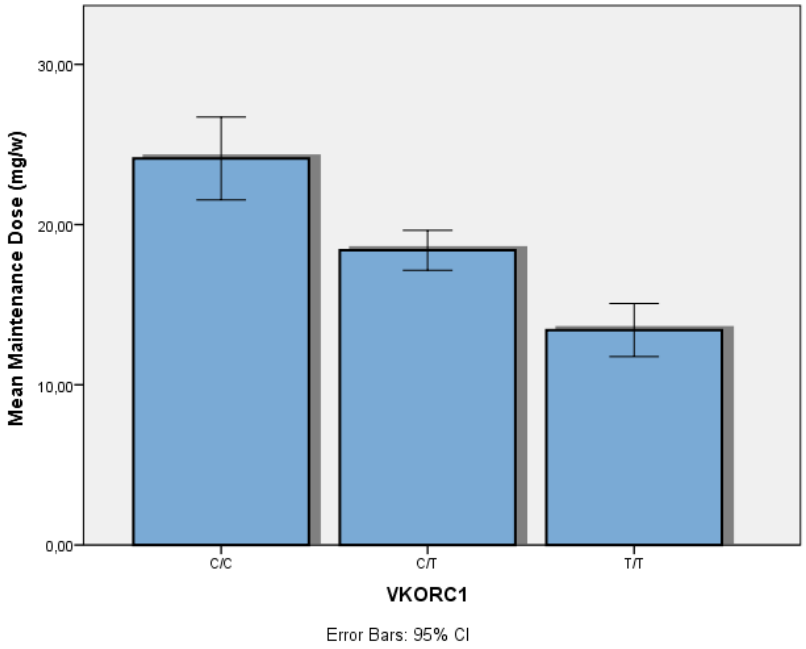



Figure 6.13 - Bar diagram of average genotype-related drug dose: the mean maintenance dose and related variances administered to different VKORC1 genotypic profileing during therapy (data obtained during KDD investigation)

Therefore the two models have been marked with the exact number of molecules involved in the process. In particular the initial marking ($M_0$), of *(Un)Healthy* HHSPN involves 12 places whose marking ranges from 75 to $3{,}01\times10^8$ molecules (the average total amount is $10{,}47\times10^8$ ) within a considered fixed plasma volume of $1\times10^{-10}$ liters. Similarly, the *Pharmacogenomic Model* has the same values for initial marking range even if the non-zero marked places are 14, adding VitK and Warf places to the marked list.

#### 6.5.4.2 Transitions Rates Estimation

The transition rates, concerning the kinetic parameter (kinetic constants), were investigated in several specialized database:
SABIO-RK [Wittig, 2012] is a reaction-kinetics database that contains *wet-lab* investigated rate laws and characteristic parameters for a large collection of bio-chemical reactions, including the values of the features

that characterize the experimental conditions (pH value, temperature) during the rate measurements;

BRENDA [Schomburg, 2013] is a comprehensive enzyme information system constituted by a collection of enzyme functional data available for free, obtained from about 5000 enzymes classified by EC number according to the Enzyme Commission list.

EzCatDB [Nagano, 2005], defined as the Enzyme Catalytic-mechanism Database, provides ligand annotation for enzymes in the PDB (Protein Data Bank) as well as literature information on structure and catalytic mechanisms.

MetaCyc [Caspi, 2012] is comprehensive and freely accessible resource for metabolic pathways and enzymes from all domains of life containing an about 1800 experimentally determined pathways derived from more than 30 000 publications.

BioModels Database [Li C. et al., 2010] is a repository included in The European Bioinformatics Institute facilities, consisting in large number of models collected from literature that are manually curated and semantically enriched with cross-references from external data resources. Additional features useful to deal with computational models such as generation of sub-models, online simulation, conversion of models into different representational formats, are included.

JWS On line [Olivier and Snoep, 2004] is a Systems Biology tool for simulation of kinetic models from a curated model database. It contains 85 curated models that can be accessed and simulated on line.

PyBioS (http://pybios.molgen.mpg.de), developed at the Max-Planck-Institute for Molecular Genetics acts as a model repository and supports the generation of large models based on publicly available information like data of the Reactome database..

In biochemistry, it is often need to describe the kinetics of complex formation (such as the dimerization of proteins, the binding of a cofactor to the enzyme, or a ligand to its receptor) described before.

In this case, the parameter less or most detailed in these DB was the *dissociation constant,* $k_{d,}$ since the common method applied to test this category of kinetic reaction consist in evaluate the inspected parameter after the system reaching of steady-state in saturation condition (i.e. only the reverse dissociation is possible).
The observed association rate constant, $k_{obs}$ is defined by the following equation:

$$K_{obs} = k_{off} + k_{on} \times [LL]$$

considering the measure units of the variables as
$K_{obs}$ in $min^{-1}$ ; $k_{on}$ in $nM^{-1}$ $min^{-1}$ ; [LL], labeled ligand concentration expressed in $nM^{-1}$.
As the concentration of LL increases, the observed rate constant or $k_{obs}$ should increase linearly (as opposed to the fixed constant $k_{off}$). Representing $k_{obs}$ as a function of the labeled ligand concentration, we can obtain $k_{on}$ and $k_{off}$ from the plot, and the $K_d$ were then calculated as

$$Kd = \frac{k_{off}}{k_{on}}$$

Thank this relation is possible to compute every omitted parameter of the double sense reaction in the web source consulted.

For Michaelis-Menten kinetic parameters estimation, we apply the more appropriated formula among which reported in sec 6.3.3 to compute every parameter not reported in the web source consulted.

Despite these sources of information are cured often happens that they have different notations for the unit of measurement (% discarge from normal values, uM / ml, and so on) was then necessary to convert all values to the unit of measurement that better satisfy the need to deal with a discrete molecular system.

To evaluate the transition rate of nodes (transitions) that mimic the enzymatic reactions we had to convert the Unit, the most prevalent standard measurement used in the cited web sources, into a more suitable for molecular scale.
The Unit of enzymatic activity, which corresponds is defined as "1 unit (U) is the amount of enzyme that catalyses the reaction of 1 nmol (nanomoles) of substrate per minute under standard conditions" [IUPAC, 2007]. For gaining usable values, we had to convert nmoles into number of molecules for our fixed volume, end transformed again into substrate molecules number per time units processed by the enzymes. The SI unit [NIST 2008] of enzyme catalytic activity is properly the katal, however, it is less commonly used in biochemical practice than the Unit. One katal is the catalytic activity that raises the rate of a chemical reaction by one mole per second otherwise 1 katal is the amount of enzyme that catalyzes the conversion of 1 mole of substrate per second (1 katal = 1 mol/s). Thanks to the conversion factor that correlates the two units of measurement, we know in fact that 1 Unit is equal to $1.667 \times 10^{-8}$ katal [NC-IUB, 1979] we were able to transform the usually used Unit into more useful values to estimate the transition rates of our model.

In those few cases where, the biochemical sources were not comply with our needs, we transform the most common continuous kinetic parameter ($k$) present in models repository or literature papers in the respective stochastic values ($c$) by the formula

$$c = k \cdot (V \cdot N_A)^{1-n}$$

where $n$ is the order of reaction an $N_A$ is the Avogadro constant.

Regarding to the transition rate of pharmacogenomic module, concerning VKORC1 and Wclr, we used again the data obtained by machine learning approach to therapy database that have revealed

pharmacodynamics properties of drug action in patients. In particular, to obtain the transition rates VKORC1 corresponding to different genotypes setting we elaborated the data showed in Figure 6.13 and, consequently, we have derived the Wclr (warfarin clearance) transition rate.
Regarding to GGCX transition, its values was extracted from the highly curated pharmacogenomic database, PharmGKB [Whirl-Carrillo et al., 2011]. It was the main web source employed to evaluate the reliability of data obtained by the KDD process. This pharmacogenomics knowledge resource encompasses clinical information including dosing guidelines and drug labels, potentially clinically actionable gene-drug associations and genotype-phenotype relationships. PharmGKB collects, curates and disseminates knowledge about the impact of human genetic variation on drug responses.

The parameters of each individual enzymatic reaction reported in the curated biochemical sources, when non-experimentally assessed, were commonly calculated with the accurate linear regression method suggested in Greco e Hakala [Greco and Hakala, 1979].

#### 6.5.4.3 Fine Parameter Tuning

The parameter fitting procedure for the complex biological system including model could be challenging, especially in a wide stochastic systems [Taflanidis and Beck, 2008]. The main effort consist in as non-linear programming problems (NLPs) where the objective is to find a set of decision variables in order to minimize or maximize a given cost function subject to a set of dynamic and algebraic constraints [Jia, 2009]. Moreover, to be effective, this requires a thorough knowledge of the system, besides the completeness of the information relating to the parameters to be tested [Ashyraliyev et al., 2009].
This procedure was performed, on account of the Snoopy capabilities to transform the stochastic net in a Continuous Petri Net and to export the corresponding network characteristics in SBML model format, whose

main feature is the portability to be used with different software. This skills allow us to perform a fine parameter tuning, with the aid of COPASI (COmplex PAthway Simulator)[Hoops et al., 2006]. COPASI (COmplex PAthway SImulator) is a software application that supports models in the SBML standard and carries out several analyses skills, including the parameter scan.

Parameter scan practice was quite similar to trial and error method, because, to be effective, it requires a thorough knowledge of the system, besides the completeness of the information relating to the parameters to be tested. So, for us, it was been not so problematic due to the extensive web data and expert knowledge of the biochemical domain. Many approach to address this issue were presented in biochemical applications, [Nobile et al., 2012][Yang et al., 2012], even directly for Petri Nets [Shaw et al., 2006], for signal transduction networks [Arisi et al., 2006][van Riel and Sontag, 2006] and even to accurately assess uncertainty in the system [Marino et al., 2008].

Most of them are, yet, performed in fairly simple systems ranging from few to a dozen reactions and their associated parameters [Yang et al., 2012][Gonzalez et al., 2007], also when more sophisticated techniques were employed [Barbuti et al., 2012][Rodriguez-Fernandez et al, 2006]. This evidence suggests that this method may be too demanding both in terms of adaptation to the model and computational cost. This evidence suggests that this method may be too applicant in terms of adaptation to the model of computational cost. This provides a certain consistency with our simple approach.

### 6.5.5 Models Configurations for Scenarios Generation

#### 6.5.5.1 (Un)Healthy Configurations

The healthy model represents a haemostatic system, based on elderly subjects, with a correct constitutive balance of the natural coagulation/anticoagulation processes. The healthy has been the reference model to compare the physiological behavior with a pre-inflamed and/or pathological time-dependent behaviors as well as the

emerging dynamics give rise to the drug induction on all changing status. The different responses to prothrombotic event exordium and pharmacologic induction were monitored by time-evolution patterns of the most relevant output molecule of the extrinsic cascade, the thrombin (IIa).

The intensification of different potential-pathological conditions on "healthy marking" , were simulated by increasing the trigger factor amount in TF places from the "normal" (healthy) molecular values represented by 1510 molecules, to ten-fold increase steps giving rise to 15100, 151000 and 1510000 different place-specific markings, mimicking the medium, large and huge thrombotic event. We apply this change based on bioclinical evidence of a local increase of Tissue Factor in cardiovascular pathologies correlated to atrial fibrillation (the prime cause to be undergone to anticoagulation therapy) which can lead to a local thrombus development, as reported by Hayashi and Wysokinski [Hayashi et al., 2011][[Wysokinski et al., 2004].

The pre-inflamed status is mainly reproduced doubling the number of "normal" molecule in VIIa_TF place (from 75 to 150), the pre-activated active forms of trigger stimulus [Monroe and Key, 2007][Girard and Nicholson, 2001], resulting as a moderate boosting factor in a prothrombotic event onset [Chu , 2006].

It is proper to remember that the thrombus formation is a local events (not a systemic event)[Furie and Furie, 2007], and is named improperly "pathological" or "disease condition", even if it is *potentially* pathological in case of its release in cardiocirculatory network

Particularly notable is the representation of the healthy *intrer-individual variability* obtained by varying within normal range the tokens number of main vitamin k dependent factors (VII, IX and X) and factor VIII, which upstream affect thrombin generation. Indeed, the variability was mainly assessed varying the factor IX and X magnitude within normal range, because, taking part to active complexes formation, they are responsible for the propagation phases (burst of thrombin generation) of coagulative signal.

#### 6.5.5.2 Pharmacogenomic Model Configuration

The overall setting according to the normal range were differently mixed for the generation of wide gamma of therapeutic scenarios employing to them up to five different amounts of blood thinner (ranging from $4.00 \times 10^7$ to $2.83 \times 10^8$ according to the patients' drug intake classes statistically (and clinically) reveled by data set analysis. The resulting fine modulation of the drug effects on several potential disease conditions have been further tested according to the three pharmacogenomics phenotypes of patients.

We have to point out, to justify the drug induction on healthy subjects (conduct that might seem ill-considered), that the PT test performed on peripheral blood sample is not able to capture the local thrombus formation. Its duty is, rather, to evaluate the systemic preventive effect of anticoagulation drug on thrombi development; therefore it measures the preventive hemorrhagic effect sorted by drug intake, consisting in a prolongation of normal PT time, (normally ranging from 12 to 14 seconds ⇒ INR values of 1.0 ± 0.1) usually around 34-38 seconds, to obtain an average INR value of 2.5 ± 0.2 [Ansell et al., 2008][Hirsh et al., 2001].

All setting employed to model simulation were extensively evaluated in the corresponding section.

It is important to note that the reaction rate constants has not been modified from healthy to unhealthy model, as well as in real system; nonetheless, the reaction firing rates change considerably between the two models because of the initial marking, promoting prothrombotic events.

In addition, it must be emphasized that both models does not represent a system in a whole, but the local effect of a procoagulant stimulus (which eventual anticoagulant drugs can counteract with their inhibitory action).

### 6.5.6 Structural Analysis for Preliminary Model Validation

Model validation aims basically at assess the reliability of the constructed model. There is no doubt that this should be a prerequisite before performing more sophisticated analysis [Baldan et al., 2010][Heiner et al., 2004].

Topological properties of the network represent the potential dynamic behavior of the systems; therefore, an effective method for model validation is to assess the compliance to main law of mass conservations, the ability of the system to perform all reactions reproducing the biochemical paths, evaluate the parallel execution of transition firing sequences, and identify functional modules. All this property and motifs have to fulfill the underlined known biochemical process and the resulting emerging static characteristics my reveal new patterns.

These model validation needs shows again how a valuable computational biology approach requires for a scientific cooperation, as for the validation process consist also in attaching a biological meaning, or provide a plausible “natural” interpretation, for each of the properties resulting from structural analysis.

Therefore, *net invariants* which may be computed using Petri Net theory have real-world meaning as the flux modes and conservation relations in the biochemical reaction network. The biological explanation should connect each of the invariants to corresponding mechanisms involved in maintaining homeostasis [Hardy and Robillard, 2008].

briefly the two informal definitions for the invariants in Petri nets can be summarized as:

P-invariant is a set of places in which the weighted sum of the tokens remains constant for all firing transition belonging to it. P-invariant may be interpreted as the law of conservation of mass or energy,

T-invariant is a set of transitions that always may fire to reproduce a given initial marking. They may represent a cycle that consistently occur originating a steady state.

To execute this invariant analysis we implemented in Snoopy the software Charlie [Rohr et al., 2010]. Charlie will detect the T-invariants and P-invariants present in our models. It directly communicates with Snoopy and its output can be imported, as “extra” task in Snoopy.
Baldan and Heiner asserted the relevance of T-invariant in model validation because they are strictly related to flux paths, which represent the potential way of evolution of the system. Similarly even the P-Invariant analysis can reveal interesting insight in the biochemical network, particularly in signal transduction pathway [Napione et al., 2009].
We extracted from the Petri Net tutorial [Blätke, 2011] the table regarding all property available from Charlye software to interpret its outputs (table 6.1)

The session analysis performed by Charlie on (Un)Healthy model has returned the following output

(Un)Healty Model Analysis:

*number of places: 36*

*number of transitions: 46*

*number of arcs: 142*

A summary of all structural analysis performed is exhibited in the txt table reported below.

----------FINAL RESULTS----------

| PUR | ORD | HOM | NBM | CSV | SCF | FT0 | TF0 | FP0 | PF0 | CON | SC | NC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N | Y | Y | Y | N | N | Y | Y | Y | N | Y | N | nES |
| RKTH | STP | CPI | CTI | SCTI | SB | k-B | 1-B | DCF | DSt | DTr | LIV | REV |
| - | - | Y | N | N | Y | Y | N | - | - | - | N | N |
| SSI | RNK | SCCS | SECS | | | | | | | | | |
| - | - | 16 | 36 | | | | | | | | | |

---

The session analysis performed by Charlie on Pharmacogenomic model has returned the following output:

## The Pharmacogenomic Model Analysis:

*number of places: 44*

*number of transitions: 52*

*number of arcs: 163*

A summary of all structural analysis performed is exhibited in the txt table reported below.

```
----------FINAL RESULTS----------

 PUR  ORD  HOM  NBM  CSV  SCF  FT0  TF0  FP0  PF0  CON  SC   NC
  N    Y    Y    N    N    N    Y    Y    N    N    Y    N   nES
RKTH  STP  CPI  CTI  SCTI  SB  k-B  1-B  DCF  DSt  DTr  LIV  REV
  N    N    Y    N    N    Y    Y    N    -    -    -    N    N
SSI  RNK  SCCS  SECS
 Y    31   18    42

-------------------------------------------------------------
```

| Main Properties | Property | Definition |
|---|---|---|
| Structural | PUR - Pure | There are no two nodes, directly connected in both arc directions. |
| | ORD - Ordinary | All arc weights are equal to 1. |
| | HOM - Homogeneous | All outgoing arcs of a given place have the same multiplicity. |
| | CON - Connected | A PN is connected if for every two nodes *a* and *b* there is an undirected path between *a* and *b*. |
| | SC - Strongly Connected | A PN is strongly connected if for every two nodes *a* and *b* there is a directed path from *a* to *b*, |
| | SCF - Static conflict free | There are no two transitions sharing the same pre-place. (Transitions involved in a *dynamic* conflict compete its tokens |

| | | |
|---|---|---|
| | CSV - Conservative | All transitions add as many tokens to their post-places as they subtract from their pre-places. |
| Behavioral | SB – Structurally bounded | Bounded in any initial marking |
| | 1-B - 1-bounded | All places are 1-bounded. |
| | k-B - k-bounded | All places are k-bounded. |
| | LIV - Liveness | Every transition is always firing |
| | REV - Reversibility | Initial marking is reachable again from each marking state |
| | DSt - Dead states | Exists a dead state if no transition can be enabled any more. |
| | DTr - Dead transition | Exists a dead transition if cannot be enabled in any reachable marking |
| Behavioral (according to structural motifs) | STP - Siphon & Trap properties | Every siphon includes an initially marked trap (Input places excluded). |
| | CPI - Covered by Place invariants | P-invariants covered if every place belongs to a P-invariant. |
| | CTI - Covered by Transition invariants | T-invariants covered if every transition belongs to a T-invariant. |

Table 6.1 – List of Petri Net properties abbreviation to interpret the Charlye software outputs according to topological analysis of Petri Net designed by Snoopy editor.

The two Petri Nets model are not covered by T-invariants, despite the assertion of Baldan and Heiner. T-invariants, indeed, can define subnets, and minimal t-invariants define self-contained subnets, which are always connected. These subnetworks are able to capture biological structural motifs.

The lack of this relevant biological property is ascribable to the specific typology of represented process and to modelers' assumptions. In fact, a closed signaling cascade, modeled to define a transient dynamic, should not have the possibility of returning to the starting point to avoid the re-induction of trigger signal, without the necessary biochemical "authorization".

Several minimal T-invariants are identified, but they correspond to macro nodes, interpreted as “micro-modular” elements of the network.
This data is compliant with general definition of signal transduction and gene-regulatory networks, such that the invariant condition cannot be interpreted as the steady state known in biology [Kestler et al., 2008][Barkai et al., 2001][Gomperts et al.,1999].

Instead, our models are both covered by P-invariants determining that the Petri net is structurally bounded and conservative. For the nature of the represented system, our pathway need therefore to be covered only by P-invariants because is a closed system, resembling a complete model of activity in a circumscribed physical space without real mechanism of mass transportation but only of signal flow.
Applied this concept to the model, it is hoped that networks containing substrates (places) that are related by a certain signal transduction pathway would emerge as the invariants. The conservation of tokens can be interpreted as a conservation of the same biological signal within the network. Hence places that are linked in a P-invariant are possibly linked by sharing the same biological signal.

# 7 Simulation Workflow

## 7.1 Adaptive Tau-Leaping Stochastic Algorithm

After excluding the possibility to perform one of the exact methods due to the elevated computing time required to solve this complex system, in particular in the highest burden phase during which each step taken about 1 picosecond to fire one transition using direct SSA method [Gillespie 1977], we consider to perform the stochastic simulation of our models with the adaptive tau-leaping method [Cao et al., 2007].
The tau-leaping method, during the simulation, consider larger time steps ( tau ) sacrificing the exactness for an imperceptible

approximation [Gillespie 2001]. The adaptive implementation of the algorithm, adaptively selects the step size based on an error tolerance, selecting sometimes a larger tau than the standard tau-leap (explicit), and reverts to the SSA method when tau-leaping could make an approximation mistake. This strategy further decreases the simulation running time, without compromising the slight degree of results approximation of the classical tau-leaping method. This algorithm is performable on StochKit2.

## 7.2 StochKit

StochKit2 is open source software for efficient stochastic simulation framework developed in the C++ language command-line driven Unix program. The software is downloadable from the URL: http://www.engineering.ucsb.edu/~cse/StochKit/
This software directly allows performing stochastic simulation, while remaining open to extension via new stochastic and multiscale algorithms and allows to developers to be incorporate into other software.
It provides command-line executables for running different stochastic simulation algorithms using the Gillespie's references method, Direct SSA algorithm, two variants of this exact method (optimized direct method and composition-rejection method) and two different tau-leaping approximated methods consisting in customizable constant leap (Fixed Stepsize algorithm) or adaptive leaping (Adaptive Stepsize algorithm) features[Sanft et al. 2011].

StochKit2 can be integrated with a java conversion tool "SBML converter" which, a command-line executable, converts a subset of SBML format text documents into the StochKit2 xml model definition format, with a structure reported below.

```
StochKit2 model definition .xml
<Model>
     <NumberOfReactions>n</NumberOfReactions>
     <NumberOfSpecies>n</NumberOfSpecies>
     <ParamenterList>
        <Parameter>
          <Id>K0</Id>
          <Expression>11.35</Expression>
        </Parameter>
        ……
     </ParamenterList>
     <ReactionsList>
        <Reaction>
           <Id>R0</Id>
           <Type>mass-action</Type>
           <Rate>K0</Rate>
          <Reactants>
           <SpeciesReference id="A1" stoichiometry="1"/>
           <SpeciesReference id="B1" stoichiometry="1"/>
          </Reactants>
          <Products>
           <SpeciesReference id="AB1" stoichiometry="1"/>
         </Products>
        </Reaction>
        ……
     </ReactionsList>
     <SpeciesList>
        <Species>
          <Id>A1</Id>
          <InitialPopulation>100</InitialPopulation>
        </Species>
        ……
     </SpeciesList>
</Model>
```

The choice to perform our research work on Stochkit software was than driven both to the selected method for stochastic simulation and for its compatibility with the Unix system, a prerequisite to run on a computer cluster.

## 7.3 High Performance Computing Aid

High Performance Computing (HPC) concerns the use of cluster computers, to improve dealing with massive sets of data to find information, make meaningful correlations, and solve very large and complex problems. The first TeraFLOP machine (ASCI Red), was introduced in 1996 for physic calculations [Christon et al. 1997].

The growing synergies between biomedical sciences and clinical technologies have generated exponential increase in the data production for the representation of the target biological systems. The knowledge extraction process from a huge amount of heterogeneous data has highlighted the need for interdisciplinary exchange among the various biomedical disciplines of clinical pathology, medical genetics, and diagnostic imaging and different computer science approaches such as data mining, digital imaging and the graph theory. The growing demand for computing power to analyze the complex systems arising from the data fusion suggest the support of supercomputers high-performance cluster and grid computing technologies. In the last years special attention is addressed to biomedical diagnostics with critical requirements in terms of processing power for knowledge discovery in field of Biomedical Modeling and Simulation (complex disease characterization, pharmacological applications and modeling of neural responses), in Computer-Aided Diagnosis (image or signals analysis and recognition) and in Clinical Decision Support for personalized medicine or pandemic management [Eugster et al. 2012][ Pereira and Freire 2010]. This technology, based on splitting of computing workload and boost processor resources through "clusters" which use redundant links to connect multiple machines therefore, can benefit researchers who wish to carry out huge amounts of calculations on large amounts of data and wish to obtain valid results in a reasonable time.

Due to the stochastic nature of the process represented in our model and the high level of interconnectivity of the reactions increase the

computational cost for running simulation. During preliminary tests, we performed on a Dell Studio XPS 7100 with AMD Phenom II X6 1035T 2.60 GHz and 8 GB RAM some simulation and the an average time, for single thread run with adaptive tau-leaping was over 8 hours. The huge number of settings to represent many different physiopathological and clinical scenarios in the two models proposed in section 6.5.2 allowed us to estimate the need to simulate at least 15000 runs, which on a single PC equipped with 6 logical cores would mean keep it fully engaged for about 20000 hours, more than 830 days. These pre-study results suggested us to exploit our computational work on a computer cluster.

The computer cluster that we have chosen is free access for institutional purposes and has no limitations in the number of computational cores CPU employed. It was developed in "ReCaS Project" (*Rete di Calcolo per SuperB e altre applicazioni*) which is a National Operational Programme for "Research and Competitiveness" 2007-2013 (NOP for R&C), PONa3_00052 Announcement 254/Ric., involving l' INFN (sections of Bari, Catania, Naples and a linked groups from Cosenza), the University of Bari "Aldo Moro" and the University of Naples "Federico II".

## 7.4 *Bc²S* Data Center

The Data Center INFN-Bari Bc²S (Bari Computer Center for Science) currently has about 250 compute nodes, overall corresponding to about 4000 CPU cores, each with 2-4 GB of ram. The node operating system is Scientific Linux 6. The network infrastructure between nodes, based on Ethernet technology, is able to guarantee about 110MB/s (1Gbits).

All the computational resources at Bc²S have a queue system installed, TORQUE (*Terascale Open-Source Resource and QUEue Manager*), and a job scheduler system, MAUI.

TORQUE is an open source resource manager providing control over batch jobs and distributed compute nodes. It is a community effort

based on the original PBS project and, with more than 1,200 patches, has incorporated significant advances in the areas of scalability, fault tolerance, and feature extensions [Staples 2006].

The MAUI Cluster Scheduler is an open source job scheduler for clusters and supercomputers. It is an optimized, configurable tool capable of supporting an array of scheduling policies, dynamic priorities, extensive reservations, and fair-share capabilities [Jackson 2001]. This scheduling manager system prevents to execute jobs interactively on any node outside the queue system without the "`qsub`" command.

The data storage subsystem is managed with Lustre, a type of parallel distributed file system, generally used for large-scale cluster computing. It provides high performance file systems for computer clusters ranging in size from small workgroup clusters to large-scale, multi-site clusters and it is frequently used in supercomputers [Weikuan et al. 2007]. Lustre is a *POSIX* file system (*Portable Operating System Interface*), a family of standards specified by the *IEEE* for maintaining compatibility between operating systems which defines the application programming interface (API), along with command line shells and utility interfaces, for software compatibility with variants of Unix and other operating systems [POSIX.1-2008].The data space available in the datacenter storage subsystem is 1650 TB SAN disk technology, shared between each of every working nodes and it is implemented with a RAID5 storage technology to increase both redundancy and performance.

Remote access to this computer farm is ensured by 3 dedicated nodes that will be addressed as “Frontend nodes”. In particular, this nodes are accessible by GARR network with *ssh* protocol at the address “`frontend.ba.infn.it`”.

Bc²S allows users to install their own software, and specifically provides 1TB of dedicated disk space for this purpose.

Furthermore, Bc²S is designed to exploit remote computational resources. To enable this feature the datacenter structure includes:

-3 gateway to export the resources to geographically distributed computing infrastructures

-2 HTTP e 2 XRootD servers to allow farm data access to geo-distributed machines.

A summary scheme of the architecture is shown in figure 7.1.

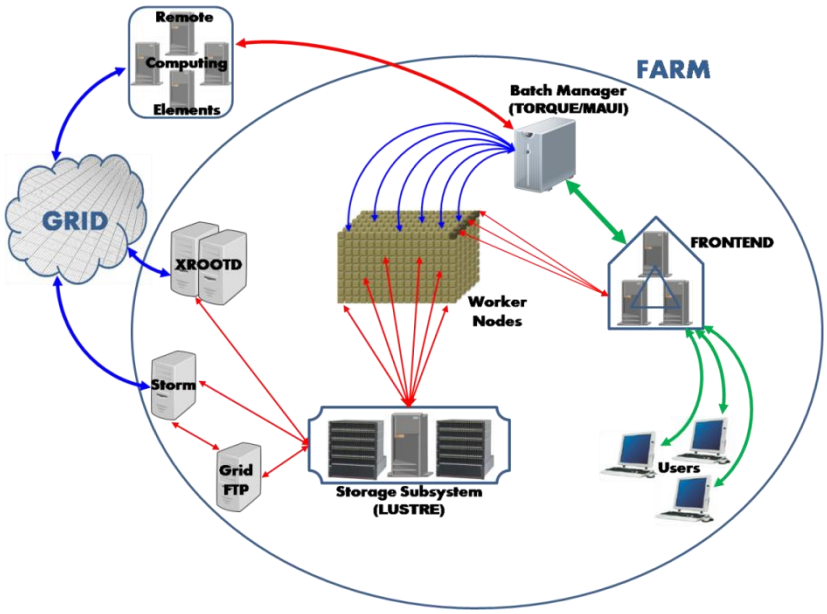



Figure 7.1- INFN BC2S Computer Cluster Architecture:. Green arrows correspond to user actions, blue ones to command instructions, and red arrows deal with data.

## 7.5 Submitting a Simulation on Bc²S

Once connected to Frontend Node using *ssh* protocol, we submitted the StochKit2 xml reference model (*Reference Model*) for each "macro-status" of patient and the two scripts, one for multiple scenarios generation and one to launch StochKit stochastic algorithm.

### 7.5.1 Scenario Generator Script

This script file provides the commands to produce the different xml model settings to represent all simulation scenarios for the patient's

“macro-status” (healthy, thrombotic event, therapy, etc.). At the end of the script was included the "qsub" command to send to the Batch Manager the $scriptfilename.sh to launch the implemented Stochkit stochastic algorithm on the specific model setting.

```
#!/bin/bash
numberRealizations=75
# change the values  you want to simulate for each species S or reaction
R
valueMol1=( ... ... ... )
valueMol2=( ... ... ... )
...
valueMolN=( ... ... ... )
valueReac1=( ... ... ... )

for Mol1 in ${valueMol1[@]}; do
  for Mol2 in ${valueMol2[@]}; do
    ...
    for MolN in ${valueMolN[@]}; do
      for Reac1 in ${valueReac1[@]}; do

      #define the model file name with different set of values

filename=Mod_Name_S1_${Mol1}_S2_${Mol2}_..._Sn_${MolN}_R1_${Reac1}.xml

      #define the launch script file name for each model

scriptfilename=Mod_Name_S1_${Mol1}_S2_${Mol2}_..._Sn_${MolN}_R1_${Reac1}
.sh

      #replace to default value (DVal), in the corresponding XML line,
the set of
      #values for each molecule or reaction to generate different XML
files to
      #represent each specific scenario of Reference Model
      sed "
      Nline s#<Express>Reac1_DVal</Express>#<Express>${Reac1}</Express>#
      Nline
s#<InitPop>Molec1_DVal</Initop>#<InitialPop>${Molec1}</InitPop>#
      ...
```

```
      Nline
s#<InitPop>MolecN_DVal</InitPop>#<InitialPop>${MolecN}</InitPop>#
      " Mod_File_Name.xml >$filename

      #generate the script name to launch the simulation of the
generated
      #scenarios adding the script file name in generic
LauchTauLeaping.sh
      sed "19 s#<filename>#$filename#g"
scriptLauchTauLeapingT55_I2500.sh
      >$scriptfilename

      #make the launch script executable giving the authorization to
execute it
      chmod 777 $scriptfilename

      #launch the script NReal times to obtain several realization of
each model
      #setting
      for i in $(seq 1 $numberRealizations); do
        qsub -q local $scriptfilename;
      done
      done
    done
   ...
  done
done
```

A comprehensive example for model file name is: "`Healty_Var_Geno_W_2.83_K46_0.00115_IX_6.73_X_10.24.xml`"
This model represent the simulation of the scenario of the variability of genotyped subject in healthy status during therapy (*References Model*: `Healthy_Var_Geno`), in which was changed the amount of FactorIX and FactorX within the normal range (`IX_6.73_X_10.24`), and was administered the highest drug dose (`W_2.83`) to a specific genotype, consisting of the specific reaction constant (`K46_0.00115`).
Therefore, for each *Reference Model* a set of scenarios could be represented as:

$$S = \{(v_1, v_2 \ldots . v_n) | v_1 \in V_1, v_2 \in V_2 \ldots . v_n \in V_n \}$$
$$= V_1 \times V_2 \times \ldots V_n$$

This implies that the overall amount of simulated settings for each *Reference Model* depend on the number of setting for each variables changed. That is given by the cartesian product of the same variables:

$$|S| = |V_1| \times |V_2| \times \ldots |V_n|$$

The diagram 7.2 below, schematizes the process to generate the whole relevant scenarios for our research study.

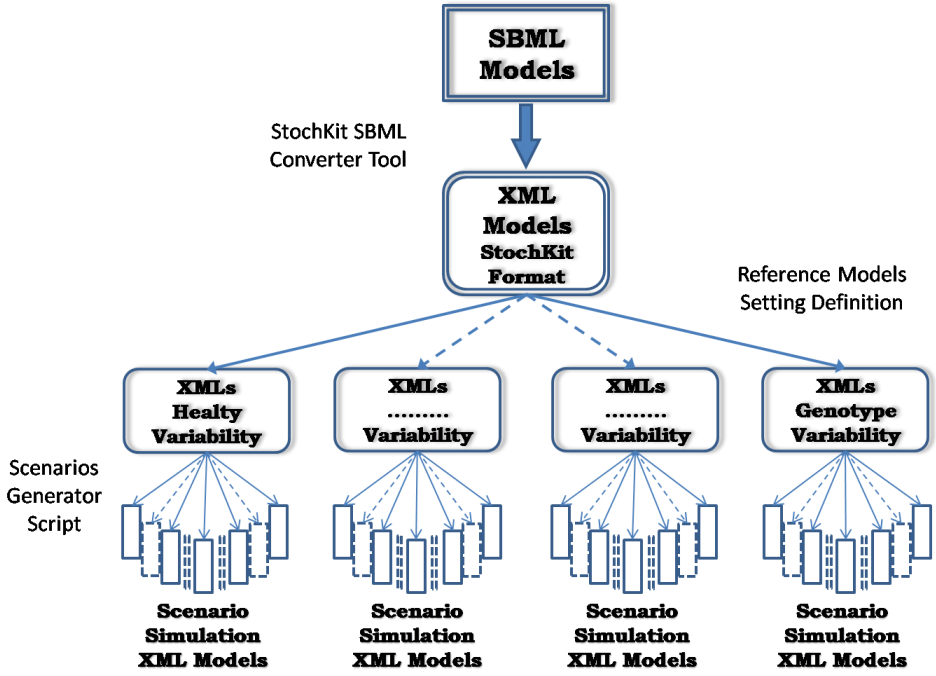



Figure 7.2- Scenarios generation process: The SBML models obtained exporting the Petri Net models is converted into StochKit xml format by the converter tool. Than they are modified in marking and transition rate to generate the Reference Model corresponding to the general health status and administered therapy. Hence, each general model was set to mimic all representative physiopathological and clinical conditions during simulations.

### 7.5.2 StochKit Launch Script

The *StochKit Launch Script* allows executing each scenario generated by *Scenario Generator Script*. This command defines the customizable parameters for simulation, outputs, and advanced features. We implement this script with a command to generate a unique random seed for algorithm initialization at the "`seed_value`" line.

```
#!/bin/sh

#PBS -N Reference Model Name

# Max execution time requested for one simulation
#PBS -l walltime=50:00:00

# Node number requested for one process
#PBS -l nodes=1

cd ${PBS_O_WORKDIR}

echo 'start'

# Job ID assignment for each process
string=${PBS_JOBID}

# Seed value definition for each stochastic simulation initialization
seed_value=${string:0:${#string}-19}

# StochKit Stochastic Algorithm Launch and Output Commands
./tau_leaping -m nomefile -r 1 -t 60 -i 2500 --keep-trajectories --out-
dir RUN_RESULTS/nomefile_${PBS_JOBID} -f --label -p 1 --seed $seed_value

echo 'end'
```

The parameters for lunching the stochastic algorithm on StochKit are listed below.

Simulation commands:

- `r <number of realizations>` specifies the number of realizations to simulate

- t <simulation time> indicates the length of the simulation in the arbitrary time units implied by the propensity functions in the model.
- i <number of intervals> specifies the number of evenly-spaced time intervals to save the states of the system, keeping the time-specific amount of all molecules.

Final data will shows a time series constituted by <number of intervals>+1 time points. The value ∅ specifies data will be saved at the end-point.

Output commands:

--keep-trajectories indicates to the solver to keep data from each individual trajectory. Will create a "trajectories" sub-directory in the output directory and one file for each realization.

--outdir<output directory name> specifies an output directory name

-f typing "--force" gives the instruction to the driver to overwrite the existing output directory with new data.

--label print column labels trajectories output files

Advanced commands:

-p <number of processors> tells the solver how many CPU to use. By default, the driver automatically selects the number of processors to use based on the number of processors on the computer. We preferred to use only one core due to both the highly interconnected nature of the system and the type of algorithm employed [Xu et al. 2010].

--seed specifies a seed for the random number generator. By default the seed is chosen randomly with random number generation Rand() from C (<arg>), but we decide to implement this features extracting the seed values from job identifier (JOBID) to obtain a unique number for each run. With the instruction ${string:0:${#string}-19} we extract the first numbers that compose the alphanumeric jobid to use as seed value.

## 7.6 Simulation Outputs

The output files structure consists of Text file with values separated by TABs: for N molecules in the model, the file contains N +1 (1 time features) columns and the rows are equal to the number of evenly-spaced time intervals saved, indicated in the `-i` parameter of the simulator launch script. The values contained in cells correspond then to the amount of each molecule in saved state of the system.
For each run (`numberRealizations` ), StochKit has generated a result folder named `filename_jobib` (`Mod_Name_S1_${Mol1}_S2_${Mol2}_..._Sn_${MolN}_R1_${Reac1}.xml_JOBID`) containing the results, in turn, the file "trajectories". Exploiting the data organization at the time of Frontend files submission to perform the simulations, it was possible to recognize the membership of each trajectory to a particular run of a particular model setting. Then, grouping the results obtained from the individual execution setting, we got the whole run results for each setting of the *Reference Model*.
Actually, to simplify further analysis, the grouping process consisted in categorize all trajectories of each molecule, for all scenarios generated from the corresponding *Reference Model*. Through pattern-recognition process based on regular expressions (`regexp`) applied to results folders, we extracted the information related to the trajectories of each scenario and we used them to index all the trajectories of each molecule in order to extract their behavior in the all simulated scenarios

# 8 Computational Results

## 8.1 The rationale of experimets pipeline

To obtain a valuable results in a such a plethora of scenarios that can be simulated, it is valuable to design an experiments' pipeline.
We will firstly describe the determinist result performed to validate the average temporal dynamics with the references works of Kanin and coworkers. Then we will firstly compare the effect of trigger signal increase that determines a procoagulant stimulus. Then we evaluate the healthily versus unhealthy configuration, to assess how the trigger factor affect the behavior of physiological pathway; next step is the representation of inflammatory profile that origins a procoagulant status. The wide variability of physiological system has been represented testing the influences of thrombin behaviors generating configuration according to the normal variability range showed by mapped coagulation factors. Joining together all this setting with the warfarin doses setting and pharmacogenomic feature we have been able to represent about 250 possible scenario that can be assumed by this system.

## 8.2 Model Validation: Deterministic Results

The next step after performing structural analysis of the network, conducted for preliminary validation on the most basic requirements for a correct representation of biological, is guided by the need for further validation of the model before starting a session of simulations particularly computationally expensive, despite the HPC.
Thus, we compared by deterministic approach the dynamics of thrombin generation and fibrin compared to the work done by Kanin et al. [Khanin et al. 1998].

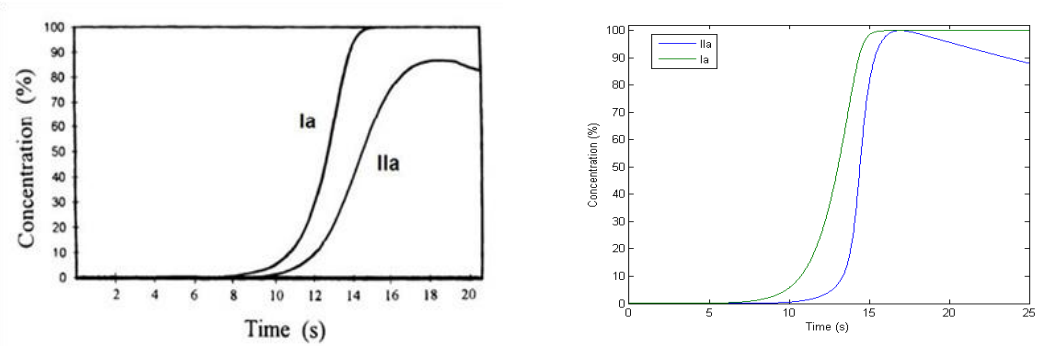
Deterministic methods for simulating the behavior of a systems biology model involve obtaining numerical solution to ODEs.

Well-developed methods for integrating ODEs, which are based on recursive procedures with adaptively adjusted step size, e.g., Runge-Kutta procedures, are fast, reliable, and applicable even to systems of high order.

As previously seen in sec. 4.3, a Continuous PN contains all the information needed to generate the system of ODE. Snoopy includes a series of ODE solvers that can be used to simulate the Continuous PN.

Most biochemical systems models are considered to be stiff, i.e. numerically unstable for simple solvers, because they contains reactions with very different rates (a part of them will be significantly faster than the others).

As the system of ODE for our model appears to be stiff, we chose a particular ODE solver, BDF (Backward Differentiation Formula), which allows correctly integrating the system and still having a quick execution time.




(a) (b)

Figure 8.1 - Deterministic result for literature model comparison: (a) Behavior of fibrin (Ia) and thrombin (IIa) in ODEs model of Khanin; (b) Deterministic solution of our (Un)Healthy Model marked with the corresponding concentrations of the reference model .

To confirm the reliability of our simulations, we compared the behavior of fibrin (Ia) and thrombin (IIa) in the mathematical model proposed by [Khanin et al. 1998]. with the results given by the deterministic ODE solver on our model (figure 8.1)

The two graphs shows clearly that both products have the same trend of Khanin's mathematical model. In our model we consider the whole

consumption of the substrates, as we simulate a closed environment comparable to what happens in an in-vitro essay during a PT test.

The deterministic approach is useful to compare the average behavior of our model to literature, but it cannot highlight other important characteristics that cannot be seen in the average behavior, such as the variability. These problems will be overcome with the stochastic approach.

## 8.3 Stochastic Simulation Approach

Deterministic approaches are very easy to use, and there are very efficient methods for numerically solving the ordinary differential equations. However, it is well known that there are situations where limitation of the deterministic approach can hamper capturing important properties of the system under study.
Most important limitations of deterministic approach to general biological modeling are as follows:
Although efficient in terms of computational cost, the deterministic solution is not accurate for systems that contain low-rate reaction, related to species occurring in small molecular quantities. If a system contains small amounts of some species, then representing these species by their mean population during molecular reactions, either can introduce significant errors in model predictions or change the model's qualitative properties.
When the behavior of a model is studied for parameter ranges close to bifurcation points that correspond to qualitative changes of systems dynamics then, as mentioned above, limiting to average values can cause overlooking important features of the system dynamics. Also for systems whose evolution significantly depends on initial conditions, such as bi-stable or multi stable systems, deterministic modeling may not adequately describe the distributions of system responses.

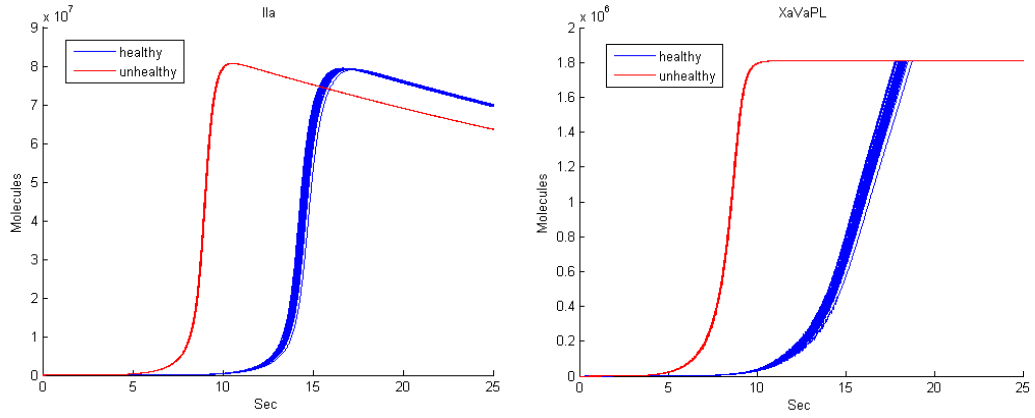
Deterministic approach ignores random variation of the system under investigation, which is given by the inability to capture much information about the molecules (such as their position, orientation and momentum) or the intervention of environmental factors, as physic-chemical changes.

Investigating stochastic variation of system trajectories is on one hand interesting by itself. On the other hand, it is well known that stochasticity in systems biology models has very important biological/evolutionary meaning [Geva-Zatorsky et al., 2006] that provides another motivation for our research.

Stochastic simulation approach is generally applied in order to study effects of random fluctuations in numbers of molecular species in systems biology models.
We conceptualize this biological variability to represent the intrinsic-randomness of tissue factor procoagulant pathway, to try to explaining at least part of the wide variability showed in bioclinical observations.

### 8.3.1 Simulation of Healthy vs. Unhealthy Configurations

After checking the validity of our method, we can employ INFN computer farm to perform many simulations. Stochastic simulation is not parallelized, but the farm allows us to run hundreds of simulations at the same time. In this way, we can obtain enough results to perform a statistical analysis and study the variability of both models.
We generated about 100 simulations of 25 seconds each, for both models, and we compared the trend of all the molecules, mainly focusing on thrombin behavior. As we can see from Figure 8.2, the amount of generated thrombin does not change from the healthy scenario to the unhealthy one (peaks y-axis values), because the total availability of its precursor (prothrombin, II) is the same.

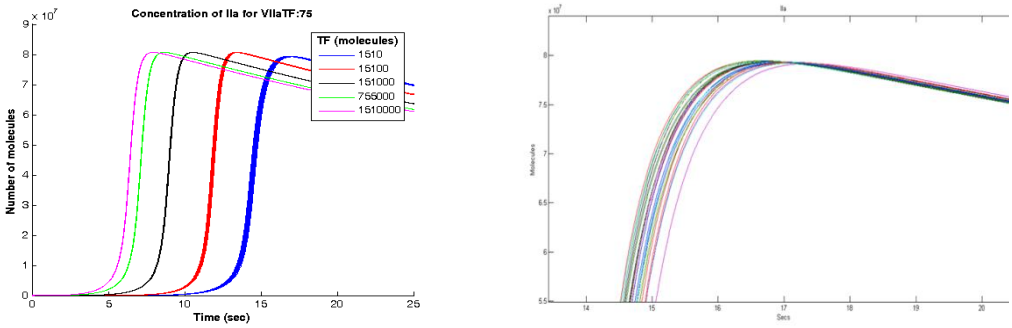



(a) (b)

Figure 8.2 - Comparison of the time-dependent behavior of generic healthy and unhealthy status: (a) Thrombin molecules (IIa) variable trends between healthy (blue) and unhealthy (red) configurations; (b) similar variability in Prothrombinase complex time-corse (XaVaPL corresponding to Xa_PL_Va places in the PN)

Differently, we can see a temporal anticipation of the growth in the unhealthy condition, which is due only to the change of initial condition. In fact, as in real system, the higher amount of Tissue Factor speeds up the cascade signal flow. We can also notice how the unhealthy setting shows a quasi-stable behavior on different simulations, as the healthy system has a higher variability (different slight time shifts). This effect is also detectable just upstream of the activation of thrombin, as in Prothrombinase, XaVaPL complex (Figure 8.2 b), which is the main responsible of the generation of active prothrombin (IIa).

To confirm the consistency of these observations we performed the same comparative experiment, testing only thrombin response to many different trigger factor values, as illustrated in Figure 8.3.




(a) (b)

Figure 8.3 - Different variability in Factor IIa production under several levels of TF induced stimulations: (a) time-series of Thrombin development (in frame are detailed the amounts of TF employed); (b) enlargement of the 100 simulation runs of healthy configuration.

We then evaluate quantitatively the observation concerning variability: the table 8.1, considering the mean peak times and their standard deviations of time series in Figure 8.3, highlights the differences (0,15 vs. 0,01 seconds for healthy and unhealthy, respectively).

| Configuration | Mean Peak Value (mol) | SD. Peak Value (mol) | Mean Peak Time (sec) | SD. Peak Time (sec) |
|---|---|---|---|---|
| **Healthy** | 7.9345e+007 | 5.9821e+004 | 15,42 | 0,15 |
| **Unhealthy** | 8.0691e+007 | 1.7668e+003 | 9,56 | 0,01 |

Table 8.1 - Means and standard deviations of peak value (expressed in number of molecules for the fixed volume) and peak time (expressed in second), in healthy and unhealthy model settings.

The lower variability showed by the simulated subjects underwent to medium prothrombotic TF stimulus, which overall augment the total amount of molecules, does not seem to be attributable to the main physics statistical law that proportionally links the fluctuation of

stochastic model to the overall amount of molecules in the systems, as in:

$$f \propto \frac{1}{\sqrt{n}}$$

Considering $n$ as the number of molecules, a higher amount of molecules allow to system to show less variability. However, in our system the maximum increment of molecules not exceeds 1/100 of the total amount; so such a slight variation does not affect the fluctuation index enough to substantiate the computationally observed differences in the variance measure.

This phenomenon it seems directly attributable to the influence of increment of the trigger factor at the beginning of the process. Immediately, in the early phases of the process, the increase in the number of molecules affects the firing rate of the few succeeding reaction (se the initiation model in sec chap). This events influence the overall probability density distribution for the entire process, which is concretized in leading to a higher rate for the procoagulant reactions. Since the rate of the inhibitory reactions is less affected, they will be disadvantaged (they will fire with a lower frequency in comparison), thus reducing the noise, and randomness, on the main cascade.

The biological meaning of this outcomes can be seen in a intrinsic phenotypic variation [Sato et al., 2003][Peng et al., 1994], which in turn fit into a wider context of phenotypic plasticity, a peculiar feature of biological systems [Kaneko k, 2012].

Besides, the intra-variability seen in Figures 8.2 and 8.3 the general behavior of thrombin formation does not differ significantly from the previous deterministic result.

### 8.3.2 Additional Feature to Unhealthy Configuration: pro-inflammatory and procoagulant status

The relationship between inflammation and coagulation has been more than extensively proven by biochemical investigations [Petäjä, 2011][Choi et al., 2006], but this aspect very often have not been taken into account by authors.

The source of tissue factor may be different in various inflammatory situations. In atherosclerotic plaques, macrophages produce abundant tissue factor at their surface; on plaque rupture, there is extensive tissue factor exposure to blood; and more recently the role of TF originated in microparticles [Garcia Rodriguez et al., 2010][Lechner and Weltermann, 2008] which arise from a inflammatory conditions, affecting further the prothrombotic stimuli [Puddu et al., 2010][Dignat-George et al., 2011]. Moreover, these literature indications provide an additional further support to the humoral initiation of the coagulation process, as assumed in our model.
Than we have repeated the simulation runs performed in previous section modifying the marking in VIIa-TF places, just in the first set of reactions at the beginning of the model, to show the proinflammatory and procoagulant status that can exist in a subject. Often they do not express a sick phenotype, i.e. they cannot be assessed a priori especially in elderly [Krabbe et al., 2004].

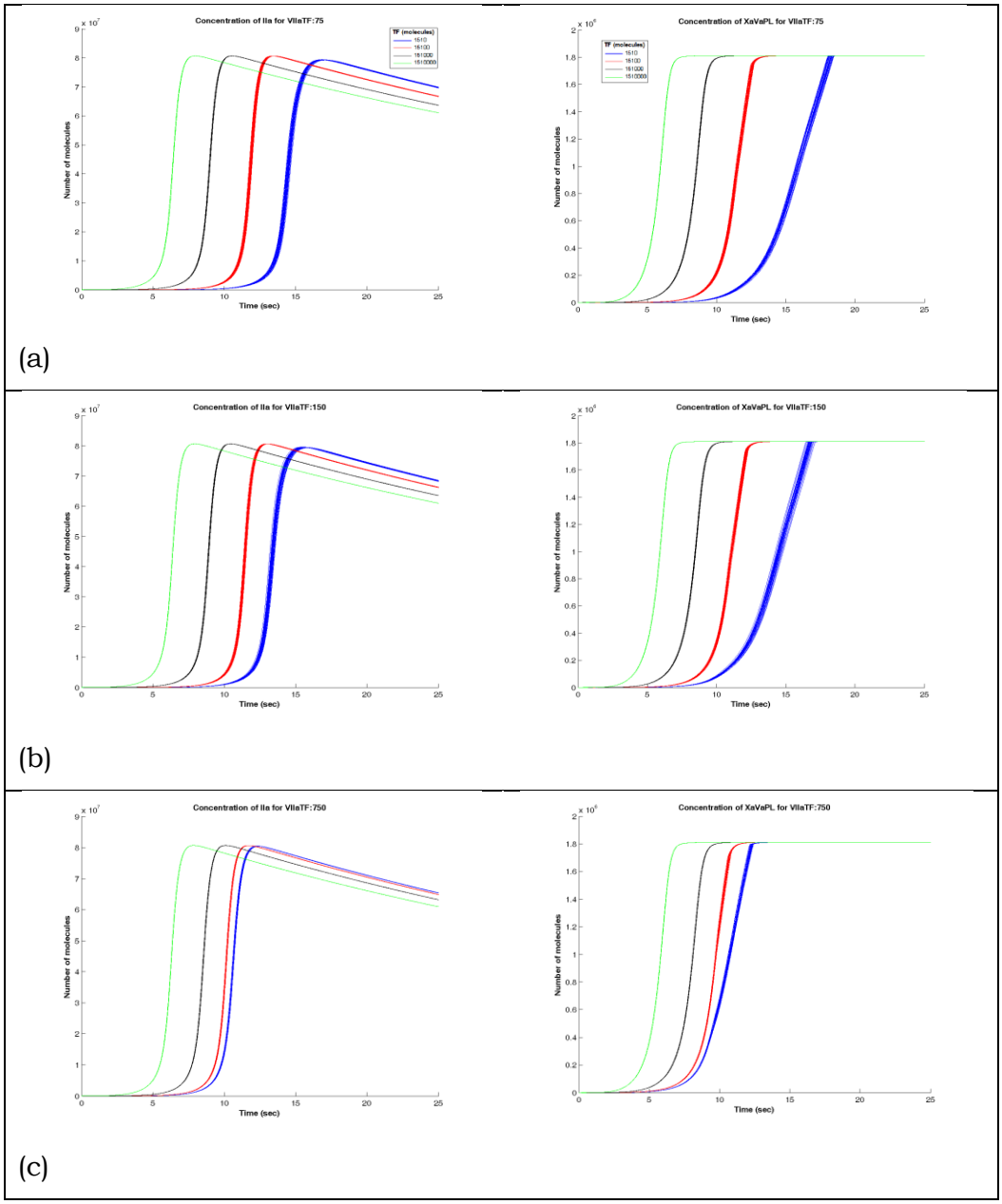



Figure 8.4 - Procoagulant induced by preexisting VIIa-TF: In the three tables' line are displayed the different tokens number set in the VIIaTF places in relation to the molecules range of the trigger stimulus (blu: 1510; red:15100; black:151000; green 15100000 as reported in legends) affecting IIa (left) and XaVaPL (right) production.

The results qualitatively exhibited in Figure 8.4, in agreement with clinical and laboratory observations about the pre-activated status of

coagulation inducible by inflammation and VIIa-TF [Thiruvikraman et al., 1996][ Berg and Scherer, 2005], the temporal shift IIa production induced by enhancement of this pre-activated form in under different quantities of Tissue Factor. In the left side of line (b) of Figure 8.4 we can observe a slight temporal shifts of peak time of IIa production corresponding to two fold enhancement 150 molecules) of VIIa-TF amount; in line (c) the this effect is boosted by a 5 fold increase in the same VIIa-TF place.

Physiologically this phenomenon mimics an increased risk of developing blood clots, proportional to previous pronounced inflammatory conditions. Alternatively, similarly, it can be interpreted as a more pronounced effect of the procoagulant signal of the trigger factor on procoagulant status, which in a therapy perspective is not an underestimating feature. Note that the temporal shift is less marked with increasing amount of TF, as the strength of this factor tends to level the pre-existing conditions.

Similar qualitative observations can also be found in the behavioral patterns formation (and activation) of the most significant complex of the coagulation cascade, the Xa_PL_Va. On the right panels of Figure 8.4, for the same configurations applied for IIa molecules development, are shown similar effect on Xa_PL_Va time courses behaviors. Note that the effect on this molecule, not only produces a contraction of time production, but also accelerates the formation process, event expressed by the change of the curves inclination corresponding to VIIa-TF amount variation (with equal TF induction).

### 8.3.3 The Extrinsic Physiological Variability Representation: The Healthy Inter-patients randomness

If the effect of the intrinsic randomness is represented by the stochastic mathematical modeling of biochemical processes, the *interpersonal* variability of this (and many others) physiological pathways must be appropriately modeled by several configurations of the described model.

It is known, or at least easily realizable, that most traits are genetically variable in most pathways, is not equally renowned that other element affecting the variability of coagulation factors are connected to a mix of personal and environmental characteristic, which manifest in the healthy status.
The causal molecular polymorphisms have been accurately identified in coagulation process, giving an insight into general understanding of the biomolecular processes that influence coagulation phenotypic variations [Boekholdt et al., 2001].
The non intuitive features that influence the coagulation status are connected to the ABO blood group, age, sex, and dietary intake which may vary, as well as genetic characteristics, both the levels of proteins that their activity [Soria, 2009][Mari et al., 2008][Pinotti et al., 2005].
These effects have been widely established both from the bioclinical applications [Yamagishi et al., 2012] both from the computational perspective [Corlan and Ross, 2001], which occasionally, as synthetic biology, is particularly helpful in the occurrence of arduous evaluations within the biochemistry framework.

We have been able to test this variability, concerning our modeled factors, perform different simulation of several healthy scenarios by means of configuring our model in most variable pre-place [Yamagishi et al., 2012]. The marking values assigned to them derived from the mean range values obtained by the investigation on cited several sources.

In the time-series, arisen from the (*Un)Healthy Model* simulation, exposed in Figure 8.5 is exhibited only a portion of such a demanding estimation, lacking in the visualization of all possible coupled configurations. We have, in effect, employed three different setting for each of the five factors modified, which, considering about 75 runs for scenarios' configuration, has returns about 18000 time-series curves ($3^5$ x 75) for monitoring the extrinsic variability (inter-patients) factor IIa development.

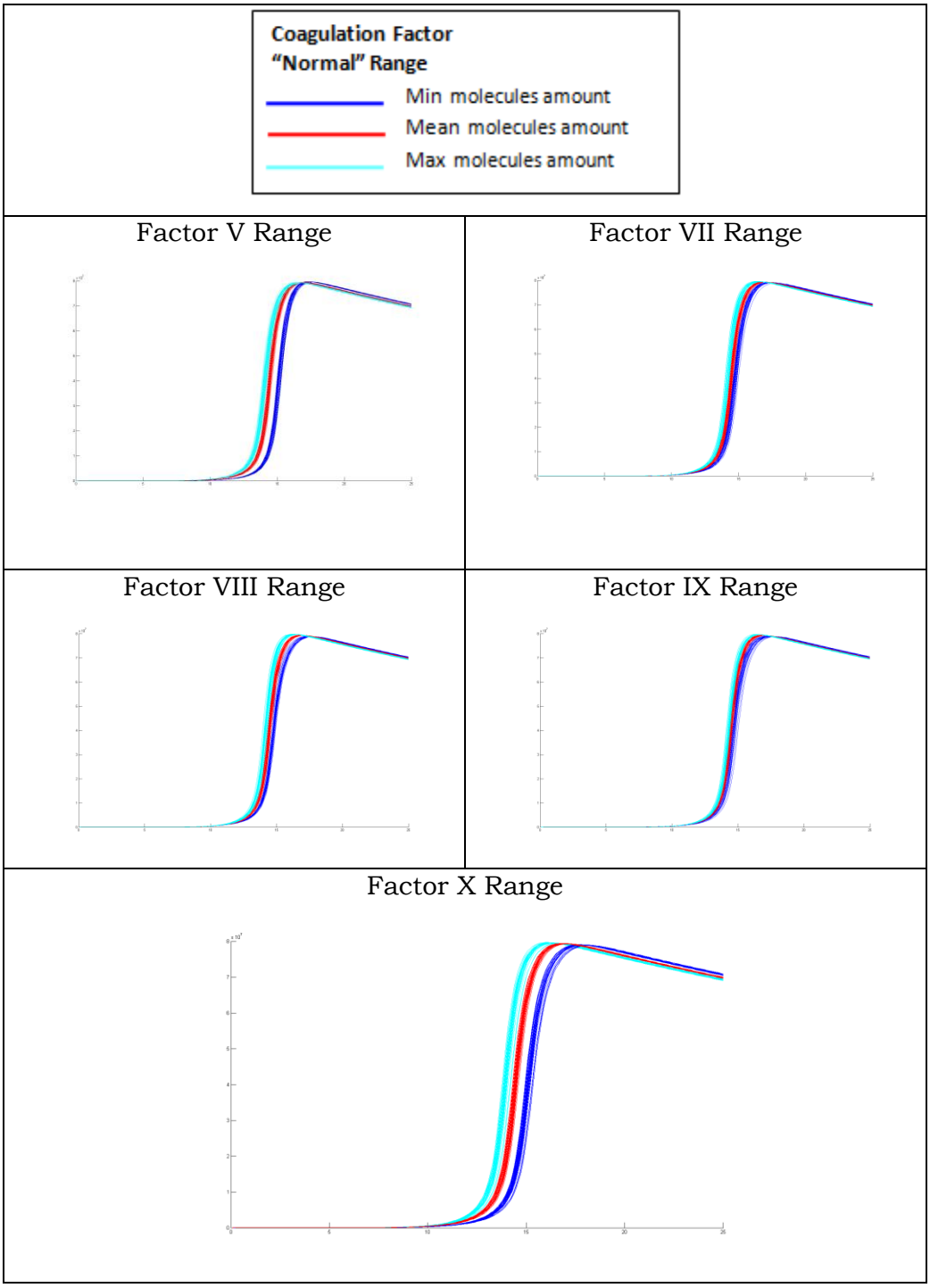



Figure 8.5 - Normal Variability in Thrombin generation investigated in (Un)Healthy Model: In each frame are reported the thrombin (IIa) time dependent trends according to the changing in corresponding procoagulant factor natural variation in healthy status set up within the values obtained from bioclinical sources (reported in subsequent table).

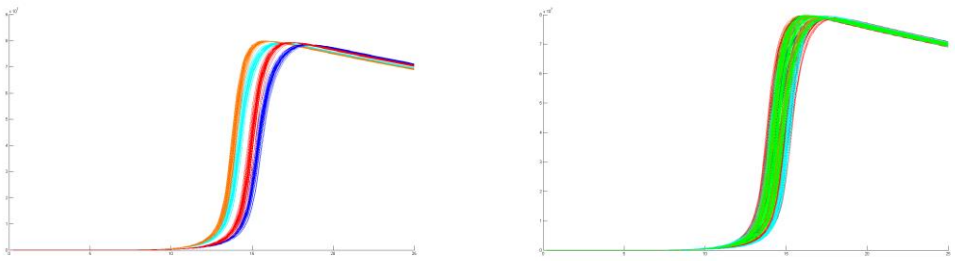
The table below explains the marking employed in corresponding pre-place marking setting.

| Factor | Minimum molecules (Blu pattern) | Mean Molecules (Red pattern) | Maximum Molecules (Light B pattern) |
|---|---|---|---|
| **V** | 1.360e+006 | 1.810e+006 | 2.260e+006 |
| **VII** | 4.515e+005 | 6.200e+005 | 7.525e+005 |
| **VIII** | 1.357e+004 | 1.810e+004 | 2.262e+004 |
| **IX** | 4.040e+006 | 5.380e+004 | 6.730e+004 |
| **X** | 6.140e+006 | 8.190e+006 | 10.240 e+004 |

Table 8.2 -Marking set ranges for coagulation factor “normal” physiological variability

Quantitatively, the result in figure show a marked variability of factor X, according to the bioclinically assigned normal ranges.

In order to decrease the huge amount of data that may be generated applying all set ups to every next scenario simulation, henceforth, will be considered only the factor IX and X normal ranges in subsequent unhealthy and therapy computational analysis.
The adequacy of choice to employ only the factors IX and X, in addition to their direct involvement in the chain of events aimed at the production of thrombin, is legitimate both by qualitative (figure  and quantitative analysis. The comparison between the variability generated by their two external range values (min and max) displayed in a single graph, respect to the overall variability generated by the remaining wide-ranging values (all left over factors together) supply the qualitative evaluation.



(a) (b)

Figure 8.6 - The thrombin generation variability produced by IX and X normal ranges (a) compared to the overall variability attributable to all remaining factors (b).

Factors IX and X are in fact able to represent the overall variability of model, without generating a significantly different panorama (t-test > 0.05) as table below reveal in a quantitative manner.

| Factors | Mean Peak Time (sec) | SD. Peak Time (sec) | *p* |
|---|---|---|---|
| **X and IX** | 15,31 | 3,55 | 0.189 |
| **Other Factors** | 14,65 | 3,41 | |

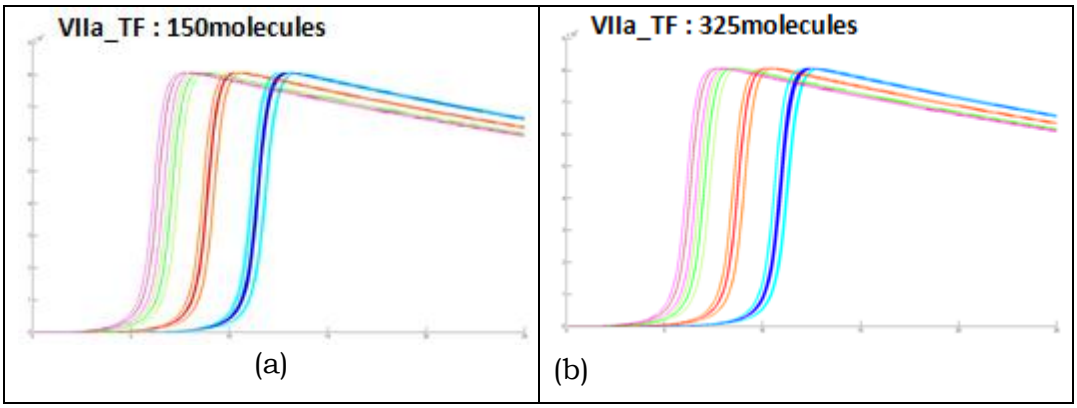
Table 8.3- The mean peck time and SD of thrombin time patterns generated from X and IX compared to the remaining factor variability

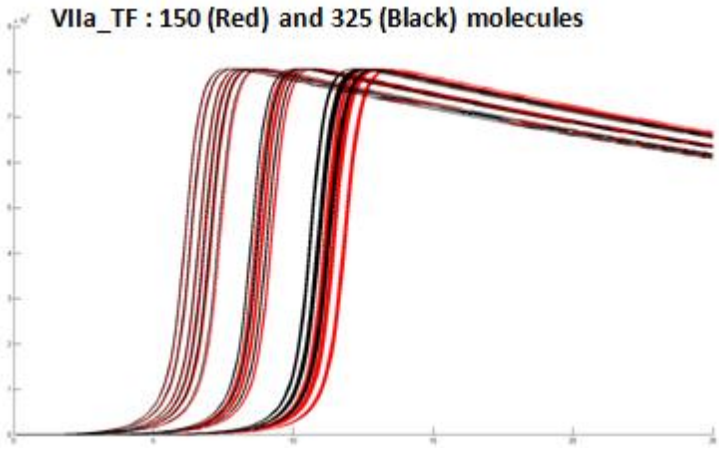
The relevance of taking into account these considerations in initiating anticoagulant therapy, has been repeatedly recommended by international journals [Ljung et al., 2013][Ichimura et al., 2012][Holmes et al., 2000], but it is very underutilized in everyday clinical practice.
The costs of tests apparently not motivated (remember that you do not show any sick phenotype) discourage this practice.

A synthetic biology approach accomplished through modeling, simulation and results validation, would provide to the medical community clinical evidence concerning the meaning of a portion of "a hidden variability" behind the physiological systems character, and is the root for a fine concept of personalized medicine.

### 8.3.4 The Unhealthy Inter-patients Randomness Assessment

The changeability of coagulation factors levels affect, obviously, the unhealthy phenotype, too. As we previously inquired into the predisposition of subject to thrombus formation, by means of the study of the VIIa-TF preexistent amount needed to disclose the prothrombotic effect, in this under section we have to demonstrate how the inherited randomness of the system rest on the patients suffering from coagulation disorders.






(c)

Figure 8.7 - Randomness in Factor IIa production with different procoagulant status and stimuli: (a) and (b) depict the effect sorted by induction with growing quantities of procoagulant trigger factor TF in different pre-conditions according to the VIIa_TF molecules number, including the potential variations of the joined factor IX and X (light color variations. (c) The overall effect of VIIa_TF IX and X variability under four different TF stimuli.

The outcomes obtained from this simulation give an insight into the different behavior of pathological condition, hardly detectable in biochemical laboratory, and almost completely unclear to clinicians.

Nevertheless this condition may be clinically supported, reflecting the hypercoagulability stimulus that arise locally during atherosclerotic plaque rupture [Reininger et al., 2010][Ardissino et al., 2000] to justify the tissue factor induction range; and the procoagulant state existing in a pre-inflammatory condition [Puddu et al., 2010][Dignat-George et al., 2011].

### 8.3.5 The Drug Therapy Perturbation on Extrinsic Coagulation Pathway

The introduction of a drug in the delicate haemostatic balance mechanism induces significant changes in the behavior of the pathway under investigation, which can monitored by the PT test that we use in this project as a virtual test to monitor the development of the major prothrombotic molecule, the active prothrombin (IIa).

We have conducted several preliminary tests on *healthy virtual* subject considering all the features up to now presented, able to provide a stochastic variability in the system behavior. The resulting IIa time courses are displayed in Figure 8.8.

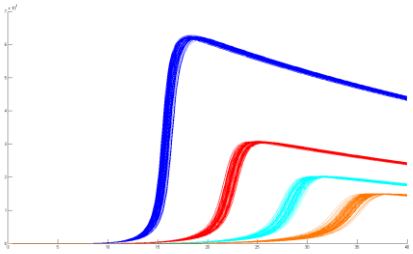
To give explanation for performing the anticoagulation drug "administration" to healthy model configurations, we recall that the PT test affords to detect only the systemic pro-hemorrhagic effect that prolongs the coagulation time of the overall circulating blood (to counteract everywhere the thrombus formation) end cannot give evidences on the extent of tissue factor coagulative signal.

**Warfarin Amounts (molecules: range mg/week)**

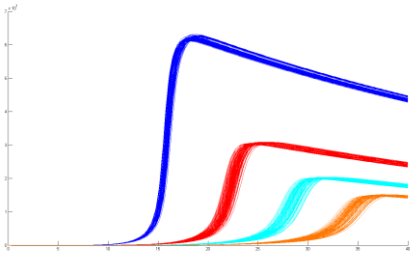
5e+007 : 5-10mg/week
10e+007 : 10-20mg/week
15e+007 : 20-40 mg/week
20e+007 : 40-80 mg/week

The triplicate lines coincide to corrisponding coagulation factor range undergoes to Warfrin dose

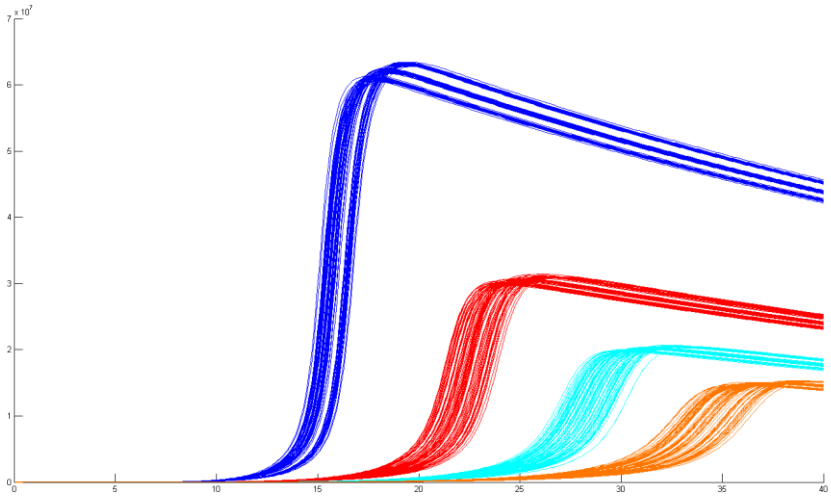
Factor VII Range



Factor IX range



Factor X range



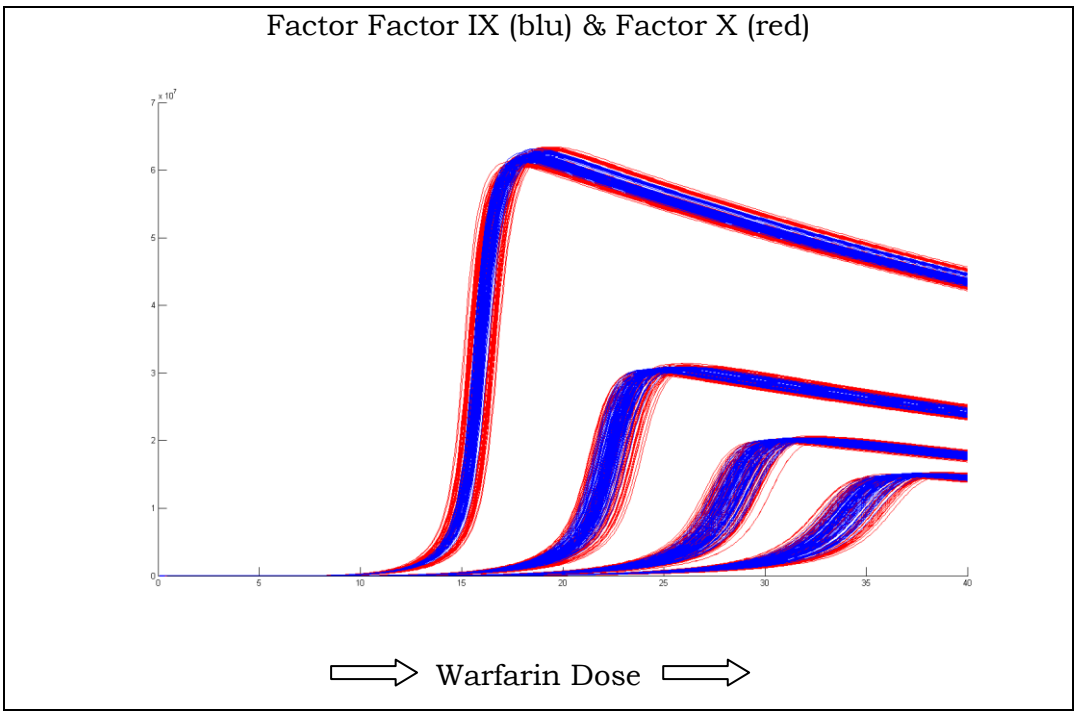



Figure 8.8 – Randomness in IIa formation according to normal coagulation factor ranges levels: Different colors match the warfarin dose markings reported in the legend; in the last graph is exhibited the cumulative variability produced by IX and X factors normal ranges undergo to drug dose increase

The prior observation that can be done is the extension of peak time range that stretches itself according to increasing warfarin dosage.
In an opposite manner to that observed in healthy vs. unhealthy comparison, in which the induction with the trigger molecule (TF) implied a decrease of the variability, the anticoagulant molecule, warfarin, leads to a noticeable increase in the time window concerning the peak time of thrombin production. This means that more anticoagulant is administered to a patient, to prevent the thrombus, more it will increased the preexisting variability of the coagulation mechanism, generating a wider spectrum of possible (random) differences in the effect of the drug on patient's blood fluidity.

The quantitative analysis performed on data relating to the cumulative variability range of IX an X reveals that this increment is quite evident and highly significant by means of ANOVA test corrected for multiple comparisons (Tab 8.4).

| Warfarin molecules #e+007 | Mean Peak Value (mol) | SD. Peak Value (mol) | Mean Peak Time (sec) | SD. Peak Time (sec) |
|---|---|---|---|---|
| **5** | 6.2694e+007 | 4.9221e+004 | 18,64 | 0,66 |
| **10** | 3.1920e+007 | 2.3360e+004 | 24,34 | 0,75 |
| **15** | 2.7293e+007 | 1.9661+004 | 30,06 | 0,82 |
| **20** | 1.4955e+007 | 1.2294e+004 | 37,52 | 0,88 |
| *P (ANOVAcorrect)* | <0.000 | | <0.000 | |

Table 8.4 -Average values and time peaks of different thrombin behaviors arisen from diverse drug stimuli

A clinical interpretation that can be given in respect of such evidence is that the specialist must face with a diverse response to therapy, especially in relation to patients undergo to intense therapy, i.e. conducted with high doses of medication.

As we investigate in our database, the simulation results are partially in accordance with the observed greater effort to remain in the therapeutic range by patients managed with medium and high doses. At low doses, as reported by the clinical experience, the challenge in adhering to therapy is due to the difficulty in self-administering small doses because tablets must be cut, and the average age of our data set is about 78 years.

### 8.3.6 The Pharmacogenomics perspective

Inherited differences in VKORC1 gene increase or decrease the amount of warfarin needed to inhibit the formation of the clotting factors in different patients, i.e. induces different responses to dose administered to patient with apparently very similar clinical and personal (age, gender, smoke habits, etc.) profile.

Therefore, the use of the genetic feature, has requested us to modify the corresponding transition rate in the model in order to represent this variable enzymatic component.

The first necessary step was to associate a phenotype to its reaction rate. Once again, the process of pattern discovery has provided us with a valid support. Considering the amount of dose required to the different phenotypes VKORC1 to stay within the assigned therapeutic range, so we were able to infer what were the ratio that were to have different reactions, considering as "normal metabolizer" the heterozigotic genotype.
The table shows the transition rate employed to configure the parameters in VKORC1 transition.

| Genotype VKORC1 | Transition Rate | Expected Effect | Assigned Color |
|---|---|---|---|
| **CC/GG** | 0.00068 | **Low Metabolizer** | Blu |
| **CT/AG** | 0.00100 | **Normal** | Red |
| **TT/AA** | 0.00115 | **Ultra Metabolizer** | Green |

Table 8.5 - Correspondence between genotype profile and transition rates of VKORC1 transition

In this section, we employed the exact dose patients' classes gained from the classification process performed during KDD exploration. It generate five class of achieved therapeutic dose (the *maintenance dose*, which elicit the patients to remain within the therapeutic INR range) as shown in table 8.6.

| M0-Warf 1,5mg/die (10,5week) Class1: 2,5-15,5 mg/w | M0-Warf 3,5mg/die (24,5week) Class2: 16-28 mg/w | M0-Warf 5mg/die (35week) Class3: 28,1-54,2 mg/w | M0-Warf 7,2mg/die (50,4week) Class4 43-54,7 mg/w | M0-Warf 9,85mg/die (69week) Class5: 54,9-84,5 mg/w |
|---|---|---|---|---|
| 4,00E+07 molecules | 1,08E+08 molecules | 1,44E+08 molecules | 1,80E+08 molecules | 2,83E+08 molecules |

Table 8.6 - Warfarin molecules values respect the patients' drug classes arisen from KDD analyses.

The explorative analysis consisted in reproduce all variable setting to
The analysis performed in this section resulting often challenging to visualize, given the complexity of multiple configurations to represent. The sources of variability included in this set of experiment are those observed in previous sections, consisting of:

- Natural Variability: we consider only the factor IX end X, counting each one for 3 values according to refined investigated sources
- Inflammation/Prothrombotic Status Variability: arisen from the VIIa_TF pre-activated molecule consisting of 3 values ranging from Normal (75 molecules); Poorly inflamed (150 molecules); and Inflamed (325 molecules)
- Prothrombotic Stimulus: represented by the amount of Tissue Factor exposed/released during the clinical event. It may assume 5 different growing values: the first four of which are clinically supported, ranging from Normal (1510 molecules) to a rupture of a wide atherosclerotic plaque (755000 molecules) while the fifth, was used mainly to assess the resilience of the system (1510000 TF molecules)
- Warfarin dose: according to the classification investigation we extract from our database 5 different class of therapeutic dose regimen, ranging from mean dose per weeks expressed in mg corresponding to: Class1=10,5 mg/week; Class2=24,5 mg/week; Class3= 35 mg/week; Class 4= 50,4 mg/week; Class 5= 69 mg/week.
- Pharmacogenomics profile: as we said consisting of 3 values as exhibit in the previous table.

We can still determine how many scenarios are needed to represent a comprehensive panorama of pharmacogenomic approach to anticoagulation: respectively $3^3 \times 5^2 \times 3 = 2025$ possible scenarios, each of which tested with 75 runs, corresponding to more than 150000 time series, taking about 8 hours each.

Due to this complexity and the feasibility of clinical interpretation we decide to split again the Healthy scenario undergoing to therapy and the Unhealthy one.

**Healthy Pharmacogenomics**

To demonstrate the influence, and the need, of genetic investigations in approaching the anticoagulation therapy, we configured firstly our Stochastic Petri Net to allow us to perform simulation to mimic the different effect of drug dosage on the three singular phenotypes. The merged resulting graph is show in Figure 8.9 below

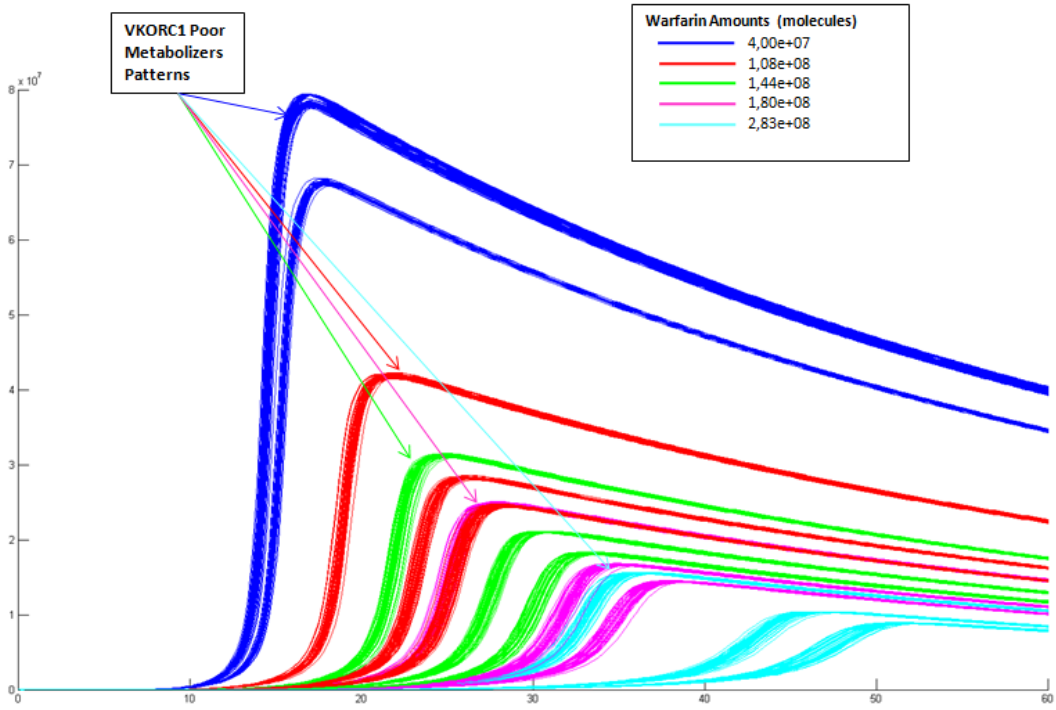



Figure 8.9 - VKORC1 influences on effective availability of warfarin to decrease the IIa production: It is shown 5 different concentration of drug corresponding to different colors and 3 joined time series patterns representing the different phenotypic profile of VKORC1 enzyme (transition). According with the Poor metabolizer indication, the resultant curves from the other genetic profile have a forwards temporal shift.

The description of such a complex chart needs a schematic approach to define the results. The observations can then be exposed as follows:

The blue pattern, representing the response to a slight dose of medication, shows the overlapping of the poor metabolizer curves with normal one. It can be interpreted as a low influence on the of the genetic component on drug effectiveness; even in the case of the curve relative to the “ultra-metabolizer”, the difference in the peak temporal average is not marked compared to the other Table 8.8. The dose employed in this simulation is actually low, and its ability counteract a thrombus formation is very slight in a typical patient (it was employed in patients suffering of kidney impairment of with particular health status profile)

The red patterns (warfarin about 24 mg/week) reveal an increment in the relevance of genetic effect on drug availability. The three different genetic assets contribute in most significant manner to the effect of warfarin. The peaks of the three curves show in fact a variation of about 7 seconds (from poor to ultra metabolizer).

Even in this case, as a reflection of the effect of blood thinner observed on the production of thrombin, genetics influences the amount of molecules.

The constant feature is the widening of the temporal shift of the time series of each setting; i.e. the repeated runs of each setting show a greater stochasticity related to the time point of maximum production of active factor II (IIa).

The growing evidence of the phenomena concerning the effect of genetic variance (inter-patient variability) and stochasticity (intra-patient variability) is appreciable according to the increase amount of blood thinner. The green pattern, magenta e ciano one, exhibit an increasing variability respect to the order presentation of, which can be quantified by the numerical representation of mean peak time for each genetic variation regarding to the factor IIa (table 8. 7).

**MEANS Pak Times -SD**

| | VKORC1 0,00068 | SD | VKORC1 0,001 | SD | VKORC1 0,00115 | SD |
|---|---|---|---|---|---|---|
| **W 0,4** | 17,1093 | 0,7389 | 17,2014 | 0,7389 | 18,0839 | 0,6773 |
| **W 1,08** | 21,9563 | 0,8838 | 26,4501 | 0,8838 | 28,5024 | 1,0638 |
| **W 1,44** | 25,144 | 0,9912 | 30,9584 | 0,9912 | 33,6139 | 1,1737 |
| **W 1,8** | 29,2735 | 1,0531 | 37,3632 | 1,0531 | 39,5931 | 1,3109 |
| **W 2,83** | 38,8294 | 1,2344 | 49,4176 | 1,2344 | 53,1960 | 1,6188 |

Table 8.7 - Quantitative evaluation of mean peak time referring to IIa evolution in warfarin therapy: Mean peak time an DS regarding to variability of active prothrombin production in compliance with VKORC1 transition rates (i.e. enzymatic reaction rate affecting the drug action)

From table 8.7 we derived the chart in Figure 8.10 that displays the general trend and relative variance index (bars).

Notably in this investigation, we identified some overlapping profiles of virtual patients that provide additional support to the reasons for the difficulty of managing this therapy. In particular the genotypic profile of low metabolizer responding to a 1,80e+08 molecules of warfarin, i.e. a patient undergo to a 43-54,7 mg/week regimen, (first magenta pattern around 25 five second of thrombin generation time) is quasi-overlapped with a very different therapy approach represented by the administration of a "hypothetical" therapeutic drug dose in 16-28 mg/week regimen on an ultra-metabolizer (last red pattern)

These two diametrically opposed genetic profiles have been proved to be able to generate the same therapeutic effect according to 25 mg of drug per week deviation. In percentage terms, this amount can be translated with a difference of about 50% on the hypothetical maintenance dose. From another point of view means that with the same dose administered, you can have a deviation of more than 10 seconds on prothrombin time test (PT test), almost one point of INR. In a context of narrow therapeutic range, such evidence is quite relevant.

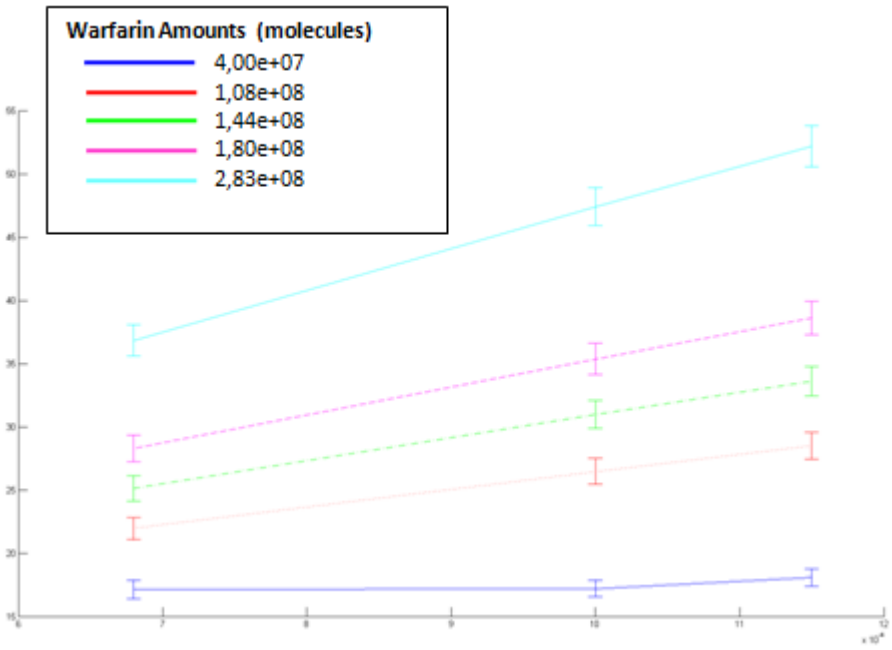



Figure 8.10- Trend representation of general impact of genetic features of VKORC1 genes, mapped into the Petri Net as the namesake transition.

**Unhealthy Pharmacogenomics**

The purpose of the administration of anticoagulants acting to the systemic level, as warfarin does, is to prevent the formation of a blood clot in any point of the cardiovascular system.

The local effect of counterbalancing carried out by the drug, designed to restore a normal "local" value of coagulation parameters, can be assessed only with a mathematical representation.

Our model is in effect able to give insight into the mechanism which guide the anticoagulation during a thrombus formation.

Therefore, the first investigation is aimed to evaluate the ability of different blood thinners to restore at normal range levels the peak time of thrombin production, facing with genetic variability. Considering an average thrombotic event consisting in the exposition/release of 151000 molecules of TF [Reininger et al., 2010] [Ardissino et al., 2000].

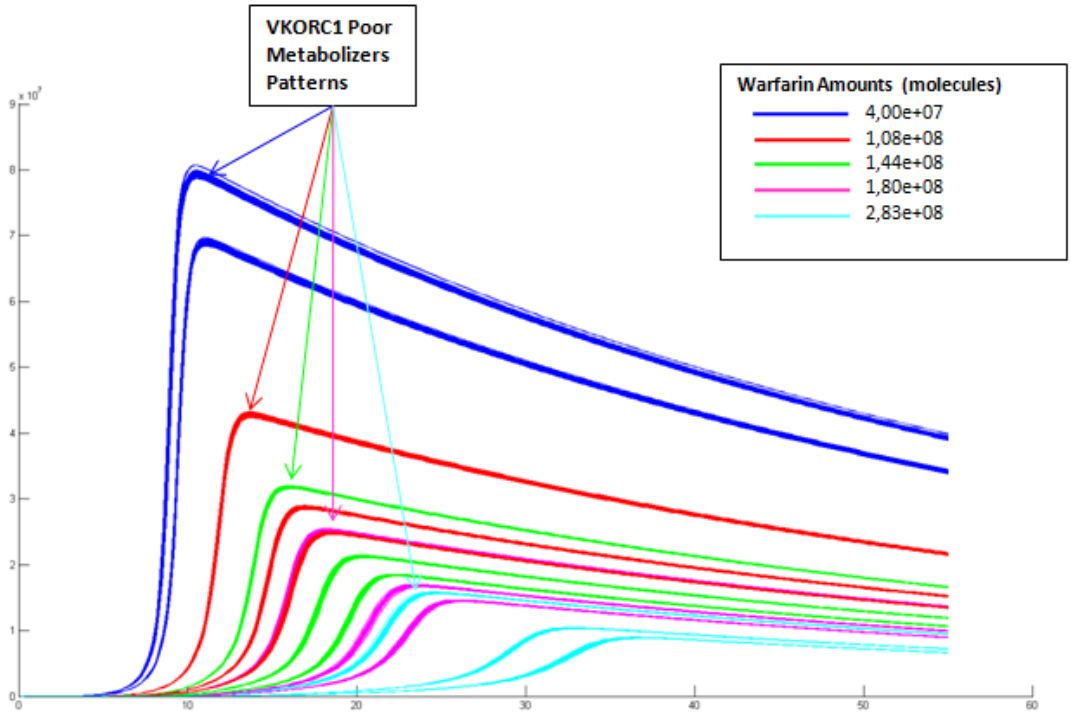



Figure 8.11 - The variability effect induced by VKORC1 pharmacogenomics on a thrombus development counteracted by different warfarin doses.

Similarly, to the healthy evaluation we can observe how different drug doses can counterbalance in the quasi-same way the thrombus development, resulting quite effective in a preventive therapy.

The experiment was generated considering a single prothrombotic stimulus, as the readability of a single graph comprising all possible stimuli of TF would have been compromised.

From the groups of time series, colored in accordance with the amount of warfarin that operates to rebalance toward coagulation times prothrombin (peak time) next to 15 seconds, it can be observed which are the most suitable concentrations to achieve this goal.

In the case represented, four configurations are capable of developing times of peak considered normal. We can observe that exist three different potential doses effective to achieve almost the same target. The average dose of 24.5 mg/week (red pattern), 35 mg/week (green), and

50 mg/week (magenta) mg a week is able to counterbalance thrombotic event in the corresponding individual genetic profiles. We can infer from chart that the 35 mg/week is suitable for a low metabolizer, even if a quite similar result can be achieved with 24.5 mg/week to a normal metabolizer. The same effect is instead produced by the "administration" to the virtual patient with 24.5 mg/week or 50 mg/week if he carries a ultra-metabolizer profile (red pattern) or low metabolizer (magenta).

Even for these set of simulation we collected the raw data and we analyzed to quantify the qualitative observation by means of table and chart representation.

**MEANS Pak Times -SD**

| | VKORC1 0,00068 | SD | VKORC1 0,001 | SD | VKORC1 0,00115 | SD |
|---|---|---|---|---|---|---|
| **W 0,4** | 10,52769 | 0,043246 | 10,538 | 0.000001 | 11,12567 | 0,102648 |
| **W 1,08** | 13,75244 | 0,10579 | 16,96994 | 0,092161 | 18,51151 | 0,094389 |
| **W 1,44** | 16,038 | 6,96E-15 | 20,37259 | 0,100555 | 22,38392 | 0,085954 |
| **W 1,8** | 18,35395 | 0,109839 | 23,71286 | 0,092514 | 26,19027 | 0,106639 |
| **W 2,83** | 24,83499 | 0,092162 | 32,97499 | 0,132855 | 36,71213 | 0,153333 |

Figure 8.12 – Quantitative evaluation of mean peak time referring to IIa evolution under warfarin administration on Tissue Factor stimulus: Mean peak time an DS regarding to variability of active prothrombin production in compliance with VKORC1 transition rates (i.e. enzymatic reaction rate affecting the drug action)

The graph diagram consisting of the data reported in the table is shown below (Figure 8.12).

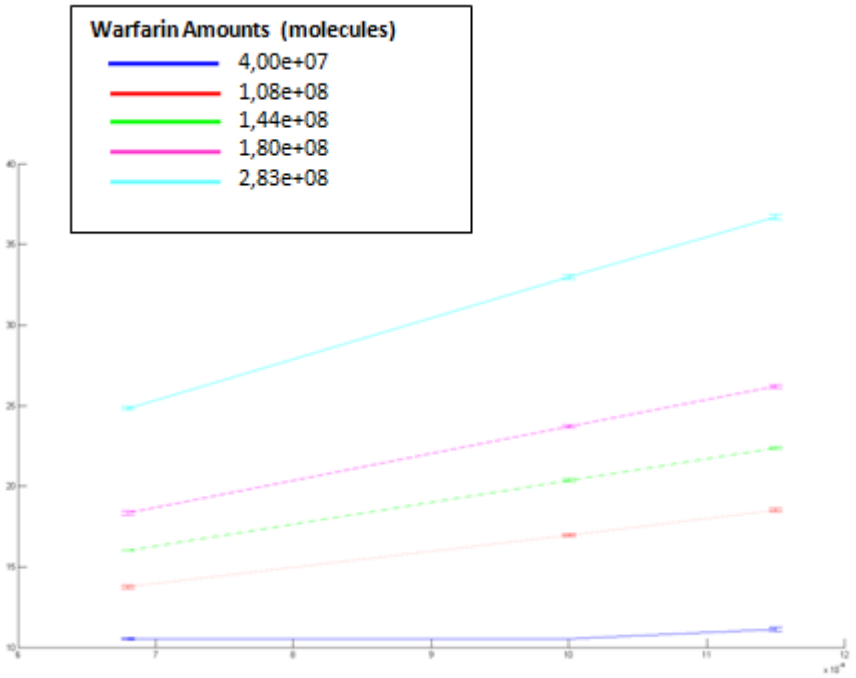



Figure 8.13 – TF induced Trend representation of general impact of genetic features of VKORC1 genes, mapped into the Petri Net as the namesake transition.

It is noticeable how the variance index (bars) narrowing under the hypercoagulation status, as we observe several times in the previous analysis.

In compliance with bioclinical observations, VKORC1 genetic polymorphism mapped in our SPN model is a valuable feature to characterize the patient profile and to manage significantly the therapy.

Moreover, detecting genetic variations in drug-metabolizing enzymes is useful for identifying individuals who may experience adverse drug reactions (ADRs) with conventional doses of certain medications. At the same time, according to the goal to safe and effective therapy by synthetic biology approach, made possible by the simulation of all the scenarios presented, we have been able to characterize the desirable

responses to therapy relating to concrete information supplied concerning the dosages to employ to achieve them.

### 8.3.7 Therapy Patient Profiling

One of the most significant results that we have gained, is the complete profile of the variables included in the model, the variability of a patient and the resulting set of views that present themselves to the clinician in the choice of treatment.
Combining all the features that give the extrinsic variability of the system, together with those highlighted by the stochastic representation (intrinsic variability) referring to coagulation Tissue factor pathway we have been able to represent the non-explained variability shown by therapy. In the figure below you can see how the time series generate a wide range of variability under our average dose of warfarin (35 mg/weak).

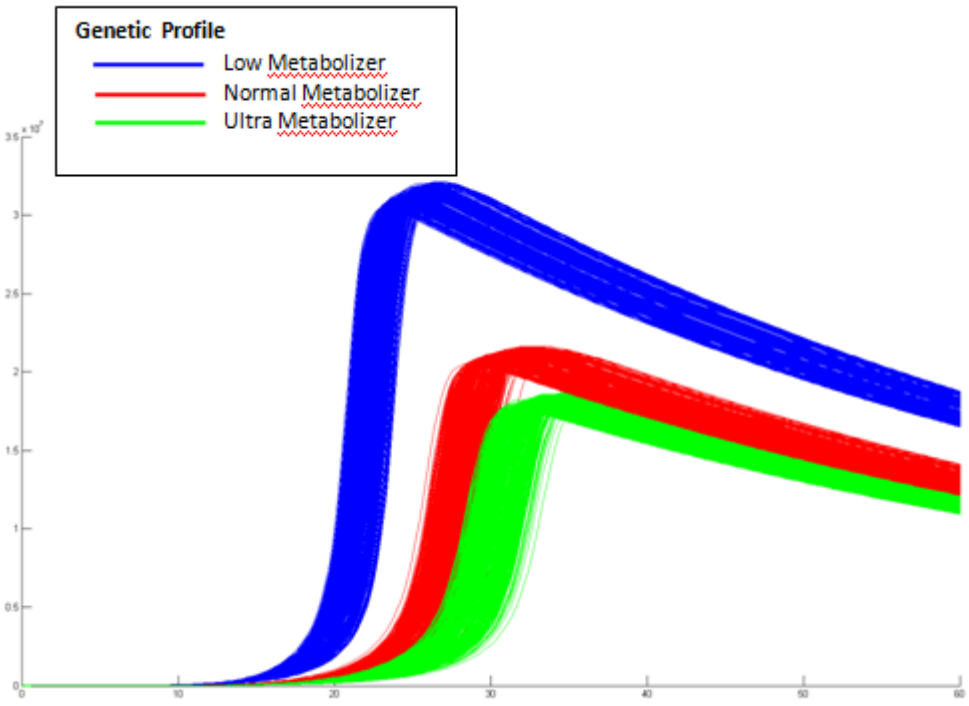



Figure 8.14 - Wide variability showed in a typical patient therapy approach: The three main curves represent the genetic profile response, included with variability induce by range coagulation factors and the potential

In the wide time courses of each genetic profile are include all possible configuration of factor IX an X, which can vary in all subjects within the "normal" range and the VIIa_TF procoagulant predisposition. In this simulation could be included even the other coagulation factor variation, but they would have been obscured by the larger variability shown by the predominant factors listed above.

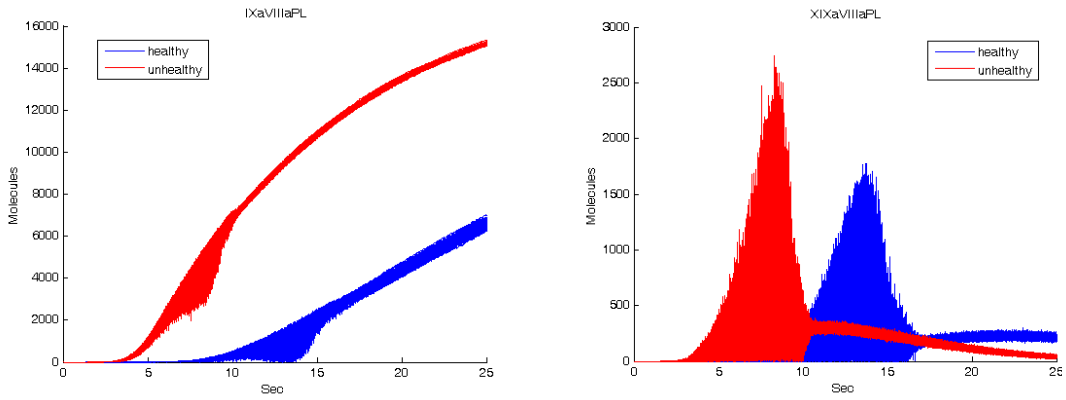
This view might be what is presented to the therapist at the start of treatment. The possible outcomes of response to the administration of the drug, so they are greatly affected by the features that can vary. Only a stochastic investigation may reveal how this scenario is almost uniformly distributed along the time axis.

Recalling that the aspirating therapy outcome is the sojourn of patient within the assigned INR range, which is affected to many factors, some

of which have been dealt in this section, our contribution is to be interpreted as a first step in advising a “clinical computational”, aimed at a gainful and mutual cooperation between the two disciplines to gain a safer and more effective management of anticoagulation application, in a translational medicine perspective.

### 8.3.8 Emerging Fluctuation Behaviors

Detailing more accurately this preliminary investigation, we noticed that there is another aspect which is captured only by the stochastic approach. The trend in molecular species involved in several central reactions (Figure 8.15) show strong fluctuability during critical phases of the simulation, when the system is at the peak of its activity, a feature that is impossible to detect using a deterministic approach, and very challenging even in a wet-lab experiment.




(a) (b)

Figure 8.15 - Stochastic fluctuability: fluctuability showed in average healthy an unhealtly vistual subjects: Trend in molecular number for IXaVIIIaPL (a) and XIXaVIIIaPL (b), shows a strong fluctuability

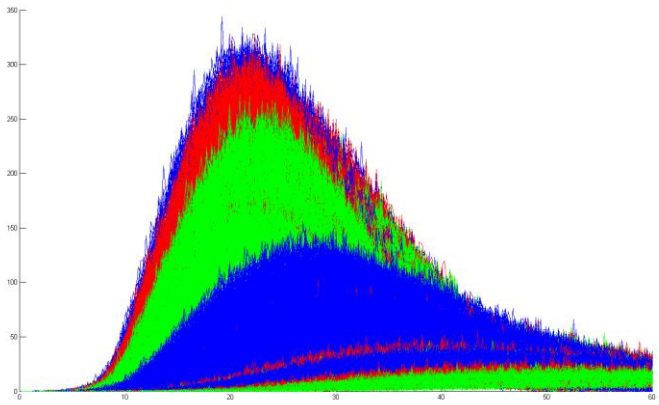


Figure 8.16 - Fluctuability in wide range variability: The same XIXaVIIIaPL molecules displayed in figure 7.15 (b)

These phenomena needs to be deeply investigated, probably by means of physics domains, but enforce the need to employ the stochastic approach both to capture them, both to evaluate the fine behavior of the system during the management of these low molecules number reactions.

# 9 Conclusions

Any computational approach oriented to investigate on a biological system, needs firstly the identification of the relevant information that compose the pathway, consisting of interacting elements and the types of relations that link them. the which can be represented as simply as possible in the computer's memory. This problem is well known in bioinformatics and has been studied extensively using existing knowledge representation techniques from computer science. Bioinformatics representations can be generally classified into three main groups as subsequently widely explained.

Database models, which define the way biological information is stored in a database. A database model may be derived from an ontology, in

which case it shares the same logical structure but must be realized in a particular database management paradigm (e.g. relational or object-oriented). Database models support simple types of query, often via a query language such as SQL (structured query language). However, they are usually employed as organized data source for performing complex network analyses such as pattern discovery.

Structural network models for analysis of the structural properties of biological networks. These models are usually based on a graph representation, with nodes standing for a certain type of entity (e.g. compounds in a metabolic network) and edges standing for relations between them (e.g. co-occurrence of two compounds in a metabolic reaction). A structural network model of a system will generally be held in the computer's memory rather than used as a database itself, and will be used for the execution of some algorithms on the network in order to discover a structural property of the system.

Quantitative analysis models for simulation and analysis of the time-dependent properties of a network. These models are traditionally based on systems of ordinary differential equations. More recent developments have focused on modeling concurrency in Petri Nets are a modeling tool widely used to map biological systems, thanks to their intuitive graphical representation which gives a valuable aid to describe and understand even systems with a complex structures. The graphical aspects of the Petri net are quite similar to biochemical network representation, and this gives superior communication ability to models and facilitates their design.

Clinical and translational medicine is expected to include scientific and regulatory investigations to translate preclinical researches to clinical application with a specific emphasis on new biotechnologies, biomaterials, bioengineering, disease-specific biomarkers, cellular and molecular medicine, -omics science, bioinformatics, applied immunology, molecular imaging, drug discovery and development, and regulation and health policy. It is believed that clinical and translational medicine will benefit and improve novel diagnostics/prognostics and therapeutics for clinical use, post-genomic knowledge and experience,

and/or new disciplines that reflect additional levels of complexity. We should clarify the bioethics at the interface and paradigms between technology and society, academies and industries, as well as publics and private models. Translational medicine should meet the demands of maintaining or expanding the biomedical workforce and education programs that attract and retain young people in the translational and biomedical sciences. In the present perspective, we collected commentaries and descriptions about clinical and translational medicine from some members of Clinical and Translational Medicine editorial board to stimulate the discussion and help the understanding better.

Moreover, modeling biological systems requires taking into account uncertainty in system behaviors, due to external interferences (molecules, energy sources, etc), intra-variability (personal features), and inter-variability (intrinsic noise and low number of molecules) of complex systems. Using the Petri Nets extension, Stochastic Petri Nets (SPNs), we can describe these random aspects.

Applying a stochastic approach for the investigation of the behavioral properties of Extrinsic coagulation pathway, we have been able to observe several feature, explaining or been unanimous with biochemical an clinical patency.

All the results obtained from simulations may only be foreseen from a deterministic modeling and the consequent solution of underlined differential equations, with the stochastic approach we can appreciate how, also only a few simulated scenarios, may confer a homogeneity of the results. The randomness expressed in the time series concerning thrombin generation is indeed able to almost completely cover all the possible intermediate cases, without simulate them.

This is one of the attractive goals for translational medicine, which can employ such applications both to manage more accurately a therapy, personalizing treatments both to identify a limiting drug administration before starting a clinical trial.

The accurate approaches to modelization presented in this dissertation , in compliance with translational personalized medicine, offer then a

strong opportunity to link a multiplicity of prothrombotic disease phenotypes to as many possible drug therapy dosages to gain valuable therapy effectiveness and decreasing the adverse effects, giving a potential advantage in challenging therapy management.

# 10 Bibliography


Albert R and Barabási AL. 2002. Statistical mechanics of complex networks. Rev Modern Phys. 74:47-97.289:1203-1211.

Albert R, Jeong H and Barabási A-L. 2002. Error and attack tolerance of complex networks. Nature 406:378–382.

Albert R. 2005. Scale-free networks in cell biology. J Cell Sci. 118(Pt 21):4947-57.

Anand M, Rajagopal K, Rajagopal KR. 2003. A model incorporating some of the mechanical and biochemical factors underlying clot formation and dissolution in flowing blood. J Theoret Med. 5: 183–218.

Anand M, Rajagopal K, Rajagopal KR. 2008.A model for the formation, growth, and lysis of clots in quiescent plasma. A comparison between the effects of antithrombin III deficiency and protein C deficiency. J Theor Biol. 253(4):725-38.

Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. 2008. Pharmacology and management of the vitamin k antagonists*: american college of chest physicians evidence-based clinical practice guidelines (8th edition).Chest. 133(6_suppl):160S-198S.

Ardissino D, Merlini PA, Arlens R, Coppola R, Bramucci E, Lucreziotti S, Repetto A, Fetiveau R, Mannucci PM. 2000.Tissue factor in human coronary atherosclerotic plaques.Clin Chim Acta. 291(2):235-40.

Arisi I, Cattaneo A, Rosato V. 2006. Parameter estimate of signal transduction pathways.BMC Neurosci. 30(7):Suppl 1:S6.

Ashton K. 2009. "That 'internet of things' thing." RFiD Journal 22:97-114.

Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. 2009. Systems biology: parameter estimation for biochemical models.FEBS J. 276(4):886-902.

Ataullakhanov FI, Krasotkina YV, Sarbash VI, Volkova RI, Sinauridse EI and Kondratov AY. 2002. Spatio-temporal dynamics of blood coagulation and pattern formation. An experimental study. Int J of Bifurc and Chaos, 12(9):1969–1983.

Bader GD, Cary MP, Sander C. 2006. Pathguide: a pathway resource list. Nucleic Acids Res. 34(Database issue):D504-6.

Baldan P, Cocco N, Marin A and Simeoni M. 2010. Petri nets for modelling metabolic pathways: a survey. Natural Computing. 9(4):955-989.

Baldwin AS and D. Basmadjian D. 1994. A mathematical model of thrombin production in blood coagulation, part I: The sparsely covered membrane case. Annals of Biomedical Engineering 22(4):357–370.

Baldwin SA and Basmadjian D. 1994. A Mathematical-Model of Thrombin Production in Blood-Coagulation . The Sparsely Covered Membrane Case. Annals of Biomedical Engineering. 22(4):357-370.

Bansal M, Belcastro V, Ambesi-Impiombato A, and di Bernardo D. 2007. How to infer gene networks from expression profiles. Mol Syst Biol. 3:78.

Barabási AL and Oltvai ZN. 2004.Network biology: understanding the cell's functional organization.Nat Rev Genet. 5(2):101-13.

Barabási AL. 2007. Network medicine-from obesity to the "diseasome". N Engl J Med. 357(4):404-7.

Barbuti R, Levi ., Milazzo P and Scatena G. 2012. Probabilistic model checking of biological systems with uncertain kinetic rates. Theoretical Computer Science. 419:2-16.

Barkai N, Alon U, Leibler S. 2001. Robust amplification in adaptive signal transduction networks. Comptes Rendus de l'Académie des Sciences - Series IV - Physics-Astrophysics

Behrends C, Sowa ME, Gygi SP and Harper JW. 2010. Network organization of the human autophagy system. Nature. 466(7302):68-76.

Beltrami E and Jesty J. 2001. The role of membrane patch size and flow in regulating a proteolytic feedback threshold on a membrane: possible application in blood coagulation. Math Biosci. 172(1):1–13.

Bender EA. An Introduction to Mathematical Modeling. Wiley, 1978 - Dover publications 2000. on ISBN-13: 978-0486411804.

Berg AH & Scherer PE. 2005. Adipose tissue, inflammation, and cardiovascular disease. Circ res 96(9):939-949.

Berg JM, Tymoczko JL, Stryer L.2002. Biochemistry. WH Freeman New York (2002),5th edition.NCBI Bookshelf

Blätke MA. 2011. Tutorial. Petri Nets in Systems Biology. 1st Edition (August 2011).Co-Authors:Heiner M, Marwan W.

Blinder P, Tsai PS, Kaufhold JP, Knutsen PM, Suhl H, Kleinfeld D. 2013. The cortical angiome: an interconnected vascular network with noncolumnar patterns of blood flow. Nature Neuroscience 16:889–897.

Boekholdt SM, Bijsterveld NR, Moons AH, Levi M, Büller HR, Peters RJ. 2001. Genetic variation in coagulation and fibrinolytic proteins and their relation with acute myocardial infarction: a systematic review.Circulation. 104(25):3063-8.

Bos W. 2008. Modeling biological systems using Petri nets. Department of Electrical Engineering, Mathematics and Comp Science, University of Twente.

Breitling R ,Gilbert D ,Heiner M ,Orton R. 2008. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. Brief Bioinform. 9(5): 404-421.

Briggs GE and Haldane JBS. 1925. A note on the kinetics of enzyme action. Biochem. J. 19:338-339

Bro C, Regenberg B, Förster J, Nielsen J.2006.In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production.Metab Eng. 8(2):102-11.

Brummel-Ziedins K, Vossen CY, Rosendaal FR, Umezaki K, Mann KG. 2005. The plasma hemostatic proteome: thrombin generation in healthy individuals. J Thromb Haemost. 3(7):1472-81

Bryant WA, Sternberg MJ, Pinney JW. 2013. AMBIENT: Active Modules for Bipartite Networks--using high-throughput transcriptomic data to dissect metabolic response.BMC Syst Biol. 7:26.

Buchholz P. 1994. Hierarchical High Level Petri Nets for complex system analysis Application and Theory of Petri Nets 1994.LNCS 815:119-138

Bungay SD, Gentry PA, and Gentry RD.2003. A mathematical model of lipidmediated thrombin generation. Math Med and Biology, 20(1):105–129.

Byrne M, Liu X, Carter CM, Zumberg MS. 2014. Myelodysplastic syndrome and associated coagulopathy: a case report and review. Blood Coagul Fibrinolysis 25(2):172-5.

Byun, Y, Jacobs HA, and Kim SW. 1996. Binding of Antithrombin III and Thrombin to Immobilized Heparin under Flow Conditions. Biotech Progress. 12(2):217-225.

Camm AJ. 2010. European Heart Rhythm Association; European Association for Cardio-Thoracic Surgery, Kirchhof P, Lip GY, Schotten U, Savelieva I, Ernst S, Van Gelder IC, Al-Attar N, Hindricks G, Prendergast B, Heidbuchel H, Alfieri O, Angelini A, Atar D, Colonna P, De Caterina R, De Sutter J, Goette A, Gorenek B, Heldal M, Hohloser SH, Kolh P, Le Heuzey JY, Ponikowski P, Rutten FH. Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). Eur Heart J. 31(19):2369-429.

Cao W, Krishnaswamy S, Camire RM, Lenting PJ, Zheng XL. 2008. Factor VIII accelerates proteolytic cleavage of von Willebrand factor by ADAMTS13. Proc Natl Acad Sci U S A. 105(21):7416-21.

Cao Y, Gillespie DT, Petzold LR. 2007. Adaptive explicit-implicit tau-leaping method with automatic tau selection. J Chem Phys. 126(22):224101

Caspi R, Altman T, Dreher K, Fulcher CA and Karp PD .The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 36 (Database issue):D623–D631.

Chaouiya C, Remy E, Thieffry D. 2008. Petri net modelling of biological regulatory networks, Journal of Discrete Algorithms. 6(2):165-77.

Chaouiya C. 2007. Petri net modelling of biological networks. Brief Bioinform. 8(4):210-9.

Choi G, Schultz MJ, Levi M, van der Poll T. 2006 .The relationship between inflammation and the coagulation system.Swiss Med Wkly. 136(9-10):139-44.

Christon M, Crawford D, Hertel, E, Peery J and Robinson, A. 1997. Asci red - experiences and lessons learned with a massively parallel teraflop supercomputer. Proceedings of the Supercomputer 1997

Chu AJ. 2006. Role of tissue factor in thrombosis. Coagulation-inflammation-thrombosis circuit.Front Biosci. 11:256-71

Chu AJ. 2011. Tissue Factor, Blood Coagulation, and Beyond: An Overview. Int J of Inflamm. vol. 2011, Article ID 367284:30pp

Cicerone S, Di Stefano G. 1999. On the extension of bipartite to parity graphs. Disc Appl Math. 95 (issue 1-3):181–195

Clauset A, Shalizi CR, Newman ME. 2009. Power-law distributions in empirical data. SIAM review. 51(4):661-703.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. 2007. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2(10):2366-82.

Coogan JS, Humphrey JD, Figueroa CA. 2013.Computational simulations of hemodynamic changes within thoracic, coronary, and cerebral arteries following early wall remodeling in response to distal aortic coarctation.Biomech Model Mechanobiol. 12(1):79-93.

Corlan AD and Ross J.2011. Canalization effect in the coagulation cascade and the interindividual variability of oral anticoagulant response. A simulation study.Theor Biol Med Model. 8:37.

Craddock RC, Jbabdi S, Yan C-G, Vogelstein JT, Castellanos FX, Di Martino A, Kelly C, Heberlein H, Colcombe S and Milham MP. 2013. Imaging human connectomes at the macroscale. Nature Methods 10:524–539

Csermely P, Korcsmáros T, Kiss HJ, London G, Nussinov R. 2013. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther. 138(3):333-408.

Dandekar T, Fieselmann A, Majeed S and Ahmed Z.2014.Software applications toward quantitative metabolic flux analysis and modeling.Brief Bioinform. 15(1):91-107

Darabos C, Harmon SH, Moore JH. 2014. Using the bipartite human phenotype network to reveal pleiotropy and epistasis beyond the gene. Pac Symp Biocomput. 2014:188-99.

Darmoni SJ, Névéol A, Renard JM, Gehanno JF, Soualmia LF, Dahamna B, Thirion B. 2006.A MEDLINE categorization algorithm.BMC Med Inform Decis Mak. 6:7.

Davizon P, Lopez JA. 2009. Microparticles and thrombotic disease. Curr Opin Hematol. 16:334–341

De Bièvre P and Peiser HS. 1992.International Union of Pure and Applied Chemistry Commission on Atomic Weights and Isotopic Abundances - Atomic Weight: The Name, Its History, Definition and Units. Pure and Applied Chemistry 64 (10):1535–43.

Dong-Hee K and Motter AE. 2009. Slave nodes and the controllability of metabolic networks. New J Phys. 11:113047

Duncan JW and Strogatz SH. 1998. Collective dynamics of 'small-world' networks. Nature 393: 440-442

Emmert-Streib F and Glazko GV. 2011. Network biology: a direct approach to study biological function. Wiley Interdiscip Rev Syst Biol Med. 3(4):379-91

Esmon CT.2003.The protein C pathway.Chest. 124(3 Suppl):26S-32S.

Eugster MJ, Schmid M, Binder H, Schmidberger M. 2013. Grid and cloud computing methods in biomedical research. Methods Inf Med. 52(1):62-4

Fayyad U, Piatetsky-Shapiro G and Smyth P. 1996. From data mining to knowledge discovery in databases. AI Magazine, 17(3):37-54

Fehling R. 1993.A concept of hierarchical Petri nets with building blocks. Advances in Petri Nets 1993 LNCS 674:148-168

Feng J, Jürgen j and Qian M (Eds.). 2007. Networks: From Biology to Theory. Springer, Berlin (2007). ISBN 978-1-84628-780-0

Finkelstein J, Gray N, Heemels MT, Marte B and Nath D. 2012. Metabolism and disease. Nature. 491:347

Forest A, Pautas E, Ray P, Bonnet D, Verny M, Amabile N, Boulanger C, Riou B, Tedgui A, Mallat Z, Boddaert J.2010.Circulating microparticles and procoagulant activity in elderly patients.J Gerontol A Biol Sci Med Sci. 65(4):414-20.

Frauwirth KA, Riley JL, Harris MH, Parry RV, Rathmell JC, Plas DR, Elstrom RL, June CH, Thompson CB. 2002. The CD28 signaling pathway regulates glucose metabolism. Immunity. 16(6):769-77.

Frawley WJ, Piatetsky-Shapiro G and Matheus CJ. 1992. Knowledge discovery in databases: An overview. AI magazine, 13(3):57.

Frawley WJ, Piatetsky-Shapiro G, and Matheus CJ. 1992. Knowledge discovery in databases: An overview. AI magazine, 13(3):57.

Fu P. 2006. A perspective of synthetic biology: assembling building blocks for novel functions.Biotechnol J. 1(6):690-9.

Furie B, Furie BC. 2007. In vivo thrombus formation. J Thromb Haemost 5(suppl 1):12–17.

Furie B, Furie BC.2008. Mechanisms of thrombus formation. N Engl J Med 359:938–949

Gaffney PJ, Edgell TA, Whitton CM. 1999. The Haemostatic Balance – Astrup Revisited. Haemostasis.29:58-71.

Gaithersburg, MD, Wang M and Sud, T. The Bio-Networking Architecture: a biologically inspired approach to the design of scalable, adaptive, and survivable/available network applications. Applications and the Internet, 2001. Proceedings. 2001 Symposium on , vol., no., pp.43,53, 2001

Gallasch R, Efremova M, Charoentong P, Hackl H, Trajanoski Z. 2013. Mathematical models for translational and clinical oncology.J Clin Bioinforma. 3(1):23.

Garcia Rodriguez P , Eikenboom HC, Tesselaar ME, Huisman MV, Nijkeuter M, Osanto S, Bertina RM.2010.Plasma levels of microparticle-associated tissue factor activity in patients with clinically suspected pulmonary embolism.Thromb Res. 126(4):345-9

Gardiner CW. 2005. Handbook of Stochastic Processes Springer, New York, 2005

Garfinkel D, Garfinkel L, Pring M, Green SB, Chance B. 1970. Computer applications to biochemical kinetics. Annu Rev Biochem. 39:473-98.

Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G, Alon U.Oscillations and variability in the p53 system. Mol Syst Biol. 2:2006.0033.

Gibson G. 2009. Decanalization and the origin of complex disease. Nat Rev Genet. 10(2):134-40.

Gibson MA and Bruck J. 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. J Phys Chem A. 104:1876–188.

Gillespie DT. 1976. General method for numerically simulating the stochastic time evolution of coupled chemical reactions. J of Comp Phys. 22:403–434

Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 81(25):2340–61

Gillespie DT. 1992. A Rigorous Derivation of the Chemical Master Equation. Physica A 188:404–425.

Gillespie DT. 2001. Approximate accelerated stochastic simulation of chemically reacting system. J of Chem Phys. 115(4):1716–1733.

Gillespie DT and Petzold LR. 2003. Improved leap-size selection for accelerated stochastic simulation. J Chem Phys. 119:8229-8234

Gillespie DT. 2007. Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. 58:35-55.

Girard TJ and Nicholson NS. 2001. The role of tissue factor/factor VIIa in the pathophysiology of acute thrombotic formation. Curr Opin Pharmacol. 1(2):159-63.

Glunde K, Bhujwalla ZM and Ronen SM. 2011. Nature Reviews Cancer. 11:835-848.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. 2007. The human disease network. Proc Natl Acad Sci U S A. 104(21):8685-90.

Gomperts BD, Kramer IM , Tatham PER. 1999. Signal transduction. Academic Press-Reed Elsevier 1st ed. 1999. ISBN-13: 978-0122896323

Gonzalez OR, Küper C, Jung K, Naval PC and Mendoza E. 2007. Parameter estimation using Simulated Annealing for S-system models of biochemical networks. Bioinformatics. 23(4):480-486.

Goss PJE and Peccoud J. 1998. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proc. Nat. Acad. Sci. USA 95 6750–6754

Gouin-Thibault I, Levy C, Pautas E, Cambus JP, Drouet L, Mahé I, Bal Dit Sollier C, Horellou MH, Golmard JL, Siguret V. 2010. Improving anticoagulation control in hospitalized elderly patients on warfarin. J Am Geriatr Soc. 58(2):242-247

Gulati A, G K Isbister GK and Duffull SB. 2014.Scale Reduction of a Systems Coagulation Model With an Application to Modeling Pharmacokinetic–Pharmacodynamic Data. CPT Pharmacometrics Syst. Pharmacol. (2014)3:e90;

Hardy S and Robillard PN. 2004. Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. J Bioinform and Comput Biology. 2(04):619-637.

Hayashi M, Takeshita K, Inden Y, Ishii H, Cheng XW, Yamamoto K, Murohara T. 2011. Platelet activation and induction of tissue factor in acute and chronic atrial fibrillation: involvement of mononuclear cell-platelet interaction. Thromb Res. 128(6):e113-8.

Heiner M, Gilbert D, and Donaldson R. 2008. Petri nets for systems and synthetic biology. In Proceedings of the Formal methods for the design of computer, communication, and software systems (SFM'08). Springer-Verlag, Berlin, Heidelberg, 215-264.

Heiner M, Koch I, Will J. 2004. Model validation of biological pathways using Petri nets--demonstrated for apoptosis.Biosystems. 75(1-3):15-28.

Hilfinger A and Paulsson J. 2011. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. Proceedings of the National Academy of Sciences. 108(29):12167-12172.

Hirsh J, Dalen J, Anderson DR, Poller L, Bussey H, Ansell J, Deykin D. 2001. Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range.Chest. 119(1 Suppl):8S-21S.

Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W. 2006. Rules for modeling signal-transduction systems. Sci STKE. 2006(344):re6.

Hockin MF, Jones KC, Everse SJ, and Mann KG. 2002. A model for the stoichiometric regulation of blood coagulation. J of Bio Chem. 277(21):18322–18333, 2002.

Holmes MB, Schneider DJ, Hayes MG, Sobel BE, Mann KG.2000.Novel, bedside, tissue factor-dependent clotting assay permits improved assessment of combination antithrombotic and antiplatelet therapy.Circulation. 102(17):2051-7.

Hoops S, Sahle S, Gauges, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P and Kummer U. 2006. COPASI—a complex pathway simulator. Bioinformatics. 22(24):3067-74.

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle J, Kitano H. 2002. The ERATO Systems Biology Workbench: enabling interaction and exchange

between software tools for computational biology.Pac Symp Biocomput. 2002:450-61.

Ichimura Y, Takahashi H, Lee MT, Shiomi M, Mihara K, Morita T, Chen YT, Echizen H. 2012. Inter-individual differences in baseline coagulation activities and their implications for international normalized ratio control during warfarin initiation therapy.Clin Pharmacokinet. 51(12):799-808.

Ideker T, Galitski T, Hood L. 2001. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet. 2:343-72.

International Warfarin Pharmacogenetics Consortium, Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, Lee MT, Limdi NA, Page D, Roden DM, Wagner MJ, Caldwell MD, Johnson JA. 2009.Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med. 360(8):753-64.

Isermann N, Weitzel M, and Wolfgang Wiechert W. 2004. Kleene's Theorem and the Solution of Metabolic Carbon Labeling Systems.German Conference on Bioinformatics. vol.53 of LNI:75-84. GI

IUPAC 2007.International Union of Pure and Applied Chemistry Quantities (2007) Units and Symbols in Physical Chemistry, third ed, RSC Publishing, Cambridge.

Jackson David B. 2001. Maui scheduler: a multifunction cluster scheduler. In Beowulf Cluster Computing with Windows, Thomas Sterling (Ed.) 2001. MIT Press, Cambridge, MA, USA pp. 345-62

Janeway CA, Travers Jr P, Walport M, and Shlomchik MJ. 2001. The Immune System in Health and Disease. Immunobiology, 5th edition. Garland Science (New York) 2001.

Jeffery IB, Higgins DG, Culhane AC. 2006. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics. 7:359.

Jeong H, Tombor B, Albert R,Oltvai ZN, Barabasi AL. 2000 The large-scale organization of metabolic networks. Nature (London). 407(6804):651-4

Jia J. 2009. Parameter Estimation for Nonlinear Biological System Model Based on Global Sensitivity Analysis. In Bioinformatics and Biomedical Engineering, ICBBE 2009. 3rd Int. Conf on IEEE:1-4

Jin Y, Turaev D, Weinmaier T, Rattei T, Makse HA. 2013 .The evolutionary dynamics of protein-protein interaction networks inferred from the reconstruction of ancient networks.PLoS One. 8(3):e58134.

Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, Anderson JL, Kimmel SE, Lee MT, Pirmohamed M, Wadelius M, Klein TE, Altman RB; Clinical Pharmacogenetics Implementation Consortium. 2011. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. Clin Pharmacol Ther. 90(4):625-9.

Johnson KA and Goody RS. 2011. The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. Biochemistry. 50(39):8264–269

Jones KC and Mann KG. 1994. A model for the tissue factor pathway to thrombin II. A mathematical simulation. J of Bio Chem. 269(37):23367–373.

Kaneko K. 2006. Life: An Introduction to Complex Systems Biology. In Understanding Complex Systems Springer, Berlin (2006). ISBN-13: 978-3540326663

Kaneko k. 2012. Evolution of Robustness and Plasticity under Environmental Fluctuation: Formulation in terms of Phenotypic Variances. J Stat Phys. 148:686-704

Kaneko K. Evolution of Robustness and Plasticity under Environmental Fluctuation: Formulation in Terms of Phenotypic Variances. J of Stat Phys. 148(4):687-705

Keener J and Sneyd J. 2009. Mathematical Physiology I: Cellular Physiology .Interdisciplinary Applied Mathematics, Vol. 8/1. Springer Book 2nd ed., XXV, 547p.

Kestler HA, Wawra C, Kracher B, Kühl M. 2008. Network modeling of signal transduction: establishing the global view. Bioessays. 30(11-12):1110-25.

Khanin MA and Semenov VV. 1989. A mathematical model of the kinetics of blood coagulation. J of Theor Biol. 136(2):127–.

Khanin MA, Rakov DV and Kogan AE. 1998. Mathematical model for the blood coagulation prothrombin time test. Thromb Res. 89(5):227-32.

Khanin MA, Rakov DV e Kogan AE. 1998. Mathematical model for the blood coagulation prothrombin time test. Thromb Res 89(5):227-32.

Kim JR, Yoon Y, Cho KH. 2008. Coupled feedback loops form dynamic motifs of cellular networks. Biophys J 94:359–365.

Kitano H, Funahashi A, Matsuoka Y, Oda K. 2005. Using process diagrams for the graphical representation of biological networks. Nat Biotechnol. 23(8):961-6

Kitano H. 2002. Computational systems biology. Nature. 420(6912):206-10.

Kitano H. 2004. Biological robustness. Nat Rev Genet. 5(11):826-37.

Koch I, Reisig W and Schreiber F (Eds.). 2010. Modeling in systems biology: the Petri Net approach (Vol. 16). Springer

Krabbe KS, Pedersen M, Bruunsgaard H.2004.Inflammatory mediators in the elderly.Exp Gerontol. 39(5):687-99.

Kuharsky AL and Fogelson AL.2001. Surface mediated control of blood coagulation: the role of binding site densities and platelet deposition. Biophysical Journal, 80(3):1050–1074.

Laplante M and Sabatini DM. 2012. mTOR signaling in growth control and disease. Cell. 149(2):274-93.

Latapya M, Magnienb C, Del Vecchio N. 2008. Basic notions for the analysis of large two-mode networks. Social Networks. 30(1):31–48

Lebiedz D, Dominik Skanda D and Fein M.2008.Automatic Complexity Analysis and Model Reduction of Nonlinear Biochemical Systems. Computational Methods in Systems Biology. LNCS 5307:123-140

Lechner D and Weltermann A. 2008.Circulating tissue factor-exposing microparticles.Thromb Res. 122(Suppl 1):S47-54.

Lee J and Ozcan U. 2014. Unfolded Protein Response Signaling and Metabolic Diseases. J Biol Chem. 289:1203-1211.

Lee J and Ozcan U. 2014. Unfolded Protein Response Signaling and Metabolic Diseases. J Biol Chem.

Lee WP and Tzou WS. 2009. Computational methods for discovering gene networks from expression data.Brief Bioinform. 10(4):408-23.

Leipold RJ, Bozarth TA, Racanelli AL and Dicker. 1995. Mathematical model of serine protease inhibition in the tissue factor pathway to thrombin. J of Bio Chem 270(43):25383–387.

Levine SN. 1966. Enzyme amplifier kinetics. Science, 152(3722):651–653, 1966.

Levy O and Netea MG. 2014. Innate immune memory: implications for development of pediatric immunomodulatory agents and adjuvanted vaccines.Pediatr Res. 75(1-2):184-8.

Li B, Shen Y, Li B. 2008. Quasi-steady-state laws in enzyme kinetics. J Phys Chem A. 112(11):2311-21

Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, Snoep JL, Hucka M, Le Novère N, Laibe C. 2010. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.BMC Syst Biol. 4:92.

Li C, Ge QW, Nakata M, Matsuno H, Miyano S. 2007. Modelling and simulation of signal transductions in an apoptosis pathway by using timed Petri nets. J Biosci.32(1):113-27.

Li C, Suzuki S, Ge Qi-Wei, Nakata M, Matsuno H, Miyano S. 2005. On modeling and analyzing signaling pathways with inhibitory interactions based on Petri net, Proc. The 2005 Internatinal Joint Conference of InCoB, AASBi and KSBI (BIOINFO2005):348-53.

Li P, Dada JO, Jameson D, Spasic I, Swainston N, Carroll K, Dunn W, Khan F, Malys N, Messiha HL, Simeonidis E, Weichart D, Winder C, Wishart J, Broomhead DS, Goble CA, Gaskell SJ, Kell DB, Westerhoff HV, Mendes P, Paton NW. 2010. Systematic integration of experimental data and models in systems biology. BMC Bioinfo. 11:582.

Lim HA. 2010. Feature-Bioinformatics in Communicable Diseases – The Way Forward. Asia-Pacific Biotech News. 14(12):12-21

Lip GYH and Blann AD.2000. Thrombogenesis and fibrinolysis in acute coronary syndromes. Important facets of a prothrombotic or hypercoagulable state? J Am Coll Cardiol 36:2044–2046.

Lisman T, Porte RJ. 2010. Rebalanced haemostasis in patients with liver disease: evidence and clinical consequences.Blood. 116(6):878-85.]

Ljung R, Auerswald G, Benson G, Jetter A, Jiménez-Yuste V, Lambert T, Morfini M, Remor E, Sørensen B, Salek SZ. 2013.Novel coagulation factor concentrates: issues relating to their clinical implementation and pharmacokinetic assessment for optimal prophylaxis in haemophilia patients.Haemophilia. 19(4):481-6.

Lorenz DM, Jeng A, Deem MW. 2011. The Emergence of Modularity in Biological Systems. Physics of Life Reviews 8:129-160

Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, Kershenbaum A, Stolovitzky GA, Blitzer RD, Iyengar R. 2005. Formation of regulatory patterns during signal propagation in a Mammalian cellular network. Science. 309(5737):1078-83

MacFarlane RG.1964. An enzyme cascade model in the blood clotting mechanism, and its function as a biochemical amplifier. Nature. 202:498–499.

Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, Rocha I. 2011. Modeling formalisms in Systems Biology. AMB Express. 5(1):45.

Madhamshettiwar PB, Maetschke SR, Davis MJ, Reverter A, Ragan MA. 2012. Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets.Genome Med. 4(5):41.

Maimon OZ and Rokach L. 2005. Data mining and knowledge discovery handbook (Vol. 1). New York: Springer.

Malý M, Vojácek J, Hrabos V, Kvasnicka J, Salaj P, Durdil V. 2003. Tissue factor, tissue factor pathway inhibitor and cytoadhesive molecules in patients with an acute coronary syndrome. Physiol Res. 52(6):719-28.

Mari D, Ogliari G, Castaldi D, Vitale G, Bollini EM,1 and Lio D.2008.Hemostasis and ageing.Immun Ageing. 2008; 5: 12.

Marino S, Hogue IB, Ray CJ, Kirschner DE. 2008.A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol. 254(1):178-96.

Martin KH, Slack JK, Boerner SA, Martin CC, Parsons JT. 2002. Integrin connections map: to infinity and beyond. Science. 296(5573):1652-3.

Marvalee HW. 2008. Integrative Biology: Science for the 21st Century. BioScience 58(4):349-353.

Mason CH and Perreault Jr, WD. 1991. Collinearity, power, and interpretation of multiple regression analysis. Journa

Masson N and Ratcliffe PJ. 2014. Hypoxia signaling pathways in cancer metabolism: the importance of co-selecting interconnected physiological pathways. Cancer Metab. 2:3

Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human genome.Annu Rev Genomics Hum Genet. 7:29-59.

Matsuno H, Doi A, Nagasaki M, Miyano S. 2000. Hybrid Petri net representation of gene regulatory network. Proc. Pacific Symposium on Biocomputing 5:341-352.

Mayer SA and Rincon F. 2005. Treatment of intracerebral haemorrhage. The Lancet Neurology. 4(10):662-672.

Miwa Y, Murakami Y, Ge QW, Li C, Matsuno H, Miyano S. 2010. Delay time determination for the timed Petri net model of a signaling pathway based on its structural information, IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences, E93-A(12):2717-2729

Molloy MK. 1982. Performance analysis using stochastic Petri nets. Computers, IEEE Transactions on, 100(9):913-917.

Monroe DM and Key NS. 2007. The tissue factor-factor VIIa complex: procoagulant activity, regulation, and multitasking. J Thromb Haemost. 5(6):1097-105

Monroe DM, Key NS.2007.The tissue factor-factor VIIa complex: procoagulant activity, regulation, and multitasking.J Thromb Haemost. 5(6):1097-105.

Nacher JC, Ochiai T, Hayashida M and Akutsu T. 2009. Mathematical model for generating bipartite graphs and its application to protein networks. J Phys A: Math Theor 42(2009):485005

Nagano N. 2005. EzCatDB: the Enzyme Catalytic-mechanism Database.Nucleic Acids Res. 33(Database issue):D407-12.

Nakamura Y, Yamane K, Murai A. 2006. Macroscopic modeling and identification of the human neuromuscular network.Conf Proc IEEE Eng Med Biol Soc. 1:99-105.

Napione L, Manini D, Cordero F, Horváth A, Picco A, De Pierro M, Pavan S, Sereno M, Veglio A, Bussolino F, Gianfranco Balbo G. 2009.On the Use of Stochastic Petri Nets in the Analysis of Signal Transduction Pathways for Angiogenesis Process. Computational Methods in Systems Biology 2009 LNCS 5688:281-295

NC-IUB. 1979. Recommendations 1978. Nomenclature Committee of the International Union of Biochemistry. Units of Enzyme Activity. Eur J Biochem. 97:319-320

Nelson DL and Cox MM. Lehninger-Principles of Biochemistry. Worth, New York, 3rd.ed. 2000.

Nesheim ME, Tracy RP,Tracy PB, Boskovic DS, and Mann KG. 1992. Mathematical simulation of prothrombinase. Methods in Enzymology. 215:316–328.

Newman SA and Rice SA. 1971. Model for Constraint and Control in Biochemical Networks. Proc Natl Acad Sci U S A. 68(1):92–96.

NIST Special Publication 330. 2008. Edition Barry N. Taylor and Ambler Thompson.National Institute of Standards and Technology

Nobile MS, Besozzi D, Cazzaniga P, Mauri G, Pescini D. 2012. Estimating reaction constants in stochastic bio-logical systems with a multi-swarm PSO running on GPUs. In Proceedings of the 14th intern.l conf. on Genetic and evolutionary computation conference companion (GECCO Companion '12), Terence Soule (Ed.). ACM, New York, NY, USA, 1421-1422.

Obraztsov IF, Popov AF, and Khanin MA. 2006. Threshold effects in kinetics of activation of contact system of hemocoagulation. Doklady Akademii Nauk (Biochemistry and Biophysics), 367(1):130–132, 1999.

Olivier BG and Snoep JL. 2004. Web-based kinetic modelling using JWS Online. Bioinformatics. 20(13):2143-44.

Palla G, Derényi I, Farkas I and Vicsek T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature. 435:814–818

Panteleev MA, Ovanesov MV, Kireev DA, Shibeko AM, Sinauridze EI, Ananyeva NM, Butylin AA, Saenko EL, Ataullakhanov FI.Spatial propagation and localization of blood coagulation are regulated by intrinsic and protein C pathways, respectively. Biophys J.90(5):1489-500.

Peacock J and Peacock P. 2010. Oxford Handbook of Medical Statistics. Oxford Medical Handbooks-OUP Oxford 1 ed.2010

Peleg M, Rubin D and Altman RB. 2005. Using Petri net tools to study properties and dynamics of biological systems. J of Americ Med Inform Assoc. 12(2):181-199.

Peng CK, Buldyrev SV, Hausdorff JM, Havlin S, Mietus JE, Simons M, Stanley HE, Goldberger AL. 1994. Non-equilibrium dynamics as an indispensable characteristic of a healthy biological system. Integrative Physiological and Behavioral Science, 29(3), 283-293.

Peng CK, Buldyrev SV, Hausdorff JM, Havlin S, Mietus JE, Simons M, Stanley HE, Goldberger AL. 1994. Non-equilibrium dynamics as an indispensable characteristic of a healthy biological system. Integrative Physiological and Behavioral Science, 29(3), 283-293.

Pereira M and Freire M. 2010. Biomedical Diagnostics and Clinical Technologies: Applying High-Performance Cluster and Grid Computing. IGI Global Publisher 2010 p.396 ISBN-13: 978-1-60566-280-0

Petäjä J. 2011. Inflammation and coagulation. An overview. Thromb Res. 127 Suppl 2:S34-7.

Petri CA. Kommunikation mit Automaten. PhD thesis, Institut für instrumentelle Mathematik, Bonn, Germany, 1962. Schriften des IIM, number 2.

Pinotti M, Bertolucci C, Portaluppi F, Colognesi I, Frigato E, Foà A, Bernardi F. 2005. Daily and circadian rhythms of tissue factor pathway inhibitor and factor VII activity.Arterioscler Thromb Vasc Biol. 25(3):646-9.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, ..*omitted*.., , Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 466(7304):368-72.

Pohl B, Beringer C, Bomhard M, Keller F. 1994. The Quick Machine—a Mathematical Model for the Extrinsic Activation of Coagulation. Haemostasis 24:325–37

Porwit A, McCullough J, Erber WN. Blood and Bone Marrow Pathology. Churchill Livingstone; 2 ed. 2011 722p. ISBN-13:978-0702031472

POSIX.1-2008. IEEE Std 1003.1™-2008 and The Open Group Technical Standard Base Specifications. Issue 7 2008

Qian H. 2011. Nonlinear stochastic dynamics of mesoscopic homogeneous biochemical reaction systems—an analytical theory. Nonlinearity 24:R19

Quon G, Lippert C, Heckerman D, Listgarten J. 2013. Patterns of methylation heritability in a genome-wide analysis of four brain regions. Nucleic Acids Res. 41(4):2095-104.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 - http://www.R-project.org.

Randhawa R, Shaffer C A, Tyson JJ. 2009. Model aggregation: a building-block approach to creating large macromolecular regulatory networks. Bioinformatics. (24)25:3289–95

Rathmell JC, Vander Heiden MG, Harris MH, Frauwirth KA, Thompson CB. 2000. In the absence of extrinsic signals, nutrient utilization by lymphocytes is insufficient to maintain either cell size or viability. Mol Cell.6(3):683-92.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. 2002. Hierarchical organization of modularity in metabolic networks. Science. 297(5586):1551-5.

Raven PH, Johnson GB, Mason KA, Losos JB. 2014. Biology . 10/e 2014. McGraw-Hill Higher Education

Reddy V, Liebman M, Mavrovouniotis M. Qualitativeanalysis of biochemical reaction systems. Comput Biol Med. 1996;26:9–24

Reddy VN, Mavrovouniotis ML, Liebman MN. 1993. Petri net representations in metabolic pathways. Proc Int Conf Intell Syst Mol Biol. 1:328-36.

Rees J. 2002. Complex Disease and the New Clinical Sciences. Science. 296(5568):698-700

Reininger AJ, Bernlochner I, Penz SM, Ravanat C, Smethurst P, Farndale RW, Gachet C, Brandl R, Siess W. 2010. A 2-step mechanism of arterial thrombus formation induced by human atherosclerotic plaques. J Am Coll Cardiol. 55(11):1147-58.

Rodriguez-Fernandez M, Egea JA and Banga JR. 2006.Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems BMC Bioinformatics. 7:483.

Rohr C,Marwan W and Heiner M. 2010. Snoopy - a unifying Petri Net framework to investigate biomolecular networks. Bioinformatics. 26(7):974-975.

Roland CG. 1968. Jargon. JAMA. 204(4):317-318

Rölke H and Heitmann F, editors. 2010 Petri nets world, website, maintained by the TGI (Theoretische Grundlagen der Informatik) group at the University of Hamburg, Germany, 2010. http://www.informatik.uni-hamburg.de/TGI/PetriNets/.

Rosen R. 2005. Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life. in Complexity in Ecological Systems series,Columbia University Press, New York (2005).ISBN-13: 978-0231075657

Sackmann A, Heiner M, Koch I. 2006.Application of Petri net based analysis techniques to signal transduction pathways.BMC bioinformatics 7(1):1-17

Samad F and Ruf W. 2013. Inflammation, obesity, and thrombosis. Blood. 122(20):3415-22

Sanchez C, Lachaize C, Janody F, Bellon B, Röder L, Euzenat J, Rechenmann F, and Jacq B. 1999. Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database.Nucleic Acids Res. 27(1):89–94.

Sanft KR, Wu S, Roh M, Fu J, Lim RK and Petzold LR. 2011. StochKit2: software for discrete stochastic simulation of biochemical systems with events. Bioinformatics. 27(17):2457-58.

Sato K, Ito Y, Yomo T and Kaneko K. 2003. On the relation between fluctuation and response in biological systems. PNAS. 100(24):14086-90.

Sauro Herbert M. 2012. Enzyme Kinetics for Systems Biology. Ambrosius Publishing 2012.pp.322.ISBN-13: 978-0982477311.

Schaff J, Fink CC, Slepchenko B, Carson JH, Loew LM. 1997. A general computational framework for modeling cellular structure and function. Biophys J. 73:1135–46

Schera F, Weiler G, Neri E, Kiefer S and Graf N.2014.The p-medicine portal—a collaboration platform for research in personalised medicine Ecancermedicalscience. 2014;8:398.

Schomburg I, Chang A, Placzek S, Söhngen C, Rother M, Lang M, Munaretto C, Ulas S, Stelzer M, Grote A, Scheer M, Schomburg D. 2013. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. 41:764-772.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering.Trends Biotechnol. 17(2):53-60.

Schwarz UI, Ritchie MD, Bradford Y, Li C, Dudek MS, Frye-Anderson A, RKim RB, Roden DM, Stein CM. 2008. Genetic determinants of response to warfarin during initial anticoagulation. N Engl J Med. 358(10):999-1008

Segel LA. 1988. On the validity of the steady state assumption of enzyme kinetics. Bull Math Biol. 50:579–593.

Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, Uetz P, Sittler T, Karp RM, Ideker T. 2005. Conserved patterns of protein interaction in multiple species.Proc Natl Acad Sci U S A. 102(6):1974-9.

Shaw O, Steggles J and Wipat A. 2006. Automatic Parameterization of Stochastic Petri Net Models of Biological Networks. Electronic Notes in Theoretical Computer Science; 151(2006):111-129.

Siguret V, Pautas E, Gouin-Thibault I.2008. Warfarin therapy: influence of pharmacogenetic and environmental factors on the anticoagulant response to warfarin.Vitam Horm.78:247-64.

Simão E, Remy E, Thieffry D, Chaouiya C. 2005. Qualitative Modelling of Regulated Metabolic Pathways: Application to the Tryptophan Biosynthesis in E.Coli. , Bioinformatics, ECCB'05 Special Issue 21: ii190-196.

Siso-Nadal F, Fox JJ, Laporte SA, Hébert TE, Swain PS. 2009. Cross-Talk between Signaling Pathways Can Generate Robust Oscillations in Calcium and cAMP. PLoS ONE 4(10): e7189.

Soria JM. 2009.The Genetic Component of Disorders of Coagulation and Thrombosis.Rev Esp Cardiol. 09(Supl.B):58-65

Staples Garrik. 2006. TORQUE resource manager. In Proceedings of the 2006 ACM/IEEE conference on Supercomputing, SC '06, Art. 8. ISBN 0-7695-2700-0

Sun JC, Ugolini S and Vivier E. 2014. Immunological memory within the innate immune system. The EMBO journal 2014.

Sun SX, Lan G, Atilgan E. 2008. Stochastic modeling methods in cell biology. Methods Cell Biol 89: 601–621

Swain PS, Elowitz MB, Siggia ED. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression.Proc Natl Acad Sci U S A. 99(20):12795-800.

Swainston N, Golebiewski M, Messiha HL, Malys N, Kania R, Kengne S, Krebs O, Mir S, Sauer-Danzwith H, Smallbone K, Weidemann A, Wittig U, Kell DB, Mendes P, Müller W, Paton NW, Rojas I. 2010. Enzyme kinetics informatics: from instrument to browser. FEBS J. 77:3769–3779.

Taflanidis AA and & Beck JL. 2008. An efficient framework for optimal robust stochastic system design using stochastic simulation. Computer Methods in Applied Mechanics and Engineering 198(1):88-101.

Thiruvikraman SV, Guha A, Roboz J, Taubman MB, Nemerson Y, Fallon JT. 1996.In situ localization of tissue factor in human atherosclerotic plaques by binding of digoxigenin-labeled factors VIIa and X.Lab Invest. 75(4):451-61.

Tiaden JD, Wenzel E, Berthold HK, Muller-Oerlinghausen B. 2005. Adverse reactions to anticoagulants and to antiplatelet drugs recorded by the German spontaneous reporting system. Semin Thromb Hemost 31: 371–80.

Tiger CF, Krause F, Cedersund G, Palmér R, Klipp E, Hohmann S, Kitano H, Krantz M. 2012. A framework for mapping, visualisation and automatic model creation of signal-transduction networks.Mol Syst Biol. 8:578

Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E, Haas J. 2006. Herpesviral protein networks and their interaction with the human proteome. Science. 311(5758):239-242.

van Riel NA and Sontag ED. 2006. Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach.Syst Biol (Stevenage). 153(4):263-74.

Vazquez A, Flammini A, Maritan A, and Vespignani A. 2003 . Global protein function prediction from protein-protein interaction networks. Nat Biotechnol. 21:697–700.

Wajima T, Isbister GK, Duffull SB. 2009. A comprehensive model for the humoral coagulation network in humans. Clin Pharmacol Ther. 86(3):290-8.

Wang J. 1998 . Timed Petri Nets, Theory and Application. The International Series on Discrete Event Dynamic Systems, Vol. 9 XI, 281 p. Springer Book ed.1998. ISBN 978-1-4615-5537-7

Wang C, Yu JT, Miao D, Wu ZC, Tan MS, Tan L. 2014. Targeting the mTOR signaling network for Alzheimer's disease therapy. Mol Neurobiol. 49(1):120-35

Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadee W. 2008. Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (*VKORC1*) affects gene expression and warfarin dose requirement. Blood. 112(4):1013–1021.

West AG, Gaszner M, Felsenfeld G. 2002. Insulators: many functions, many mechanisms.Genes Dev. 16(3):271-88.

Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, Klein TE. 2011. Pharmacogenomics knowledge for personalized medicine.Clin Pharmacol Ther. 92(4):414-7.

Wilkinson DJ. 2009. Stochastic modelling for quantitative description of heterogeneous biological systems. Nat Rev Genet 10:122–133.

Williams SR , Yang Q, Chen F, Liu X, Keene KL, Jacques P, Chen WM, Weinstein G, Hsu FC, Beiser A, Wang L, Bookman E, Doheny KF, Wolf PA, Zilka M, Selhub J, Nelson S, Gogarten SM, Worrall BB, Seshadri S, Sale MM, Genomics and Randomized Trials Network - Framingham Heart Study. 2014. Genome-Wide Meta-Analysis of Homocysteine and Methionine Metabolism Identifies Five One Carbon Metabolism Loci and a Novel Association of ALDH1L1 with Ischemic Stroke. PLoS Genet. 10(3):e1004214.eCollection 2014.

Wittig U, Kania R, Golebiewski M, Rey M, Shi L, Jong L, Algaa E, Weidemann A, Sauer-Danzwith H, Mir S, Krebs O, Bittkowski M, Wetsch E, Rojas I, Müller W. 2012. SABIO-RK-database for biochemical reaction kinetics. Nucleic Acids Res. 40:D790–D796.

Wysokinski WE, Owen WG, Fass DN, Patrzalek DD, Murphy L, McBane RD 2nd. 2004. Atrial fibrillation and thrombosis: immunohistochemical differences between in situ and embolized thrombi.J Thromb Haemost. 2(9):1637-44.

Xu L, Taufer M, Stuart Collins S, Vlachos DG. 2012. Parallelization of tau-leap coarse-grained Monte Carlo simulations on GPUs. In IEEE International Parallel & Distributed Processing Symposium IPDPS 2010 pp.1-9

Yamagishi K, Aleksic N, Hannan PJ, Folsom AR; ARIC Study Inverstigators. 2012. Coagulation factors II, V, IX, X, XI, and XII, plasminogen, and alpha-2 antiplasmin and risk of coronary heart disease.J Atheroscler Thromb. 17(4):402-9.

Yang X, Guo Y and Bradley J. 2012. An iterative parameter estimation method for biological systems. In Proceedings of the 3rd international workshop on Emerging computational methods for the life sciences (ECMLS '12). ACM, New York, NY, USA, 65-74.

Yen HC. 2006. Introduction to Petri Net Theory. Recent Advances in Formal Languages and Applications 25:343-373.

Yu W, Vetter J, Canon, RS, Jiang S. 2007. Exploiting Lustre File Joining for Effective Collective IO. In Cluster Computing and the Grid, CCGRID 2007. Seventh IEEE International Symposium on pp.267-274. ISBN 0-7695-2833-3

Zinman GE, Zhong S, and Bar-Joseph Z. 2011. Biological interaction networks are conserved at the module level.BMC Systems Biology. 5:134

CASTALDI DAVIDE FABIO

University of Milano-Bicocca - Italy

July 2014

Department of Informatics, Systems and Communication -DISCo