

Systrip: a visual environment for the investigation of time-series data in the context of metabolic networks

Jonathan Dubois, Ludovic Cottret, Amine Ghozlane, David Auber, Frédéric Bringaud, Patricia Thebault, Fabien Jourdan, Romain Bourqui

▶ **To cite this version:**

Jonathan Dubois, Ludovic Cottret, Amine Ghozlane, David Auber, Frédéric Bringaud, et al.. Systrip: a visual environment for the investigation of time-series data in the context of metabolic networks. 16th International Conference on Information Visualization (IV'12), Jul 2012, Montpellier, France. pp.4771a204. hal-00742176

HAL Id: hal-00742176
https://hal.science/hal-00742176

Submitted on 16 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Systrip: a visual environment for the investigation of time-series data in the context of metabolic networks

J. Dubois[1,2], L. Cottret[3], A. Ghozlane[1], D. Auber[1,2], F. Bringaud[4]
P. Thébault[1,5], F. Jourdan[3] and R. Bourqui[1,2]
[1] Université de Bordeaux CNRS UMR5800
[2] INRIA Bordeaux Sud-Ouest,
[3] INRA UMR1089 Xénobiotiques Toulouse,
[4] Université Bordeaux Segalen CNRS UMR 5536,
[5] Université Bordeaux Segalen CBiB
{jonathan.dubois, amine.ghozlane, david.auber, patricia.thebault, romain.bourqui}@labri.fr,
{Ludovic.Cottret, Fabien.Jourdan}@toulouse.inra.fr, bringaud@rmsb.u-bordeaux2.fr

Abstract

Technological advances in biology lead to a profusion of quantitative data, raising analytical challenges. Visual analytics is particularly well suited to address these difficulties. It helps to interactively move through the different levels of analysis and to simultaneously investigate data with different point of views. It is especially the case when dealing with biological networks that can contain hundreds of elements. In these studies biologists generally follow the same analytic process which consists in first getting an overview of the data before focussing on a few relevant subnetworks. In this article we present, Systrip, a visual environment for the analysis of time-series data in the context of biological networks. In particular we focus on the study of metabolism. Systrip gathers bioinformatics and graph theoretical algorithms that can be assembled in different ways to help biologists in their visual mining process. This framework had been used to analyse various real biological data. In this article we describe how it helped in understanding drug effects on the metabolism of the parasite of the tsetse fly causing sleeping sickness.

1 Introduction

These last years, a technological gap has been filled in molecular biology with the development of high sequencing technologies. While it took years to sequence the genome of a single organism (e.g. fourteen for the human genome), it can now be done within a few days (and soon a few hours). It opens a new area in terms of biological investigation since the so called *omics* technologies can generate a large amount of high throughput data for the newly sequenced organisms. This richness and complexity is leading to the creation of many bioinformatics methods. These approaches have to be combined into efficient methodological pipelines in order to go from raw data to meaningful biological interpretations. Setting up one of these pipelines is a challenge as the output of each step can be critical for the next ones. Visualisation is therefore required to efficiently monitor the assembling of such complex analysis pipelines. In this article we will focus on the analysis of biological networks and in particular metabolic networks.

The metabolism is the set of biochemical reactions that occur in a living system. Each reaction transforms some *metabolites* (small molecules like *glucose*) called substrates into other metabolites called products. For a long time, the information about metabolism was partial and fragmented into well known metabolic functions (e.g. *glycolysis* or *Krebs cycle*) called *metabolic pathways* or *metabolic maps*. The metabolism of an organism is therefore generally modelled as a collection of disconnected graphs. However the data that are currently generated cover many of these pathways and thus requires the simultaneous analysis of all of them (generally more than a hundred). The best way to perform such a global study is to integrate all the metabolic pathways of a given organism into a single network called *metabolic network*. According to Gehlenborg *et al.* [21], the representation of the complete metabolic network and the data associated to the reactions, metabolites or enzymes is one the greatest challenge today in bioinformatics. It would in particular help in understanding biological shifts induced on an organism by environmental stresses.

In this article we present Systrip, a visual environment for the analysis of time-series data in the context of bi-

ological networks. While Systrip provides a large range of visual data mapping methods to handle metabolic networks and related high throughput data, it completes them with a large range of computational methods. The originality of Systrip is to combine of these classical bioinformatics methods (e.g. scope detection [23], chokepoint detection [41]) with well-known theoretical -still unexploited in biology- graph ones (e.g strahler [1], strength [3]). Systrip also supports exploration awareness as it enables the biologists to manually create views on the data at each important filtration step of their analysis process.

The remainder of this paper is structured as follows. Section 2 reviews related work on the analysis of time-series data in the context of metabolic networks. In section 3, we present how Systrip supports the biologists' visual analytics process. Section 4 refers to the detailled description of Systrip software and its main features. We then present in section 5 case studies on *Trypanosoma brucei* metabolism. Finally, we draw a conclusion and give directions for future work in section 6.

2 Related work

Metabolic pathways and networks: Before the outburst of high throughput data, metabolic investigations were performed at the metabolic pathway scale. For instance, many study focus on central carbon metabolism in order to decipher the different kind of energy intakes. This part of the metabolism only gathers few pathways resulting in graphs of relatively small size (dozens of nodes and edges). Visualisation efforts were thus put on the drawing of these subnetworks [7, 11, 40, 20, 17, 27, 29]. These approaches usually combine a cycle detection [7, 40, 20] step and a layered [11, 20] or a force-directed [7, 40] drawing algorithm. Recently, Meyer *et al.* [29] proposed a new visualization method called linearized metabolic pathway views that represents pathways as segments corresponding to sequences of elements computed during a linearization process.

Even if these methods can successfully represent several metabolic pathways (up to five according to [7]), they cannot be used to draw the entire metabolic network. To represent the entire network, tools such as *SBML Viewer* [1] or *Cytoscape* [35] use classical drawing algorithms (e.g. force-directed ones). However, these layouts do not respect the biological drawing conventions (see [30]) and do not therefore fulfill biologists' and bioinformaticians' expectations [34]. Very few tools try to respect these drawing conventions [25, 32, 8, 33, 28]. While in [25, 32, 33] nodes shared by several pathways are duplicated and each pathway is laid out independently, Bourqui *et al.* [8] introduced a method avoiding node duplication in order to preserve a realistic network topology. Recently, Lambert *et al.* [28] proposed a trade-off between the approach of Bourqui *et al.* [8] and the approach of Rohrschneider *et al.* [33] that enables the user to define whether or not node duplication should be performed.

The choice of the drawing algorithm has to be adapted to the purpose of the visualization. To allow the user to explore different options, Systrip provides many classical layout algorithms (e.g. layered, force-directed, circular) but also metabolism oriented drawing algorithms like the one proposed in [8]. In addition to the node-link diagram visualization of metabolic networks, Systrip also integrates parallel coordinates, spreadsheet, scatterplot and histogram views (see section 4.2).

Contextual visualization: Global biological studies are generally designed to investigate the adaptation of the organism metabolism to an environmental stress (e.g. drug treatment). Samples are collected all along the adaptation process. The output data will be, for each biological entity (e.g. metabolite), a vector of numerical values corresponding to measures made at different time points. These kinds of data are called time series data. However, most general biological network visualization software only allows to superimpose biological measurements on nodes for one time point.

Among them, *Cytoscape* [35], *Pathway Tools* [32], *BiologicalNetworks* [4] and *Cerebral* [6] support the visualization of a single time point from a time-series experiment. This is achieved by coloring nodes according to experimental data value for a particular measurement. In a more dynamic way, *Pathway Tools* [32] and *Cerebral* [6] allow to visualize the whole time-series by respectively using smooth animation between time points and using linked views, called *small multiples*, for each time point. Other tools, such as *VANTED* [26] and *VisANT* [24], support the visualization of the whole experimental data visualization by embedding an expression profile plot, a bar chart or line chart inside the nodes. In GENeVis [9], the authors combined gene expression levels and their statistical significances into a single glyph.

In Systrip we decided to integrate visualization of metabolite measurement data (called metabolomics data) as both size- or color-coding for a particular time point and heatmap or curvemap [29] for whole time-series. This allows to take advantage from both techniques: high (resp. low) values are emphasized by large (resp. small) node sizes while heatmap (or curvemap) shows evolution of these values inside each node.

[1] SBMLViewer, http://sbw.kgi.edu/layout/

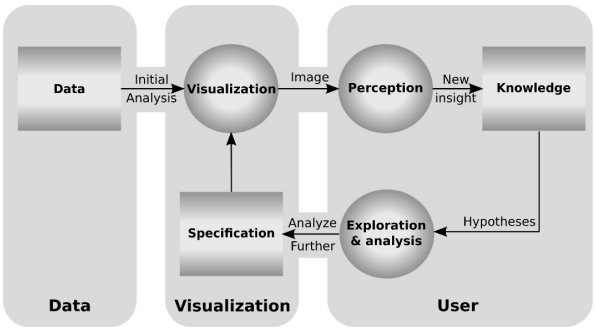


Figure 1: The "sense making loop" adapted from [38].

3 Mining biological data

Information Visualization exploits human visual capabilities -according to Ware [39], 40% of our cortical activities are dedicated to processing visual signals- to support visual exploration and analysis. It has now been established as a fruitful strategy to tackle the problem posed by the abundance of information [36]. It is particularly suitable in biology to visually identify original patterns in the data like subnetworks where molecule concentrations are affected when the system is perturbed. The Visual Analytics initiative promotes the use of Information Visualization to support analytical reasoning through a sense-making loop (see figure 1). That sense-making loop allows, through the visualization, the human perception and knowledge and the exploration to define new specifications (or requirements) in order to incrementally build but also to validate hypotheses, and therefore leads to information discovery.

The visual analytic process designed in Systrip follows that sense-making loop as it combines -semi-automated data processing together with human analytical and perceptual capabilities. Primary aim of Systrip is to support visual analysis of time-series data in the context of the complete metabolic network of an organism. To achieve that goal, Systrip provides 5 classes of functionalities (that we call *bricks*): input, drawing, graph measure, filtration and views (see figure 2). These elementary bricks can also be classified into network related (e.g. betweenness centrality [10] or excentricity [12]), bioinformatics related (e.g. scope selection [23] or chokepoint selection [41]) and multi-dimensional data related (e.g. parallel coordinates or scatterplot). By combining several simple but also complex bricks in a meaningful pipeline, the user can define his own specifications and therefore build an image that answers the original biological question.

Most biologists' and bioinformaticians' analysis process follows the so-called *Schneiderman's Visual Information-Seeking Mantra*: "Overview first, zoom and filter, then details-on-demand". Once the metabolic network and related series of values have been loaded, the main challenge is to find the chemical and biological links between them. However, the number of monitored nodes is often much smaller than the total number of nodes. In that case, the number of interesting biological scenarios connecting them quickly becomes huge. To overcome that issue, biologists and bioinformaticians use to filter out irrelevant nodes and edges. The aim is to iteratively reduce the size of the network to reach an interpretable subnetwork (for example, see figure 3). At each important filtration step, the user can create a subnetwork of interest that will be considered as a breakpoint in his exploration. That feature is important since the analysis process does not necessarily follow a linear flow and the user may have to restart his exploration from a previous step of the analysis. Moreover it may help the user to remind the main steps of an analysis that was started during a previous session. When designing Systrip, we defined an intuitive data and view management. Unlike tools like Cytoscape, Systrip disconnects the original data from the focussed subnetworks and related views. As the subnetworks are fully disconnected from the original network, the user can easily but also safely modify them (e.g. adding/removing elements, computing graph measures or applying metabolic-dedicated algorithms) without altering the original data. That feature is particularly suited to the top-down analysis process used in biology and bioinformatics. It also enables to perform simultaneously several analytics pipelines on the same data by applying different combinations of elementary bricks.

4 Software System

4.1 System Architecture

Systrip is developed in C++ and is based on Tulip [2]. Tulip is an information visualization framework dedicated to the analysis and visualization of relational data. It aims at providing to the developer a complete library supporting the design of interactive information visualization applications. Tulip is made of three independent sub-libraries Tulip, Tulip-OGL and Tulip-Qt that are respectively in charge of graph management, OpenGL rendering and user interaction. The current framework also enables the development of plug-ins to enrich the set of available algorithms, visual encodings, interaction tools and also domain-specic visualizations. The plug-ins management system can connect to a distant server to check for newer versions of installed plug-ins or to download new ones. Following Munzner's nested model [31], Tulip also supports the implementation of user interface overlays, called *perspectives*.

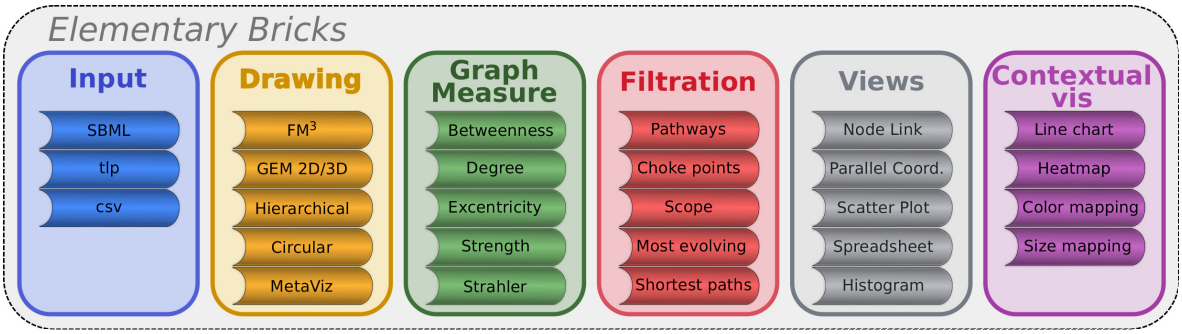


Figure 2: Systrip provides an environment gathering input facilities, graph drawing, graph measure and filtration algorithms, and supports several visualization techniques. These elementary (and computational) bricks can also be classified as network, bioinformatics and multi-dimensional data related.

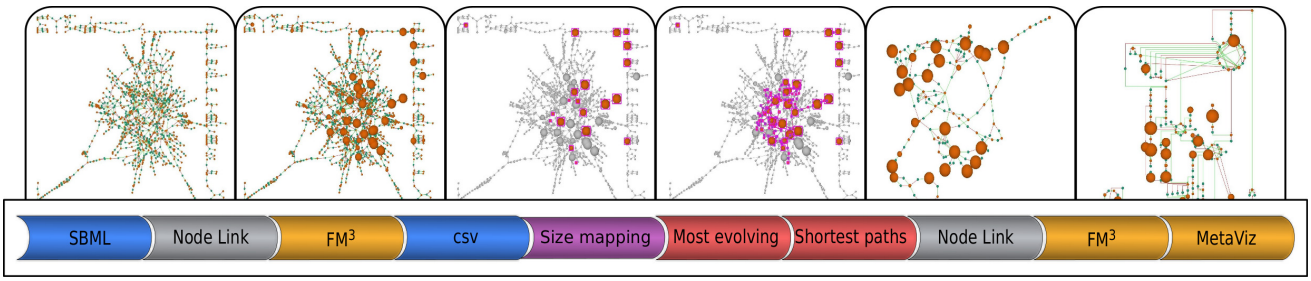


Figure 3: Example of analysis pipeline. A metabolic network is loaded into a node-link diagram view and drawn with FM^3 algorithm [22]. During the next step, metabolomic data is loaded and a size mapping is computed. To focus his analysis, most relevant element are selected and the graph induced by the shortest paths between these elements is extracted. The user then builds a node-link view on that network and tries both FM^3 and MetaViz layout algorithms.

Such an overlay integrates dedicated features in order to fulfil end-user's expectations. Since Systrip is a perspective of the Tulip framework, it takes full advantage. Relying on a graph structure makes Systrip flexible as it allows to support both relational and multi-dimensional data. It also eases the future integration of other biological networks such as gene regulatory or protein-protein interaction networks.

The *Model-View-Controller* architecture of Tulip allows Systrip to support multiple and synchronized views. Any interaction on a view (e.g. selection of an element) implies the automatic update of all views displaying this data. Finally, in addition to the algorithms and views developed specifically for Systrip (see section 4.2), its users can access to all Tulip plug-ins (included in Tulip released or via Tulip plug-ins server).

4.2 Main Features

In this section, we present Systrip's user interface, its main widgets and views to support both metabolic networks and time-series data visualization and exploration. Figure 4 shows a screenshot of Systrip. In this figure, one can see on the left side a panel containing (top-left corner) the elements of the metabolic network (and its decomposition into metabolic pathways) and (bottom-left corner) all subnetworks created by the user (also containing the related list of elements and pathways). The middle part of the tool contains the different views created by the user on the metabolic network, the time-series data but also database queries results. Since all the data are related to elements of the metabolic network, these views (except the information views) can be synchronized. We support a linking and brushing technique that allows to highlight selected elements in all views.

Visualizing metabolic network: In figure 5, one can see two node-link diagram views on the *Trypanosoma brucei* metabolic network. In this representation, a node is either a metabolite or a reaction, and, a metabolite and a reaction are linked if and only if the metabolite is consumed or produced by the reaction. Some of these reactions are reversible (they can occur in both ways), we use double arrow to render this information. As a reaction may consume

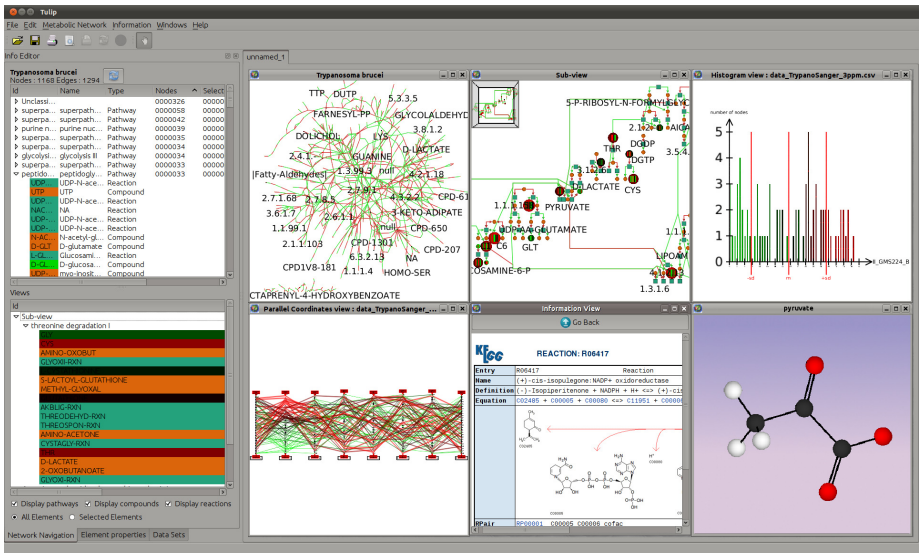


Figure 4: Screenshot of Systrip software. (Left) A panel containing (top-left corner) the elements of the metabolic network (and its decomposition into metabolic pathways) and (bottom-left corner) all subnetworks created by the user (also containing a list of elements and pathways); (Right) The different representations of the entire metabolic network, of the subnetworks, the time-series data and also database queries results.

and/or produce several metabolites, we also use different colors to encode incoming (green edges) and outgoing (red edges) sides of the reaction. Systrip also takes advantage of Tulip plug-ins system and offers many graph related algorithms. Among them, Systrip supports various drawing algorithms (hierarchical [2], 2D/3D force-directed [19, 22] and MetaViz [8] layout algorithms), bioinformatics algorithms (scope selection [23], chokepoints selection [41]) but also graph theoretical measures (betweenness centrality [10], eccentricity [12], strahler [1], strength [3]).

Visualizing time-series data: Systrip supports different visualization methods to mine time-series -or any stamped- data (figure 6 illustrates these visualization methods). When focussing on individual time points, the user may use the spreadsheet to visualize the values associated to the monitored elements (e.g. concentration or expression levels). Histograms can also be used to emphasize, for each time point, the distribution of the their values (double clicking allows to focus on a single histogram). If the user is interested in comparing values over several time points, Sytrip offers a scatterplot matrix view (again, double clicking allows to focus on a single scatterplot) or parallel coordinates one. While scatterplot supports pairwise comparison and therefore evaluation of correlations between two dimensions (or time points), parallel coordinates allows to compare simultaneously elements over an arbitrary number of dimensions. The former representation can also help

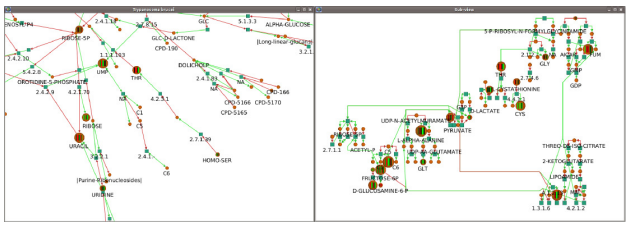


Figure 5: Node-link diagram representation of the metabolism. (Left) Zoomed view on the whole metabolic network of *Trypanosoma brucei* drawn with FM^3 [22] ; (Right) View on a subnetwork of interest drawn using the metabolic network dedicated algorithm [8]. Size mapping and heatmap glyph are used to emphasize time-series data.

the user to identify clusters of elements having similar behaviour during the entire experiment.

Contextual visualization: The user can also visualize time-series data in the context of the metabolic network. In that case, one can again distinguish two tasks in the analysis (i) the comparison of values of the elements at a single time point (ii) their evolutions over the entire experiment. To solve the first task, Systrip supports two visual encodings that emphasize the values of the element at a given time point. Both node colors and node sizes can be used to encode the node values for that particular time point. In figure 5, a size mapping has been performed to highlight ele-

ments with high values, while in figure 8, node colors were also computed according to the time-point 3. Concerning the second task, first solution supported by Systrip is to visualize the whole time-series of each element by rendering an heatmap or curvemap [29] glyph inside it. Another option is to navigate through the time points and to follow the evolution of the node sizes and/or colors. Systrip can "play the movie" of the time-series by automatically animating transition between time points. To preserve the user mental map, a smooth animation is performed during each transition between a time point to the next one by interpolating the node sizes and/or colors.



Figure 6: Expression data visualizations supported by Systrip software (from top-left corner to bottom-right): histograms, scatterplot matrix, parallel coordinates and spreadsheet views.

Visualizing database queries: The elements of the metabolic network are related to a large range of databases. It is important for a visual framework in biology to give access to this shared knowledge. In Systrip, the user can query both KEGG [27] and PublicHouse [2] databases. These databases contain detailed information about metabolites and reactions such as their name, molecular weight, chemical formula, or 2D molecular geometry. To visualize database results and also to allow navigation through the databases, Systrip offers an information view embedding a web browser. For instance, figure 4 shows the result of a KEGG query on *NADP+ oxidoreductase*. Another feature is the 3D molecular geometry visualization obtained from the 2D geomtry and the OpenBabel library [3] (figure 4 shows the 3D geometry of *pyruvate*).

[2] PublicHouse, http://biowarehouse.ai.sri.com/PublicHouseOverview.html
[3] OpenBabel, http://openbabel.sourceforge.net/

4.3 Simple vs Advanced user interface

Systrip integrates many features that can be complex to correctly set for a non-expert in graph visualization and analysis. For instance, Tulip's plug-ins installed on the computer are all available in Systrip, and, parameters related to drawing issues can be difficult to interpret. According to the biologist feedbacks, we decided to set a simple user interface as default, and, on demand, to display advanced options and algorithms.

The simple interface contains the most useful algorithms from a biological point of view, i.e. mostly algorithms dedicated to biological networks. To make the interface even simpler, the parameters are automatically set to default values when applying an algorithm (these values were either set according to their biological meaning or according to empirical tests). The advanced user interface is designed for expert users and offers more functionalities than the simple one. First of all, an "Algorithms" menu, containing Tulip plug-ins installed on the computer, is added in the menu bar. These plug-ins covers a large variety of algorithms classified into "Size", "Layout", "Color" or graph "Measure". Using that advanced interface, the user can set manually the needed parameters when applying an algorithm. The left panel (see figure 4) is also enriched: a new tab to manage the views and interaction tools settings is inserted and a configuration widget is added in the "Data Sets" tab (to enable the user to modify animation parameters). Both simple and advanced user interfaces allow to edit preferences. These preferences include rendering parameters (e.g. default node and edge colors, sizes, shapes, etc), information displayed in the left panel of the tool (e.g. properties displayed when an element is selected) and database settings.

5 Case studies

In this section, we propose to demonstrate some measurable benefits when using Systrip to monitor the metabolism of *Trypanosoma brucei*, which is the causative agent of human African trypanosomiasis [5] (sleeping sickness). These trypanosomes possess a complex life cycle that includes two replicating stages: the bloodstream form in the blood of mammals, and the insect (procyclic) form in the midgut of the vector. Like all parasites with an obligate host change, the parasitic protozoan *T. brucei* has to adapt its metabolism to widely varying conditions during its life cycle. By exploiting experimental data and using Systryp, we propose to further investigate three analysis of interest: (i) the metabolic impact of different carbon sources (proline and glucose), (ii) the metabolic effects of a drug (DFMO) and (iii) the distribution of metabolic fluxes in

the different branches of the glucose metabolic network. These different analysis were carried out by exploiting different datasets (metabolite concentrations) coming from several metabolomic experiments that were designed to capture the metabolic concentration profile in a given environmental context and were not focused on specific biological functions. We therefore investigated these datasets by exploiting the analysis framework provided by Systrip, which enable us to visualize the experimental data through their global context given by the metabolic network. In the following user cases, we propose several analysis, that exploits three different metabolomic datasets where each sample were obtained with some variable conditions. For the two first analysis, the metabolic network was built and processed by respectively using the MetExplore Trypanocyc database and platform [14, 15]; and for the third one, the model was based on the topology of the metabolic network given in [13]. The three metabolic models were then exported in the widely used Systems Biology Markup Language (SBML [18]) that can be imported into Systrip.

5.1 Glucose proline dataset

The first analysis explored the global topological impacts on metabolic networks in trypanosomes associated with responses to environmental change. We focussed on the procyclic forms of the parasite with metabolomic data obtained with different carbon sources. For each monitored metabolites, a ratio is provided between its measured concentration in the proline and glucose sample experiments (close to one means no change, close to zero means an over production in the glucose experiment while higher values corresponds to an over production in the proline experiment). We mapped the differential data onto the metabolic network elements with the objective to identify a potential partition of the graph that may be related to the shift that occurred between the two environmental conditions. To do so, we used a combined visualization of the distribution of the differential concentrations and the global metabolic netowrk (see figure 7).

Thanks to this analysis, we were able to propose two subnetworks related to each extremities of the histogram, in other words related to the two differential conditions given by the two carbon sources. Selecting the lower values of the histogram allows the identification in the network of the metabolites which are related to the glucose experiment (see figure 8). In order to connect these nodes and therefore to extract a relevant subnetwork, we then computed all the biological shortest paths between them by taking into account the direction and reversibility of reactions. The same operation was repeated for the proline experiment with the highest values of the histogram. At

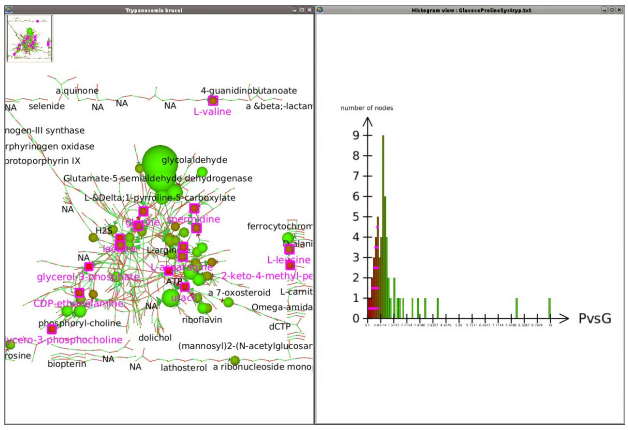


Figure 7: Linked view of network and data. The selection obtained on the network (Left) was obtained by the selection on the histogram of data (Right).

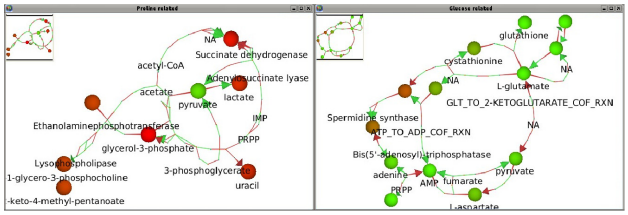


Figure 8: The two subnetworks extracted using the histogram. (Left) The glucose related subnetwork and (Right) the proline related subnetwork.

the end, the two subnetworks corresponding to the two carbone sources were analysed thanks to Systryp and we visually identified the pathways that are mainly involved in one or the other conditions. We confirmed that the compounds and reactions of the glycolysis pathway are mainly involved in the subnetwork related to the glucose condition. On the other hand, the proline degradation pathway is only present in the proline related subnetwork. At last, our analysis allowed to identify the backbone of the metabolic network that corresponds to the unchanged metabolites and may not be involved in the glucose/proline catabolism.

5.2 DFMO dataset

Some metabolites in the network may not have been monitored during the experiment. So a first way to narrow the search space consisted in computing, based on the detected metabolites, all the metabolites that can be synthesized. This definition corresponds to a bioinformatics method called the scope [23]. It is possible to compute this scope with Systrip, we then obtained a meaningful metabolic subnetwork.Even with the selection facility, the

view still contained too much nodes to be navigated. To deal with this problem, the aim was then to focus on the backbone of the network. In biology the notion of flux is central when studying the metabolism. The Strahler metric on graphs is thus particularly adapted to compute the potential flux through the reactions. By coloring the edges according to these values the user can already detect a backbone that can be extracted by applying a threshold on the edges. Since this subnetwork contained only tenth of elements it was, as a second step, possible to apply a graph drawing algorithm dedicated to metabolic drawings [8].

The applied filtrations guaranty that we focussed on a relevant subnetwork in terms of biological and topological meanings. In fact, this pathway contains the enzyme targeted by the drug. The question was then to understand the link between the inhibition of this enzyme and the metabolomics data. In that case, the animation provided by Systrip is particularly adapted to go through the different time points as it allowed to understand how the pathway is modified. The conclusion raised by the observation of the animation were consistent with the hypothesis of the biologists on the drug effects.

5.3 Glucose metabolomic network dataset

The third analysis was based on the procyclic form of *T. brucei*. That form converts glucose, the major carbon source used in rich-medium, into the excreted end products acetate and succinate. This parasite has the particularity to perform the first six glycolytic steps in the glycosomes, which are peroxisome-like organelles [13]. This unique metabolic strategy implies a tightly controlled use of metabolic cofactors (ATP and NAD^+) inside the organelle. Interestingly, the proportion of acetate and succinate excreted from glucose metabolism varies depending on the geographical context of experimental analyses, with acetate production ranging between 28% and 55% of the major excreted end products from glucose [37, 16]. This is suggestive of significant flexibility in the flux distribution among the various branches. While the metabolic network leading to the production of acetate and succinate in the procyclic trypanosomes is well known, the flux distributions of metabolites in the different branches of the network were not investigated so far. To address this question and analyze the extent of the metabolic flexibility of glucose metabolism in this parasite, we used Metaboflux [4]. This software, by running simulations based on a multiple objective optimization framework, gave as results the flux distributions (percentage of flux throw the enzymes) and metabolite ratios when constraining the system by the different constraints known from experimental data or assumed from the literature on *T. brucei* glycolytic metabolism. Proportion of end products and fluxes were predicted for fixed proportion of acetate excretion ranging from 1 to 99%, with a 5% increment.

We used the csv import of Systrip to visualize the flux and metabolite distribution in the metabolic network context. The aim is to identify the underlying process of this flexibility. We have computed the graph states for each constraint proportion on acetate. By coloring and adjusting the size of the edges according to flux and metabolite proportions, the animation revealed the central role of the malic enzymes in flexibility process. The animation clearly demonstrated the direct correlation between the rate of acetate production and the metabolic flux through the malic enzyme. The study presented here, is consistent with further analysis reported on malic enzymes with Metaboflux. This conclusion was not raised from our initial visual analysis of the Metaboflux data set, which only displayed a limited number of parameters. A subsequent deep analysis of the Metaboflux data set was required to observe the correlation described above.

6 Conclusion and future work

We presented Systrip, a visualization environment dedicated to metabolic networks. Selection and extraction of subnetworks based on the classification of the nodes in reference to metabolic pathways or to biological features such as biosynthetic capability have been facilitated. One of the main contributions of Systrip for the metabolic network visualization is the powerful interface to support contextual visualization of time-series data. The entire time series can be visualized as heatmap and curmap glyphs while single time points can be mapped to node size and/or color. Additional ways to explore time-series data are proposed: spreadsheets, histograms, scatterplots and parallel coordinates. Each way to explore data can be used at the same time and mapped onto the metabolic network and eventually the extracted subnetworks. Selections on any of the representations are reflected in the other ones. Moreover, by combining biological and graph measures, the Systrip user can reduce the data to offer a visualize by restricting the scope of the study. Systrip is made of elementary bricks that can be combined into sophisticated pipelines to select subnetworks that follow topological or biological criteria. At last, links to external databases enable to enrich the analysis.

Regarding the first analysis that were carried out with Systrip and thanks to their relavance in biological analysis, an interesinting list of additional improvements started to come up for a further investigation of metabolism in a dynamic context for analysing the fluxes. It would be also of

[4] Metaboflux, http://www.cbib.u-bordeaux2.fr/metaboflux/

interest to offer the abality of loading metabolic networks directly from public databases. Moreover, an actual challenge in system biology is to integrate various sources of ∼omics data with the objective to improve the realistic significance of bioinformatics analyses. Regarding Systrip, the integration of new methods dedicated to the visualisation of specific and various omics data would be on purpose and may be carried by extending the actual well defined framework. A first stage would concern the dimensionality reduction and may lead to interesting works to define relevant analysis environments (e.g. principal component analysis or multidimensional scaling).

Acknowledgements

This work was partially done under the Systryp project, supported by the ANR (France) and BBSRC (UK); and under the EVIDEN project (ANR 2010 JCJC 0201 01), supported by the ANR (France).

References

[1] D. Auber. Using Strahler numbers for real time visual exploration of huge graphs. In *ICCVG*, pages 56–69, 2002.

[2] D. Auber. Tulip : A huge graph visualisation framework. In P. Mutzel and M. Jünger, editors, *Graph Drawing Softwares*, Mathematics and Visualization, pages 105–126. Springer-Verlag, 2003.

[3] D. Auber, Y. Chiricota, F. Jourdan, and G. Melançon. Multiscale Visualization of Small-World Networks. In *Proc. of IEEE Information Visualization Symposium*, pages 75–81, 2003.

[4] M. Baitaluk, M. Sedova, A. Ray, and A. Gupta. BiologicalNetworks: Visualization and analysis tool for systems biology. *Nucleic Acids Res.*, 34:W466–W471, 2006.

[5] M. P. Barrett, D. W. Boykin, R. Brun, and R. R. Tidwell. Human african trypanosomiasis: pharmacological re-engagement with a neglected disease. *Br J Pharmacol*, 152(8):1155–1171, Dec 2007.

[6] A. Barsky, T. Munzner, J. L. Gardy, and R. Kincaid. Cerebral: Visualizing multiple experimental conditions on a graph with biological context. *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2008)*, pages 1253–1260, 2008.

[7] M. Becker and I. Rojas. A Graph Layout Algorithm for Drawing Metabolic Pathways. *Bioinformatics*, 17:461–467, 2001.

[8] R. Bourqui, V. Lacroix, L. Cottret, D. Auber, P. Mary, M.-F. Sagot, and F. Jourdan. Metabolic network visualization eliminating node redundance and preserving metabolic pathways. *BMC Systems Biology*, 1(29), 2007.

[9] R. Bourqui and M. A. Westenberg. Visualizing temporal dynamics at the genomic and metabolic level. In *13th Int. Conf. Information Visualization 2009*, pages 317–322. 2009.

[10] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Methmatical Sociology*, 25:163–177, 2001.

[11] U. Brandes, T. Dwyer, and F. Schreiber. Visualizing Related Metabolic Pathways in Two and Half Dimensions. In *Proc. Graph Drawing 2003*, pages 11–122, 2003.

[12] U. Brandes and D. Wagner. Visone: Analysis and visualization of social networks. In *Graph Drawing Softwares*, Mathematics and Visualization, pages 321–340. 2003.

[13] F. Bringaud, C.E. Ebikeme, and M. Boshart. Acetate and succinate production in amoebae, helminths, diplomonads, trichomonads and trypanosomatids: common and diverse metabolic strategies used by parasitic lower eukaryotes. *Parasitology*, 137:1315–1331, 2010.

[14] B. Chukualim, N. Peters, C. Fowler, and M. Berriman. Trypanocyc - a metabolic pathway database for trypanosoma brucei. *BMC Bioinformatics*, 9(Suppl 10):P5, 2008.

[15] L. Cottret, D. Wildridge, F. Vinson, M. P. Barrett, H. Charles, M.-F. Sagot, and F. Jourdan. Metexplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res*, 38(Web Server issue):W132–W137, Jul 2010.

[16] V. Coustou, M. Biran, S. Besteiro, L. Riviere, T. Baltz, J.-M. Franconi, and F.Bringaud. Fumarate is an essential intermediary metabolite produced by the procyclic trypanosoma brucei. *J Biol Chem*, 281:26832–26846, 2006.

[17] Dogrusoz, Erson, Giral, Demir, Babur, Cetintas, and Colak. Patikaweb: a web interface for analyzing biological pathways through advanced querying and vi-

sualization. *Bioinformatics*, 22(3):374–375, November 2005.

[18] A Finney and M Hucka. Systems biology markup language: Level 2 and beyond. *Biochem Soc Trans*, pages 1472–3, 2003.

[19] A. Frick, A. Ludwig, and H. Mehldau. A Fast Adaptive Layout Algorithm for Undirected Graphs. In *Proc. Graph Drawing 1994*, pages 388–403, 1994.

[20] S. D. Gabouje and E. Zimányi. Generic visualization of biochemical networks: A new compound graph layout algorithm. In *Poster Proc. of WEA 05*, 2005.

[21] Nils Gehlenborg, Sen I O'Donoghue, Nitin S Baliga, Alexander Goesmann, Matthew A Hibbs, Hiroaki Kitano, Oliver Kohlbacher, Heiko Neuweger, Reinhard Schneider, Dan Tenenbaum, and Anne-Claude Gavin. Visualization of omics data for systems biology. *Nat Methods*, 7(3 Suppl):S56–S68, Mar 2010.

[22] S. Hachul and M. Jünger. Drawing large graphs with a potential-field-based multilevel algorithm. In *Proc. Graph Drawing 2004*, pages 285–295, 2004.

[23] T. Handorf, O. Ebenhöh, and R. Heinrich. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J Mol Evol*, 61(4):498–512, Oct 2005.

[24] Z. Hu, D. M. Ng, T. Yamada, C. Chen, S. Kawashima, J. Mellor, B. Linghu, M. Kanehisa, J. M. Stuart, and C. DeLisi. VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. *Nucleic Acids Res.*, 35:W625–W632, 2007.

[25] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research*, 33:D428–D432, 2005.

[26] B. H. Junker, C. Klukas, and F. Schreiber. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7:109, 2006.

[27] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research*, 38:D355–D360, 2010.

[28] A. Lambert, J. Dubois, and R. Bourqui. Pathway preserving representation of metabolic networks. *In Proc. of EuroVis 2011*, 2011.

[29] M. Meyer, B. Wong, M. Styczynski, T. Munzner, and H. Pfister. Pathline: A Tool for Comparative Functional Genomics. *Computer Graphics Forum*, 29, 2010.

[30] G. Michal. On representation of metabolic pathways. *BioSystems*, 47:1–7, 1998.

[31] T. Munzner. A Nested Process Model for Visualization Design and Validation. *IEEE TVCG*, 15:921–928, 2009.

[32] S. Pailey and P. Karp. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Research*, 34(13):3771–3778, 2006.

[33] M. Rohrschneider, C. Heine, A. Reichenbach, A. Kerren, and G. Scheuermann. A Novel Grid-Based Visualization Approach for Metabolic Networks with Advanced Focus&Context View. In *Proc. Graph Drawing 2009*, pages 268–279, 2010.

[34] P. Saraiya, C. North, and K. Duca. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Information Visualization*, 4:1–15, 2005.

[35] P. Shannon, A. Markiel, O. Ozierand, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13:2498–2504, 2003.

[36] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2006.

[37] S.W. van Weelden, B. Fast, A. Vogt, P. van Der Meer, J. Saas, J.J. van Hellemond, A.G. Tielens, and M. Boshart. Procyclic trypanosoma brucei do not use krebs cycle activity for energy generation. *J Biol Chem*, 278:12854–12863, 2003.

[38] J. J. van Wijk. The value of visualization. In *IEEE Visualization*, pages 79–86, 2005.

[39] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2000.

[40] K. Wegner and U. Kummer. A new dynamical layout algorithm for complex biochemical reaction networks. *BMC Bioinformatics*, 6(212), 2005.

[41] I. Yeh, T. Hanekamp, S. Tsoka, P. Karp, and R. Altman. Computational analysis of plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res*, 14(5):917–924, May 2004.