



Tutorial PM9

6 August 2006

## FROM PATHWAYS DATABASES TO NETWORK MODELS

### *Presenters*

***Baltazar D. Aguda***

Mathematical Biosciences Institute  
Ohio State University, USA

***[bdaguda@mbi.osu.edu](mailto:bdaguda@mbi.osu.edu)***

***Andrew B. Goryachev***

Centre for Integrative Systems Biology,  
University of Edinburgh, UK

***[andrew\\_goryachev@yahoo.com](mailto:andrew_goryachev@yahoo.com)***

## **CONTENTS**

<b>I. INTRODUCTION.....</b>	<b>3</b>
<b>II. PATHWAYS DATABASES AND KNOWLEDGBASES [ AGUDA ].....</b>	<b>4</b>
II.1 Pathguide .....	4
II.2 Pathway Data Standards .....	6
II.3 A Modeling-focused use of Pathways Databases .....	7
Gene Ontology (GO) .....	9
Kyoto Encyclopedia of Genes and Genomes (KEGG) .....	10
Reactome .....	12
GenMAPP .....	12
Biocarta .....	13
II.4 Repositories of Models .....	15
<b>III. NETWORK VISUALIZATION AND ANALYSIS [ GORYACHEV ].....</b>	<b>17</b>
III.1 Some examples of network graphs .....	18
Qualitative network graphs, Metabolic network graphs .....	18
Gene regulatory network graphs .....	19
Towards a general graphical network representation .....	20
III.2 Methods and tools for network analysis and modelling .....	23
Stoichiometric kinetic modelling .....	25
Petri Nets, Boolean Networks .....	26
Bayesian Networks .....	27
Analysis of network topology .....	27
Network stability through circuit analysis .....	31
<b>IV. EXTRACTING AND ANALYZING A BIOLOGICAL MODEL [ AGUDA ].....</b>	<b>33</b>
IV.1 The G1-S transition in the mammalian cell cycle .....	33
IV.2 From a qualitative network to a kinetic model .....	34
IV.3 Computer simulation of the model .....	40
<b>REFERENCES .....</b>	<b>42</b>

## I. INTRODUCTION

The excitement in today's biology is driven by the huge amounts of information generated by high-throughput data-acquisition technologies, and by the expectation that these datasets will soon provide detailed understanding of life's processes. Ultimately, these datasets have to be integrated into a system-theoretic framework that should allow the study of the dynamics arising from networks of physico-chemical interactions orchestrating the physiology of a biological cell. The bioinformatics community is actively responding to this call for integration with the creation of a wide array of pathways databases. This tutorial will first provide an overview of these databases and existing graphical pathway representations. The underlying objective of the tutorial is to motivate the development of methods for extracting network models from databases. Models come at different resolutions, and pathways databases often provide only information on the connectivity (topology) of the interactions involved in a biological process. Thus, a unique feature of the tutorial is a discussion of a method of qualitative network analysis that the presenters think are appropriate for the treatment of uncertain or incomplete pathway datasets. Also summarized in the tutorial are existing methods and tools for network visualization, analysis, and simulation. Model extraction from databases cannot be automated at this time; however, we will explain how a modelling-focused utilization of pathways databases can be carried out. The modelling problem that is treated in this tutorial involves a switching behaviour of an enzymatic activity at the G1-S transition in the mammalian cell cycle.

## II. PATHWAYS DATABASES AND KNOWLEDGBASES

### II.1 PATHGUIDE

PATHGUIDE provides a list of more than 210 web-accessible biological pathways and networks databases. It is located at <http://www.pathguide.org> . The most recent paper describing this resource is the following:

Bader GD, Cary MP, and Sander C. (2006) "Pathguide: A Pathway Resource List," Nucleic Acids Research 34: D504-D506 (Database Issue)

As of April 2006, the number of databases under the following categories used in PATHGUIDE are as follows (some databases are in listed in more than one category):

<u>Categories</u>	<u>Number of databases</u>
1. Protein-protein interactions	86
2. Metabolic pathways	45
3. Signaling pathways	45
4. Pathway diagrams	23
5. Transcription factors/Gene regulatory networks	30
6. Protein-compound interactions	16
7. Genetic interaction networks	5
8. Protein-sequence focused	12
9. Other	13

Below are brief descriptions of the above categories as quoted from the reference given above (Bader, Cary & Sander, 2006):

1. Protein-protein interaction databases "mainly store pairwise interactions or complexes between proteins and sometimes other molecular interaction types."
2. Metabolic pathways databases "generally store a series of biochemical reactions in pathways involved in metabolite conversions."

3. Signaling pathways databases “generally collect sets of molecular interactions and chemical modifications (such as post-translational protein modifications) as regulatory pathways.”
4. Pathway diagrams databases “generally store hyperlinked pathways images.”
5. Transcription factors/Gene regulatory networks databases “capture transcription factors and the genes they regulate.”
6. Protein-compound interactions are interactions of proteins with non-protein compounds.
7. Genetic interaction networks databases are “composed of genetic interactions, such as epistasis and synthetic lethality, which occur when two mutations have a combined phenotypic effect that is not simply the sum of the effects caused by either mutation alone.”
8. Protein-sequence focused databases are “protein-sequence databases that store pathway information as secondary information.”
9. ‘Other’ databases refer to those that are uncategorized.

According to its creators, PATHGUIDE was designed to be “complementary to existing database link resources, such as Michael Galperin’s Molecular Database Collection

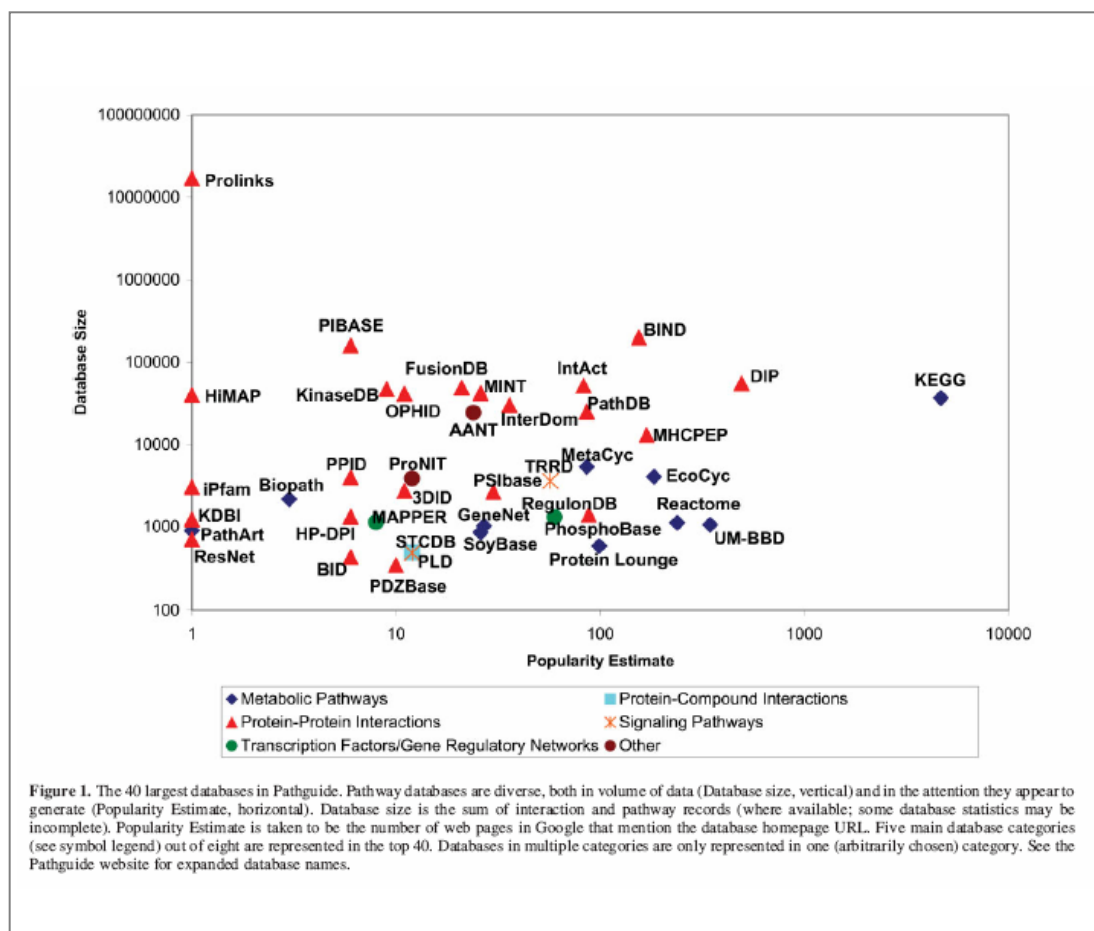
<http://www.oxfordjournals.org/nar/database/cap/>

and the UBC Bioinformatics Links Directory:”

[http://bioinformatics.ubc.ca/resources/links\\_directory](http://bioinformatics.ubc.ca/resources/links_directory)

PATHGUIDE highlights databases that are “free to all users and can be downloaded in a standard format such as the Proteomics Standards Initiative Molecular Interaction (*PSI-MI*) and *BioPAX* pathway data exchange standards, and the Systems Biology Markup Language (*SBML*) and *CellML* pathway simulations model exchange standards.” (Bader, Cary & Sander, 2006)

Figure 1 shows the 40 largest databases in PATHGUIDE plotted in a database size-popularity plane (Bader, Cary & Sander, 2006).



**Figure 1.** (from Fig 1 of Bader, Cary & Sander, 2006)

## II.2 PATHWAY DATA STANDARDS

A brief review of pathway data standards is given in the following reference:

Cary, M. P., Bader, G. D., and Sander, C. (2005) "Pathway information for systems biology," FEBS Letters 579: 1815-1820.

Shown in Fig. 3 of this reference (reproduced below) are the data coverage of the following pathway data formats:

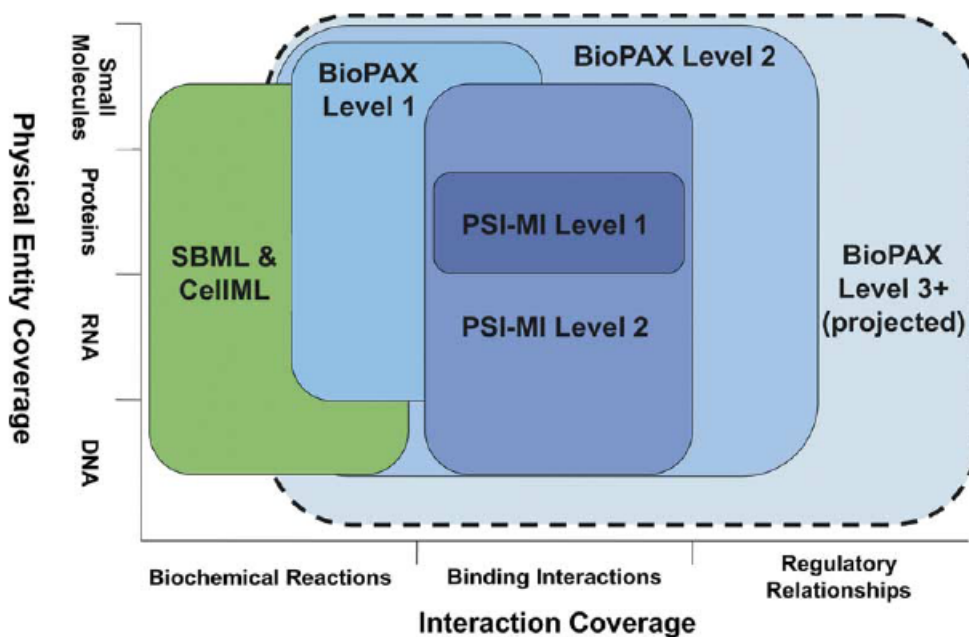
PSI-MI (Proteomics Standards Initiative's Molecular Interaction) is "a data exchange format for protein-protein interactions

(<http://psidev.sourceforge.net/mi/xml/doc/user/>)

SBML (Systems Biology Markup Language) is “a computer-readable format for representing models of biochemical reaction networks. SBML is applicable to metabolic networks, cell-signaling pathways, regulatory networks, and many others.” (<http://sbml.org/index.psp>)

CellML (Cell Markup Language) : stores and exchange mathematical models even if different model-building software were used. (<http://www.cellml.org/>)

BioPAX (Biological Pathways Exchange) (<http://www.biopax.org>) is “being developed by various database groups...Because many less-detailed data types that exist the pathway data space are difficult to represent in a highly detailed format, the BioPAX ontology allows representation of multiple levels of data resolution using an abstraction hierarchy.” (quoted from Cary et al., 2005).

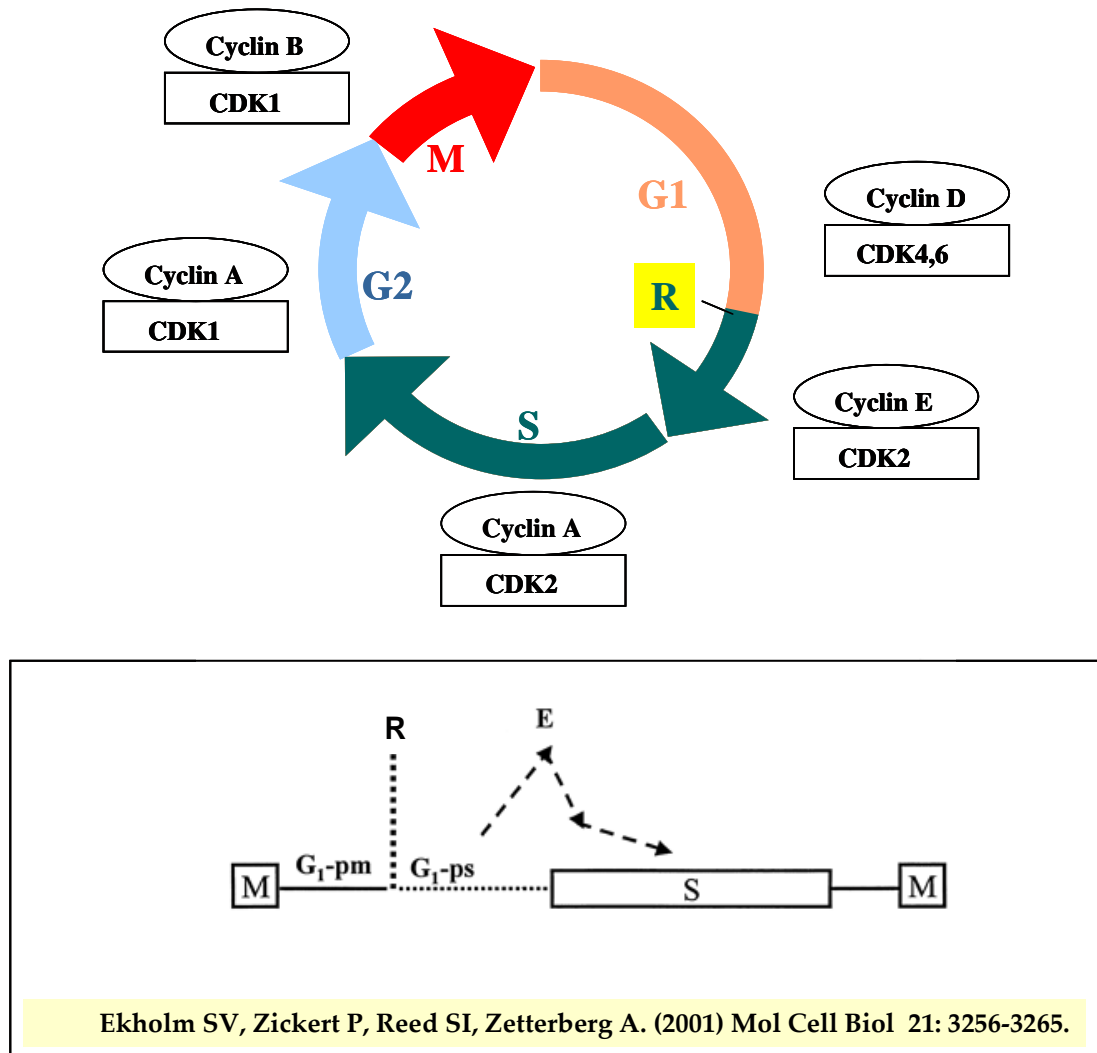


**Figure 2.** Coverage of pathway data formats (figure from Cary *et al.*, 2005)

### II.3 A MODELING-FOCUSED USE OF PATHWAYS DATABASES

Rather than enumerating and discussing a long list of pathways databases, we will consider a specific modelling problem to illustrate how one can extract relevant network information. The biological process we consider is the G1-S transition in

the mammalian cell cycle, and the specific modelling problem is to account for the switching behaviour of the kinase activity of Cyclin E/CDK2, a marker for entry into S phase (see Fig. 3 below).



**Figure 3.** The mammalian cell cycle showing the G1, S, G2, and M phases along with the predominant cyclin-CDK activities associated with each phase (top panel). The lower panel shows the position of the restriction point (R) which subdivides the G1 phase into G1-pm (post-mitosis) and G1-ps (pre-S-phase). After R and a finite induction period, cyclin E/CDK2 activity increases (shown by the dashed curve labelled 'E') as reported in the reference given below the graph (Ekholm et al., 2001).

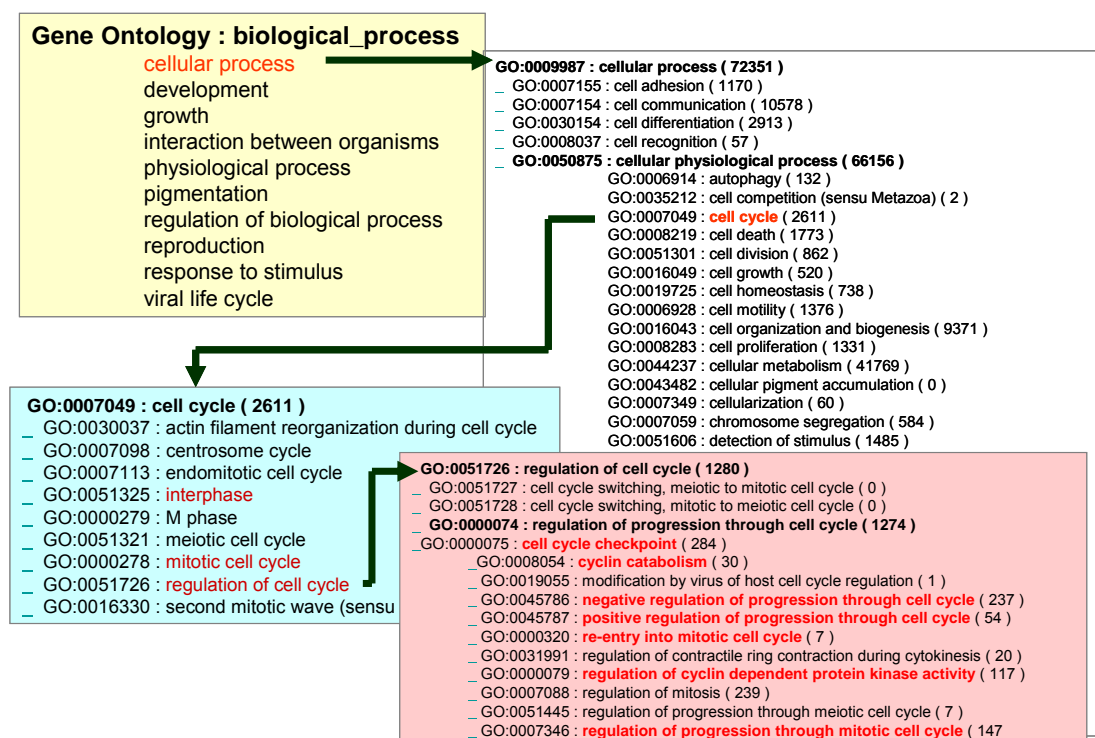
We will discuss the mammalian cell cycle in more detail later (in the last section of this tutorial). In this section, we will only show what databases are relevant and what problems are encountered before arriving at a working network model that has the potential to answer the biological question posed – i.e. *What is the mechanistic*



and kinetic origins of the switching behaviour associated with the restriction point? A few of the useful relevant pathways databases are discussed next.

## Gene Ontology (GO)

The omics revolution is providing a comprehensive parts list of biological cells. The Gene Ontology (GO) project aims for a controlled vocabulary for describing genes and gene products (<http://www.geneontology.org>). GO's classification and hierarchy of *biological processes* can be used as a starting point for identifying the parts list of the G1-S molecular machinery. Figure 4 below illustrates the search sequence used to generate a list of genes involved in the G1-S process. Unfortunately, the GO hierarchy is not a tree, and a GO term (e.g. cell cycle) could have many parents.



**Figure 4.** A search sequence to extract a list of genes involved in the regulatory network of the G1-S transition in the mammalian cell cycle.

## Kyoto Encyclopedia of Genes and Genomes (KEGG)

The preceding GO search will not give information on the structure of the G1-S network. One can begin to learn about the network by visiting pathways databases such as KEGG. Its URL is <http://www.genome.jp/kegg/> . There are 4 constituent databases in KEGG, but we will only mention two of them: PATHWAY and BRITE. KEGG PATHWAY is a collection of manually drawn pathway maps on

1. Metabolism
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Human Diseases
6. Drug Development (drug structure maps)

Items 1-5 represent the first level of the KEGG Orthology (KO), a pathway-based classification of orthologous genes. '*Drug Development*' (item 6 above) includes chronology of drug development, target-based structure classification, and skeleton-based structure classification.

KEGG BRITE is "a collection of hierarchical classifications representing our knowledge on various aspects of biological systems. In contrast to KEGG PATHWAY, which is limited to molecular interactions and reactions, KEGG BRITE incorporates many different types of relationships. Thus, the mapping of genomic and molecular data to KEGG BRITE (by the KO system) supplements the KEGG PATHWAY mapping for inferring higher-order functions....The KEGG Orthology (KO) system is the backbone of KEGG BRITE. It is a pathway-based classification of orthologous genes, including orthologous relationships of paralogous gene groups. The KO identifier, or the K number, is a common identifier for linking the gene and the pathway node, enabling automatic generation of organism-specific pathways." (quotes from <http://www.genome.jp/kegg/brite.html> ).

Figures 5A & 5B illustrate what pathway information on the cell cycle is obtained from KEGG.

## NETWORK HIERARCHY IN KEGG

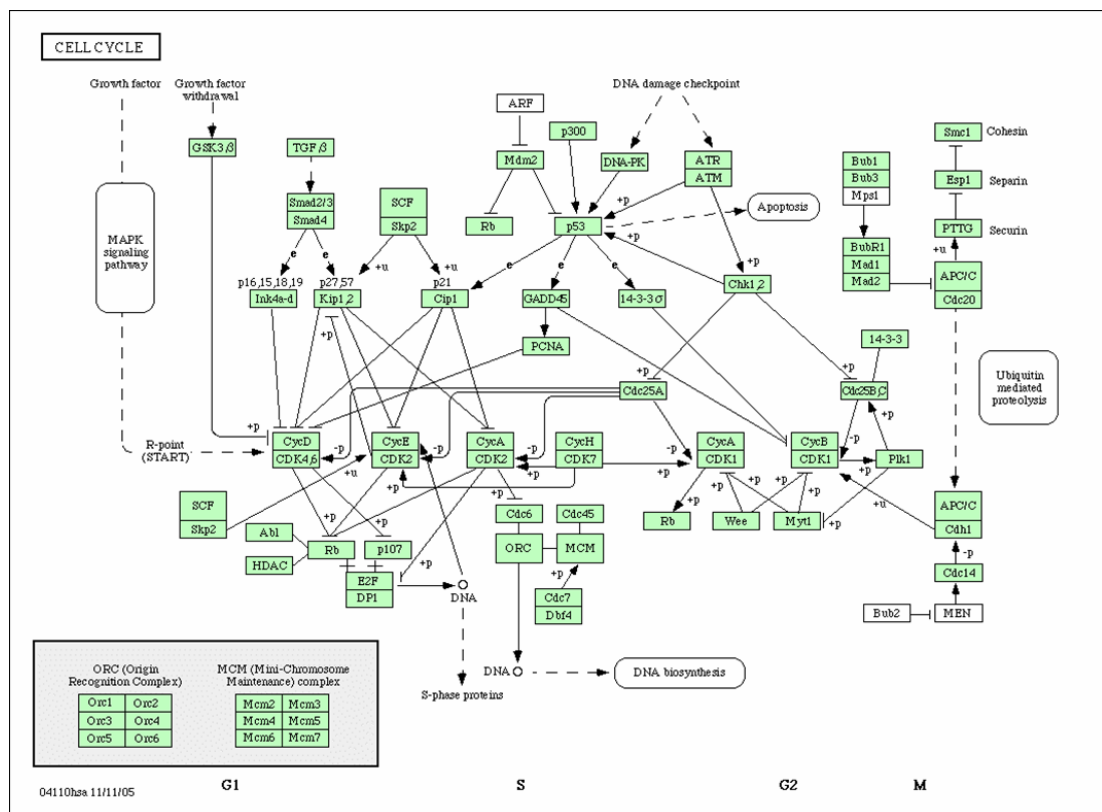
Metabolism  
Genetic Information Processing  
Environmental Information Processing  
**Cellular Processes**  
Human Diseases

### 01400 Cellular Processes

01410 Cell Motility  
01420 Cell Growth and Death  
04410 Cell division  
04420 Sporulation [GO:0030435 0030436]  
04430 Germination [GO:0009847]  
04110 **Cell cycle** [PATH:ko04110hsa]  
04210 Apoptosis [PATH:ko04210] [GO:0006915]

[CLICK TO SEE PATHWAY](#)

**Figure 5A.** Finding a cell cycle pathway map from the network hierarchy link in BRTE.



**Figure 5B.** Pathway diagram of the cell cycle (H. sapiens) from KEGG.

The boxes in the pathway map are clickable if there is information stored in them in the database. Clicking on other processes or modules (e.g. 'MAPK signalling pathway' or 'Apoptosis') will open up the corresponding pathway map. Clicking on

the 'Help' button located at the top right of the map will show the legend for the graphical objects (boxes and edges) used. Beyond this legend, no links are provided for further information on the interactions (edges). More detailed information on a particular interaction may be found in binary interaction databases such as BIND and DIP (go to the PATHGUIDE list to link to these protein-protein interaction databases). An unsatisfactory feature of the map given in Fig. 5B is the assignment of parts of the network to the G1, S, G2, and M phases of the cell cycle (see bottom of map). One must always remember that, at least at this point in time, pathway maps such as those in KEGG embody the curators' interpretation of available literature information and are tentative.

## Reactome

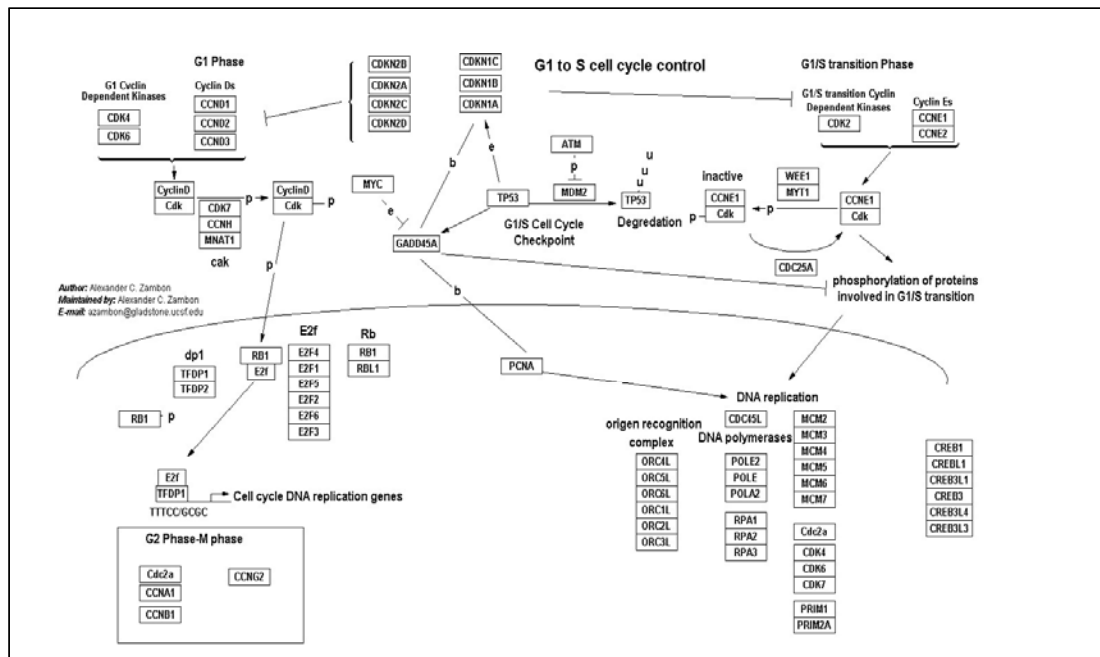
*Reactome* is a curated knowledgebase of human biological pathways which operates like a scientific journal in the sense that specialists in particular biological topics (i.e. *biological processes selected by the editors*) are invited to provide experts' reviews which are subsequently substantiated with bioinformatic weblinks by in-house curators. The URL is <http://www.reactome.org>. Reactome is a collaborative project among the Cold Spring Harbor Laboratory (USA), European Bioinformatics Institute, and the Gene Ontology Consortium.

The Table of Contents (TOC on the main menu bar) and the Pathway Topics List on the home page of Reactome give listings of the curated biological pathways. For our G1-S modelling problem, clicking on *Cell Cycle, Mitotic Hs* leads to another link called *G1/S transition [homo sapiens]* which can be perused to learn more about the process. A useful Reactome tool is *Pathfinder* which can be used to identify or discover pathways between a starting molecule, gene, or event and a terminating molecule, gene, or event.

## GenMAPP

Unfortunately, *Reactome* does not provide good pathway maps that integrate the interactions described in detail under each biological process. GenMAPP (<http://genmapp.org>) contains a database of pathway maps contributed by users, including some that are translated from *Reactome* (see Fig. 6). A clickable listing of human pathway maps could be found at

[http://genmapp.org/HTML\\_MAPPs/Human/MAPPIndex\\_Hs\\_Contributed.htm](http://genmapp.org/HTML_MAPPs/Human/MAPPIndex_Hs_Contributed.htm)

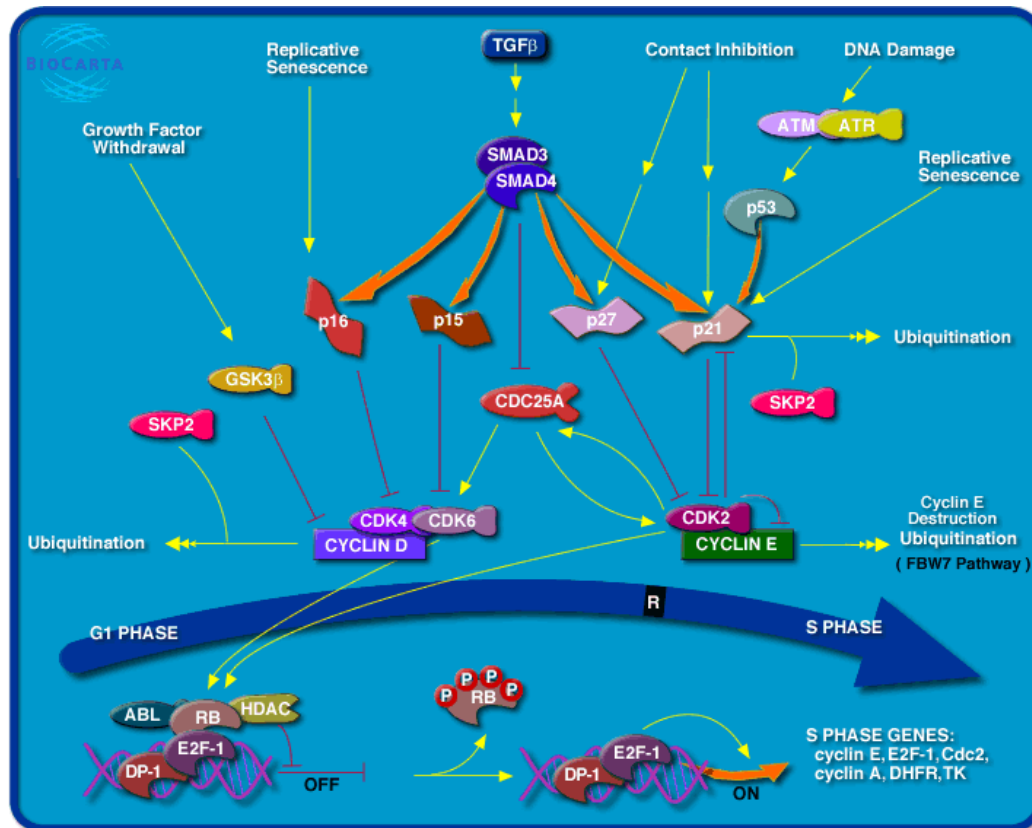


**Figure 6.** The Hs\_G1\_to\_S\_cell\_cycle\_Reactome.mapp from GenMAPP (see URL preceding this figure). This map was translated from *Reactome* by A. C. Zambon.

## Biocarta

The company *BioCarta* is a “developer, supplier and distributor of uniquely sourced and characterized reagents and assays for biopharmaceutical and academic research” (quoted from Biocarta’s URL : <http://biocarta.com>). This website contains maps of pathways that are subjects of active research. For the subject on G1-S cell cycle transition, one finds at least 6 relevant pathway modules, namely

1. Cyclins and Cell regulation
2. Cell Cycle: G1/S checkpoint (see Figure 7 below)
3. Regulation of p27 phosphorylation
4. CDK regulation of DNA replication
5. Cyclin E destruction pathway
6. Influence of Ras and Rho proteins on G1 to S transition



**Figure 7.** The pathway module called *Cell Cycle: G1/S checkpoint* in Biocarta (contributed by *Cell Signaling Technology*). A legend for the meaning of the symbols (edges and molecules) is provided on the same page. The genes/proteins can also be clicked to open windows of information on genes, proteins, Biocarta products, references, etc.

*Pathguide* lists other sources of pathway maps which should be consulted for more detailed information. The small number of pathways databases discussed above already provides a good start for sketching a network model focusing on the regulation of cyclin E/CDK2 which we assume to be the primary marker for the G1-S transition. Admittedly, the extraction of the G1-S network model discussed in the last section of this tutorial was largely guided by published review papers on the subject ; one can carry out a *Pubmed* search to search for these papers at:

<http://www.ncbi.nlm.nih.gov>

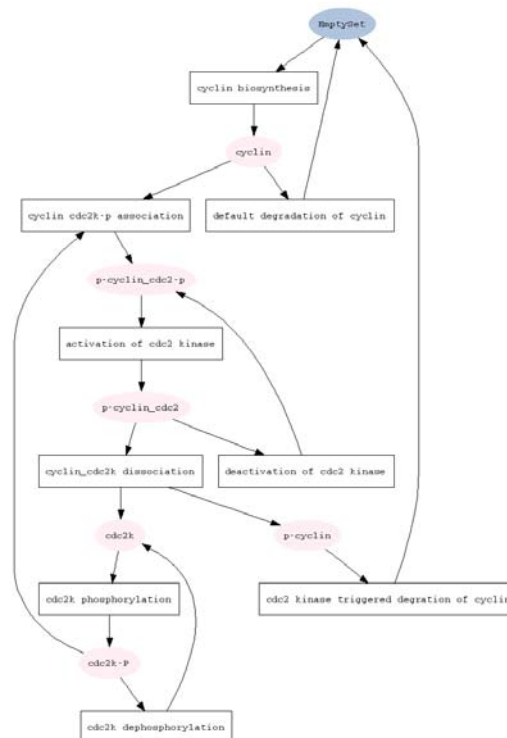
## II.4 REPOSITORIES OF MODELS

Mathematical and computational modelling of biological pathways is, of course, not new but this activity has recently been stimulated by the availability of large amounts of data generated by omics technologies. The recent creation of online repositories of models is a welcome development because they promise to gather and standardize model representation so that models can be conveniently shared and interpreted by members of the modelling community. The URLs of these model repositories are given below.

Biomodels Database at EBI: <http://www.ebi.ac.uk/biomodels/>

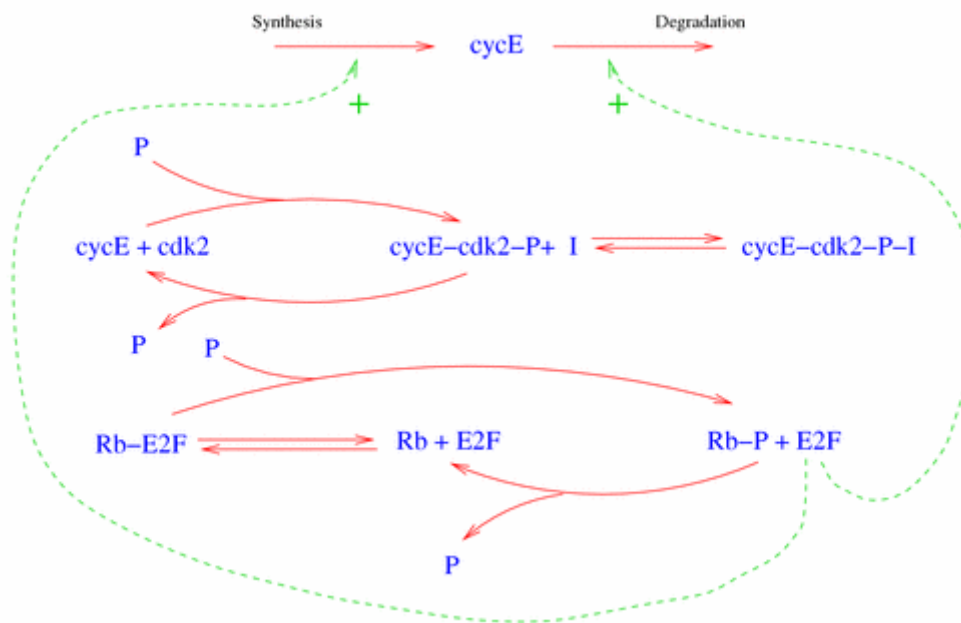
CellML at the Univ Auckland: <http://www.cellml.org>

The Biomodels Database is a “free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems” (quote from the Biomodels URL above). CellML aims to “store and exchange computer-based mathematical models” (quote from the CellML URL above). An example of a model graph from the Biomodels Database is the cell cycle model shown in Fig. 8. This model (kinetic equations and parameters) can be downloaded from the website using various formats such as SBML, CellML, SciLab, and XPP.



**Figure 8.** Graph view of model Tyson1991\_CellCycle\_6var from Biomodels Database.

Figure 9 below is a G1-S model network from CellML.



**Figure 9.** The Hatzimanikatis model of the G1-S network in the mammalian cell cycle (downloaded from the CellML database).



### III. NETWORK VISUALIZATION AND ANALYSIS

By ‘network visualization’ we mean the graphical representation of networks of molecular processes and interactions. Ideally, a network graph would contain all the details (or at least clickable links to them) of the individual interactions and processes comprising the network. Beyond these local details, a network graph is essential in understanding the associated biological process because the graph as a whole embodies system-level properties emerging from the coupling (or topology) of the interactions among the network components. It is these non-intuitive emergent global properties that are often the object of computer-based mathematical modeling. In this section, we summarize the activities in the systems biology community that are geared towards the development of standards of graphical representation of networks; we also survey existing methods and tools of network analysis and computer simulation of models.

With the view of developing a kinetic model, a network graph is used to extract the two essential model components, namely, a set of *dynamical variables* and a set of *interaction functions* corresponding to the network topology.

Every kinetic model assumes a set of dynamical variables that sufficiently describes the state of the system. A network graph is composed of nodes and edges where the nodes correspond to entities (molecules, genes, proteins, complexes, even pathway modules) that are connected by edges (often directed to signify causality of the interaction). In other graphs, interactions themselves are considered nodes and an edge between an entity node and an interaction node could mean either ‘the entity is a substrate or reactant of the interaction’ or ‘the interaction gives rise to or affects the entity’. Only entity nodes may correspond to dynamical variables in kinetic models. Depending on the resolution of the model, a dynamical variable may correspond to one entity node or a set of nodes (modules).

A directed interaction edge signifies that the state of the target node is a function of the state of the source node. In a dynamical model, the instantaneous state of a given target node would be equal to the algebraic sum of the interaction functions associated with all source nodes. These interaction functions are referred to as the ‘kinetics’ or ‘rate expressions’ of the interactions.

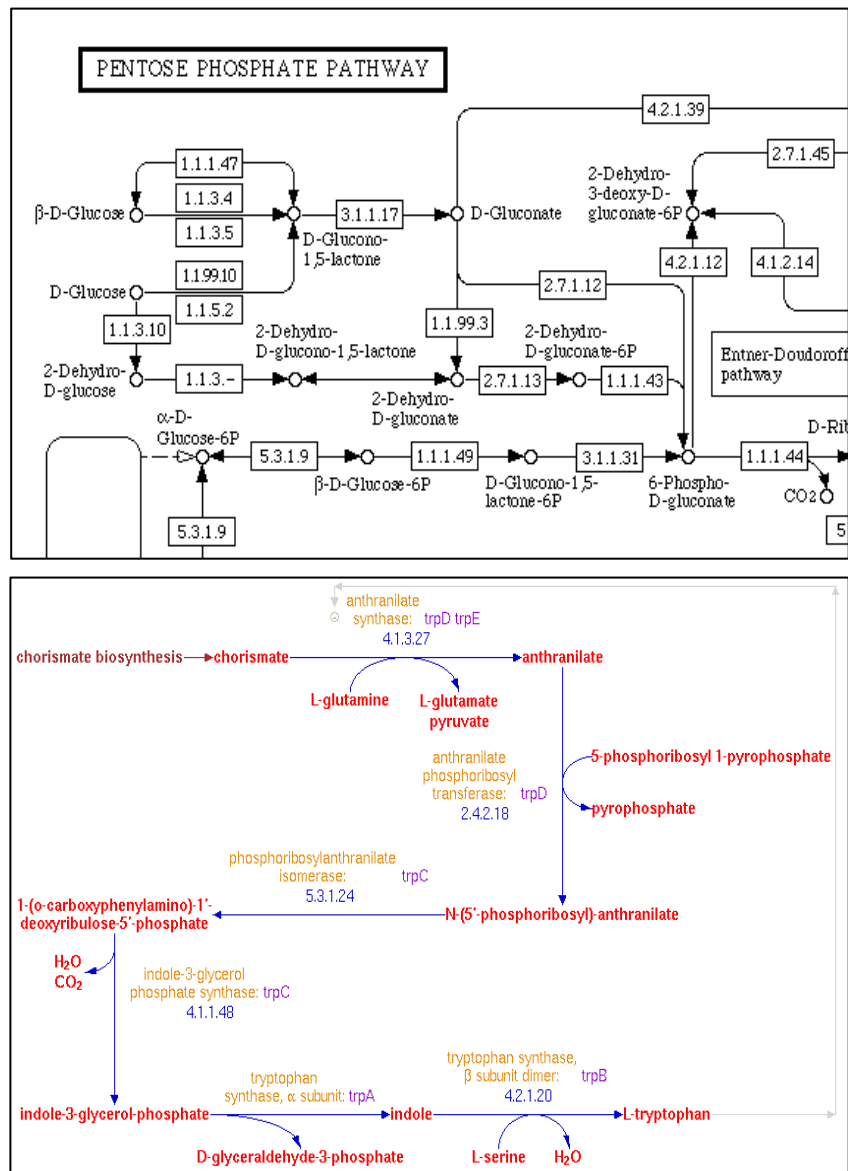
### III.1 SOME EXAMPLES OF NETWORK GRAPHS

#### Qualitative network (qNET) graphs

In the primary literature, most molecular biologists and geneticists present pictures of pathways by using what we call qualitative network graphs where only arrows and hammerheads are shown (see, for example, Fig 7 above). These qNET graphs are directed binary interactions where *arrows* could mean any of the following: ‘activates’, ‘induces’, or some positive influence that increases the level of the target node; and *hammerheads* could mean ‘inhibits’ or some other negative influence that decreases the level of the target node. Admittedly, the meaning of these arrows and hammerheads is not well defined. In the last section of this tutorial, we will give clear definitions and say more about the utility of these qNET graphs in assessing network stability. Majority of the available information on biological pathways and networks is at the qNET level – this is really the motivation why we must find ways to analyze qNET graphs to generate valid conclusions despite the incompleteness and uncertainty of the data.

#### Metabolic network (MBN) graphs

Sample graphs of metabolic networks from the two most popular databases KEGG and EcoCyc are shown in Fig 10 below. The nodes are low molecular metabolites and the reactions are characterized by functional classes of enzymes which are abstracted to standardized EC numbers. The MBN graph has a clear and simple semantics that is easily amenable to mathematical analysis and, in fact, co-evolved with a number of analytical methods and tools, such as metabolic control analysis and stoichiometric network analysis.

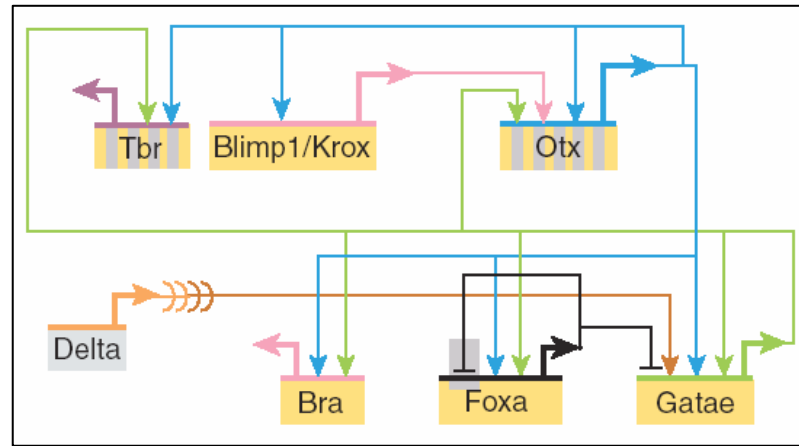


**Figure 10.** Top panel: a portion of the pentose phosphate pathway from KEGG. Bottom panel: tryptophan synthesis pathway from EcoCyc.

### Gene regulatory network (GRN) graphs

Gene Regulatory Network (GRN) notation is another example of a specialized representation developed to capture regulatory relationships specifically within gene networks, or even more precisely, among various transcription factors. The notation, originally developed in Davidson's group at Caltech (Yuh et al., 1998), is centered on cis-regulatory elements of genes and their positive or negative regulation by other genes. Several extensions of the GRN notation were provided by Bolouri (Longabaugh et al., 2005) and Arkin (McAdams and Arkin, 1997) to include

description of proteins and their interactions. However, the GRN notation is a DNA-centric approach which does not cover signal transduction or metabolic networks. Presently the GRN notation is supported by the *BioTapestry* software being developed by the Bolouri group (Longabaugh et al., 2005).



**Figure 11.** An example of a GRN graph from (Davidson and Erwin, 2006).

### Towards a general graphical network representation

**MIM.** *Molecular Interaction Maps* (Kohn, 1999; Kohn, 2001) represent one of the first attempts to develop a general graphical notation suitable for description of any molecular network but specifically geared towards signaling networks. MIMs are also based on nodes and edges; however complexes are denoted differently from “single” molecules (see Figure 12 for an example). The semantics of notation is rich and allows representation of enzymatic activity, molecular modifications, formation of large molecular complexes, etc. The set of notations was designed to represent both qualitative and detailed mechanistic interactions. The development of the MIM notation also uncovered issues and pitfalls that are blocking the way to a universal network notation. Thus, to provide representational richness the notations become relatively complex and their correct and unambiguous interpretation may require prior domain knowledge which, in turn, aggravates the problem of being machine readable and writable.



have multiple states. To conserve graphical space, all states of the biomolecule contained in the same object. EPN is also meant to provide a more convenient interface for machine readability and writability. Less human-readable but more compact EPN can be developed into the more human-readable but much more verbose PDN on the per-module basis to enable both perception and efficiency of information presentation. At the moment this notation is developed and supported by the Edinburgh Pathway Editor.

**PATIKA.** The acronym stands for *Pathway Analysis Tools for Integration and Knowledge Acquisition* (Demir et al., 2004; Demir et al., 2002). Patika represents another example of a conceptual platform and graphical notation that provides modularization, encapsulation and hierarchical representation of interaction networks. The primary semantic elements are “state”, “complex”, “transition”, “compartment” and “abstraction”. “Abstractions” of states and transitions serve to implement modularization of networks as well as to incorporate the uncertainty of biological data.

**SBGN.** Presently, the need to develop a standard notation for representing biomolecular networks became well appreciated by the systems biology and bioinformatics community. The *Systems Biology Graphical Notation* (SBGN) consortium was formed to discuss problems, identify contradictions in notation and find solutions. A pathbreaking first workshop of SBGN took place in Tokyo in February 2006 to set the course for the much-anticipated unification of various graphical notations. A complementary effort is currently being undertaken by the XML language communities represented by SBML, CellML and BioPAX projects (Gauges et al., 2006). The goal of this effort is to enable a next-generation of SBML and other exchange standards for graphical representation of networks in a clear and unambiguous format, independent of particular platforms and graphical notations. Convergence of the SBGN and the SBML-CellML-BioPAX efforts, long awaited by the systems biology community, should result in a free and effortless sharing of pathway and network information among scientists from multiple disciplines and heterogeneous backgrounds.

### III.2 METHODS AND TOOLS FOR NETWORK ANALYSIS AND MODELLING

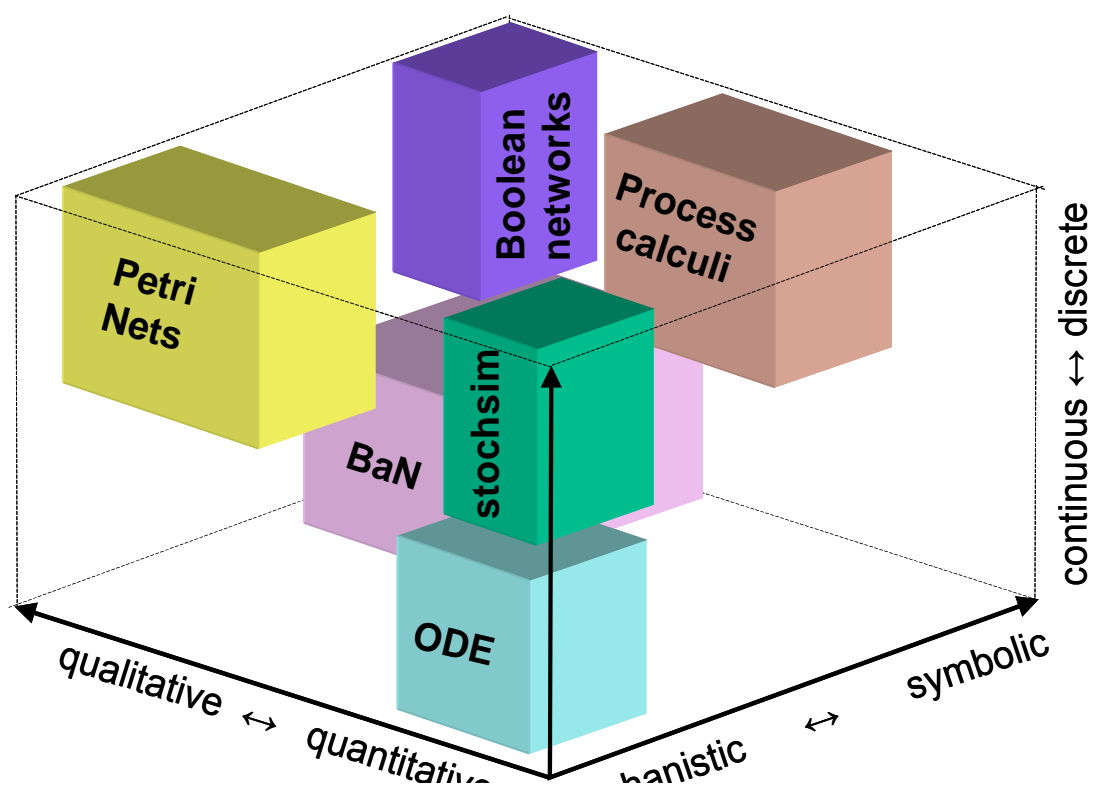
Once the elements of the network have been identified and the relationships between them have been established, the dynamical behavior of the network can be analyzed by various methods through the construction of a network model. Typically, “simulation” of the model implies some computational process that attempts to imitate the temporal (and sometimes spatial) dynamics of the actual network. Simulation is often conceptually straightforward, and multiple tools are available. In contrast, “analysis” usually implies that some qualitative conclusions are derived about the behavior of the network without explicitly simulating its dynamics. In this part of the tutorial we start by surveying the space of modeling methods and exploring this space beginning with the conventional simulation methods to the more qualitative, symbolic, and analytic techniques. We briefly cover foundations of the formalism of stoichiometric network analysis and metabolic control analysis, with emphasis on available tools and potential applications rather than on the details of their mathematical formalism.

The choice of method or approach for analyzing a network depends on the levels of certainty and detail on network data, as well as on one’s research objective. Two complementary approaches can be formulated as follows. The more “brute force” approach relies on numerical simulation to arrive at a qualitative insight. Typically one would begin by formulating a detailed mechanistic mathematical model that is equipped with kinetic parameter values. The model is then numerically simulated over a wide range of parameters. These computer simulations sometimes suggest ways to reduce the model to smaller, more manageable size through, for example, elimination of fast and slow variables. Qualitative insights can be also obtained from numerically computed bifurcation and phase diagrams. Alternatively, one can start by applying methods of qualitative analysis to reduce the complexity of the model prior to embarking on numerical simulations. For example, one may analyze the topology of the network to identify the “elementary” or “principal” fluxes, and identify which among these are responsible for a particular qualitative phenomenon, such as bistability. A reduced or minimal model can sometimes be generated by identifying the smallest number of these fluxes that could still retain the qualitative behavior being modeled.

A large variety of modeling methods is presently available for researchers interested in biological pathways and networks (Alves et al., 2006; Pettinen et al.,

2005). Before delving into the details of some of these methods, we first give an overview of the full spectrum of methods. An early approach to modeling biological systems - dating back to the works of Jacob, Monod, Volterra and Lotka - are almost entirely based on the use of ordinary and partial differential equations. This approach is deterministic. More recently, significant interest has been paid to the use and development of stochastic methods which allow the investigation of noise in biological systems.

In the past decade, significant efforts have also been put toward the development of qualitative and symbolic methods of modeling. In the early 1970s Kaufmann and Glass developed the formalism of Boolean networks to predict qualitative behaviors of gene networks. Later, several methods borrowed from computer science, such as Petri Nets and Bayesian networks, were adopted for the analysis of biological systems. Process calculi is a more recent addition to this collection (see Fig 13).

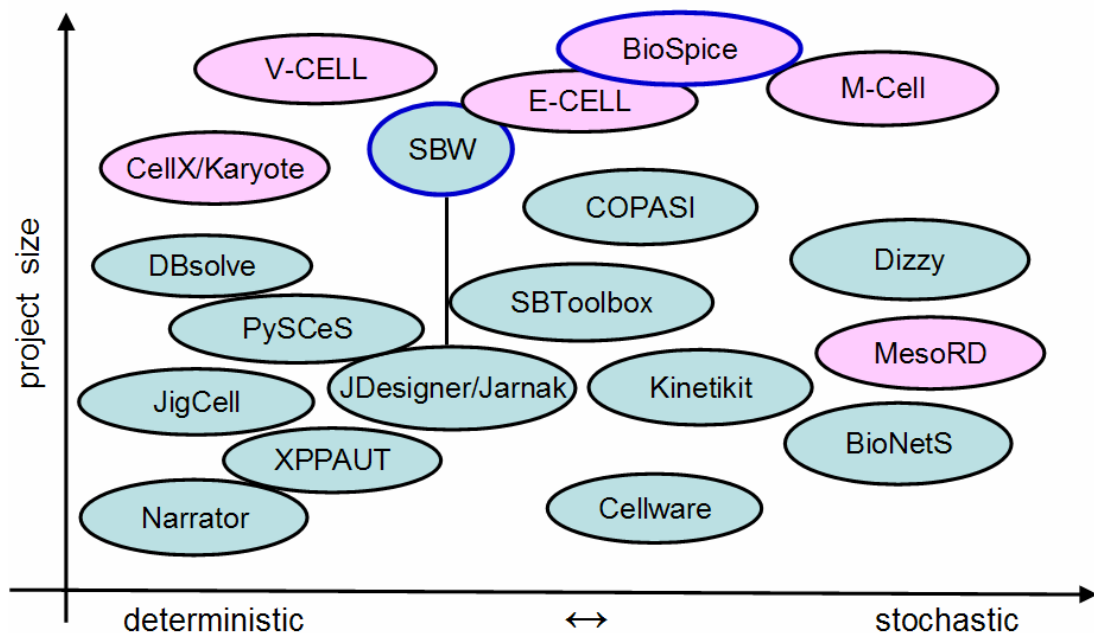


**Figure 13.** Space of modeling methods



## Stoichiometric Kinetic Modeling (SKN)

To perform a stoichiometric or chemical kinetic modeling, a network should be “translated” into the stoichiometric kinetic notation. Many modern simulation packages allow the user to enter diagrammatic information while some older packages require direct inputs of equations. To define a model, the user needs to supply connectivity information (graphically), define the compartments (volume), species (names, concentrations or copy numbers) and reactions (kinetic law, reaction rate constants). Once the model is fully defined, its simulation can be performed either deterministically or stochastically and the simulation output will be generated usually as a time series of variable concentrations or molecule copy numbers. It should be kept in mind that for stochastic simulations, the steps in the mechanism must be elementary steps (usually first or second order reactions) that correspond to mass-action rate laws, as opposed to composite rate expressions such as Michaelis-Menten or Hill-type functions. The example given on the slide represents the classical Michaelis-Menten kinetic diagram for the transformation of substrate S into product P catalyzed by an enzyme E. The mechanism, as defined here consists of only mass-action rate laws so it can be simulated either deterministically or stochastically.



**Figure 14.** Various software tools used to simulate kinetic models.

### **Petri Nets (PNs)**

The Petri Net formalism was initially introduced as an analytical tool for testing concurrent processes in computer engineering in the early 1960s. Only recently was it realized that this formalism can also be used for modeling biological systems (Goss and Peccoud, 1998). In their basic formulation, PNs are equivalent to the SKNs with “places” equal to “species” and “transitions” equal to “reactions”; however, PNs assume discrete values for species numbers (“marking”, “tokens”). This concept was extended in hybrid functional Petri nets (HFPNs) to admit both discrete and continuous values for place markings. This flavor also introduced ‘test’ and ‘inhibitory’ arcs to represent closer features found in signaling networks. To achieve more quantitative approximation of chemical kinetics, the formalism of stochastic Petri Nets can be applied to essentially mimic the behavior of the Gillespie algorithm (Peleg et al., 2002). More recently, colored Petri Nets (CPNs) (Mandel et al., 2004) have been applied to introduce hierarchical representation of biological systems with increasingly more fine-grained description. A variety of computational tools exist for the three major flavors described; however, none of these has been designed specifically to model biological networks and pathways. Thus expert-level knowledge in the PN domain is generally needed to operate these tools.

### **Boolean Networks (BNs)**

Boolean networks were proposed by Kauffman and Glass as a ‘simple’ model for gene interaction networks (see e.g., (Kauffman, 2004; Perkins et al., 2006) and references therein). An advantage of Boolean networks is that they offered a biologically plausible and computationally tractable model when virtually nothing is known about the details of gene expression control. BNs are cellular automata with simple transition rules and are close relatives of neural networks. Original BNs, heavily influenced by cybernetics, had only Boolean variables (0,1) and logical transition rules. Later transition rules have been generalized, e.g., as shown on the slide, and variables were allowed to take a finite range of values. Another recent development in the area is the introduction of probabilistic Boolean networks (PBN) that allow many transition functions for each node of the network which are chosen at random with a given probability (Shmulevich et al., 2002). The main proposed use of various BNs is the inference of gene regulation networks from genomic data.

## Bayesian Networks (BaNs)

In Bayesian networks (BaN) the edges represent causality between the nodes through the statistical dependence. Usually, many networks will fit the experimental data. Further analysis is required to narrow down the selection of potential networks. As such, BaN have been used as an analytical (Friedman et al., 2000) rather than simulation tool.

## Analysis of network topology

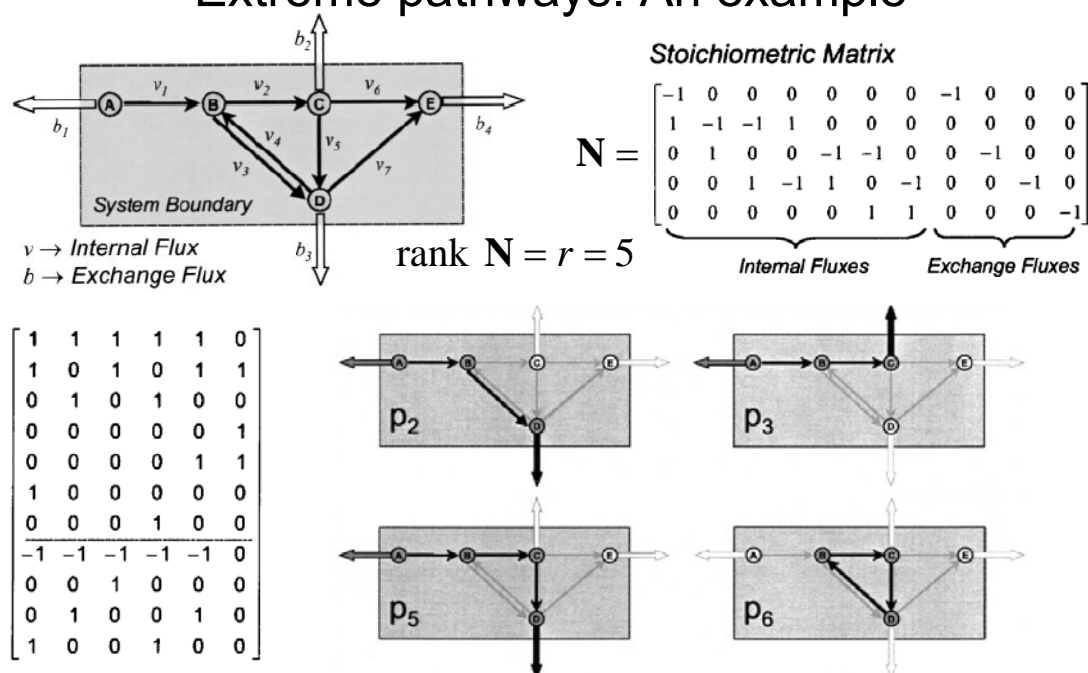
**Graph-theoretic Analysis.** Application of graph-theoretic approaches to large-scale networks, regardless of the nature of the component interactions, resulted in the modern theory of “network biology” mainly developed by Barabasi and colleagues (Barabasi and Oltvai, 2004; Ravasz et al., 2002). The main topological characteristics of the networks are the distributions of connectivity  $P(k)$  and modularity  $C(k)$  or clustering coefficient. The majority of biological networks show scale-free distribution characterized by exponential dependence of  $P$  and  $C$  on  $k$ . While scale-free networks are not necessarily modular, many natural networks also show modularity which requires that  $C(k)$  drops with  $k$ , for hierarchical networks  $C \sim 1/k$ . The scale-free property of biological networks, which has been demonstrated for some networks, such as metabolic and protein-protein interaction, endows them with some structural robustness properties but also leaves them open to catastrophes upon removal of the “hubs”, the highly connected nodes that are responsible for the connectivity of the network. Biological networks with exponent  $2 < \gamma < 3$  are also ultrasmall in a sense that their average node connecting path scales as  $L \sim \log \log N$  while for random networks it is only  $\log N$ . Perhaps the most popular general-purpose tool specifically to work with large networks, such as protein-protein interaction networks, is *Cytoscape* from Ideker’s group. Among many “plug-ins” that extend the functionality of this tool is *Network Analyzer* which computes a number of network topology statistics and probability distributions.

**SNA.** This is the acronym that Bruce Clarke gave for the formalism he developed and called *stoichiometric network analysis* (Clarke, 1988). Central to the use of stoichiometric methods is the concept of the stoichiometric matrix (SM)  $\mathbf{N}$  which describes connectivity and stoichiometry of the reaction network. The power of the approaches based on the analysis of the SM lies in the fact that a number of qualitative conclusions can be made regarding the properties of the network

regardless of the detailed definition of reaction fluxes which are usually complex nonlinear functions of the species concentrations. Introduction of the SM results in a linear relationship between the time derivative of the vector of  $m$  species concentrations and the vector of  $n$  reaction fluxes. To ensure that all reaction fluxes are non-negative, reversible reactions are often represented as two opposite reaction fluxes following the work of Clarke.

In the stationary state, the vector of fluxes satisfies a very simple equation. If the rank of the SM is  $r$  then all possible stationary flux vectors are found within the so-called null space of the SM with dimension  $n - r$ . It is convenient to introduce a so-called kernel matrix  $\mathbf{K}$  with dimensions  $n \times n - r$  that consists of the null-space basis vectors. Incidentally, the kernel matrix allows to express all reaction fluxes through the independent reaction fluxes as shown on the slide. However, arbitrarily chosen basis vectors of the null-space are not unique and do not have any biochemical meaning. Fortunately, using convex analysis it is possible to overcome this problem. It can be shown that all admissible flux vectors lie within a  $(n - r)$ -dimensional convex cone within the positive orthant of the  $n$ -dimensional flux space. The edges of this cone are so-called “*extreme currents*” and their number is generally higher than the dimension of the null-space. Interestingly, although this means that algebraically they are dependent, biochemically they are still independent because they cannot be expressed through each other using only **non-negative** coefficients. On the network diagram, such extreme currents correspond to characteristic pathways with none of them being a subset of another. Schuster and co-workers relaxed the requirement for non-negativity of fluxes introducing ‘*elementary modes*’ as the cone edges in this situation (see for review (Papin et al., 2003)). Finally, Schilling (Schilling et al., 1999) and colleagues introduced an intermediate concept of ‘*extreme pathways*’ by assuming non-negativity for internal fluxes and arbitrary sign for the exchange reactions with the ‘environment’ of the system. This concept is illustrated in Fig 15.

## Extreme pathways: An example



**Figure 15.** An example of how to determine the extreme pathways in a network.

Mathematical analysis of stoichiometric matrices and especially extraction of the extreme currents, elementary modes and pathways is a nontrivial task and should be performed with dedicated tools. Publicly available software for SNA has become available since the end of 1990s. The most recent developments are *CellNetAnalyser* and *SNA toolbox* which are toolboxes for Matlab and Mathematica respectively.

What can be done once the extreme currents are computed? Originally SNA was developed by Clarke specifically to analyze the stability of stationary states of large reactions networks. The mathematical apparatus of this approach is beyond the scope of this tutorial. Importantly, SNA was applied to derive a reduced model of a complex network so that it preserves certain characteristic behavior. Thus Aguda and Clarke (Aguda and Clarke, 1987) used SNA to derive a reduced model for the bistability behavior exhibited by the peroxidase-oxidase reaction. This approach involved identification of those extreme currents that are necessary to preserve the bistable behavior of the whole system. More recently SNA was applied to identify criteria for the emergence of calcium oscillations in olfactory cilia (Reidl et al., 2006). Papin and Palsson (Papin and Palsson, 2004) presented an example of

how SNA can be used to analyze the structure and properties of a generic signal transduction pathway on the example of JAK-STAT pathway.

**Metabolic Control Analysis (MCA).** MCA (Hofmeyr et al., 2002) is another theoretical approach based on the analysis of the stoichiometric matrix. The main question addressed by MCA is how the stationary state of the network described by the set of all stationary concentrations  $\mathbf{S}$  and reaction fluxes  $\mathbf{J}$  is influenced by the perturbation introduced into the system parameters. In the classical metabolic context, these parameter perturbations are normally achieved by altering concentrations or molecular properties of the enzymes that catalyze the corresponding reactions. To reach its goal, MCA introduces a number of differential characteristics which define sensitivities of stationary quantities to the change in concentrations and parameters. “Local” properties that describe how the individual reaction rates  $\mathbf{v}$  depend on the reactants and parameters that are directly involved in these reactions are called elasticities and are expected to be experimentally measured. Indeed for a great number of biochemical systems these parameters can be relatively easily measured *in vitro*. Of interest, however, are the “global” characteristics, response and control coefficients that represent systemic properties of the entire network. Experimental measurements of these characteristics would require *in vivo* experiments which are normally difficult. The main advantage of the MCA is the derivation of the algebraic relationships between hard to obtain coefficients and elasticities based only on the topology of the network as encoded in the stoichiometric matrix.

MCA relates global properties to local properties through the so-called *summation and connectivity theorems* that relate matrices of control coefficients with the stoichiometric matrix through the kernel matrix  $\mathbf{K}$  and link matrix  $\mathbf{L}$ . These matrices describe the linear dependence between the network fluxes (columns of  $\mathbf{N}$ ) and the species (rows of  $\mathbf{N}$ ). In the example shown on the slide a simple system of 3 species  $\mathbf{S}$  is connected by 4 reactions. Four additional “boundary” species  $\mathbf{X}$  are assumed to be kept constant by external processes are thus the parameters of the system together with the reaction rate constants. The final outcome of the MCA analysis here is the system of linear equations that relates all control coefficients with the stationary fluxes, concentrations and elasticities which are assumed to be measured experimentally. For the systems of practical size, all the above calculations, of course, are performed using software tools.

A number of software tools provides support for MCA analysis, usually by calculating elasticities, response and control coefficients numerically. The majority

of these tools are also kinetic modeling/simulation engines. Among the available tools the most prominent are PySCeS (Olivier et al., 2005) which evolved from the first MCA tool MetaMod, JDesigner/Jarnac and Copasi (the latest, and improved version of Gepasi). Importantly, Jarnac can be run as a module of the broker applications SBW and BioSpice, providing additional functionality and application data exchange. Some other tools like BioSens and MetaFluxNet are not MCA tools in the strict sense but provide additional and complimentary functionality such as sensitivity analysis and metabolic flux balance analysis, respectively.

Sensitivity coefficients computed using MCA can be useful for the understanding of the organization of the network. An example (Goryachev et al., 2005) shown on the slide presents a bacterial quorum sensing network which operates as a bistable switch that is flipped by accumulation of the communication molecule, termed autoinducer, in the extracellular environment. Calculation of sensitivity of the transcription factor concentration to variation in the reaction rates revealed parts and submodules of the network which are responsible for the maintenance of the transcription factor in the “on” and “off” states. Interestingly, the analysis showed that most of the submodules have non-overlapping functions as they control either “on” or “off” states and rarely both. While in a relatively small network, as in this example, the function of network components can be inferred directly from the simulations, in larger networks, analytical approaches, such as MCA, could potentially offer an advantage over brute force simulation approaches.

### **Network stability through circuit analysis**

Further generalization of the network topology analysis assumes that only the signs of relationships between the network nodes are known in the qualitative “activates” – “inhibits” terms. Since this method is considered in detail in the next section of our tutorial, only a brief exposition of the early results is given here. The approach is based on the qualitative analysis of the system’s Jacobian matrix. The major observation mentioned in works of Clarke (Clarke, 1988), Thomas (Thomas et al., 1995) and others is that only the network cycles contribute to the characteristic equation and therefore only cyclic paths influence the network stability. Defining closed feedback loops in the network and classifying them as “positive” or “negative” it is often possible to make conclusions about the potential instability of a stationary state of the network without knowing the kinetic details. Thus, it has been shown that a positive circuit is a necessary condition for multistationarity and a negative circuit is a necessary condition for stable oscillations. In the examples shown on the lecture slides (borrowed from Tyson’s paper (Tyson, 1975)) two mutually repressing

pathways are characterized by the Jacobian with both positive and negative circuits. Indeed, at various values of network parameters it exhibits oscillations and bistability.



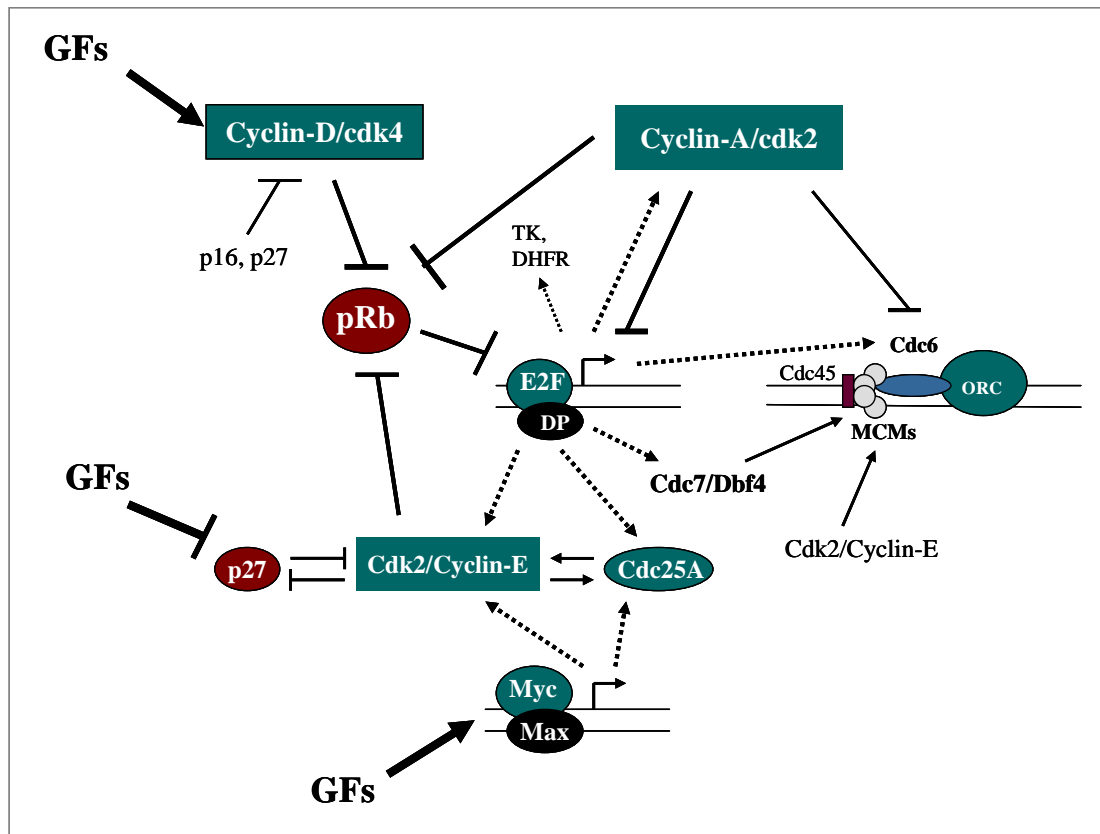
## IV. EXTRACTING AND ANALYZING A BIOLOGICAL MODEL

In this section, we give a detailed example of a biological model to illustrate the use of pathways databases and, more importantly, to show what valid conclusions can be generated despite the incompleteness and uncertainties of the information currently available; it is this latter aim that we believe is crucial in exploiting the current state of pathway information in databases.

### IV.1 THE G1-S TRANSITION IN THE MAMMALIAN CELL CYCLE

The modelling process usually begins with a question that focuses on specific phenomenon. We already stated our biological question in Section II.3 : *What is the mechanistic and kinetic origins of the switching behaviour associated with the restriction point?* The different phases of the eukaryotic cell cycle are shown in Figure 3. The restriction (R) point is 'located' in mid- to late G1 phase, and is often described as a commitment point for another round of DNA replication. The significance of studying R point regulation is underlined by the fact that almost all human cancers involve dysregulation of this G1 checkpoint (it is considered a checkpoint in the sense that if there is something wrong, such as DNA damage, then the cell cycle is arrested to give time for some DNA damage repair machinery to operate).

From consulting literature reviews and pathways databases (such as Biocarta, see Figure 7), a consensus qualitative network for G1-S regulation can be drawn, as shown in Figure 16. A brief description of this network is given in this figure's caption. Since the G1-S marker to be used for our model is cyclin E/CDK2, we focus on the interactions involving this kinase and find out how the switching behaviour of its activity is generated. Very often, such switches originate from some intrinsic instability of the network. We will make the assumption that a network instability causes the cyclin E/CDK2 switching behaviour, and then identify from Fig 16 a core subnetwork that exhibits this instability. We summarize in the next subsection the theoretical basis for the method we used in identifying this subnetwork.



**Figure 16.** The regulatory network of the G1-S transition in the mammalian cell cycle. Growth factors (GFs) trigger certain signalling cascades that lead to the activation of cyclin D/CDK4 complexes and to the inhibition of CDK inhibitors such as p27. Active CDK4 phosphorylates (thereby deactivating) the retinoblastoma protein (pRb) which inhibits entry into S phase due mainly to inhibitory binding with E2F transcription factors; these factors induce many of the genes required for S phase (such as members of the pre-replication complex, cyclin E, cyclin A, Cdc25A, etc.). The dashed arrows signify gene expression. Synthesis of cyclins E and A leads to activation of CDK2 which further phosphorylates (thereby deactivates) pRb. Another transcription factor, namely Myc, also contributes to the G1-S transition but this protein's regulation is not shown in the figure. Arrows mean 'activate' and hammerheads mean 'inhibit'.

## IV.2 FROM A QUALITATIVE NETWORK TO A KINETIC MODEL

The major steps we take in arriving at a kinetic model for R point regulation are the following:

1. Start with a qualitative network (qNET) that contains the core subnetwork you are interested in; this requires that you know a set of markers and processes that describes the phenomenon you are modelling. We have done this in Figure 16. The marker would be cyclin E/CDK2 and the

process would be growth-factor stimulation that leads to the activation of the marker.

2. Identify destabilizing cycles that involve the set of markers and processes. We will define what we mean by ‘destabilizing cycles’ below. This step is required to find an instability that we assume (hypothesize) to cause the switching behaviour in the activity of cyclin E/CDK2. If information on mechanisms involved in these destabilizing cycles is available, one can check what kind of instabilities are involved (as we will show below).
3. A minimal qNET model is formed from the destabilizing cycles involving the marker and other interactions encompassing the process involved (this is growth-factor stimulation in our example).
4. From the minimal qNET model a kinetic model is generated by using available information on the mechanisms and rate expressions for the interactions involved.

### Destabilizing cycles in a qNET graph

A qNET graph is a directed binary interaction graph. A qNET graph corresponds to the algebraic signs of the elements of the Jacobian matrix **M** associated with the dynamical equations (which are assumed to be ODEs) that are linearized about the steady state. The correspondence between the edges of a qNET graph and the signs of a matrix element  $m_{ij}$  is as follows:

qNET edge	meaning	sign of $m_{ij}$
$X_j \rightarrow X_i$	$X_j$ ‘activates’ $X_i$	+
$X_j \dashv X_i$	$X_j$ ‘inhibits’ $X_i$	-
$X_j \dashv\bullet X_i$	$X_j$ ‘influences’ $X_i$	$\neq 0$

‘Activates’ in the table above should be generally interpreted as ‘increases the rate of growth of’ while ‘inhibits’ would mean ‘decreases the rate of growth of’.

The local stability of the steady states is determined by the eigenvalues  $\lambda$  of the matrix **M**. These eigenvalues are the roots of the characteristic polynomial

$$\lambda^n + \alpha_1 \lambda^{n-1} + \alpha_2 \lambda^{n-2} + \alpha_3 \lambda^{n-3} + \dots + \alpha_{n-1} \lambda + \alpha_n = 0.$$

assuming that the size of  $\mathbf{M}$  is  $n \times n$ . The steady state is unstable if any of the eigenvalues has a positive real part.

It turns out that the coefficients  $\alpha_i$  in the characteristic polynomial above can be expressed as follows:

$$\alpha_1 = \sum_i [-C_1(i)]$$

$$\alpha_2 = \sum_{i,j} [-C_1(i)][-C_1(j)] + \sum_{pq} [-C_2(pq)]$$

$$\alpha_3 = \sum_{i,j,k} [-C_1(i)][-C_1(j)][-C_1(k)] + \sum_{t,pq} [-C_1(t)][-C_2(pq)] + \sum_{vws} [-C_3(vws)]$$

etc.

where  $C_k$  is a  $k$ -cycle in the qNET graph examples of which are given below:

$$C_1(i) = m_{ii}$$

$$C_2(pq) = m_{pq}m_{qp}$$

$$C_3(vws) = m_{vw}m_{ws}m_{sv}$$

etc.

We sometime also refer to the  $C_k$  expressions above as the ‘strengths’ of the cycles determined by the magnitudes of the  $m_{ij}$ ’s. Since the eigenvalues depend on the coefficients  $\alpha_i$ ’s which in turn depend on the cycles, we conclude that *only cycles in the qNET graph can influence the local stability of the network*. We say that a cycle is destabilizing if any eigenvalue increases towards a more positive direction when the strength of the cycle is increased. The following theorem is useful for linear stability analysis:

**Routh-Hurwitz Theorem.** The number of eigenvalues  $\lambda_i$  with  $\text{Re } \lambda_i > 0$  is equal to the sum of the number of changes of sign in the sequences

$$\{1, \Delta_1, \Delta_3, \Delta_5, \dots\} \text{ and } \{1, \Delta_2, \Delta_4, \Delta_6, \dots\}.$$

The  $\Delta_i$ ’s are called Hurwitz determinants; they come from the Hurwitz array which is defined below:

Hurwitz array

$$\begin{bmatrix} \alpha_1 & \alpha_3 & \alpha_5 & : \\ 1 & \alpha_2 & \alpha_4 & : \\ 0 & \alpha_1 & \alpha_3 & : \\ : & : & : & : \end{bmatrix}$$

Hurwitz determinants

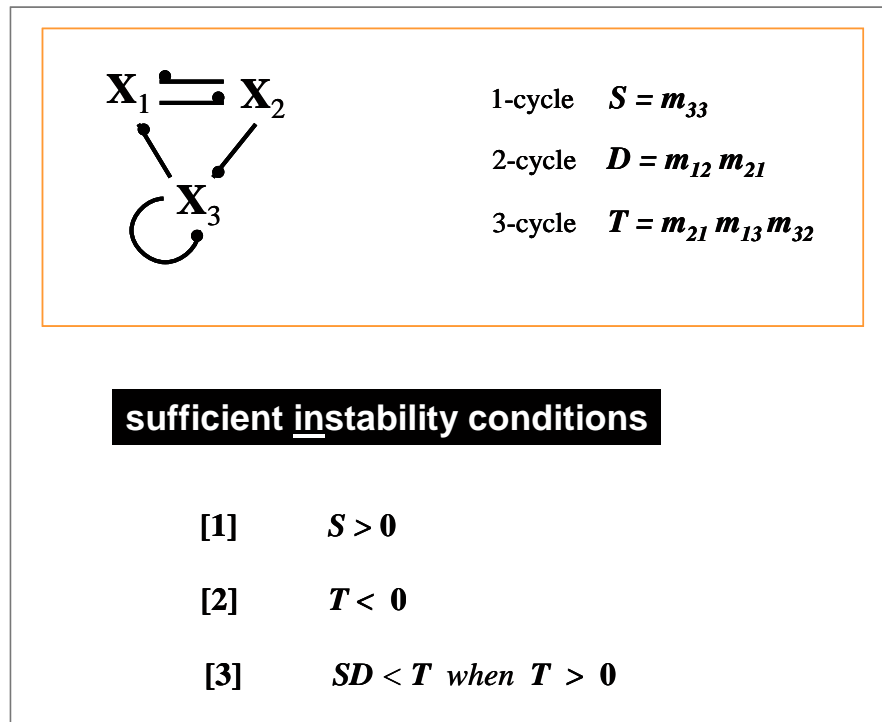
$$\Delta_1 = \alpha_1$$

$$\Delta_2 = \alpha_1 \alpha_2 - \alpha_3$$

$$\Delta_3 = \alpha_3 \Delta_2 - \alpha_1(\alpha_1 \alpha_4 - \alpha_5)$$

etc.

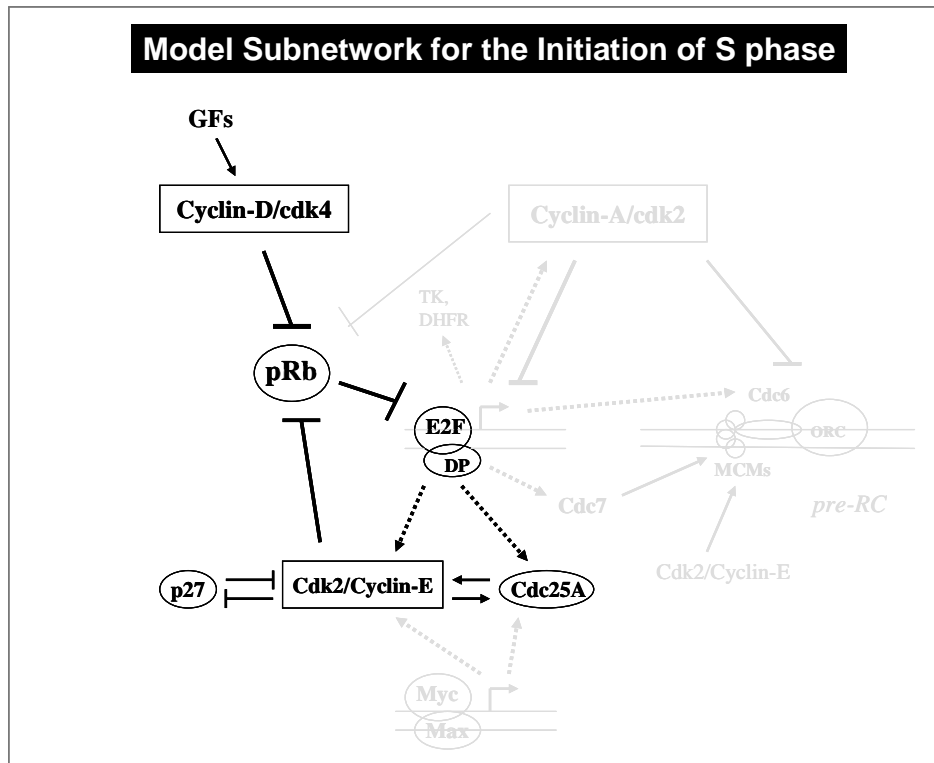
The stability analysis presented in Fig 17 is an example of how to use the Routh-Hurwitz Theorem. This figure also illustrates that the mere topology of the qNET could already allow some conclusions on the stability of steady states.



**Figure 17.** A 3-node qNET with the strengths of the component 1-, 2-, and 3- cycles labeled as S, D, and T, respectively. Application of the Routh-Hurwitz Theorem gives the sufficient instability conditions listed.

### A minimal qNET model

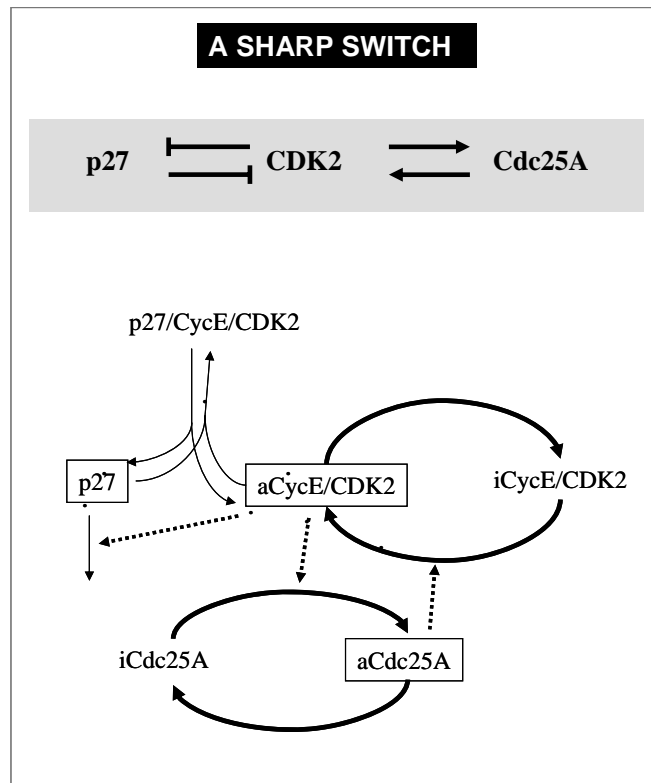
From Fig 16, one can then extract a minimal qNET model that includes destabilizing cycles that directly involve cyclin E/CDK2 and interactions that link these cycles to growth-factor stimulation. This minimal qNET model is shown in Fig 18.



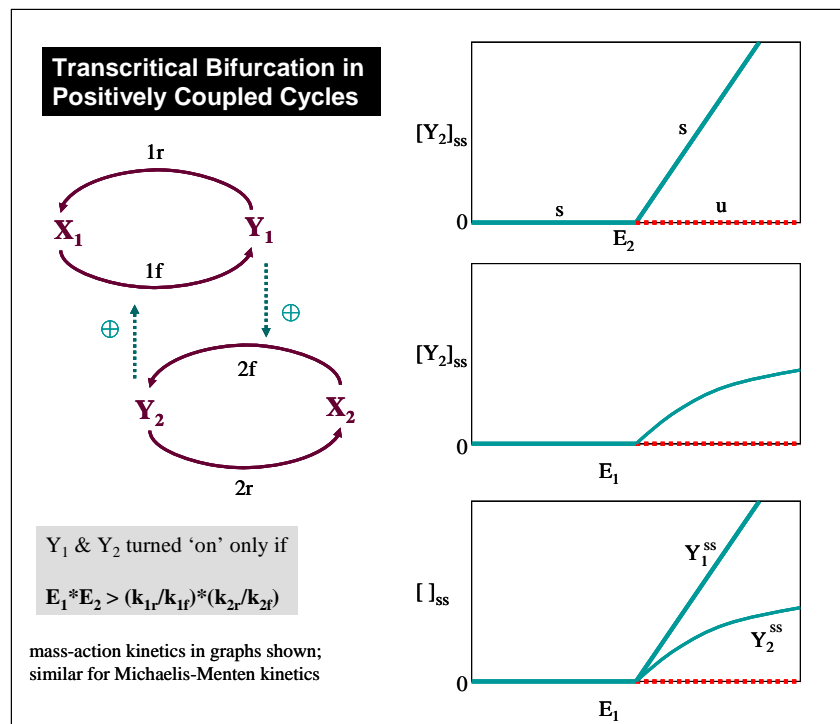
**Figure 18.** The proposed minimal qNET model for the initiation of S phase from which the control machinery of the R point is analyzed (see text).

### The nature of the instability

Two of the destabilizing cycles that involve cyclin E/CDK2 are shown in Fig 19. This mutual-activation-mutual-inhibition topology is expected to generate a sharp switch (as was shown by Aguda & Tang, 1999). CDK2 and Cdc25A are locked in a pair of positively coupled phosphorylation-dephosphorylation (PD) cycles which exhibits transcritical bifurcation (see Fig 20, and Aguda, 1999).



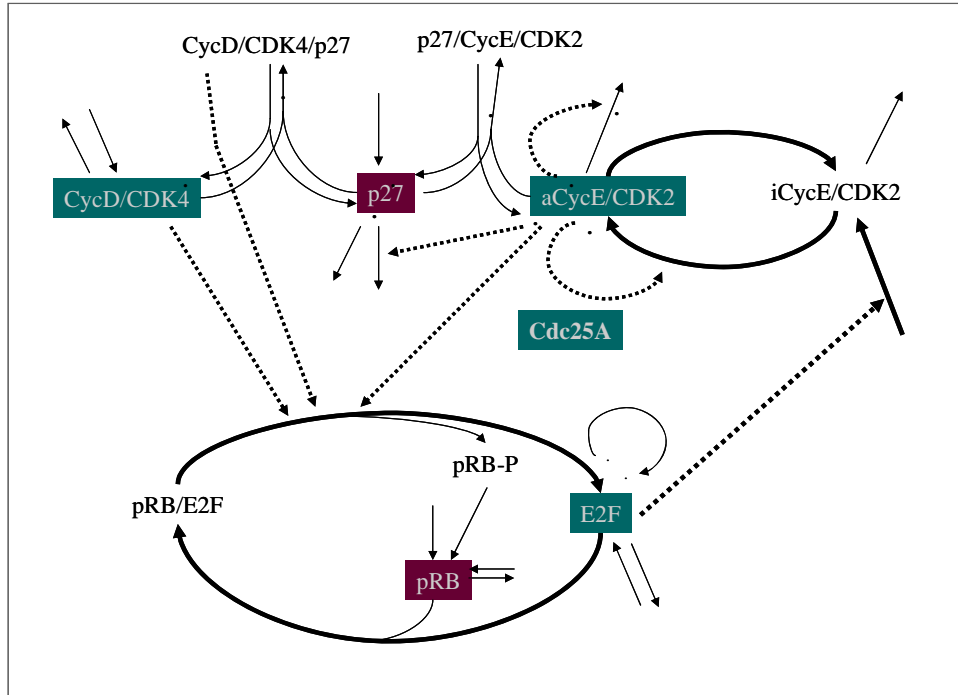
**Figure 19.** A sharp switch is expected from the mutual-activation and mutual-inhibition topology involving CDK2. Also shown are the known detailed mechanistic steps corresponding to the qNET. “a” refers to active, and “i” to inactive.



**Figure 20.** The instability (transcritical bifurcation) involved in positively coupled cycles.

### The kinetic model

From the minimal qNET model (Fig 18) and known details of the molecular mechanism, one can then set up a kinetic model with the associated ODEs. A summary of the kinetic model is shown in Fig 21 (see Aguda & Tang (1999) for details).



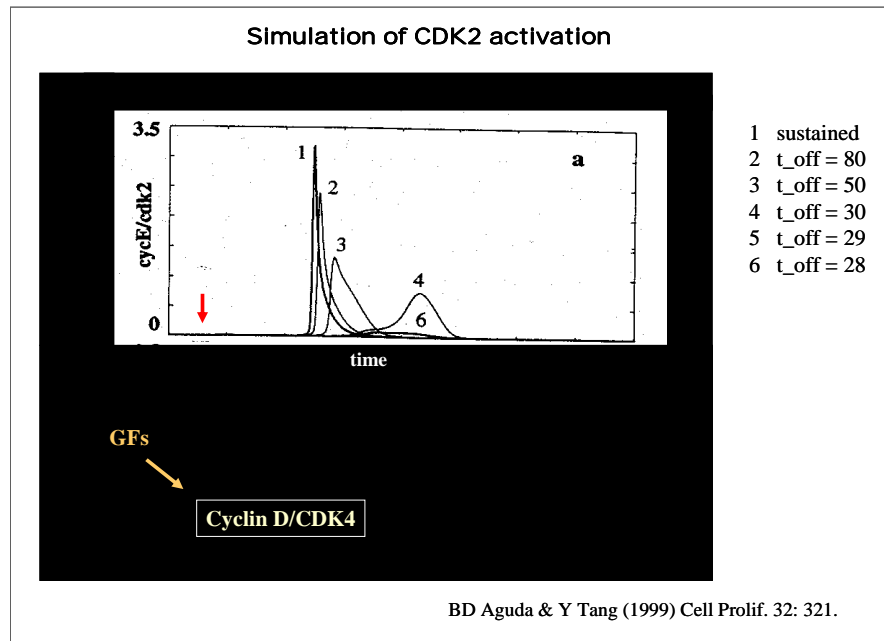
**Figure 21.** The kinetic model of the R point based on the minimal qNET shown in Fig 18. See Aguda & Tang (1999) for details.

### IV.3 COMPUTER SIMULATION OF THE KINETIC MODEL

Experimentally, the operational definition of the restriction (R) point is the following: for a quiescent (non-dividing) cell exposed to growth-factor stimulation, the R point is the point in time after which withdrawal of growth factors does not prevent entry into S phase (in our model, this would correspond to the activation of cyclinE/CDK2). The computer simulation shown in Fig 22 shows different times at which growth-factors are cut off (this is implemented by setting the synthesis/activation of cyclin D – i.e. the left-most arrow pointing towards cyclin D/CDK4 in Fig 21). Qualitatively, the results of this set of simulations agree well with the experimental observation



shown in the lower panel of Fig 3, including a conspicuous lag period prior to activation of cyclin E/CDK2. The conclusion of this work is that, within the known regulatory network of the G1-S transition, one can identify a subnetwork that reproduces the behaviour of the R point (at least qualitatively).



**Figure 22.** Simulation of R point behaviour using the kinetic model given in Fig 21. The red arrow indicates the point in time after which cutting off growth factor stimulation can be done without preventing the activation of CDK2. See Aguda & Tang (1999) for details.


## REFERENCES

- Aguda, B.D. (1999) Instabilities in Phosphorylation-Dephosphorylation Cascades and Cell cycle Checkpoints. *Oncogene* **18** : 2846-2851.
- Aguda, B.D. and Clarke,B.L. (1987) Bistability in chemical reaction networks: Theory and application to the peroxidase-oxidase reaction. *J. Chem. Phys.* **87**: 3461-3470
- Aguda, B.D. & Tang Y. (1999) The kinetic origins of the restriction point in the mammalian cell cycle. *Cell Proliferation* **32**: 321-335.
- Alves,R., Antunes,F., and Salvador,A. (2006) Tools for kinetic modeling of biochemical networks. *Nat. Biotechnol.* **24**: 667-672
- Bader, G.D., Cary, M.P., and Sander, C. (2006) Pathguide: A Pathway Resource List. *Nucleic Acids Research* **34**: D504-D506 (Database Issue)
- Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**: 101-113
- Cary, M. P., Bader, G. D., and Sander, C. (2005) Pathway information for systems biology. *FEBS Letters* **579**: 1815-1820.
- Clarke,B.L. (1988) Stoichiometric network analysis. *Cell Biophys.* **12**: 237-253
- Davidson,E.H. and Erwin,D.H. (2006) Gene regulatory networks and the evolution of animal body plans. *Science* **311**: 796-800
- Demir,E., Babur,O., Dogrusoz,U., Gursoy,A., Ayaz,A., Gulesir,G., Nisanci,G., and Cetin-Atalay,R. (2004) An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics.* **20**: 349-356
- Demir,E., Babur,O., Dogrusoz,U., Gursoy,A., Nisanci,G., Cetin-Atalay,R., and Ozturk,M. (2002) PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics.* **18**: 996-1003
- Ekholm, S.V., Zickert, P., Reed, S.I., Zetterberg A. (2001) Accumulation of cyclin E is not a prerequisite for passage through the restriction point. *Mol Cell Biol.* **21**:3256-65.
- Friedman,N., Linial,M., Nachman,I., and Pe'er,D. (2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**: 601-620

- Gauges,R., Rost,U., Sahle,S., and Wegner,K. (2006) A model diagram layout extension for SBML. *Bioinformatics*.
- Goryachev,A.B., Toh,D.J., Wee,K.B., Lee,T., Zhang,H.B., and Zhang,L.H. (2005) Transition to quorum sensing in an agrobacterium population: a stochastic model. *PLoS. Comput. Biol.* **1**: e37
- Goss,P.J. and Peccoud,J. (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. U. S. A* **95**: 6750-6755
- Hofmeyr,J.H., Snoep,J.L., and Westerhoff,H.V. (2002). Kinetics, control and regulation of metabolic systems. In Third International Conference on Systems Biology 2002, tutorial, (Stockholm.
- Kauffman,S. (2004) A proposal for using the ensemble approach to understand genetic regulatory networks. *J. Theor. Biol.* **230**: 581-590
- Kitano,H., Funahashi,A., Matsuoka,Y., and Oda,K. (2005) Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* **23**: 961-966
- Kohn,K.W. (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* **10**: 2703-2734
- Kohn,K.W. (2001) Molecular interaction maps as information organizers and simulation guides. *Chaos.* **11**: 84-97
- Kohn,K.W. and Aladjem,M.I. (2006) Circuit diagrams for biological networks. *Mol. Syst. Biol.* **2**: 2006
- Longabaugh,W.J., Davidson,E.H., and Bolouri,H. (2005) Computational representation of developmental genetic regulatory networks. *Dev. Biol.* **283**: 1-16
- Mandel,J., Palfreyman,N.M., Lopez,J.A., and Dubitzky,W. (2004) Representing bioinformatics causality. *Brief. Bioinform.* **5**: 270-283
- McAdams,H.H. and Arkin,A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U. S. A* **94**: 814-819
- Oda,K. and Kitano,H. (2006) A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.* **2**: 2006
- Olivier,B.G., Rohwer,J.M., and Hofmeyr,J.H. (2005) Modelling cellular systems with PySCeS. *Bioinformatics.* **21**: 560-561

- Papin,J.A. and Palsson,B.O. (2004) The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. *Biophys. J.* **87**: 37-46
- Papin,J.A., Price,N.D., Wiback,S.J., Fell,D.A., and Palsson,B.O. (2003) Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28**: 250-258
- Peleg,M., Yeh,I., and Altman,R.B. (2002) Modelling biological processes using workflow and Petri Net models. *Bioinformatics.* **18**: 825-837
- Perkins,T.J., Hallett,M., and Glass,L. (2006) Dynamical properties of model gene networks and implications for the inverse problem. *Biosystems* **84**: 115-123
- Pettinen,A., Aho,T., Smolander,O.P., Manninen,T., Saarinen,A., Taattola,K.L., Yli-Harja,O., and Linne,M.L. (2005) Simulation tools for biochemical networks: evaluation of performance and usability. *Bioinformatics.* **21**: 357-363
- Ravasz,E., Somera,A.L., Mongru,D.A., Oltvai,Z.N., and Barabasi,A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297**: 1551-1555
- Reidl,J., Borowski,P., Sensse,A., Starke,J., Zapotocky,M., and Eiswirth,M. (2006) Model of calcium oscillations due to negative feedback in olfactory cilia. *Biophys. J.* **90**: 1147-1155
- Schilling,C.H., Schuster,S., Palsson,B.O., and Heinrich,R. (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **15**: 296-303
- Shmulevich,I., Dougherty,E.R., Kim,S., and Zhang,W. (2002) Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics.* **18**: 261-274
- Sorokin,A., Paliy,K., Selkov,A., Demin,O.V., Dronov,O.V., Ghazal,P., and Goryanin,I. (2006) The Pathway Editor: a tool for managing complex biological networks. *submitted*
- Thomas,R., Thieffry,D., and Kaufman,M. (1995) Dynamical behaviour of biological regulatory networks--I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.* **57**: 247-276
- Tyson,J.J. (1975) Classification of instabilities in chemical reaction systems. *J. Chem. Phys.* **62**: 1010-1015
- Yuh,C.H., Bolouri,H., and Davidson,E.H. (1998) Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**: 1896-1902

14th Annual International Conference On Intelligent Systems For Molecular Biology



## FROM PATHWAYS DATABASES TO NETWORK MODELS

**Baltazar D. Aguda**  
Mathematical Biosciences Institute  
Ohio State University, USA  
[bdaguda@mbi.osu.edu](mailto:bdaguda@mbi.osu.edu)

**Andrew B. Goryachev**  
Centre for Integrative Systems Biology  
University of Edinburgh, UK  
[andrew\\_goryachev@yahoo.com](mailto:andrew_goryachev@yahoo.com)

I/MB 2006 Fortaleza, Brazil August 6-10, 2006

### I. Pathways databases and knowledgebases


Pathguide  
Pathway data standards  
A modeling-focused use of pathways databases  
Repositories of models

### II. Network visualization and analysis

Graphical representation of pathways and networks  
Methods and tools for network analysis and modelling


### III. Extracting and analyzing a biological model

The G1-S transition in the mammalian cell cycle  
From a qualitative network to a kinetic model  
Computer simulation of the model



**OUTLINE** 178

## I. Pathways databases and knowledgebases



179


*D504-D506 Nucleic Acids Research, 2006, Vol. 34, Database issue  
doi:10.1093/nar/gkj126*

## Pathguide: a Pathway Resource List

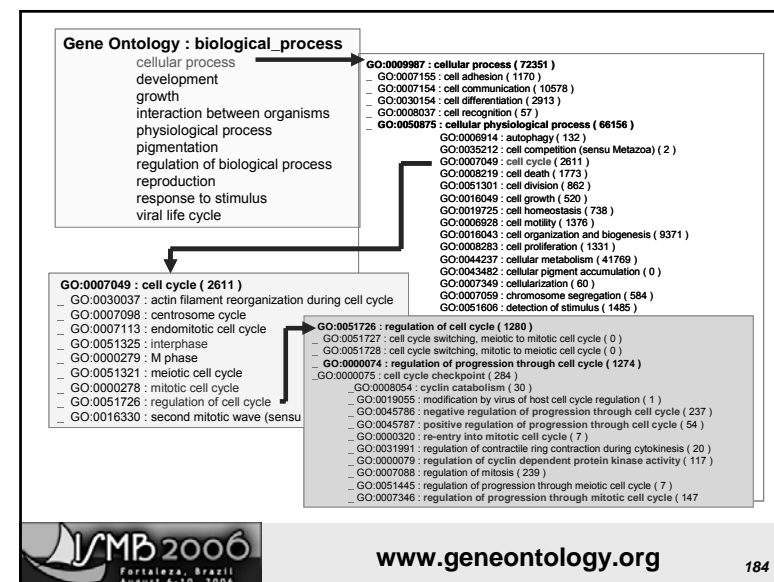
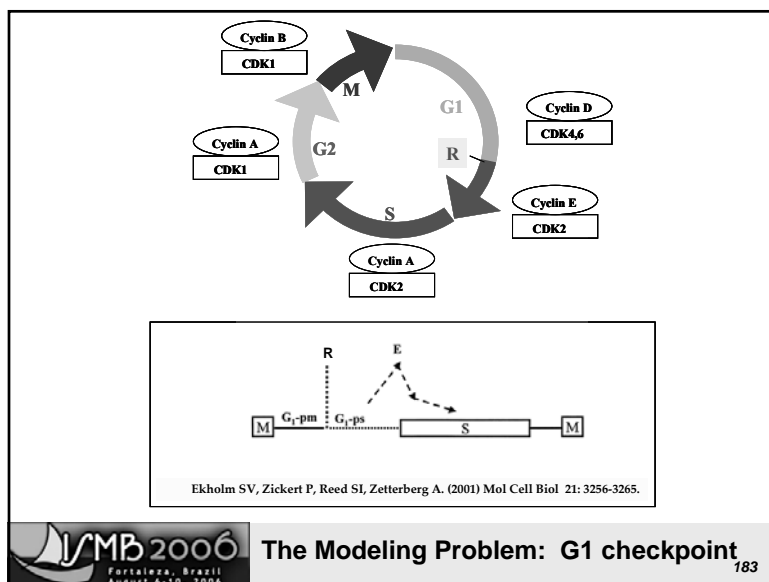
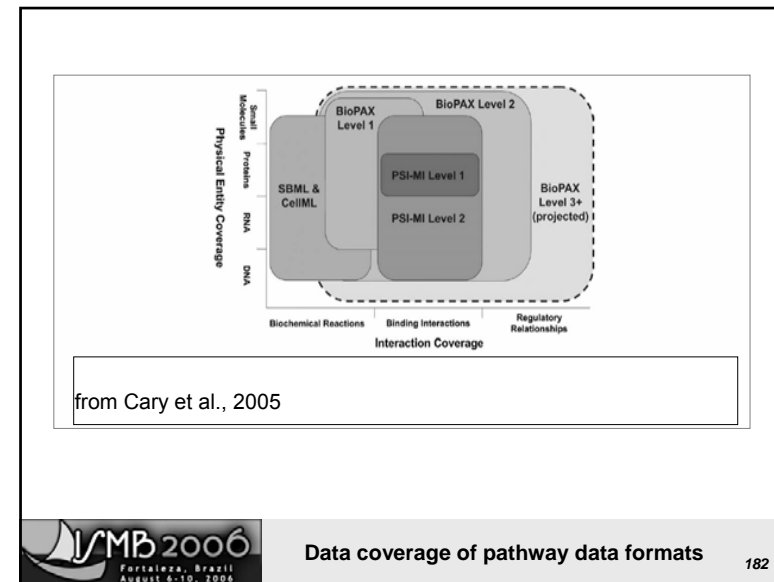
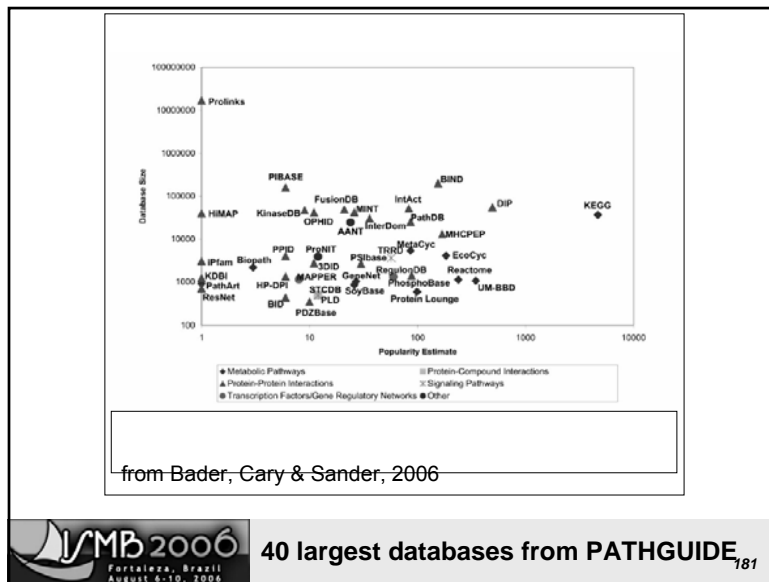
Gary D. Bader, Michael P. Cary and Chris Sander\*

Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue,  
Box 460, New York, NY 10021, USA

1. Protein-Protein Interactions (86)
2. Metabolic Pathways (45)
3. Signaling Pathways (45)
4. Pathway Diagrams (23)
5. Transcription Factors/Gene Regulatory Networks (30)
6. Protein-Compound Interactions (16)
7. Genetic Interaction Networks (5)
8. Protein Sequence Focus (12)
9. Other (13)



<http://www.pathguide.org> 180



## NETWORK HIERARCHY IN KEGG

Metabolism  
Genetic Information Processing  
Environmental Information Processing  
**Cellular Processes**  
Human Diseases

Kyoto Encyclopedia of  
Genes and Genomes

### 01400 Cellular Processes

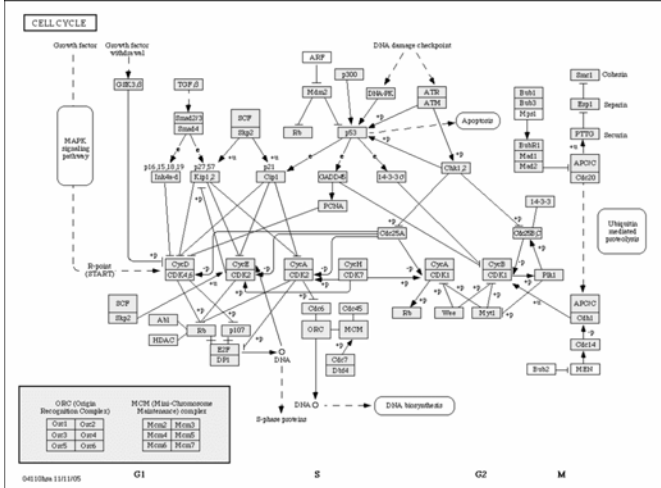
- 01410 Cell Motility
- 01420 Cell Growth and Death
- 04410 Cell division
- 04420 Sporulation [GO:0030435 0030436]
- 04430 Germination [GO:0009847]
- 04110 Cell cycle [PATH:ko04110hsa]
- 04210 Apoptosis [PATH:ko04210] [GO:0006915]

CLICK TO SEE PATHWAY



www.kegg.jp

185



Cell cycle network from KEGG

186

D428-D432 Nucleic Acids Research, 2005, Vol. 33, Database issue  
doi:10.1093/nar/gk072

## Reactome: a knowledgebase of biological pathways

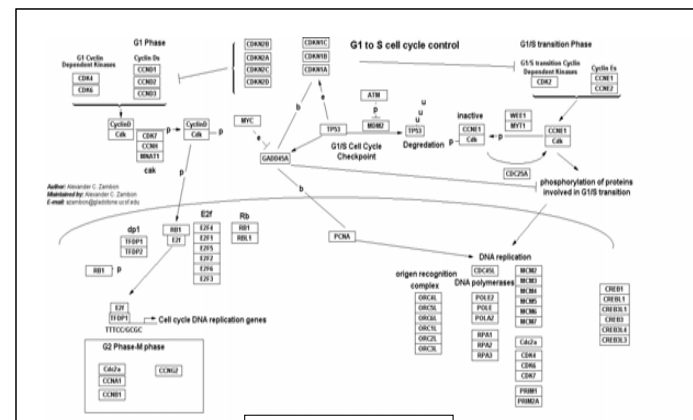
G. Joshi-Tope<sup>1,\*</sup>, M. Gillespie<sup>1,3</sup>, I. Vastrik<sup>2</sup>, P. D'Eustachio<sup>1,4</sup>, E. Schmidt<sup>2</sup>, B. de Bono<sup>2</sup>,  
B. Jassal<sup>2</sup>, G.R. Gopinath<sup>1</sup>, G.R. Wu<sup>1</sup>, L. Matthews<sup>1</sup>, S. Lewis<sup>5</sup>, E. Birney<sup>2</sup> and L. Stein<sup>1</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA, <sup>2</sup>European Bioinformatics Institute, Hinxton, Cambridge, UK, <sup>3</sup>St Johns University, NY, USA, <sup>4</sup>New York University School of Medicine, NY, USA and <sup>5</sup>University of California, Berkeley, CA, USA

Received August 19, 2004; Revised and Accepted October 6, 2004



187

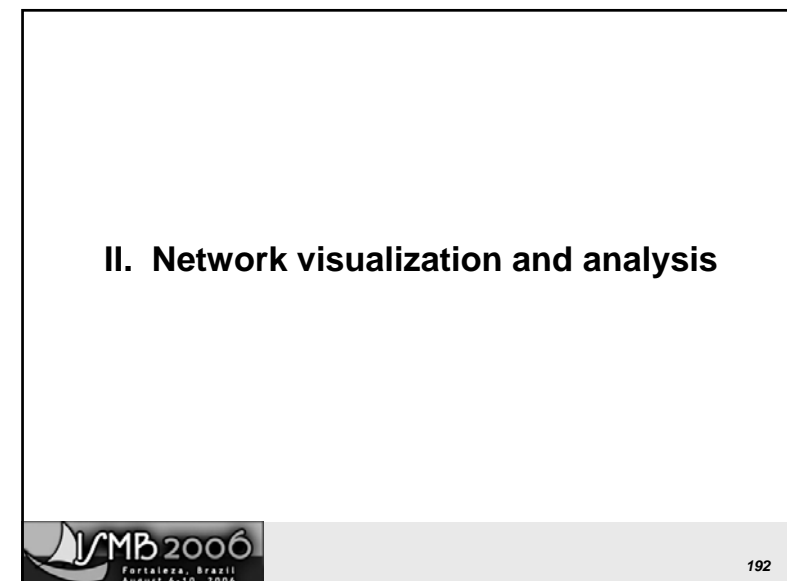
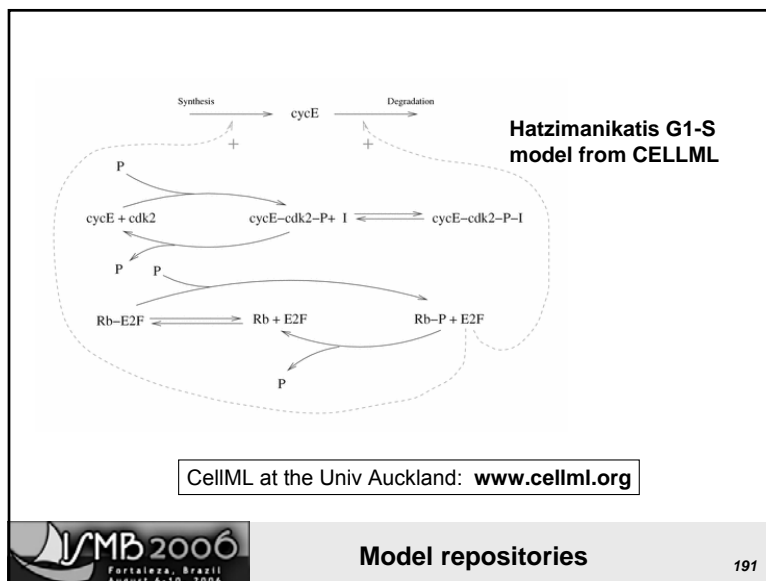
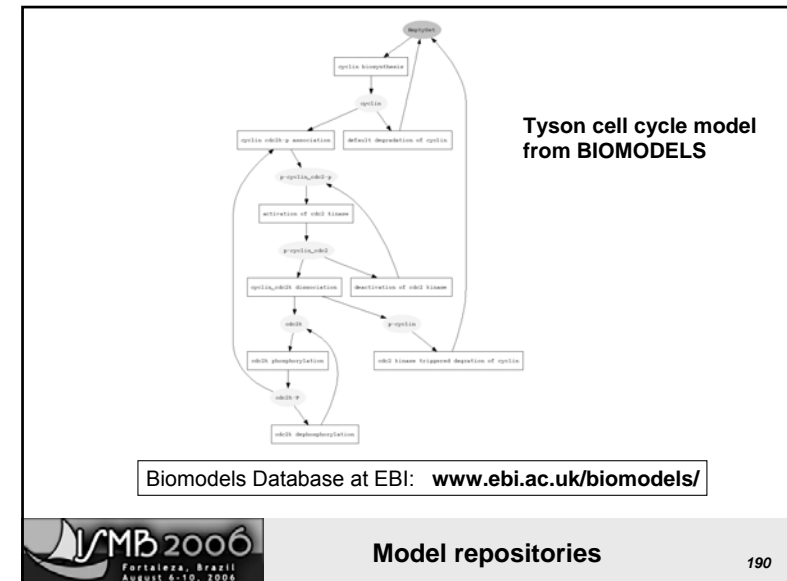
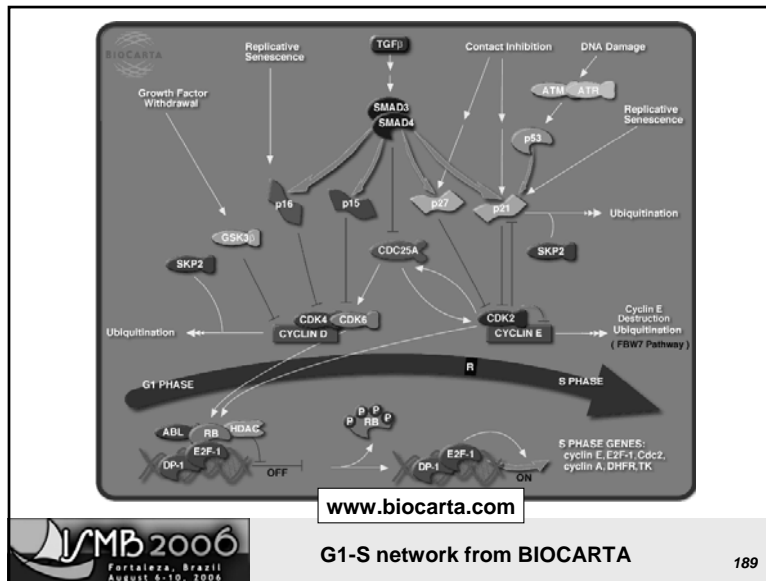


www.reactome.org  
www.genmapp.org



G1-S network from REACTOME and GENMAPP

188



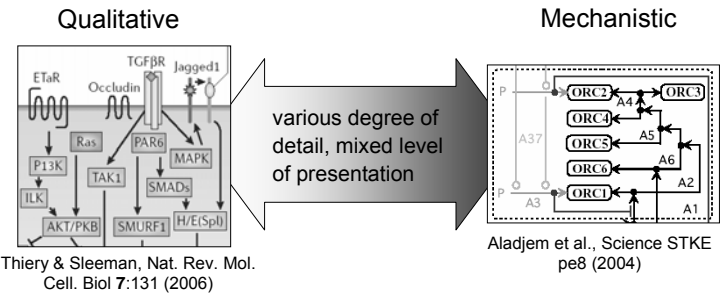


## Graphical Representation of Pathways and Networks

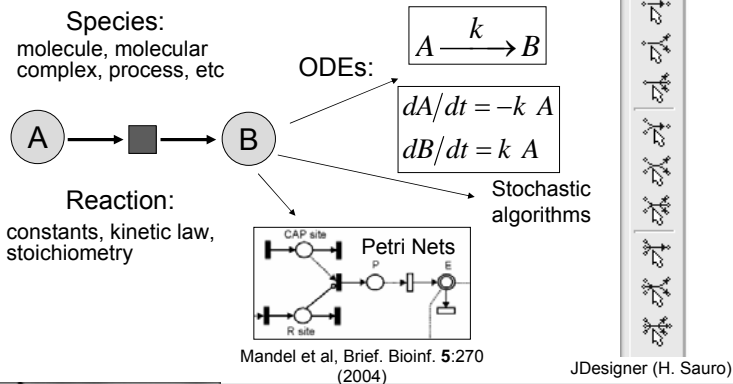
### Problem definition and challenges

Math perspective : General kinetic notation  
 "Metabocentric" view : Biochemical/metabolic notation  
 "Genecentric" view : "Caltech" notation  
 Signalling views : Molecular Interaction Maps  
 Process Diagram Notation  
 Edinburgh Pathway Notation  
 Modular perspective : Patika  
**Future: unification and standardization**

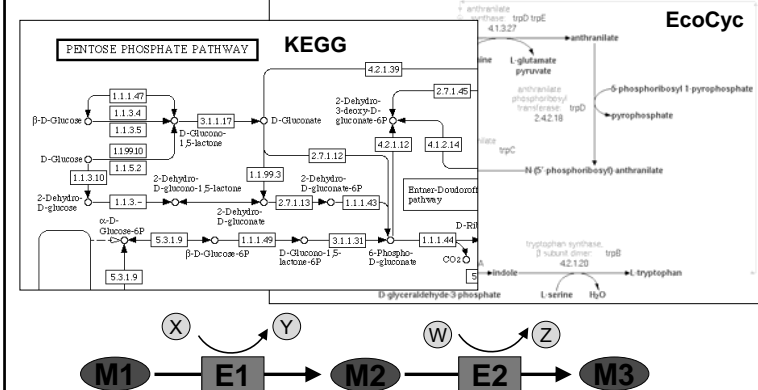
## Graphical Notation: a necessity for the conceptual representation of biopathways



## Stoichiometric Kinetic notation: language of mathematical models (almost standard)

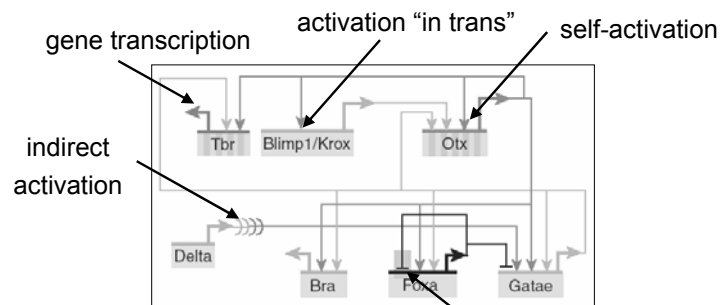


## Notations accepted in the field of metabolic biochemical pathways



## (E. Davidson, H. Bolouri, A. Arkin, H. MacAdams)

(E. Davidson, H. Bolouri, A. Arkin, H. MacAdams)

Davidson & Erwin, Science **311**:796 (2006)

self-inhibition

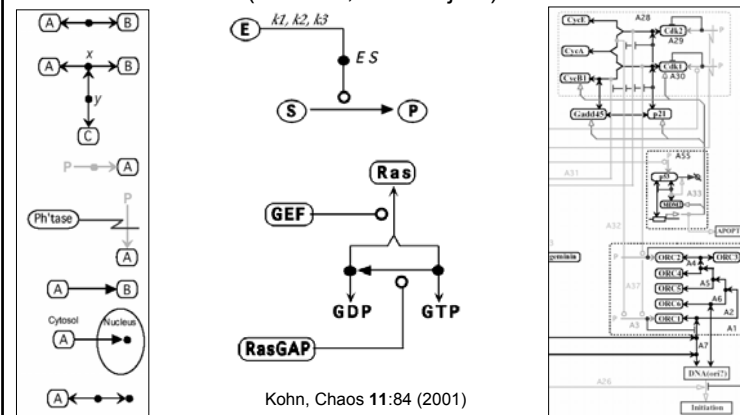


*Supported and extended by BioTapestry (H. Bolouri)*

197

## Molecular Interaction Maps

(K. Kohn, M. Aladjem)



Kohn, Chaos 11:84 (2001)

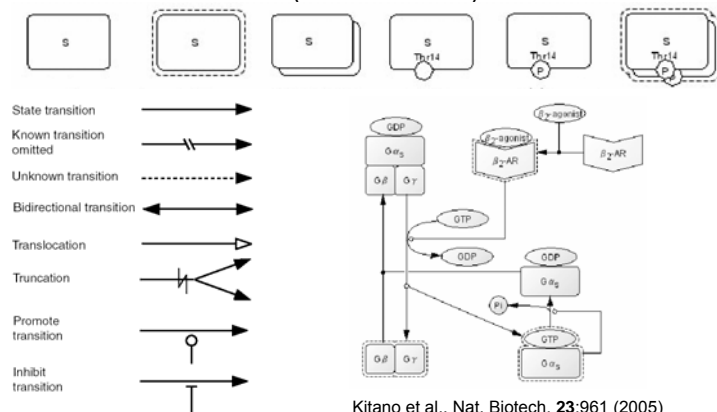
Aladjem et al., Science STKE pe8 (2004)



198

## Process Diagram Notation

(H. Kitano et al.)



Kitano et al., Nat. Biotech. **23**:961 (2005)

*Supported by CellDesigner (SBI)*

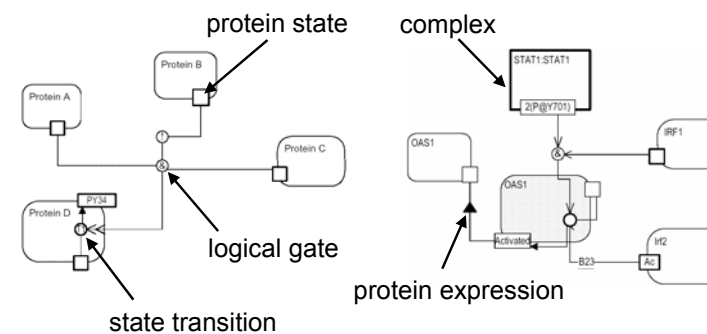


199

## Edinburgh Pathway Notation

(I.Goryanin, P. Ghazal et al.)

## Meta-level notation



Sorokin et al., ?. in press (2006)

Supported by Edinburgh Pathway Editor (UofE)

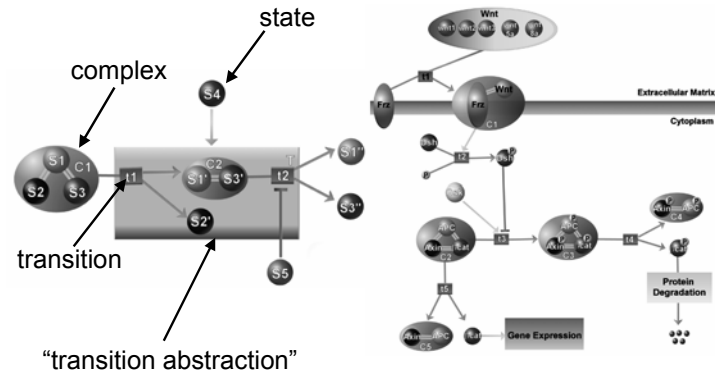


200

## PATIKA: Abstract Pathway Notation

(U. Dogrusoz, E. Demir et al.)

complex



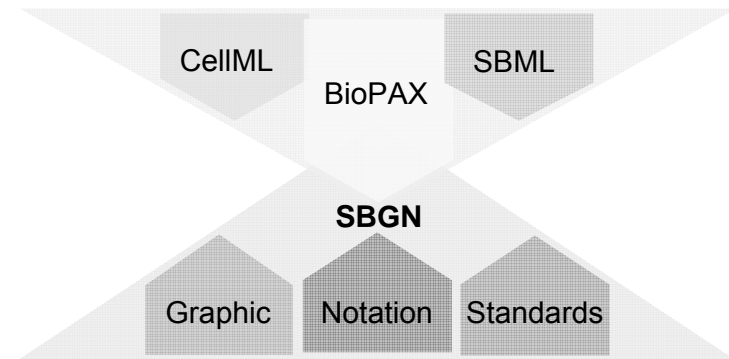
Demir et al., Bioinf. 20:349 (2004)



Supported by PATIKA (Bilkent University, Turkey)

201

## SBGN: towards the unified graphics standard



202

## Methods and Tools for Network Analysis & Modelling

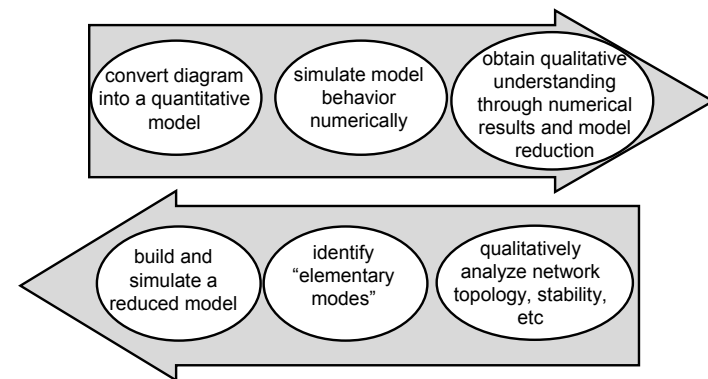
Simulation versus analysis: choice of strategy and methods  
 Multidimensional space of modeling techniques  
 Kinetic modeling with ODEs and stochastic methods  
 Petri Nets, Boolean and Bayesian Networks  
 Topological analysis of large networks based on graph theory  
 Stoichiometric Network Analysis  
 Metabolic Control Analysis  
 Qualitative stability analysis



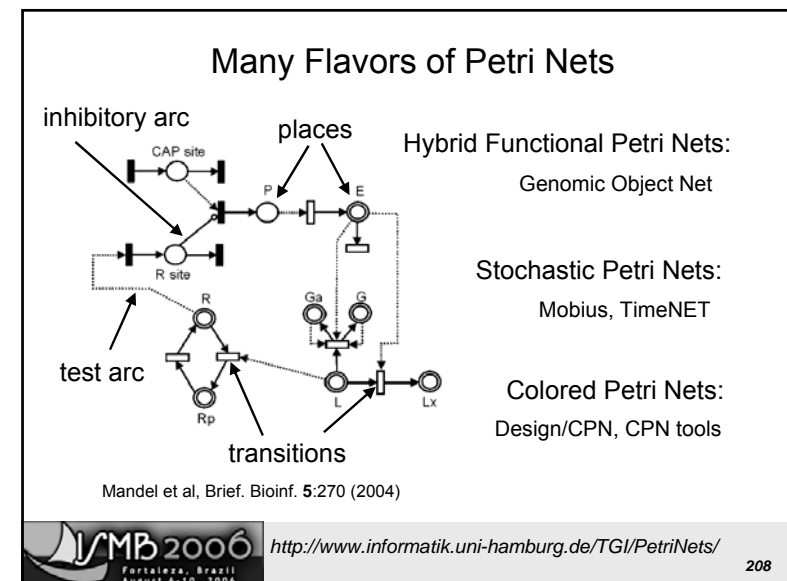
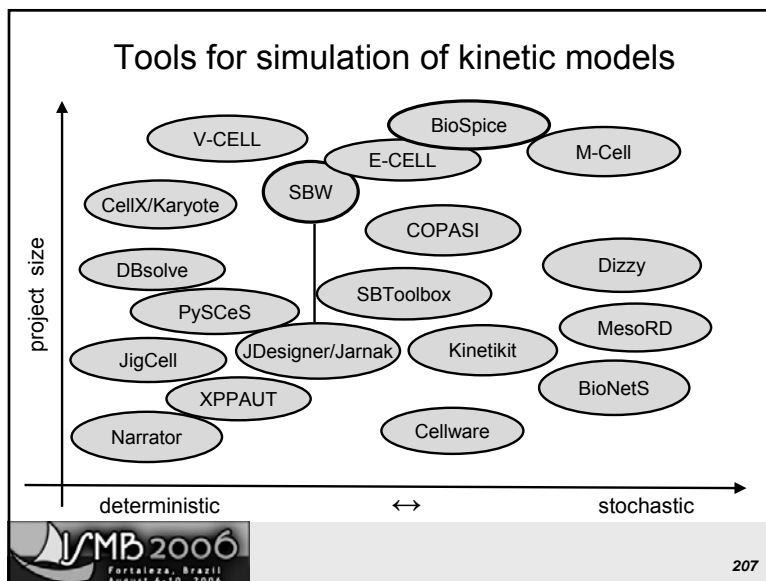
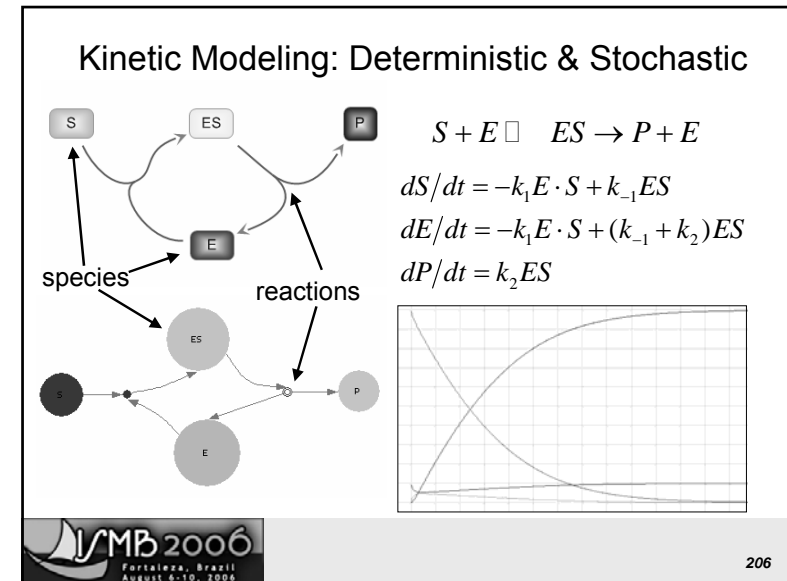
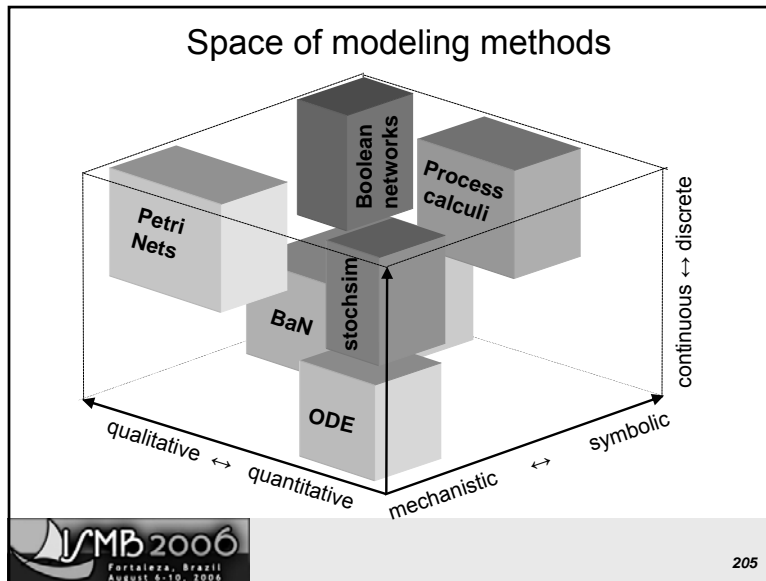
203

## Strategies: simulate or analyse?

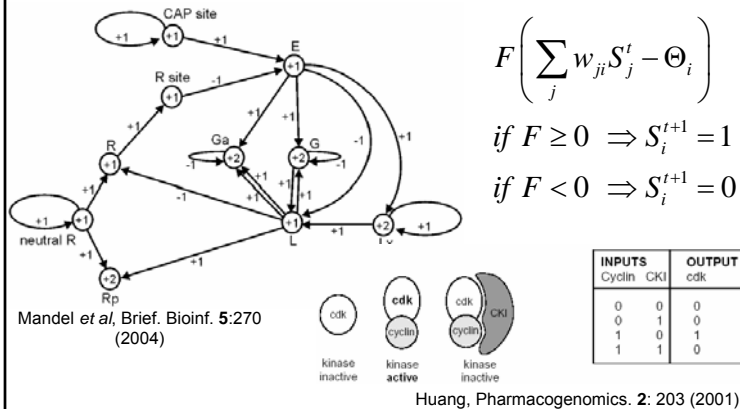
(or rather what to do first)



204



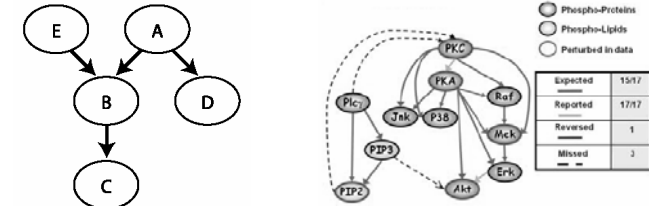
## Boolean networks



Genetic Network Analyzer, Biocham

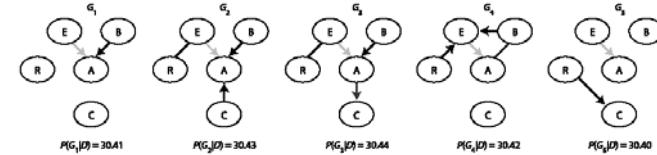
209

## Bayesian Networks



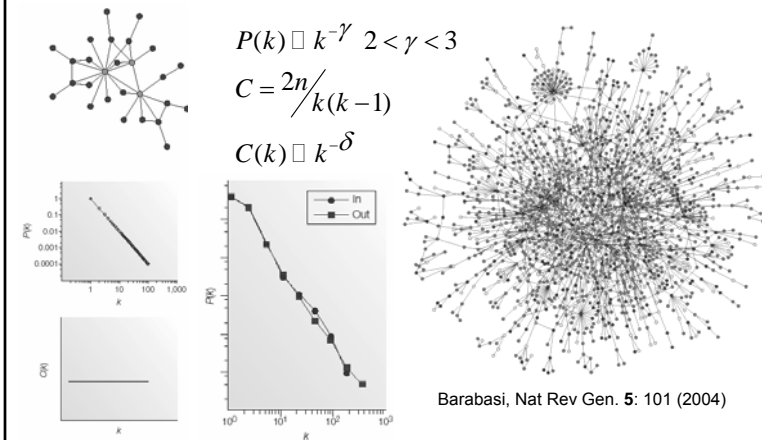
Pe'er, Sci. STKE. pl4 (2005)

Sachs, Science. 308: 523 (2005)



210

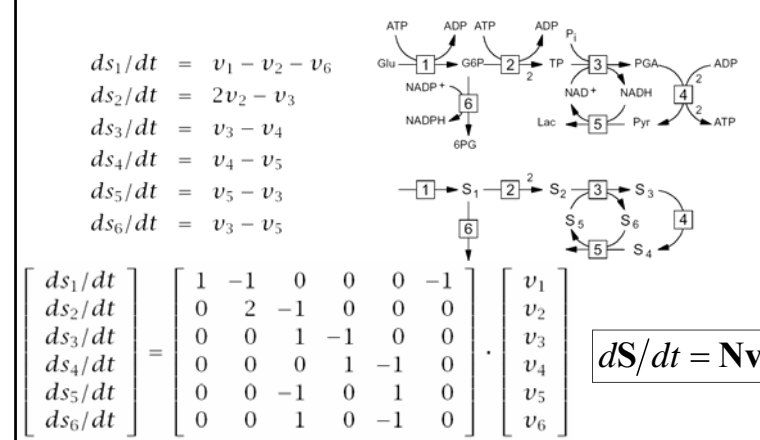
## Topological analysis of network connectivity



Cytoscape/NetworkAnalyzer

211

## Stoichiometric Matrix



Hofmeyr *et al.*, Kinetics, Control and Regulation of Metabolic Systems. ICSB02. (2002)

212

## Stoichiometric Network Analysis

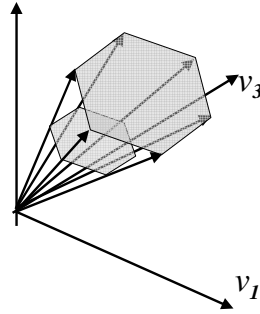
$$\mathbf{N}\mathbf{v} = 0$$

$$\dim \text{Nul } \mathbf{N} + \text{rank } \mathbf{N} = \dim \mathbf{v} = n \quad v_2$$

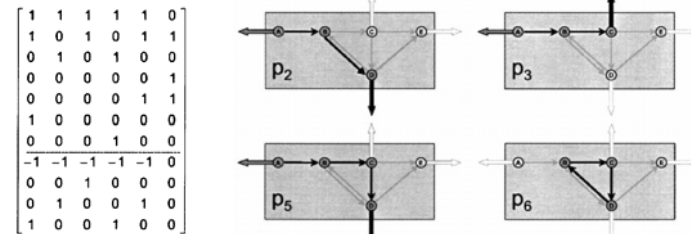
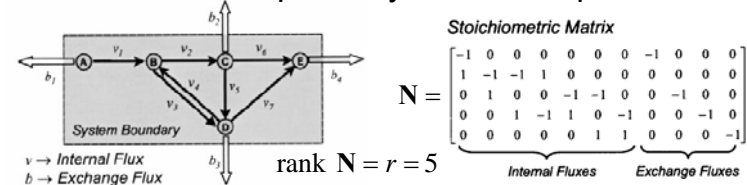
$$\text{rank } \mathbf{N} = r \Rightarrow \dim \text{Nul } \mathbf{N} = n - r$$

$$\mathbf{N}\mathbf{K} = 0 \quad n \times n - r$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 2 & 0 \\ 2 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_2 \\ v_6 \end{bmatrix}$$

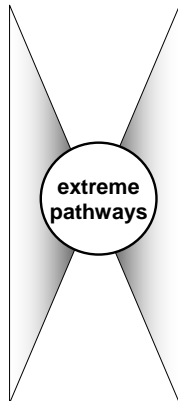


## Extreme pathways: An example



## SNA: Tools and Uses

- **METATOOL**  
Pfeiffer et al. Bionf. **15**:251 (1999)
- **FluxAnalyzer**  
Klamt et al. Bionf. **19**:261 (2003)
- **CellNetAnalyzer**  
Klamt et al. BMC Bionf. **7**: 56 (2006)
- **SNA toolbox**  
Urbaniczik. BMC Bionf. **7**: 129 (2006)



- **Network stability analysis**  
Clarke, Adv. Chem. Phys. **43**:1 (1980)
- **Extraction of reduced models**  
Aguda & Clarke. J. Chem. Phys. **87**: 3461 (1987)
- **Signal pathway analysis**  
Papin & Palsson. Bioph. J. **87**: 37 (2004)
- **Analysis of Ca oscillations**  
Reidl et al. Bioph. J. **90**:1147 (2006)

## Metabolic Control Analysis

Local properties:

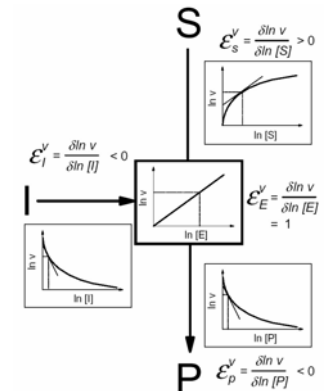
Elasticities:  $\epsilon_s^v = \frac{\partial \ln v}{\partial \ln S}$ ,  $\epsilon_p^v = \frac{\partial \ln v}{\partial \ln p}$

Global properties:

Response coefficients:  $R_p^Y = \frac{\partial \ln Y}{\partial \ln p}$ ,  $Y = S, J$

Control coefficients:  $C_v^Y = \frac{\partial \ln Y}{\partial \ln v}$ ,  $Y = S, J$

$$R_p^Y = C_v^Y \epsilon_p^v, \quad Y = S, J$$



## MCA relates global to local properties

Summation theorems: Connectivity theorems: Control-matrix equation:

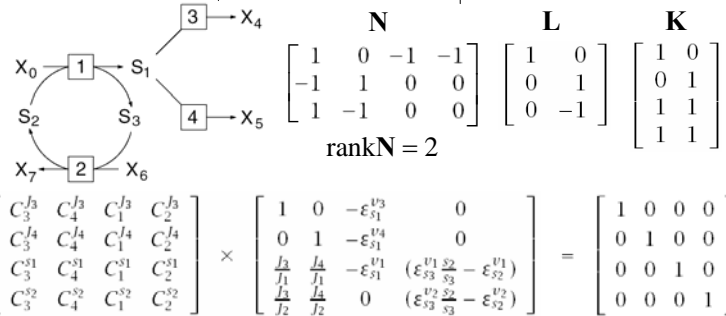
$$C^S K = 0$$

$$C^J K = K$$

$$C^S \varepsilon_s L = -L$$

$$C^J \varepsilon_s L = 0$$

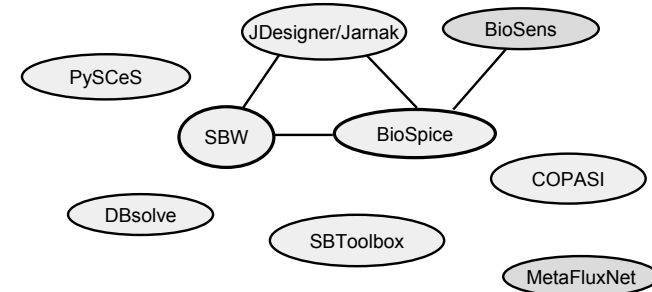
$$\begin{bmatrix} C^J \\ C^S \end{bmatrix} [K - \varepsilon_s L] = I_n$$



Hofmeyr et al., Kinetics, Control and Regulation of Metabolic Systems. ICSB02. (2002)

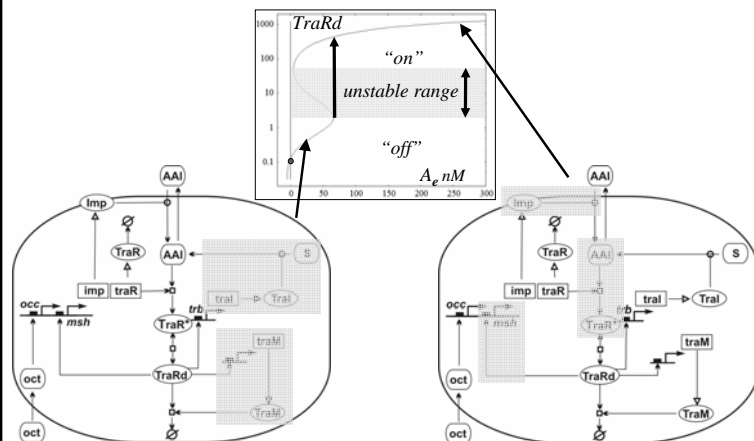
217

## MCA-MFA enabled tools



218

## MCA: understanding the network function

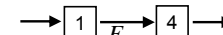


Goryachev et al., PLOS Comp. Biol. 1: 265 (2005)

219

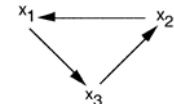
## Analysis of circuits and network stability

$$\frac{dS_i}{dt} = F_i(\vec{S}, \vec{p}) \quad A = [a_{ij}] = \frac{\partial F_i}{\partial S_j}$$

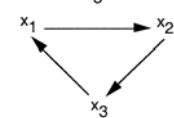


$$A = \begin{bmatrix} - & + & 0 & 0 \\ 0 & - & + & - \\ 0 & 0 & - & + \\ + & - & 0 & - \end{bmatrix}$$

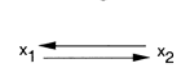
$$\begin{bmatrix} . & a_{12} & . \\ . & . & a_{23} \\ a_{31} & . & . \end{bmatrix}$$



$$\begin{bmatrix} . & . & a_{13} \\ a_{21} & . & . \\ . & a_{32} & . \end{bmatrix}$$



$$\begin{bmatrix} . & a_{12} & . \\ a_{21} & . & . \\ . & . & . \end{bmatrix}$$



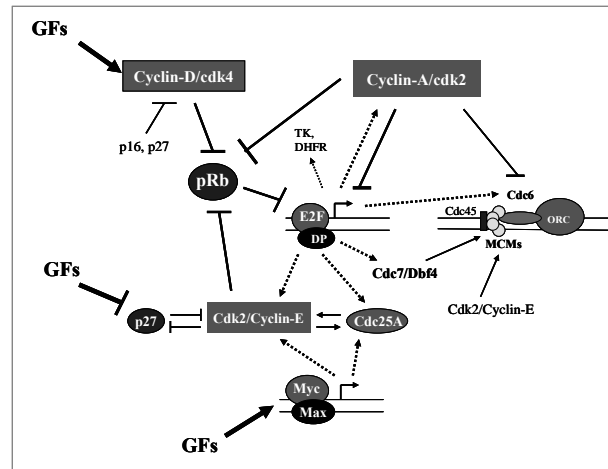
Tyson, J. Chem. Phys. 62: 1010 (1975)

Thomas et al., Bul. Math. Biol. 57: 247 (1995)



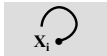
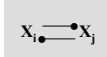
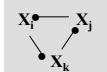
220

### III. Extracting and analyzing a biological model

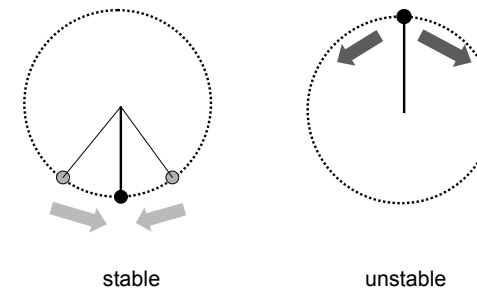


$$m_{ij} = [\partial \dot{x}_i / \partial x_j]_o$$

$m_{ij} > 0$	$X_j$	activates	$X_i$	$(X_j \longrightarrow X_i)$
$m_{ij} < 0$	$X_j$	inhibits	$X_i$	$(X_j \dashv X_i)$

Cycle	strength	graph
1-cycle	$m_{ii}$	
2-cycle	$m_{ij}m_{ji}$	
3-cycle	$m_{ij}m_{jk}m_{ki}$	

### STABILITY OF A STEADY STATE





## eigenvalues are functions of cycles only

$$\lambda^n + \alpha_1 \lambda^{n-1} + \alpha_2 \lambda^{n-2} + \dots + \alpha_{n-1} \lambda + \alpha_n = 0$$

where

$$\alpha_1 = \sum_i [-C_1(i)]$$

$$\alpha_2 = \sum_{i,j} [-C_1(i)][-C_1(j)] + \sum_{j,k} [-C_2(jk)]$$

$$\alpha_3 = \sum_{i,j,k} [-C_1(i)][-C_1(j)][-C_1(k)] + \sum_{i,j,k} [-C_1(i)][-C_2(jk)] + \sum_{i,j,k} [-C_3(ijk)]$$

...

where

$C_1(i) = m_{ii}$	(1-cycles)
$C_2(jk) = m_{jk}m_{kj}$	(2-cycles)
$C_3(ijk) = m_{ij}m_{jk}m_{ki}$	(3-cycles)
...	...

Hurwitz determinants

$$\Delta_1 = \alpha_1$$

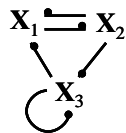
$$\Delta_2 = \alpha_1 \alpha_2 - \alpha_3$$

$$\Delta_3 = \alpha_3 \Delta_2 - \alpha_1 (\alpha_1 \alpha_4 - \alpha_5)$$

etc.

## Routh-Hurwitz Theorem

The number of eigenvalues  $\lambda_i$  with  $\text{Re } \lambda_i > 0$  equals the sum of the number of changes of sign in the sequences  $\{1, \Delta_1, \Delta_3, \Delta_5, \dots\}$  and  $\{1, \Delta_2, \Delta_4, \Delta_6, \dots\}$ .



1-cycle  $S = m_{33}$

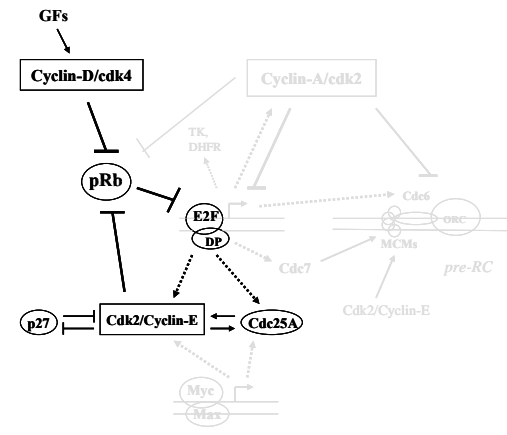
2-cycle  $D = m_{12} m_{21}$

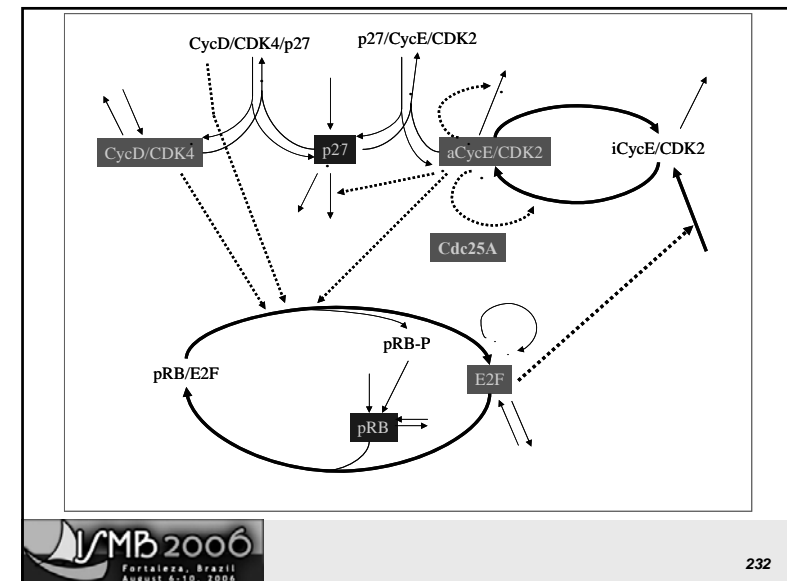
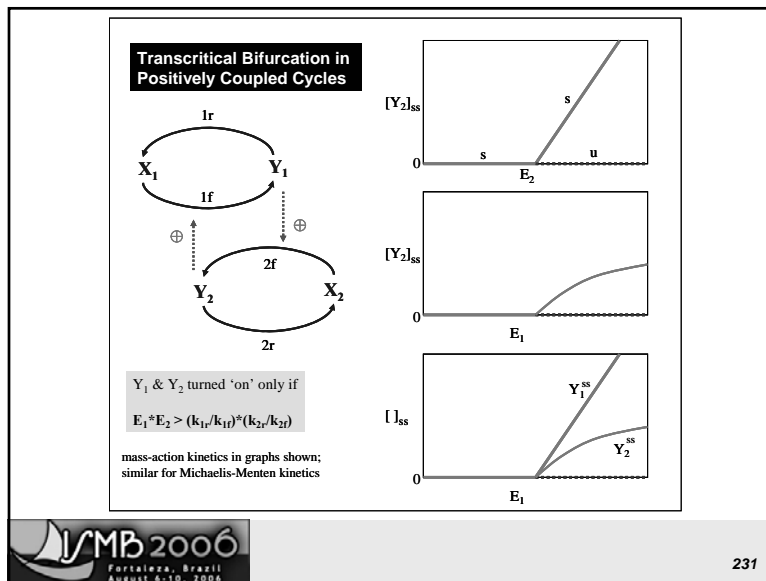
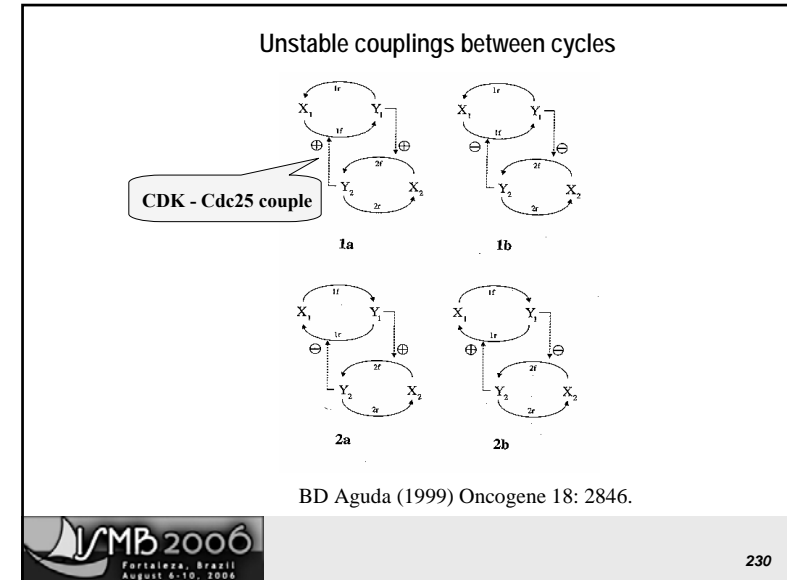
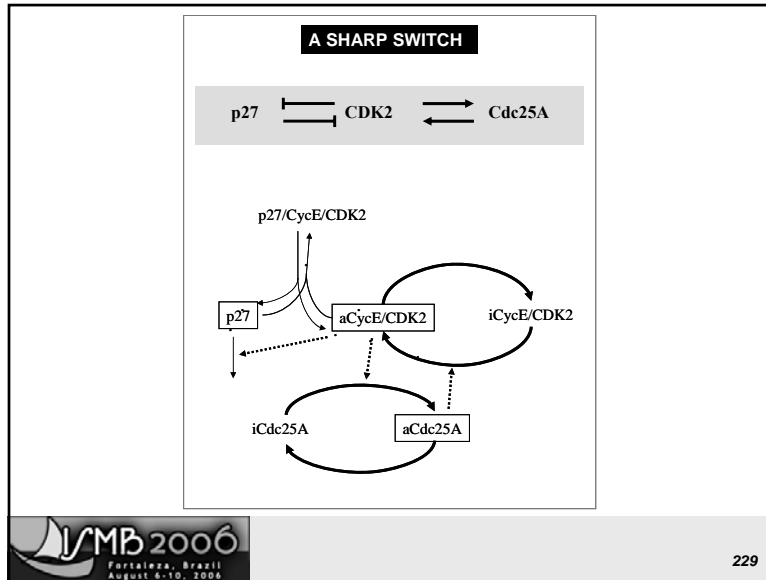
3-cycle  $T = m_{21} m_{13} m_{32}$

## sufficient instability conditions

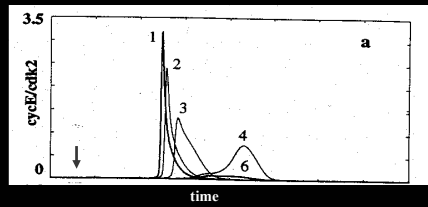
- [1]  $S > 0$
- [2]  $T < 0$
- [3]  $SD < T$  when  $T > 0$

## Model Subnetwork for the Initiation of S phase





### Simulation of CDK2 activation



- 1 sustained
- 2  $t_{\text{off}} = 80$
- 3  $t_{\text{off}} = 50$
- 4  $t_{\text{off}} = 30$
- 5  $t_{\text{off}} = 29$
- 6  $t_{\text{off}} = 28$

GFs

Cyclin D/CDK4

BD Aguda & Y Tang (1999) Cell Prolif. 32: 321.