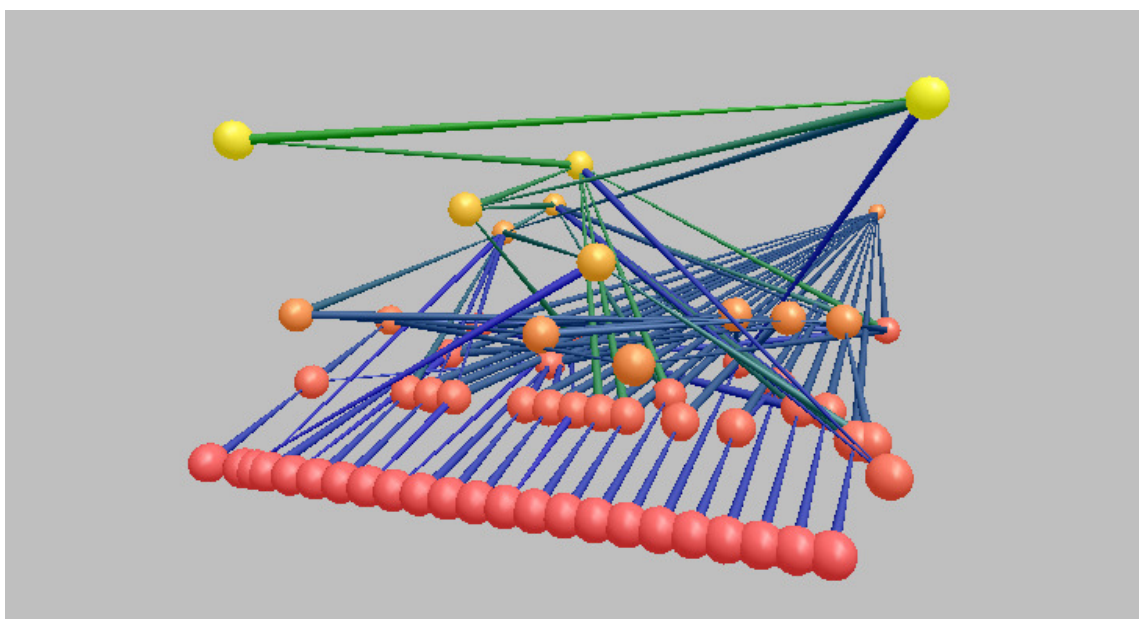# Analysis and simulation of metabolic networks: Application to HEPG2

J. M. Zaldívar and F. Strozzi[1]

[1]Carlo Cattaneo University, Engineering Faculty, Castellanza, Italy

The mission of the JRC-IHCP is to protect the interests and health of the consumer in the framework of EU legislation on chemicals, food, and consumer products by providing scientific and technical support including risk-benefit assessment and analysis of traceability.

***Europe Direct is a service to help you find answers
to your questions about the European Union***

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/

Cover. Un-weighted network representation of metabolic fluxes. Figure generated with the FoodWeb3D software (Wiliams, 2003).

EXECUTIVE SUMMARY

The first step in the toxicity mechanism involves the interaction between the chemical and one or more macromolecular targets which may consists on genetic material, enzymes/proteins, transport molecules, receptors, etc. Therefore, to progress in this direction, it is necessary to develop approaches able to link adverse effects (or measurable physiological changes) at high level to initiating events at molecular level through a complete pathway of biochemical reactions. Chemical toxicity in most of the cases has a metabolic component and metabolic pathways are the most studied biochemical reactions. For these reasons we have started to analyze modelling approaches for studying toxicity at molecular level using the metabolism as a case study. In addition, to facilitate the study, we have focused on cell-based assays –in this case HepG2 cell lines.

Two approaches have been investigated. The first consists on the analysis of metabolic fluxes in HepG2 under different toxicants concentrations using complex network analysis tools. This is important when a correlation between dose-response curves and metabolomic data has to be established. The second approach is based on the implementation and application of a simplified dynamic model for the HepG2 metabolism. Both approaches are complementary, in the first case only metabolic fluxes are necessary and therefore a more detailed description of HepG2 metabolic network can be developed; whereas the second approach requires not only the specification of the biochemical reactions, but also their kinetic parameters and, therefore, at the moment, the data necessary to describe the HepG2 metabolism with the same level of detail is not available. The results suggest that a combined approach, starting from metabolic fluxes and then solving the kinetic equations as our experimental data and knowledge of metabolic pathways increases is the most promising avenue to develop a quantitative approach of toxicity at molecular level. In addition, there is the necessity to link the molecular level approach with the cell growth and mortality with an intermediate level that translates metabolism changes into measurable parameters in cell-based assays for the experimental validation.

CONTENTS

1. INTRODUCTION

Systems Toxicology integrates the methodologies of conventional toxicology, of high throughput molecular genomics, proteomics, metabonomics with bioinfomatics to analyse and predict the effects of environmental stressors or toxicants from chemical exposure to adverse effects. The final objective of Systems Toxicology is to be able to describe the response of a functioning organism to toxicants at all different levels of biological organization and complexity (NRC, 2007).

The basic idea is to combine the information from several sources to gain a deeper understanding on the mechanisms of toxic action and to develop approaches allowing a more sensitive and earlier detection of adverse effects in toxicity studies. This constitutes a big challenge since it requires to study interactions between genes, proteins and metabolites and to integrate the results from different theoretical (EPA, 2003; Kitano 2002a,b) and experimental set-ups. However, a better understanding of the molecular mechanisms of toxicity is necessary to improve the extrapolation between experimental results, from low dose to high dose, from short term to chronic exposure, etc; therefore, reducing the animal testing as well as the dose levels needed for toxicity elucidation.

The first step in the toxicity mechanism normally involves the interaction between the chemical and one or more macromolecular targets which may consists on genetic material, enzymes/proteins, transport molecules, receptors, etc. The general view is that the differences between chemicals acting with the same mode of action depend on the differences of interaction with the targets (Rabinowitz et al. 2008). However, to progress beyond the mechanism or mode of action concept, it is necessary to develop approaches able to link adverse effects to initiating events through a complete pathway of biochemical and physiological changes, i.e. from receptor activation, to molecular discrete events, to cellular effects, to target organ responses, to apical end points, to physiological responses of the organism. This detailed knowledge would facilitate the use of models to treat certain chemicals as a family rather than individuals as well as to understand and predict minimal dose (Thresholds of Toxicological Concern –TTC- paradigm, see Kroes et al. 2005).

In general terms, metabolism can be defined as the overall set of chemical reactions that occur in a living organism and the chemical reactions may be separated in two categories: catabolic –breakdown and produce energy- and anabolic –use energy to build up essential components-. Chemical toxicity in most of the cases has a metabolic component (Patterson and Idle, 2009), which in most cases are detoxification products (the conversion of the toxic compound in more soluble and easy to excrete). Therefore one way to gain insight into the toxicity mechanisms is to characterize the differences in metabolism when biological matrices are exposed to a certain substance that exerts a toxic effect.

The qualitative and quantitative study of the metabolome is the subject of metabolomics where experimental techniques such as NMR (Nuclear Magnetic Resonance) and LC-MS (Liquid

chromatography-mass spectrometry) are employed in combination with advanced statistical and data modelling approaches (Patterson and Idle, 2009). A complement to this approach which may help to elucidate toxic effects at molecular level and formulate hypothesis that can be further validated using metabolomic approach consists on the development and implementation of mathematical models at different complexity levels.

In this work, we have started to analyze several modelling approaches proposed for studying toxicity at molecular level using HepG2 cell lines. Two approaches have been investigated. The first approach consists on the analysis of metabolic fluxes in HepG2 (Metabolic Flux Balance, MFB) under different toxicants concentrations by complex network analysis tools. In addition, to the understanding of the results provided by MFB, it is important to check which parameters can be useful to characterize metabolomic data obtained from real experiments with the corresponding dose-response curves. The second approach is based on the development and application of a simplified dynamic model of metabolism. Whereas in the first case only metabolic fluxes are necessary and a quasi steady-state condition is assumed for internal metabolites, i.e. constant concentrations during time; the second approach requires the specification not only of the most important biochemical reactions –as in the first case-, but also their kinetic parameters, which, with our actual level of knowledge, implies considering lumped metabolic reactions, e.g. only one reaction for protein production. The preliminary results suggest that a combined approach starting from metabolic fluxes and then developing the metabolic kinetic equations, as our knowledge and experimental data increases, is the most promising avenue to develop a quantitative approach of toxicity at molecular level. The results obtained using complex network analysis suggest that it is also necessary to incorporate signalling and regulatory networks to have a more complete description of molecular pathways (Lee et al., 2008). In addition, there is the necessity to link the molecular level approach with the cell growth and mortality with an intermediate level that translates metabolism changes into measurable parameters in cell-based assays.

2. ANALYSIS AND MODELLING APPROACHES TO METABOLIC NETWORKS

The analysis of complex metabolic networks has a static and a dynamic component. In the first case the metabolic network, see Fig. 1 as an example, is studied, using techniques developed from complex network analysis, to characterize its properties (Albert and Barabási, 2002; Newman, 2003). There have been a large number of studies on the statistical properties of such networks, which are usually considered as directed graphs in which the nodes correspond to metabolites and the connections between them correspond to reactions (edges), see Jeong et al., (2000); Fell and Wagner, (2000); Podani et al., (2001); Wagner and Fell, (2001); Stelling et al., (2002); Gulmerà and Nunes Amaral, (2005); amongst others. For example, Jeong et al. (2000) showed that metabolic networks were Small World networks; however, in their analysis they included connections through current metabolites, e.g. ATP, ADP, NADH, NAD$^+$, etc. In a subsequent analysis Ma and Zeng (2003) showed that even after the removal of these connections, the networks still showed a power law distribution, but with longer pathways and diameters than previously calculated by Jeong et al. (2000). In addition, they identified several hub metabolites that were common in all organisms.
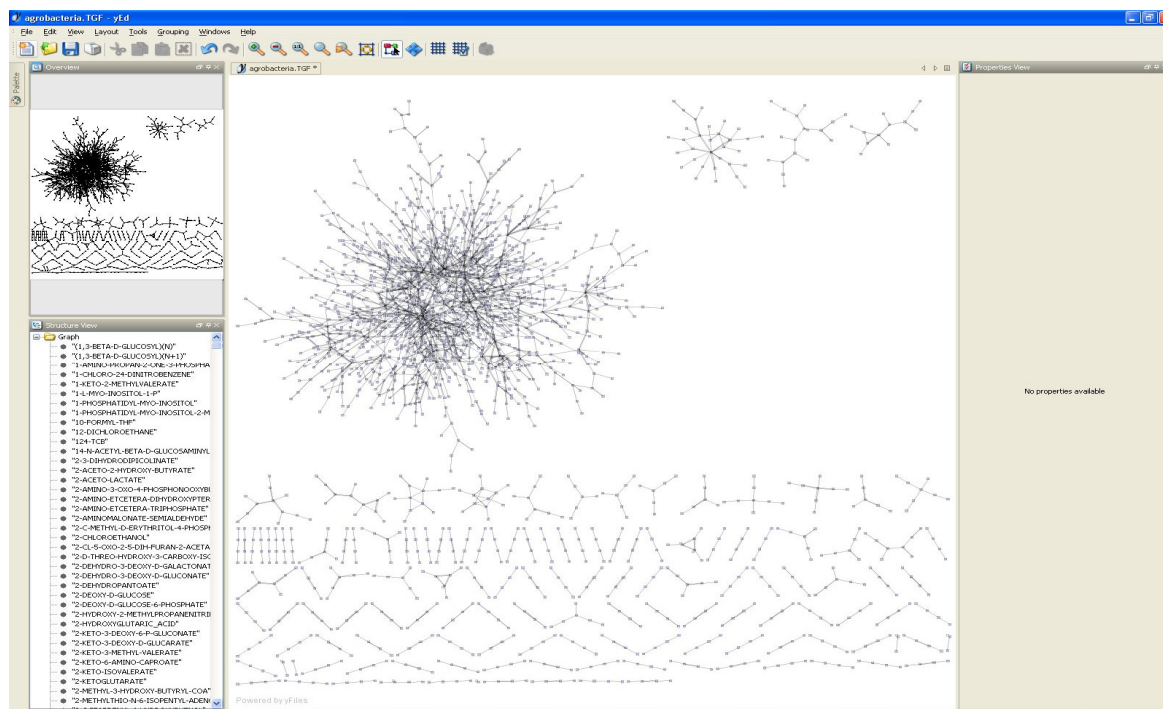

Figure 1. Example of a reconstructed metabolic network.

In the second case, i.e., the dynamic analysis, the objective is to obtain the time dependent changes of concentration of metabolites, proteins or other cellular components and reaction fluxes. An important consideration is the different time constants, the rate of change of a variable, that occur normally in

biochemical reaction networks that may led to a separation between slow and fast dynamics and, therefore, a reduction of the complexity of the problem. Dynamic mass balance models are typically formulated as a set of ordinary differential equations. This formulation has several important assumptions: the system is deterministic, i.e. the effects due to thermal noise or stochasticity due to few molecules present in the cell are not considered, cells are homogeneous with constant volume, and temperature is assumed constant.

The compact form of the mass balance equations is:

$$\frac{dx}{dt} = S \cdot v(x) \tag{1}$$

where S is the stoichiometic matrix, v is the vector of reaction fluxes and x is the vector of the concentration of metabolites. Figure 2 shows an example of a metabolic network, and the stoichiometric matrix.
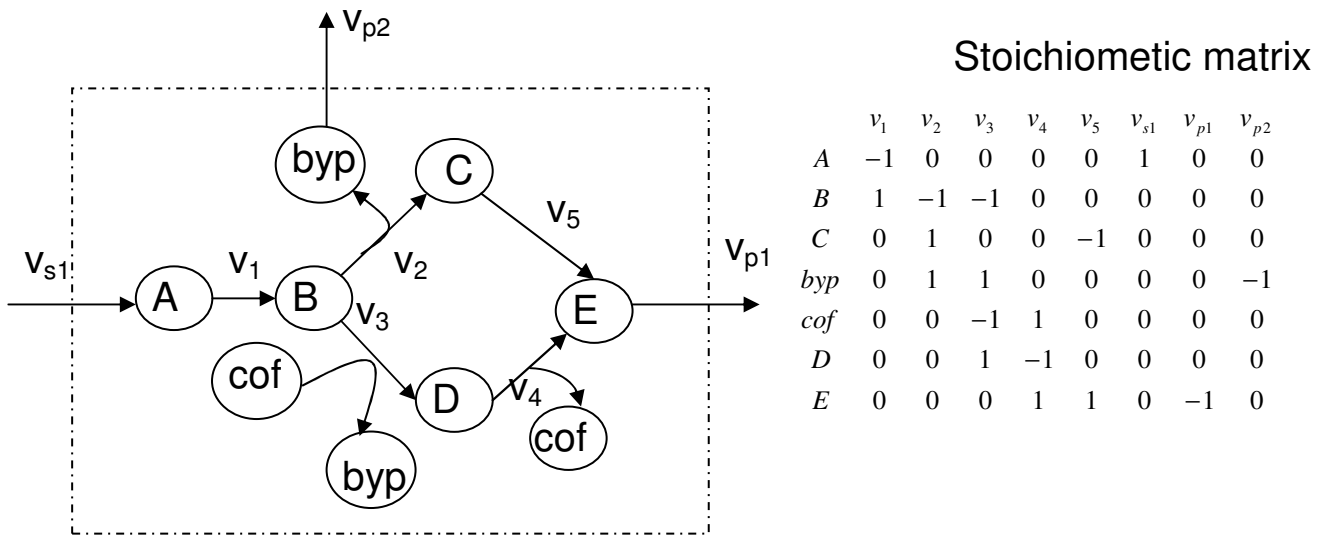

Stoichiometic matrix

	v_1	v_2	v_3	v_4	v_5	v_{s1}	v_{p1}	v_{p2}
A	−1	0	0	0	0	1	0	0
B	1	−1	−1	0	0	0	0	0
C	0	1	0	0	−1	0	0	0
byp	0	1	1	0	0	0	0	−1
cof	0	0	−1	1	0	0	0	0
D	0	0	1	−1	0	0	0	0
E	0	0	0	1	1	0	−1	0

Figure 2. Example of a metabolic network and the corresponding stoichiometic matrix S. The v_i 's correspond to internal fluxes, whereas v_{sj} and v_{pk} are extracellular fluxes and refer to substrates and products, respectively.

2.1. Metabolic networks characterization

- Definitions and characterization of complex networks

A graph $G = (V, E)$ is defined with a *node* set V where n denotes the number of vertices, and a *connection* set E, where m denotes the number of connections. The topology of the graph can be presented by using the $n{\times}n$ *adjacency matrix A*, where $a_{ij} \neq 0$ if $(i, j) \in E$ and 0 otherwise. In the case of *undirected graphs* the adjacency matrix is symmetrical, $a_{ij} = a_{ji}$. It is possible also to assign certain numerical values, or weights, to the connections. In this work, the fluxes between metabolites are the the weights that connect two metabolites, $a_{ij}=v_{ij}$. In addition, we have also used the square root of their inverses, $a_{ij}=1/(v_{ij})^{1/2}$, to mimic a distance matrix, i.e., higher the flux smaller the distance.

There are several measures to characterise complex networks. Here we briefly define those that we have used in this work to characterize the metabolic networks.

- Local Measures: centrality measures

There are a number of different measures of *centrality*. The *degree centrality* of a node in an undirected network is the number of ties a node has and does not take into account the connections weights. The *betweenness centrality* can be defined on a node or a connection, and it measures how many times a node or a connection occur on the shortest paths (geodesic) between other vertices in comparison with the total number of shortest paths. Node *efficiency centrality* is based on the idea that the most important vertices are most efficient in the connection with the other vertices where efficiency among two nodes is equal to inverse of the shortest path length (Latora and Marchiori, 2001).

- Global measures

The *distance* between two vertices is the number of connections in a shortest path connecting them. The *diameter* of a network is the maximum length of all shortest paths between couples of vertices; whereas the *characteristic path length* is the average shortest path length. The *largest component* of a network is the largest subnetwork with all the nodes connected to each other directly or indirectly. In the case of social networks, the spread of information may be measured by the *connectedness*, R_0, (Heffernan et al., 2005). The expected transmission for a node with centrality degree k and with a probability r to transfer, information will be $r(k-1)$, the node at the origin is excluded. The *connectedness* number is given by the weighted average over the vertices:

$$R_0 = r \frac{\sum_i k_i (k_i - 1)}{\sum_i k_i} = r \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \qquad (2)$$

If $R_0 > 1$ then the transmission grows exponentially; if $R_0 < 1$ the information spread will die out. The dependence by mean square degree means that a node with high degree has a higher probability to receive and/or to transmit information. The spread of information occurs if and only if $r > L$, where $L = \langle k \rangle / (\langle k^2 \rangle - \langle k \rangle)$; when $\langle k^2 \rangle$ is very large in comparison with $\langle k \rangle$, L is small and then r does not need to be large to spread information. Instead of checking if $R_0 > 1$, it is possible to check for $\kappa = \langle k^2 \rangle / \langle k \rangle > 2$. κ is called the *connectivity parameter* (Newman, 2008) and the value of 2 is a percolation threshold when the giant component (i.e., the subset of the original network which is connected and largest) disappears and the network becomes a set of small clusters.

- Network robustness

Different strategies have been employed in the literature to study the robustness of complex networks. In this work we have considered attacks based on the removal of vertices which can be divided into two categories: random and selective attacks (Holme et al., 2002). In the first case, the removal does

not require a new calculation of the network parameters, whereas in the second case the ranking can be recalculated for the new network after the node removal. In this work the selective attacks are performed based on node degree, efficiency, and betweenness centrality. This means that nodes with the highest degree or with the highest betweenness are first removed, after which the ranking is recalculated in order to select, in the next step, again the nodes with highest values of degree or betweenness (i.e., dynamic updating). The consequences of these attacks are measured by observing the changes of the global properties of networks i.e., size of the largest component (and the mean size of the others), diameter, characteristic length and connectivity.

- Network conversion into time series

The algorithms developed in Strozzi et al. (2009) to convert a time series into a network or vice versa have also been applied to the metabolic networks. The approach consider the adjacency matrix of a network as a recurrence plot, which consists on the computation of the distance between points in state space (Eckmann et al., 1987), of a time series and then tries to reconstruct a time series that would generate this recurrence plot. As showed by Hirata et al. (2008) these methods are able to reproduce the topological features of the original time series. However, as Xu et al. (2008) underlined, one drawback of this approach is that if we change the order in which the network nodes are visited, then the reconstructed time series can be different, even if the network is the same; but once we have established an order in which the nodes are visited, the reconstruction of the time series is unique.

We have analysed the data with Pajek (Nooy et al., 2005) and Matlab®. Pajek has been used mostly for network visualization, whereas the network analyses have been performed with procedures written in Matlab using the functions of the MatlabBGL library (Gleich, 2006). MatlabBGL uses the Boost Graph Library to efficiently implement the graph algorithms on unweighted or weighted graphs and is designed to work with large sparse graphs with hundreds of thousands of nodes.

2.2. Metabolic flux analysis

Metabolic Flux Analysis (MFA) is a methodology developed for modelling metabolic networks by the determination of metabolic pathway fluxes (Stephanopoulos et al., 1998). In this methodology, intracellular fluxes are calculated by solving a linear system provided by a stoichiometric model (metabolic pathway map or stoichiometric matrix) and applying mass balances of the selected metabolites at steady-state conditions. As input, the linear system requires a set of measured rates, typically uptake and excretion extracellular fluxes of metabolites (substrates and products). The output is a steady state set of fluxes that solve the set of linear equations together with defined constraints and, to avoid the high number of possible solutions typical of an undetermined linear system, an objective function to maximize or minimize is also defined.

The pseudo-steady state assumption requires that the intracellular metabolites do not accumulate in the cell, which implies that Eq. (1) converts into:

$$S \cdot v = 0 \qquad (3)$$

where S is the *mxn* stoichiometric matrix and v is the flux distribution vector. Assuming that only extracellular fluxes of substrates (s) and products (p) are measurable, we can write the extended system as:

$$\begin{pmatrix} S \\ S_s \\ S_p \end{pmatrix} \cdot v = \begin{pmatrix} 0 \\ v_s \\ v_p \end{pmatrix} \qquad \text{with } v \geq 0 \qquad (4)$$

where v_s and v_p are the measured uptake and excretion rates/fluxes.

As explained above, normally the number of measurements is too small to make Eq. (3) determined. This means that the solution of Eq. (3) consists of a set of solutions Λ. Each solution of Eq. (3) can be written as a sum of nonnegative vectors, i.e. $w_i \geq 0 \ \forall_i$, as:

$$v = \sum_i c_i \cdot w_i \qquad (5)$$

which can be written in compact notation as $v = Cw$ with $w \geq 0$. C is a *nxk* matrix with $k \geq null(S) = n-m$. This nonnegative combination is called the convex basis of the solution space (Rockafellar, 1970). The convex basis vectors are also called Elementary Flux Mode (EFM) and they define a pointed convex polyhedral cone in the k-dimensional nonnegative quadrant and are essentially a collection of various pathways alternatives for converting the external substrates into internal biomass and metabolic products .

Combining Eqs. (3)-(5), we obtain

$$\begin{pmatrix} S_s \cdot C \\ S_p \cdot C \end{pmatrix} \cdot w = \begin{pmatrix} v_s \\ v_p \end{pmatrix} \qquad (6)$$

Usually $k > s + p$ and Eq. (7) provides a set of solutions Λ. This solution set can, afterwards, be refined using additional knowledge, such as necessary products to feed the cell lines that cannot have zero fluxes as solution or biosynthesis products that we know are formed but not measured.

In addition to the maximization of biomass yield, other goals have been defined such as the optimal metabolic regulation (Young et al., 2008) aiming at optimizing the distribution of uptake rates to maximize the carbon uptake by controlling the synthesis of enzymes that catalyze metabolic reactions. Furthermore, discussion on when MFA is admissible has also been the object of a considerable amount of work. For example, Song and Ramkrishna (2009) have proposed to examine the trajectories of extracellular measurements in flux space as a diagnostic tool to determine whether quasi-steady state assumption is valid.

A number of tools have been developed over the years to solve this constraint-based linear problem. A review of the different approaches and techniques may be found at Lewis et al. (2008) and software is freely available on Internet (Becker et al., 2007).

2.3. Dynamic modelling of metabolic networks

In this case, and taking into account the same assumptions that with MFA, a dynamic metabolic model for cell cultures can be described by the following set of ordinary differential equations:

$$\frac{1}{c}\frac{dx_s}{dt} = S_s \cdot v \tag{7}$$

$$\frac{1}{c}\frac{dx_p}{dt} = S_p \cdot v \tag{8}$$

$$\frac{dx_m}{dt} = S \cdot v - \mu \cdot x_m \tag{9}$$

$$\frac{1}{c}\frac{dc}{dt} = \mu \tag{10}$$

where x_s, x_p and x_m represent the extracellular (n_s substrates and n_p products) and the intracellular components (n_x), respectively; v is the vector of fluxes (n_r), μ is the growth rate of biomass and c is the biomass weight defined per unit volume of the culture. The matrices S_s (n_sxn_r), S_p (n_pxn_r) and S (n_mxn_r) represent the stoichiometic matrices relating v to the exchange and intracellular fluxes, respectively.

The growth rate μ can be obtained from the relation with the intracellular fluxes and the contribution of each flux normally available from literature (Stephanopoulos, et al., 1998) as:

$$\mu = h^T \cdot v \tag{12}$$

Conversely to MFA, in this case, it is necessary to write the kinetic equation that normally for enzyme catalyzed reactions can be written following the Michelis- Menten equation:

$$v_i = \frac{x_m^i}{K_m^i + x_m^i}V_{max} \tag{13}$$

Where K_m^i is a function of two consecutive reactions: the binding of the ligand and the enzyme which is an equilibrium reaction (k_1 and k_{-1}) and the formation of the product (k_2) to produce:

$$K_m^i = \frac{k_2^i + k_{-1}^i}{k_1^i} \tag{14}$$

and V_{max} in the maximal velocity which is directly proportional to the enzyme concentration and k_2.

The main problem with this approach, when compared to FBA (flux balance analysis), is the fact that one has to consider a high number of pathways for which a detailed reaction scheme and kinetic parameters are unknown and therefore lumped reactions are normally used.

3. APPLICATION OF COMPLEX NETWORK ANALYSIS TO METABOLIC FLUXES OF HEPG2 WHEN EXPOSED TO SUBTOXIC CONCENTRATIONS

In this Section we are going to study, using complex network analysis techniques, the effects of several toxicants on the metabolic fluxes of a HepG2 metabolic network. The metabolic fluxes were obtained by Niklas et al. (2009). They determined the effects on the cellular metabolism at subtoxic concentrations of several compounds: DMSO –to check the effects of this solvent-, amidarone, diclofenac and tacrine. To perform this calculation, they developed a metabolic network model (see figure 1) that comprised an HepG2 metabolic network with 62 fluxes: 23 extracellular fluxes (transport of compounds from outside the cell 21 substrates and 2 products), 34 intracellular reactions and 5 lumped reactions for the synthesis of carbohydrates, DNA, RNA, lipids and proteins, respectively. For a detailed description of the metabolic reactions as well as the fluxes obtained assuming pseudo-steady state, the reader is referred to the *Supplementary material* of the original publication. Figure 2 shows the un-weighted network representation of the metabolic network considered by Niklas et al. (2009).
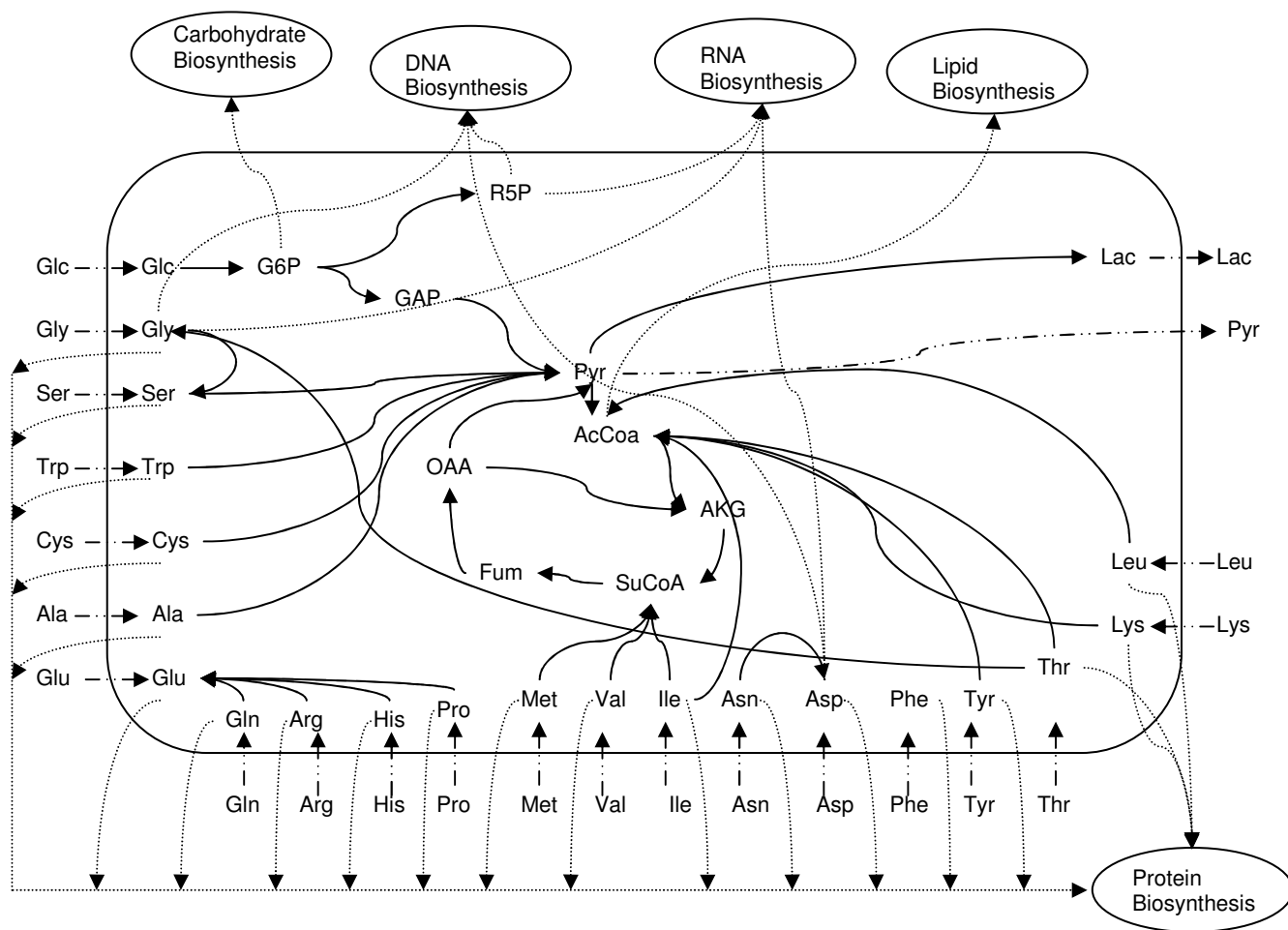
Figure 1. Schematic representation of the metabolic fluxes considered by Niklas et al. (2009) for HepG2.
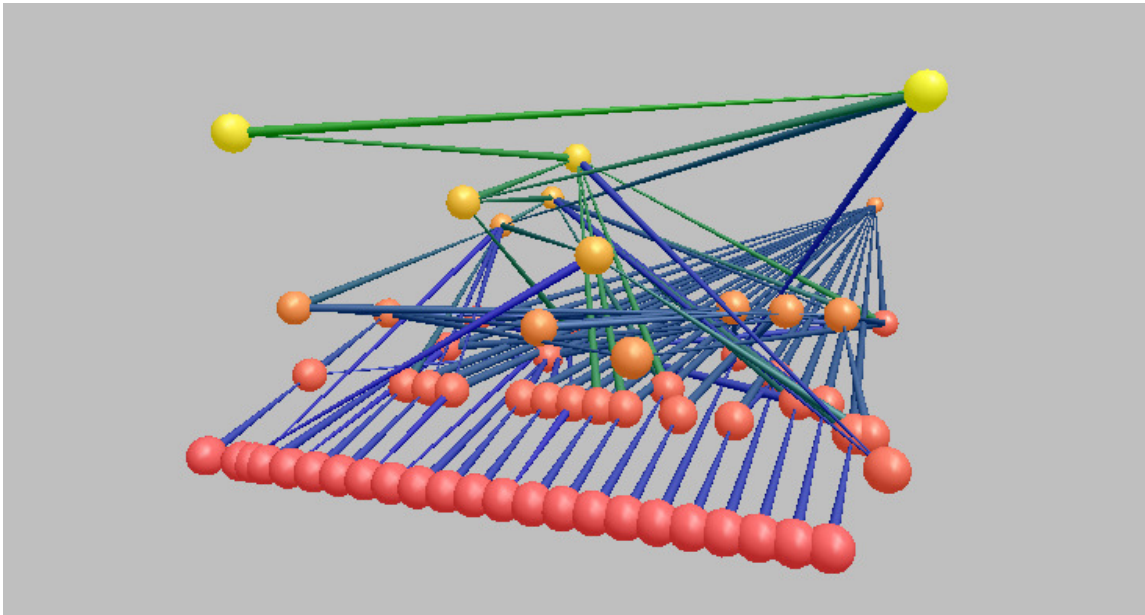
9

Figure 2. Un-weighted network representation of metabolic fluxes, ordered by distance from inside the cell, i.e., bottom: extracellular fluxes, top: Krebs's cycle. Figure generated with the FoodWeb3D software (Wiliams, 2003).

Our objective in this part of the work is to try to understand the relationship between toxic effects and the metabolic fluxes. We were expecting that, as the toxicant concentration increases, the vector of metabolic fluxes changes in some continuous form indicating that the functionalities of the metabolic network are decreasing. In this case, the main problem is how to find a way to convert the metabolic fluxes into a global measure of the performance of the network and then be able to compare this measure with a dose-response curve. In principle, one should expect that a correct metrics characterization of the changes in metabolic fluxes as a function of toxicant concentrations will provide an increasing or decreasing value, which is not the case of the results obtained calculating the simple Euclidean distance between the control (no compound) and the different treatments as it can be observed in Figure 3. In this case, the logical result would be that distance from the control should increase as a function of the dose.

Since this simple distance calculation does not seem able to characterize the changes in the metabolism, we have applied several network characterization measures. *A priori*, it is not clear which measure will be the most appropriate to characterize these datasets. For this reason we have analyzed typical global static measures, but we have also analysed the use of network attack algorithms that test the robustness of the network and, finally, the conversion of the metabolic networks into time series (Strozzi et al., 2010) and the calculation of the distances once converted into them.
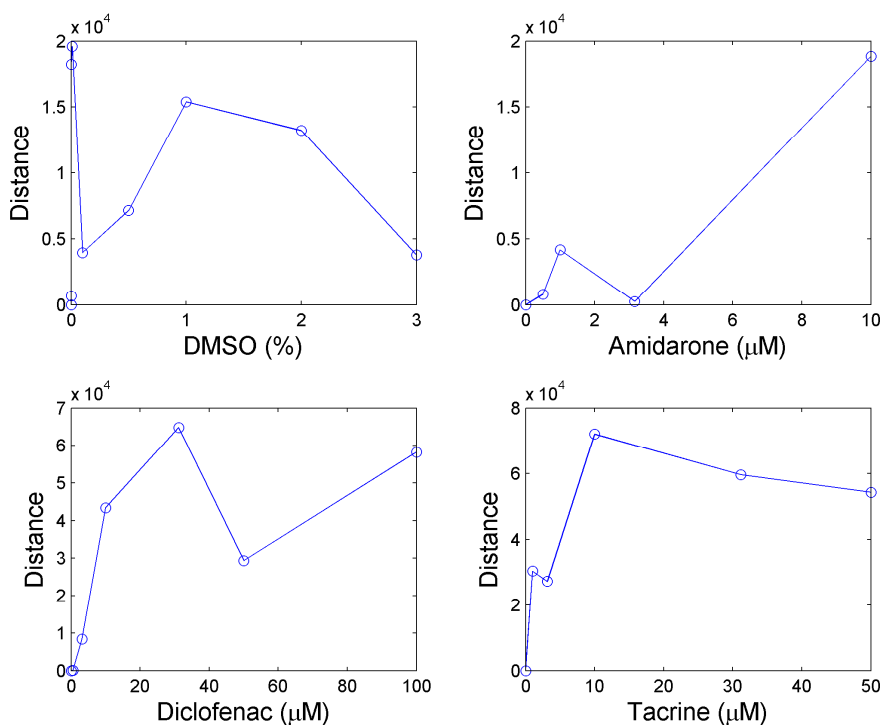
Figure 3. Sum of the Euclidean distances between fluxes differences with respect to the control experiment (without toxicant) as a function of compound concentrations.

3.1. Global static measures

We have calculated the global network measures defined in Section 2.1. There are some measures (e.g. network diameter which is based in the shortest path calculation) and algorithms (conversion to time series) that only work when the network is connected and undirected. Only in these cases, we have converted the adjacency matrix into a symmetric matrix.

As an example, we have plotted the results obtained after measuring the network efficiency, mean clustering coefficient and mean betweeness centrality as a function of the doses of chemicals, for the flux's matrices, Figures 4-6, and for the inverse of the fluxes, Figures 7-9. Even though in general terms the values follows what one should expect; for example, the efficiency of the network should decrease as the toxicant dose increases, there is no single global network measure that shows the same trend for all compounds as a function of the dose which could provide a metrics able to link fluxes with toxic effects.
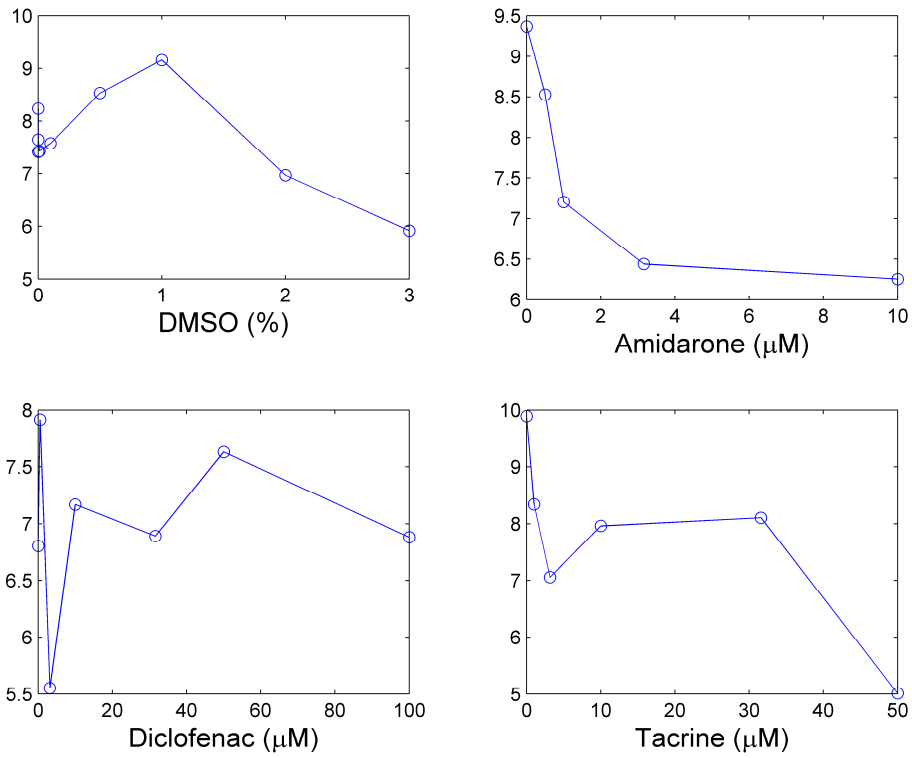
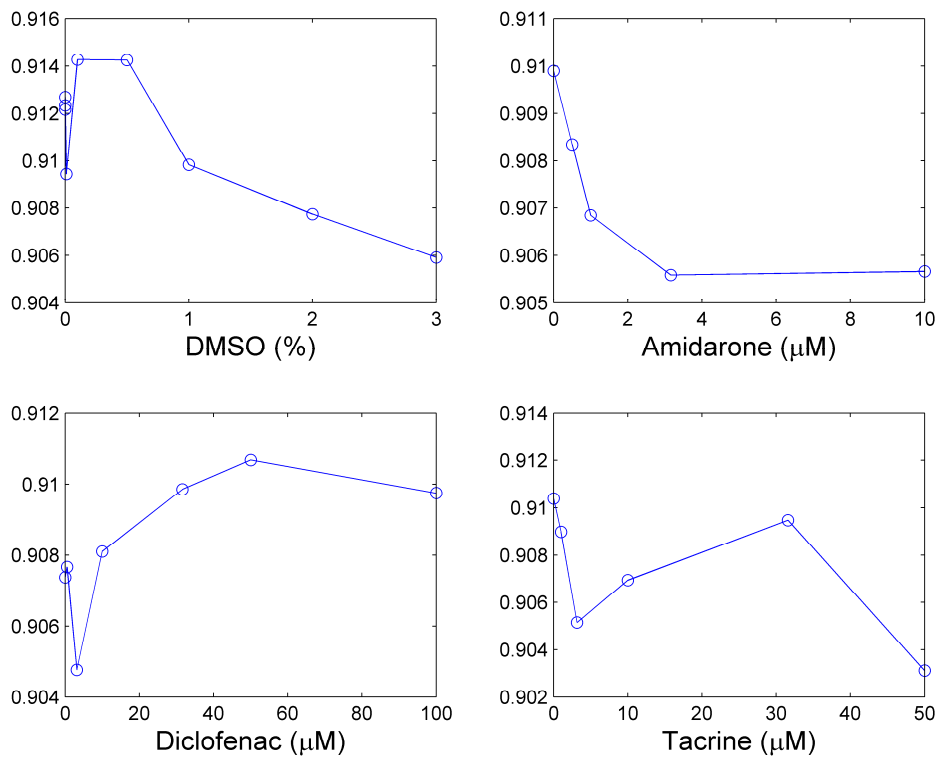Figure 4. Evolution of the efficiency of the fluxes networks as function of chemical dose

Figure 5. Evolution of the mean clustering coefficient of the fluxes networks as function of chemical dose.
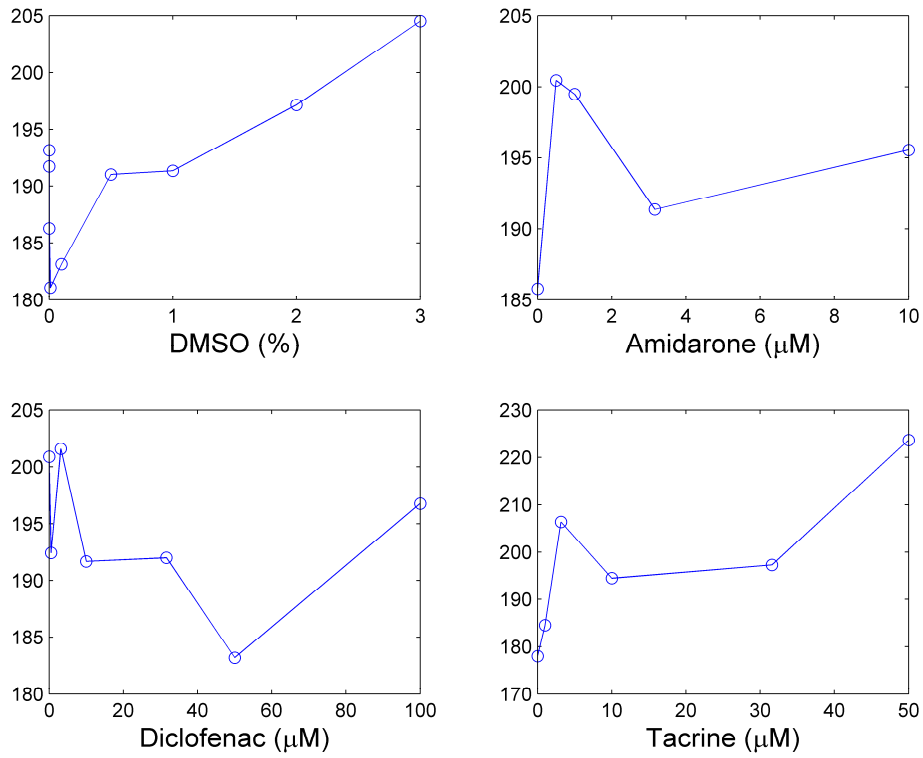
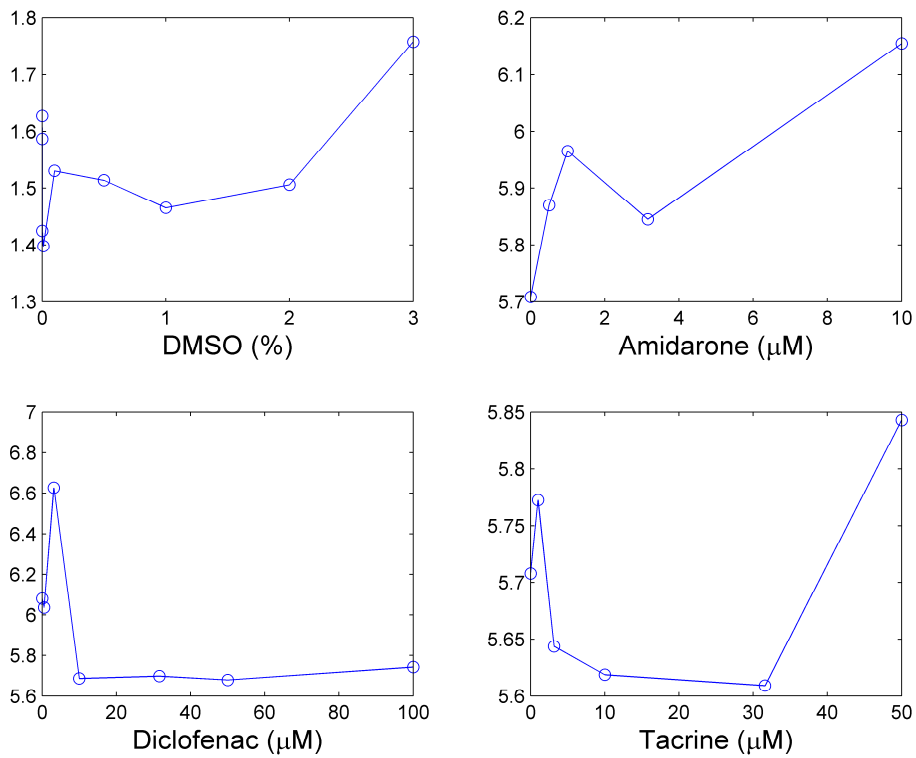Figure 6. Evolution of the mean betweeness centrality coefficient of the fluxes networks as function of chemical dose.

Figure 7. Evolution of the efficiency of the inverse of the fluxes networks as function of chemical dose.
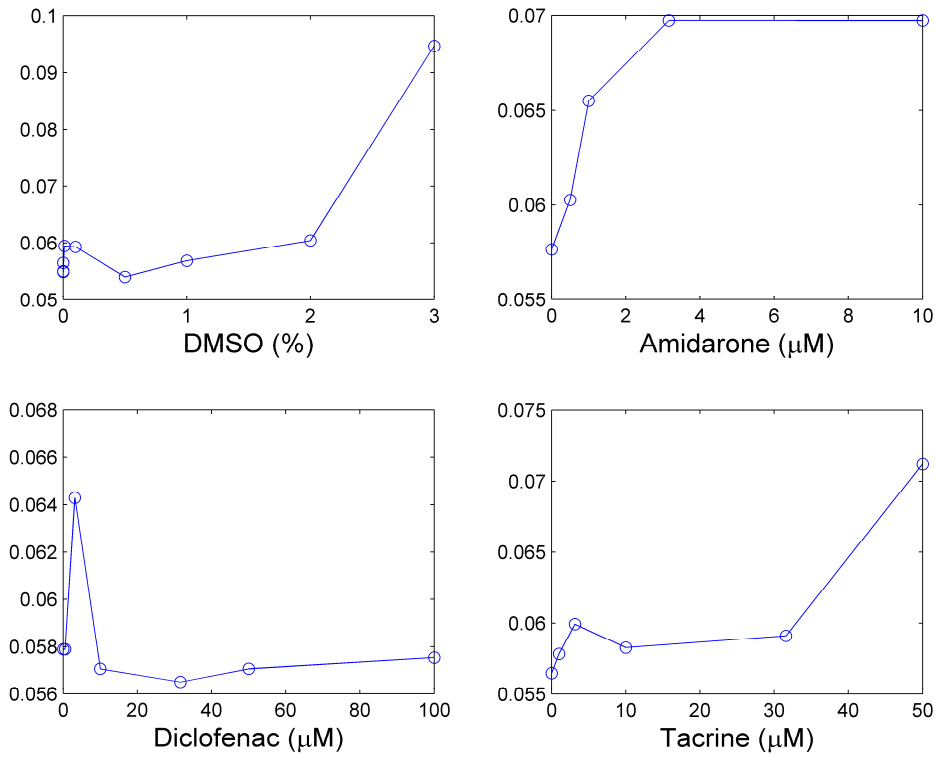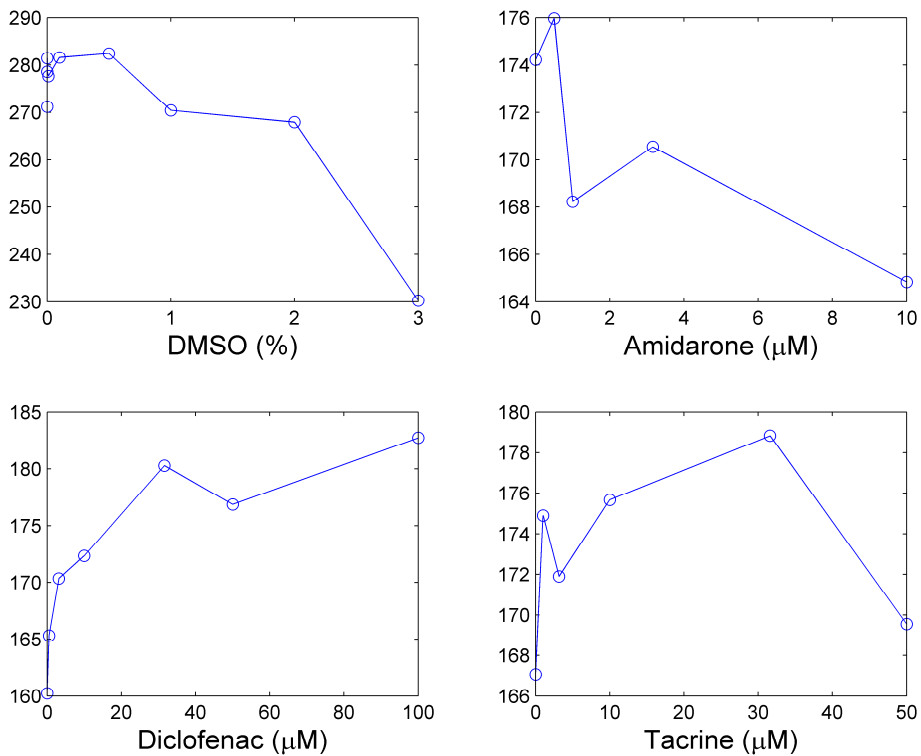
13

Figure 8. Evolution of the mean clustering coefficient of the inverse of the fluxes networks as function of chemical dose.

Figure 9. Evolution of the mean betweeness centrality coefficient of the inverse of the fluxes networks as function of chemical dose.

3.2. Network Robustness

Since the global network measures do not provide a measure that links the toxic dose with the metabolic network functioning, we have tested measures based on attacking the network and measuring how the network is able to resists these attacks that are normally divided between random attacks, i.e. a node is removed randomly from the network, and selective attacks, i.e. the node that has the higher degree is removed and the network performance evaluated. Then the degree is re-calculated for the rest of the nodes and the process is repeated until the value of the measure reaches a certain threshold or until all nodes have been removed.

Also in this case there is no a clear correlation between network robustness under random and selective attacks and toxic dose. As an example, we have plotted in Figures 10-13 the results for selective attacks based on the connectivity of the network for the four compounds, but similar results have been obtained for random or other types of selective attacks.
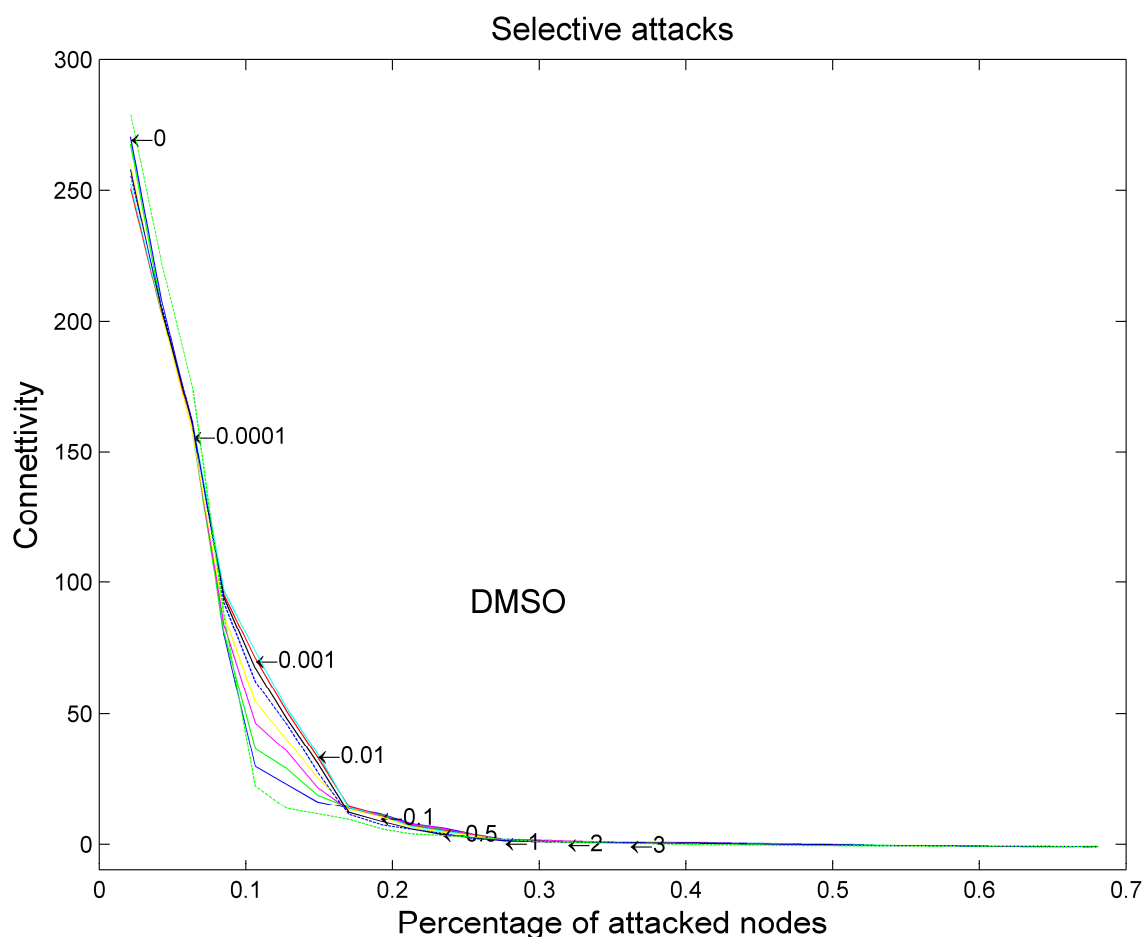

Figure 10. Evolution of the connectivity of the DMSO networks as a function of the percentage of nodes attacked (0 % blue line, 0.0001 % green line, 0.001 % red line, 0.01% cyan line, 0.1 % magenta line, 0.5 % yellow line ,1 % black line, 2% discontinuous blue line, 3 % discontinuous green line).
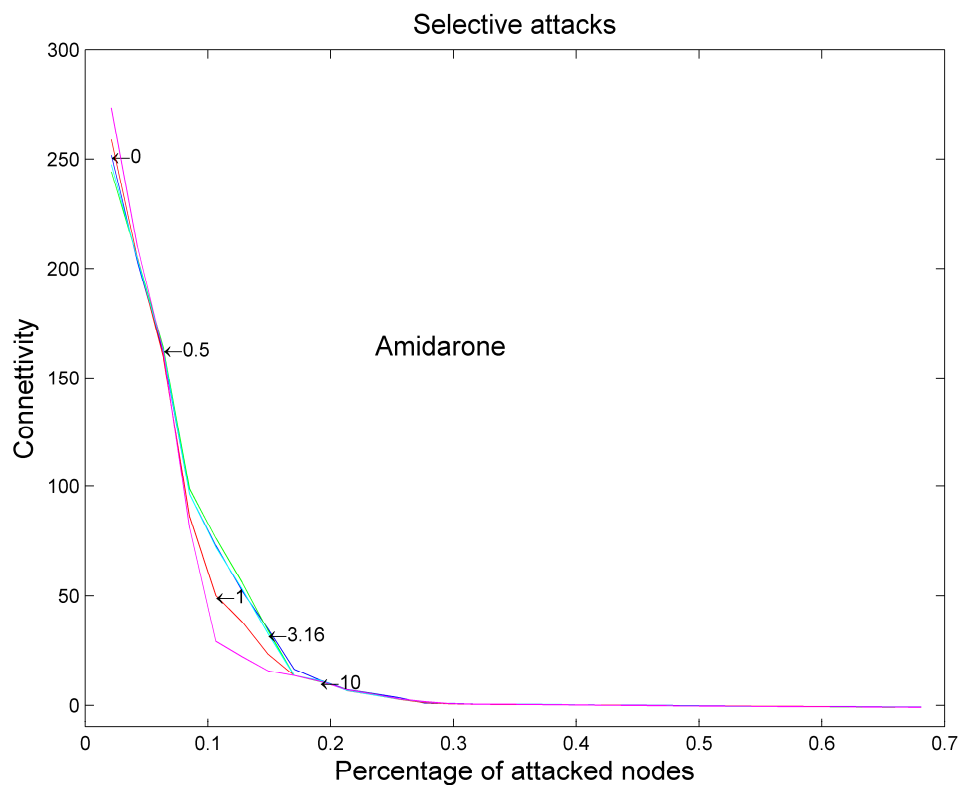
Figure 11. Evolution of the connectivity of the Amidarone networks as a function of the percentage of nodes attacked (0 μM blue line, 0.5 μM green line, 1 μM red line, 3.16 μM cyan line, 10 μM magenta line).
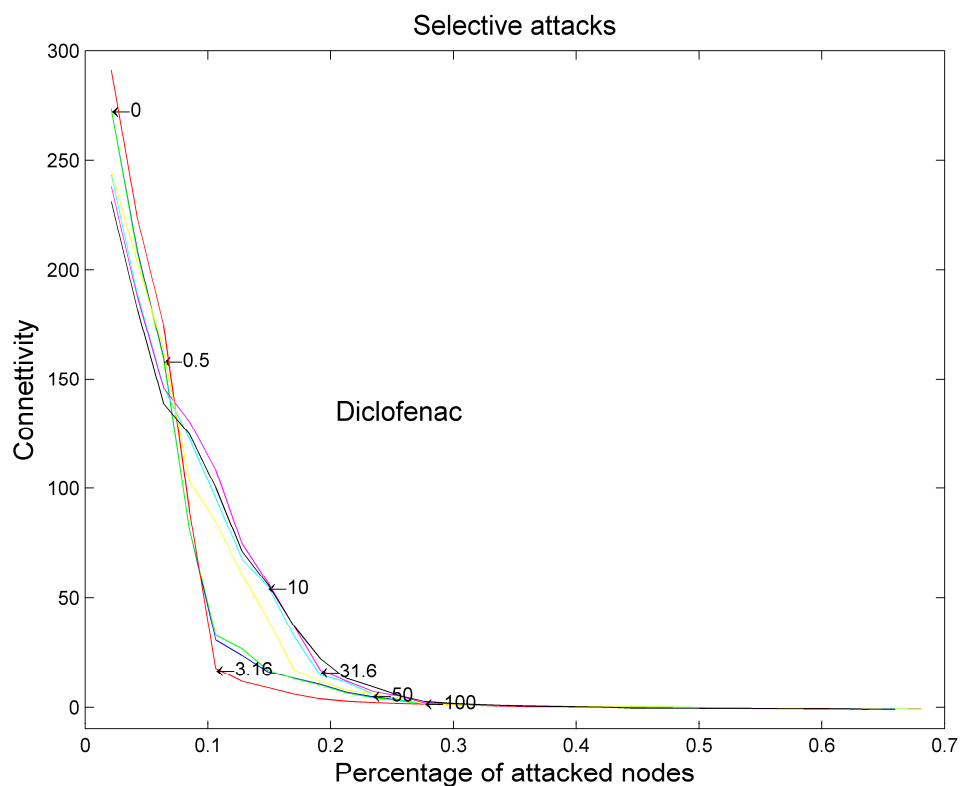

Figure 12. Evolution of the connectivity of the Diclofenac networks as a function of the percentage of nodes attacked (0 μM blue line, 0.5 μM green line, 3.16 μM red line, 10 μM cyan line, 31.6 μM magenta line, 50 μM yellow line ,100 μM black line).
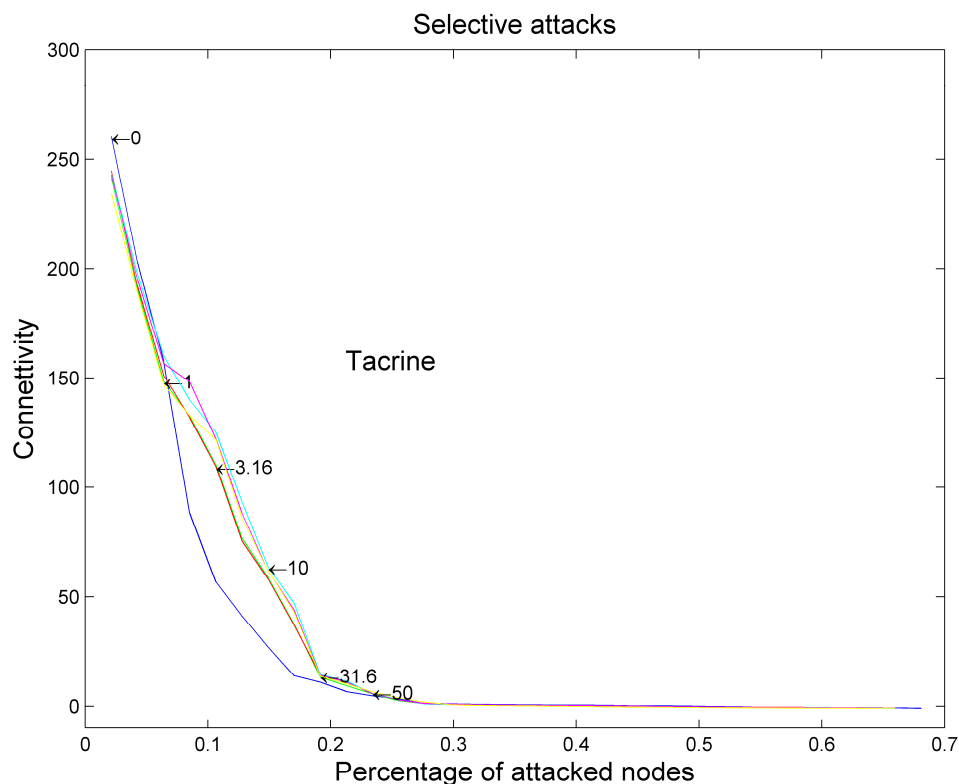
Figure 13. Evolution of the connectivity of the Tacrine networks as a function of the percentage of nodes attacked (0 µM blue line, 1 µM green line, 3.16 µM red line, 10 µM cyan line, 31.6 µM magenta line, 50 µM yellow line).

3.3. Time series conversion

The nine time series obtained for the case of the fluxes at several DMSO percentages are shown in Fig. 14. Since the ordering of the metabolites and fluxes is the same in all matrices and the calculated fluxes does not change appreciably the generated time series are quite similar. However several differences can be observed around the point number 10 and between points 41-50. Point 10 corresponds to OAA, whereas points 41-50 to several amino acids that change direction in the fluxes.

In this case, we can also calculate a Euclidean distance between the control (no toxicant) and the different concentrations introduced. The results of this calculation are shown in Figure 15. As it can be observed also in this case, there is no clear trend that can correlate the increase in the concentration of the chemical with an increase/decrease of some property of the network.
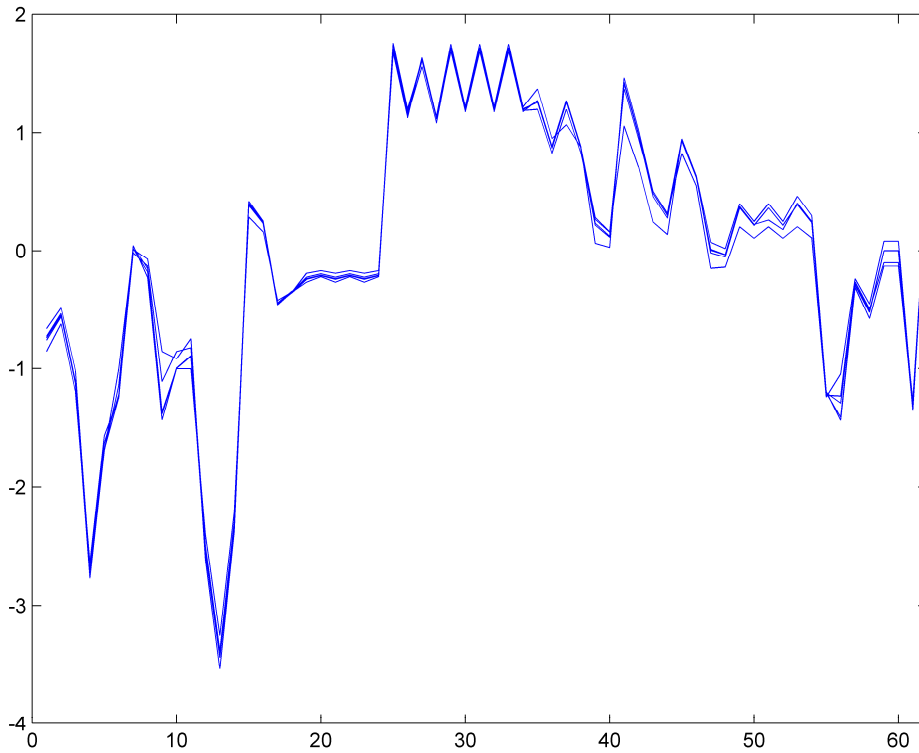
17

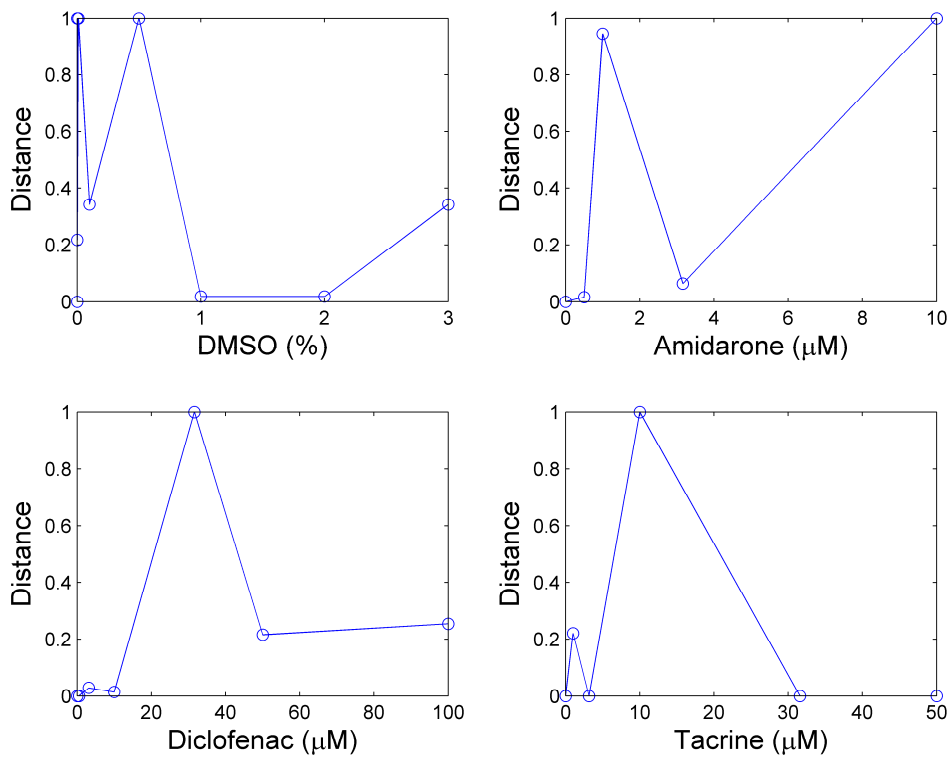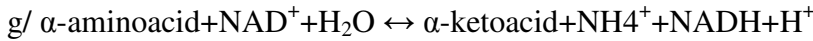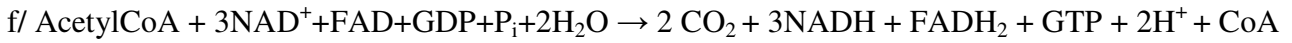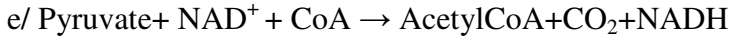Figure 14. Time series obtained from the metabolic flux matrices generated at nine different percentages of DMSO.
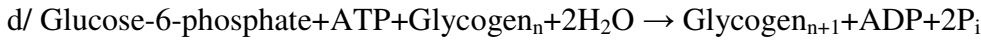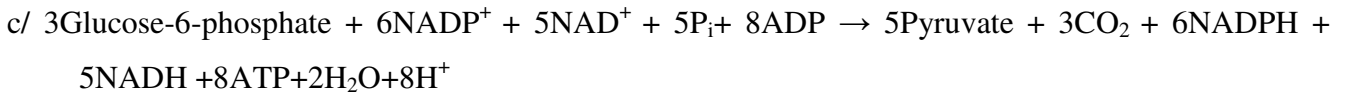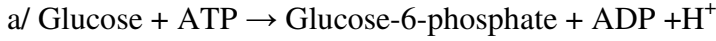

Figure 15. Normalized sum of the Euclidean distances between the points in the generated time series with respect to the control (no compound) as a function of tested concentrations

4. DYNAMIC MODELLING OF HEPG2

The dynamic model of HepG2 metabolism is based on the HEMET (HEpatocyte METabolism) model developed by De Maria et al. (2008). We describe briefly the equations here. However, for a detailed description the reader is referred to the original paper.

4.1. Implementation of a dynamic model for HepG2 metabolism

The model considers a schematic metabolic network represented by the following lumped biochemical reactions:

a/ Glucose + ATP → Glucose-6-phosphate + ADP +H^+

b/ Glucose-6-phosphate+$2P_i$+3ADP+$2NAD^+$ → 2Pyruvate+3ATP+2NADH+H^++$2H_2O$

c/ 3Glucose-6-phosphate + $6NADP^+$ + $5NAD^+$ + $5P_i$+ 8ADP → 5Pyruvate + $3CO_2$ + 6NADPH + 5NADH +8ATP+$2H_2O$+$8H^+$

d/ Glucose-6-phosphate+ATP+$Glycogen_n$+$2H_2O$ → $Glycogen_{n+1}$+ADP+$2P_i$

e/ Pyruvate+ NAD^+ + CoA → AcetylCoA+CO_2+NADH

f/ AcetylCoA + $3NAD^+$+FAD+GDP+P_i+$2H_2O$ → 2 CO_2 + 3NADH + $FADH_2$ + GTP + $2H^+$ + CoA

g/ α-aminoacid+NAD^++H_2O ↔ α-ketoacid+$NH4^+$+NADH+H^+

h/ 8acetylCoA+7ATP+14NADPH+ $6H^+$ → palmitate+$14NADP^+$+8CoA+$6H_2O$+7ADP+$7P_i$

and 22 state variables reported in Table 1.

Table 1. State variables in the model of HepG2 metabolism.

State variable	Symbol	State variable	Symbol
Glucose	*glu*	Fumarate	*fumarate*
Glucose-6-phosphate	*glu6P*	L-malate	*Lmalate*
Pyruvate 1	*pyruv₁*	Ossalacetate	*Ossalacetate*
Pyruvate 2	*pyruv₂*	NADH	*NADH*
Glycogen	*n_residual*	ATP	*ATP*
Acetyl CoA	*AcetylCoA*	Aminoacids	*a.a.*
Citrate	*citrate*	Urea	*urea*
isocitrate	*isocitrate*	Acetyl CoA eq.	*A_CoA_eq*
αKetogl	*αKetogl*	Palmitate	*palmitate*
Succinyl CoA	*Succinyl CoA*	NADPH	*NADPH*
Succinate	*Succinate*	Albumin	*album*

Following De Maria et al. (2008), the mass balance equations for each state variable may be written as:

$$\frac{d[glu]}{dt} = -\alpha_1[glu] \tag{15}$$

$$\frac{d[glu6P]}{dt} = \beta_2[glu] - \alpha_2[glu6P] \tag{16}$$

$$\frac{d[pyruv_1]}{dt} = \beta_3[glu6P] - \alpha_3[pyruv_1] \qquad \text{if } [ATP] < \sigma_1 \tag{17}$$

$$\frac{d[pyruv_2]}{dt} = \beta_4[glu6P] - \alpha_3[pyruv_2] \qquad \text{if } [ATP] > \sigma_1 \text{ and } [NADPH] < \theta_1 \tag{18}$$

$$[pyruv] = [pyruv_1] + [pyruv_2] \tag{19}$$

$$\frac{d(n_{residual})}{dt} = \alpha_4 \left| \frac{d[glu6P]}{dt} \right| \qquad \text{if } [ATP] > \sigma_1 \text{ and } [NADPH] > \theta_1 \tag{20}$$

$$\begin{cases} \dfrac{d[AcetylCoA]}{dt} = \beta_5[pyruv] - \alpha_5[AcetylCoA] & \text{if } [ATP] < \sigma_2 \text{ or } [NADPH] < \theta_2 \\[3mm] \dfrac{d[AcetylCoA]}{dt} = \beta_5[pyruv] - \alpha_6[AcetylCoA] & \text{if } [ATP] > \sigma_2 \text{ and } [NADPH] > \theta_2 \end{cases} \tag{21}$$

$$\frac{d[citrate]}{dt} = \tau_1[AcetylCoA] - \lambda_1[citrate] \tag{22}$$

$$\frac{d[isocitrate]}{dt} = \tau_2[citrate] - \lambda_2[isocitrate] \tag{23}$$

$$\frac{d[\alpha Ketogl]}{dt} = \tau_3[isocitrate] - \lambda_3[\alpha Ketogl] \tag{24}$$

$$\frac{d[succinylCoA]}{dt} = \tau_4[\alpha Ketogl] - \lambda_4[succinylCoA] \tag{25}$$

$$\frac{d[succinate]}{dt} = \tau_5[succinylCoA] - \lambda_5[succinate] \tag{26}$$

$$\frac{d[fumarate]}{dt} = \tau_6[succinate] - \lambda_6[fumarate] \tag{27}$$

$$\frac{d[Lmalate]}{dt} = \tau_7[fumarate] - \lambda_7[Lmalate] \tag{28}$$

$$\frac{d[ossalacetate]}{dt} = \tau_8[Lmalate] - \lambda_8[ossalacetate] \tag{29}$$

$$\frac{d[NADH]}{dt} = \sum_{i=1}^{6} \gamma_i [prod.NADH]_i - \alpha_7[NADH][ATP] \tag{30}$$

where $[prod.NADH]_1 = [\alpha Ketogl]$; $[prod.NADH]_2 = [fumarate]$; $[prod.NADH]_3 = [AcetylCoA]$; $[prod.NADH]_4 = [pyruv_1]$; $[prod.NADH]_5 = [pyruv_2]$; $[prod.NADH]_6 = [A\_CoA\_eq]$ if $[ATP] < \sigma_1$ or $[NADPH] > \theta_1$ (Gianni Orsi , *personal communication*).

$$\frac{d[ATP]}{dt} = \sum_{i=1}^{4} \xi_i [prod.ATP]_i - \rho[ATP] \sum_{j=1}^{6} \psi_j [cons.ATP] \tag{31}$$

where $[prod.ATP]_1 = [pyruv_1]$; $[prod.ATP]_2 = [pyruv_2]$; $[prod.ATP]_3 = [succinate]$; $[prod.ATP]_4 = [NADH]$; and $[cons.ATP]_1 = \alpha_8 \eta [a.a.]$; $[cons.ATP]_2 = [glu6P]$; $[cons.ATP]_3 = [ATP][NADPH]|d[glu6P]/dt|$; $[cons.ATP]_4 = [AcetylCoA]/8$;

$[cons.ATP]_5 = \left| d[album] / dt \right|$ if $[ATP] < \sigma_3$ and $[palmitate] > \omega^0$.and $[cons.ATP]_6 = 1$ for maintenance (Gianni Orsi , *personal communication*).

$$\frac{d[a.a.]}{dt} = -\alpha_8[a.a.] \tag{32}$$

$$\frac{d[urea]}{dt} = \eta \left| \frac{d[a.a.]}{dt} \right| \tag{33}$$

$$\frac{d[A\_CoA\_eq]}{dt} = \beta_6[a.a.] - \alpha_9[A\_CoA\_eq] \tag{34}$$

$$\frac{d[palmitate]}{dt} = \beta_7[AcetylCoA] \qquad \text{if } [ATP] > \sigma_2 \text{ and } [NADPH] > \theta_2 \tag{35}$$

$$\frac{d[NADPH]}{dt} = \sum_{i=1}^{2} \phi_i [prod.NADPH]_i - \alpha_{10}[NADPH] \tag{36}$$

where $[prod.NADPH]_1 = [A\_CoA\_eq]$ if $[ATP] > \sigma_2$ and $[NADPH] > \theta_2$ and $[prod.NADPH]_2 = [pyruv_2]$ (Gianni Orsi , *personal communication*).

$$\frac{d[album]}{dt} = \beta_8[palmitate] - \mu[album] \qquad \text{if } [ATP] > \sigma_3 \text{ and } [palmitate] > w^0 \tag{37}$$

When conditions are not met the differential equations are set to zero.

The parameters of the model taken from De Maria et al. (2008) are summarized in Table 2.

Table 2. Metabolic HepG2 model parameters, from De Maria et al. (2008).

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
α_1	1/24	β_8	1/27	λ_8	1.80	ψ_4	7.00
α_2	3.01	τ_1	1.36	λ_9	1.80	ψ_5	2330
α_3	2.03	τ_2	1.36	γ_1	1.00	ψ_6	1.00
α_4	1.01	τ_3	1.36	γ_2	1.00	φ_1	6/5
α_5	1.04	τ_4	1.47	γ_3	1.00	φ_2	1.00
α_6	8.31	τ_5	1.48	γ_4	1.00	μ	1.00
α_7	1.08	τ_6	1.48	γ_5	0.60	η	-0.20
α_8	1/24	τ_7	1.76	γ_6	1.00	ω^0	0.008
α_9	2.11	τ_8	1.80	ξ_1	3/2	σ_1	4.98
α_{10}	1	λ_1	1.36	ξ_2	8/5	σ_2	5.00
β_2	1.03	λ_2	1.36	ξ_3	1.00	σ_3	4.98
β_3	2.03	λ_3	1.47	ξ_4	2.50	θ_1	1.00
β_4	5/3	λ_4	1.48	ρ	0.1	θ_2	0.98
β_5	1.04	λ_5	1.48	ψ_1	3.00		
β_6	1.09	λ_6	1.76	ψ_2	1.00		
β_7	1/8	λ_7	1.80	ψ_3	1.00		

4.2. Simulated results

As an example, the model was run for eight days, the initial conditions in ng/(ml·cell) were (Gianni Orsi , *personal communication*): [*glu*]=3, [*a.a.*]=2.6, [*AcetylCoA*] = 0.05, [*pyruv₁*] = 0.05, [*pyruv₂*] = 0.05, [*ATP*] = 2.0; [*NADH*] = 0.05 and at the fourth day the medium was changed and therefore the initial concentration of [*glu*] and [*a.a.*] were added again to the system. Figures 16-21 show the simulated results.

These results indicate the existence of a dynamic behaviour in the concentrations of the different metabolites when performing in vitro experiments. This behaviour should be considered if we are interested in assessing toxicity at molecular level. In addition, the change in the medium has a strong influence on the dynamics of the metabolism and subsequently on the toxicity effects if the compound has a direct or indirect influence on metabolism.
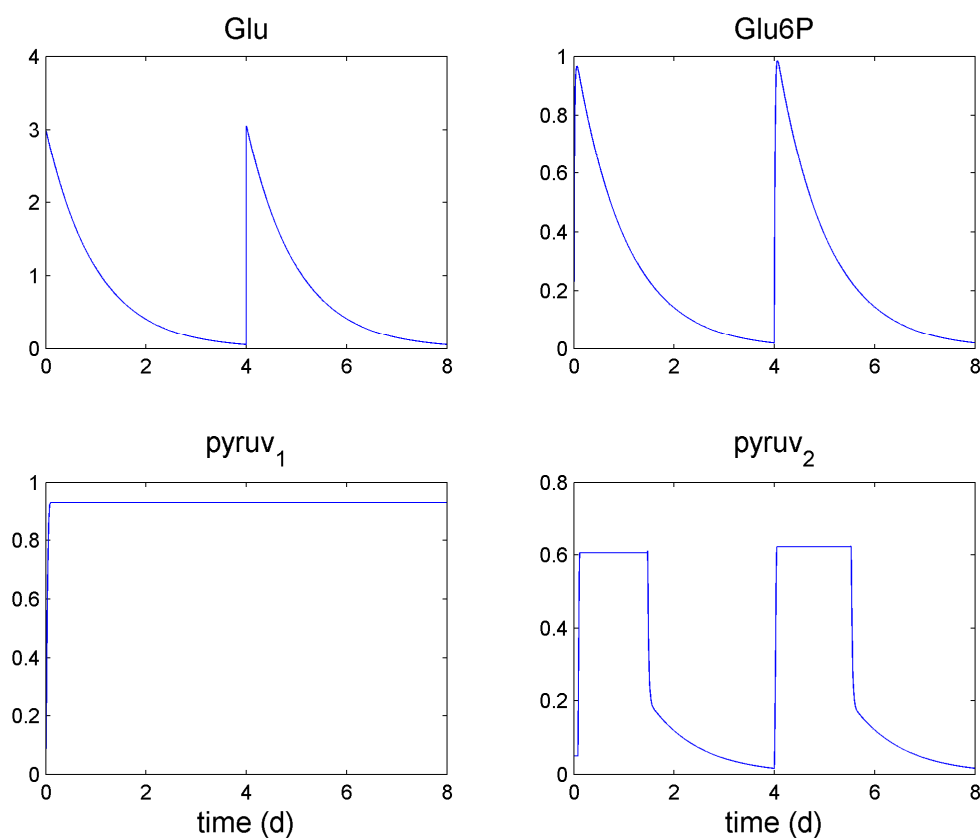

Figure 16.Concentration profiles (ng ml^{-1}cell^{-1}) for [*glu*], [*glu6P*] and [*pyruvate*]=[*pyruv₁*]+[*pyruv₂*].

The simulated results point out that the initial glucose and amino acids doses disappear practically after 4 days, whereas the production of glycogen (carbohydrate biosynthesis), palmitate (fatty acids biosynthesis) and albumin (protein biosynthesis) stops before the second day and starts again only after the introduction of fresh medium in the system. In addition there are several time scales involved as it can be seen in the NADH and ATP dynamics.
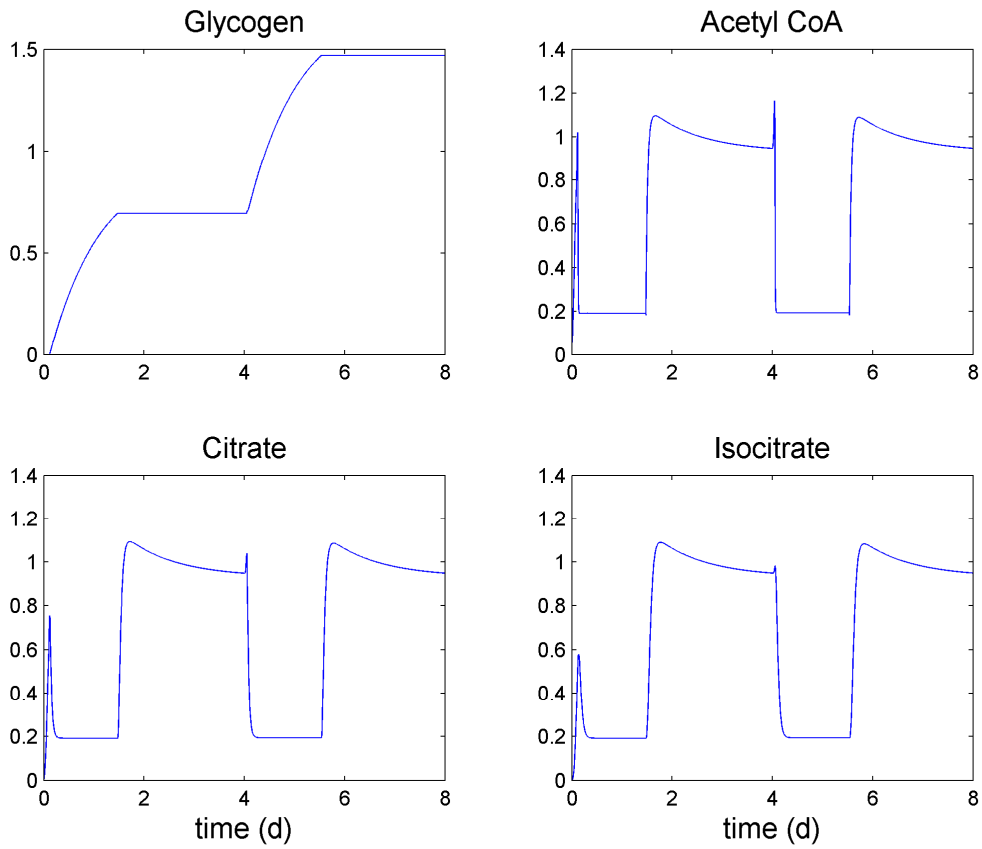
Figure 17. Concentration profiles (ng ml^{-1}cell^{-1}) for $n_{residual}$ (Glycogen), [*AcetylCoA*], [*citrate*] and [*isocitrate*].
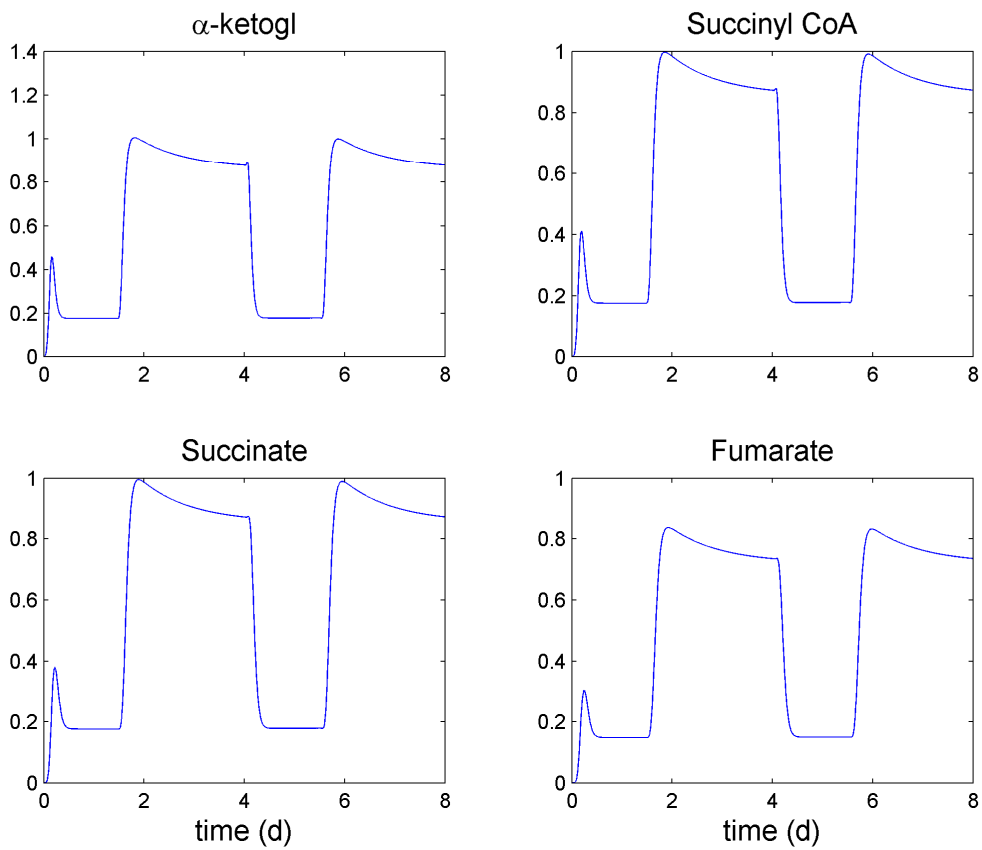
Figure 18. Concentration profiles (ng ml^{-1}cell^{-1}) for [*αKetogl*], [*succinylCoA*], [*succinate*] and [*fumarate*].

Figure 19. Concentration profiles (ng ml^{-1}cell^{-1}) for [*Lmalate*], [*ossalacetate*], [*NADH*] and [*ATP*].

Figure 20. Concentration profiles (ng ml^{-1}cell^{-1}) for [*a.a.*], [*urea*], [*ACoAeq*] and [*palmitate*].

24

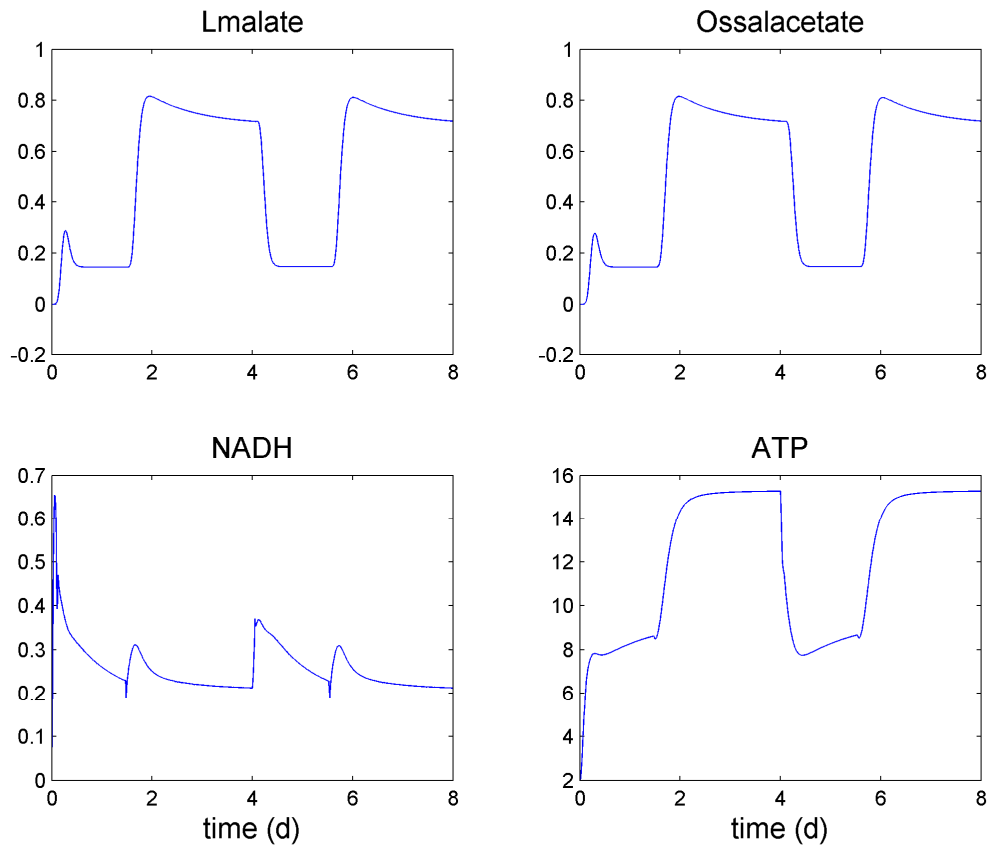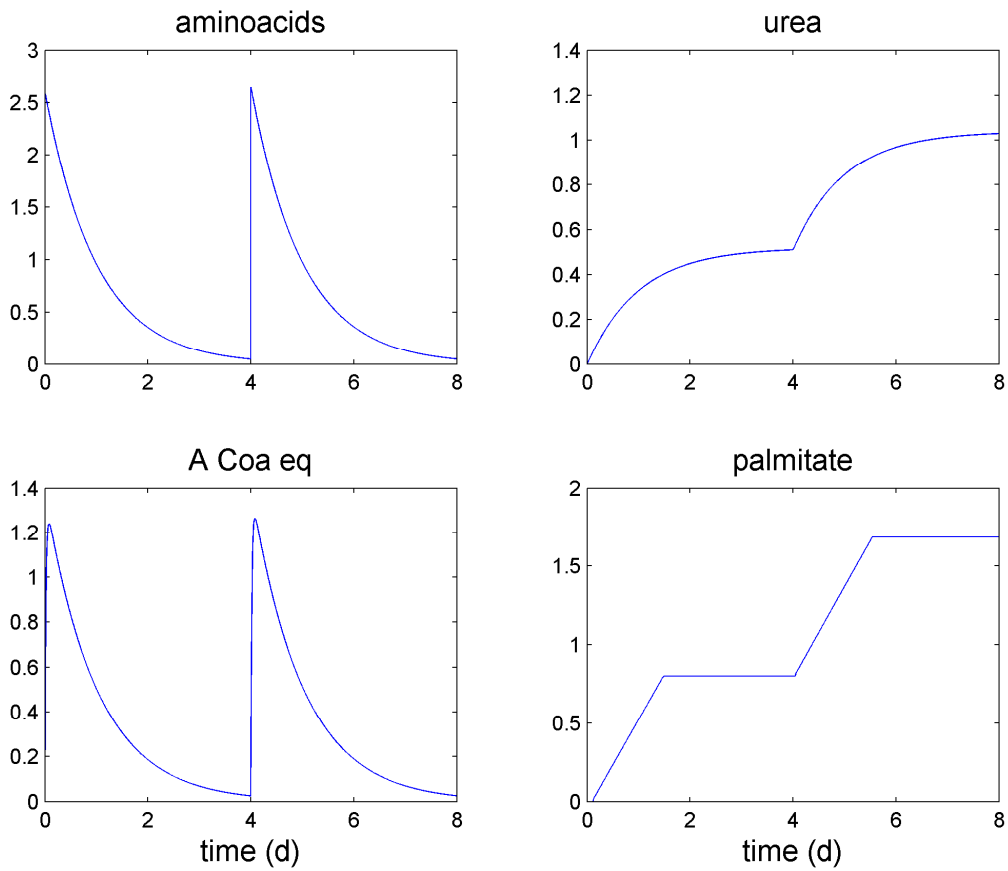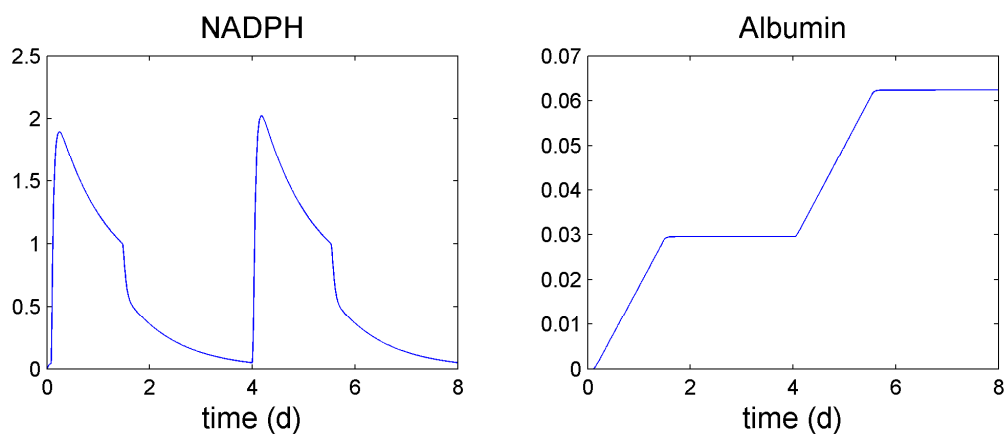Figure 21. Concentration profiles (ng ml^{-1}cell^{-1}) for [*NADPH*] and [*albumin*].

Probably the model could be improved by defining continuous sigmoid function for the start/stop of several metabolic processes, instead of an on/off switch as it is now implemented. This effect can be observed in the flat curves (derivatives set to zero) when some conditions related to [*ATP*] and [*NADH*] are met. However, it would be first necessary to obtain experimental data to assess the validity of the parameters used as well as to develop further the model as a function of measured metabolites concentrations.

4.3. Simulation of toxic effects in the metabolism

According to Niklas et al. (2009) experimental data, one of the most affected metabolic pathways, in absolute value, by the addition of a toxicant is the conversion of pyruvate to acetyl CoA and the subsequent Krebs cycle. To simulate this effect we have run several simulations decreasing the rate constant β_5 which control the pyruvate conversion, see Eq. (21). There is a linear decrease in the production of palmitate, whereas a similar behaviour is observed in the albumin, i.e. protein biosynthesis, production until $\beta_5/250$ at which albumin production stops because the concentration of palmitate, i.e., fatty acids, drops the threshold value ω^0, Eq. (37).

Other changes observed by Niklas et al. (2009) which involve the fluxes of aminoacids into the cell or the DNA and RNA biosynthesis can not be simulated since the metabolic pathways have not been included into the model due to the absence of experimental data to develop the kinetic expressions.

25

5. PRELIMINARY CONCLUSIONS AND GENERAL RECOMMENDATIONS FOR FUTURE ACTIVITIES

Network analysis provides an alternative approach to link metabolic data, in the form of metabolic fluxes, with toxicity data resulting form dose-response experiments. However, this implies that the important fluxes (metabolic or not) have been identified and considered in the developed molecular model. In our case only the main metabolic fluxes have been considered, but no signaling and regulatory fluxes have been considered. Probably, this is the reason why for certain compounds – amidarone – there are a relationship between several network measurements, e.g. network efficiency, and the chemical dose, whereas for other compounds this relationship is not evident – diclofenac –. To analyze further this problem, it would be interesting to use chemicals that affect a well-know metabolic pathway and to check if in this case more clear correlations emerge.

The dynamic model results have shown that the metabolic status of a cell line in a cell-based assay could be important for the results obtained during a dose-response experiment. For example if a certain chemical affects the protein biosynthesis, but this process is slowed due to the consumption of amino acids in the medium, this may slow down toxic effects when compared to another experiment where amino acids are readily available. For this reason, it would be interesting to check if the dose-response curves are affected by the metabolism by changing the starting of time of the experiment, which normally is fixed in the experimental procedure. This exercise will provide a way to quantify the importance of metabolic activity on the toxic response and provide an explanation for one source of variability in cell-based toxicity assays.

These preliminary results suggest that a combined approach starting from metabolic fluxes obtained using experimental data from the metabolomics facility and then, as our experimental data and knowledge of metabolic pathways increases, solving the kinetic equations to develop a dynamical model of the behaviour of the cell lines, is the most promising avenue to develop to reach a quantitative description of toxicity at molecular level. Furthermore, not only metabolic pathways should be considered, but it is important to develop an integrated approach combining these networks with signalling and regulatory networks (Lee et al., 2008).

In addition to experimental data necessary to validate and extend the metabolic network pathways, it is necessary to link metabolic performance to cellular behaviour in terms of growth and mortality. This is necessary to link data at this molecular scale with typical experimental data obtained during cell-based assays experiments. Our research is continuing in this direction (Zaldívar et al., 2010).

REFERENCES

Albert, R. & Barabási A.L. 2002. Statistical mechanics of complex networks, *Rev. Mod. Phys*. **74**, 47-97.

Becker S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.O., and Herrgard, M.J. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols* **2**, 727-738.

De Maria, C., Grassini, D., Vozzi, F., Vinci, B., Landi, A., Ahluwalia, A. and Vozzi,G. 2008. HEMET: Mathematical model of biochemical pathways for simulation and prediction of HEptocyte MEtabolism. *Computer Methods and Programs in Biomedicine* **92**, 121-134.

Eckmann, J.P., Kamphorst, J.O. & Ruelle, D. 1987. Recurrence plots of dynamical systems. *Europhys. Lett.* **4**, 973-977.

EPA US. 2003. A framework for a computational toxicology research program. Washington D.C. EPA600/R-03/65.

Fell, D.A. & Wagner, A. 2000. The small world of metabolism, *Nature Biotechnology* **18**, 1121-1122.

Gleich, D. 2006. MatlabBGL," http://www.mathworks.com/matlabcentral/fileexchange/10922.

Gulmerà, R & Nunes Amaral, L.A. 2005. Functional cartography of complex metabolic networks. *Nature* **433**, 895-900.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L. 2000. The large-scale organization of metabolic networks. *Nature* **407**, 651-654.

Heffernan JM, Smith RJ, Wahl LM .2005. Perspectives on the Basic Reproductive Ratio. *Journal of the Royal Society Interface* **2**, 81–93.

Hirata, Y., Horai, S., Kzuyuki,A. 2008. Reproduction of distance matrices and original time series from recurrence plots and their applications. *European Physical Journal Special Topics* **164**, 13-22.

Holme, P., Kim B. J, Yoon, C. N., Han, S. K. 2002. Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109.

Kitano, H. 2002a. Systems Biology: A brief overview. *Science* **295**, 1662-1664.

Kitano, H. 2002b. Computational systems biology. *Nature* **420**, 206-210.

Kroes, R., Kleiner, J., Renwick, A. 2005. The threshold of toxicological concern concept in risk assessment. *Toxicological Sciences* **86**, 226-230.

Latora, V. & Marchiori, M. 2001. Efficient behavior of small-world networks. *Phy. Rev. Lett.* 87, 198701.

Lee, J.M., Gianchandani, E.P., Eddy, J.A. and Papin J.A. 2008. Dynamic analysis of integrated signalling, metabolic and regulatory networks. *PLoS Computational Biology* **5**, e1000086.

Lewis, N.E., Thiele, I, Jamshidi,N. Palsson, B.O. 2008. Metabolic Systems Biology: A Constraint-Based Approach, In Encyclopedia of Complexity and Systems Science, Meyers.

Ma, H. and Zeng, A.-P. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19, 270-277.

National Research Council .2007. Toxicity testing in the twenty-first century: A vision and strategy.

Newman, M.E.J. 2003. The structure and function of complex networks" *SIAM Review* **45**, 167-256.

Newman, M.E. 2008. The mathematics of networks. The New Palgrave Encyclopedia of Economics, 2nd edition, L.E. Blume and S.N. Durlauf (eds.), Palgrave Macmillan, Basingstake.

Niklas J., Noor F., Heinzle E. 2009. Effects of drugs in subtoxic concentrations on the metabolic fluxes in human hepatoma cell line HepG2. *Toxicology and Applied Pharmacology* **240**, 327-336.

Nooy W de, M. A., Batagelj V. 2005. Exploratory Social Network Analysis with Pajek. Cambridge University Press.

Patterson, A. D. and J. R. Idle. A metabolomic perspective of small molecule toxicity. In General and Applied Toxicology. Marrs, T. C., B. Ballantyne, and T. Syversen. Wiley. Chichester. 2009.

Podani, J., Oltvai, Z.N., Jeong, H., Tombor, B., Barabasi, A.L., Szathmary, E. 2001. Comparable system-level organization of Archaea and Eurokaryotes. Nature Genetics 29, 54-56.

Rabinowitz, J.R., Goldsmith, M.-R., Little, S.B. and Pasquinelli, M. A. 2008. Computational molecular modelling for evaluating the toxicity of environmental chemicals: Prioritizing bioassay requirements. *Environmental Health Perspectives* **116**, 573-577.

Rockafellar, T. 1970. Convex Analysis. Princenton University Press. Princenton.

Song, H.-S., Ramkrishna, D. 2009. When is the quasi-steady-state approximation admissible in metabolic modelling? When admissible, what models are desiderable? Ind. Eng. Chem. Res. 48, 7976-7985.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190-193.

Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J. 1998. Metabolic Engineering: Principles and Methodology. Academic Press, San Diego.

Strozzi F, Zaldívar JM., Poljansek K, Bono F, Gutierrez Tenreiro E. 2009. From Complex Networks to Time Series Analysis and Viceversa: Application to Metabolic Networks. EUR 23947 EN.

Strozzi F., Poljansek K, Bono F, Gutierrez Tenreiro E., Zaldívar JM., 2010.International Journal of Bifurcations and Chaos (submitted)

Wagner, A., & Fell, D. 2001. The small world inside large metabolic networks. Proc. Roy. Soc. London Ser. B 268, 1803-1810.

Williams R.J. 2003. FoodWeb 3D instructions manual. pp 4.

Xu, X., Zhang, J., Small, M. 2008. Superfamily phenomena and motifs of networks induced from time series. *Proc. Nat. Acad. Sci. USA* **105**, 19601-19605.

Young J.D., Henne K.L., Morgan, J.A., Konopka, A.E., Ramkrishna D. 2008. Integrating cybernetic modelling with pathway analysis provides a dynamic system-level description of metabolic control. *Biotechnol Bioeng* 100, 542-559.

Zaldívar J.M., Mennecozzi M, Marcelino Rodrigues R, Bouhifd M. 2010. A biology-based dynamic approach for the modelling of toxicity in cell-based assays. Part I: Fate modelling . EUR 24374 EN.

Abstract.

Chemical toxicity in most of the cases has a metabolic component and metabolic pathways are the most studied biochemical reactions. For these reasons we have started to analyze modelling approaches for studying toxicity at molecular level using the metabolism and focusing in HepG2 cell lines. Two approaches have been investigated. The first consists on the analysis of metabolic fluxes in HepG2 under different toxicants concentrations by complex network analysis tools. This is important when a correlation between dose-response curves and metabolomic data has to be established. The second approach is based on the development and application of a simplified dynamic model of metabolism. This approach requires the specification of the biochemical reactions as well as their kinetic parameters and, therefore, at the moment, the degree of detail in the description of HepG2 metabolism is reduced. The results suggest that a combined approach starting from metabolic fluxes and then solving the kinetic equations as our experimental data and knowledge of metabolic pathways increases is the most promising avenue to develop a quantitative approach of toxicity at molecular level. In addition, there is the necessity to link the molecular level approach with the cell growth and mortality with an intermediate level that translates metabolism changes into measurable parameters in cell-based assays.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-24553-EN-C

EUROPEAN COMMISSION

■ Publications Office