SimWiz3D - Visualising biochemical simulation results

Katja Wegner
EML Research, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany
katja.wegner@eml-r.villa-bosch.de

Abstract

Thanks to modern experimental technologies the knowledge about processes in living cells has been increasing rapidly in recent years. This is resulting in masses of data that need to be handled, analysed and interpreted. In addition, simulation and modelling which is more and more frequently used to support experimental investigations also leads to large amounts of numerical data, e.g. times series. Therefore, it becomes more and more important to develop visualisation tools which faciliate the analysis of the respective data. This paper introduces the novel tool SimWiz3D which visualises time series data in a concise way using a three-dimensional representation and provides the user with a lot of different ways to interact with and exploit the simulation data.

1 Introduction

During the last years experimental technologies, e.g. in genomics and proteomics have enabled researchers to explore the processes in living cells. These processes consist mostly of (bio)chemical reactions which produce, rebuild or break down molecules in the cell. The participants of these reactions are called reactants. The experiments in laboratories which are time consuming and expensive are more and more supported by modelling and simulating techniques. These techniques help experimentalists to complete their knowledge and explain experimental observations. In simulations reactions are modelled by mathematical equations, e.g. differential equations which describe concentration or particle changes of reactants over time.

Time series are typically represented in concentration-time diagrams (x-axis = time, y-axis = concentration/number of particles), but the diagrams are only suitable for a small number of reactants. On the one hand the more curves are present in the same diagram, the harder it is to follow each curve fluently, although different colours are used. On the other hand the representation is inappropriate when the range of concentration values is very different, e.g. curves with a smaller range could look like a flat line, although they exhibit oscillations.

Out of these reasons it is necessary to develop visualisation techniques to faciliate the handling and visual analysis of multiple time series. This will improve the understanding of the studied processes (e.g. reactions) and will hopefully help to develop new medicine and cure diseases in the future.

At the moment there only exists one approach of Dwyer et al. [2] for a more sophisticated visualisation of complete time series data in biochemistry. They visualise the reaction network as graph in the x,y-plane and uses the third dimension z to represent experimental results. It is possible to switch between a disc or a box (similar to a histogram) representation. For each time step they create a single disk or box and arrange them in a consecutive manner. Unfortunately, in this representation the continuity of time series is lost and it is hard to compare single time steps or complete time series. Arrows displaying the flow in the network for each step are also disturbing. Furthermore the user has to write a GML file (Graph Modelling Language [4]) including the network structure and the corresponding time series.

Out of these reasons and on the basis of our SimWiz project [7, 8] I developed a new visualisation tool: SimWiz3D. SimWiz is a Java package that visualises the network of processes (e.g. reaction network) as graph and the time series data is mapped onto the network structure by modifying the shape of nodes in the graph. It is a step by step animation in which the size of the nodes increases or decreases symbolising the concentration changes per time step. SimWiz3D on the other hand uses this information to represent the whole data in a single 3D picture and lets the user explore this representation in various ways as described in the following.

2 SimWiz3D

SimWiz3D is written in Java and combines the animation features of the SimWiz package with a new 3D view of the time series data and further data exploration techniques. The implementation follows the visualisation pipeline in figure 1. The next sections will describe the different views and features.

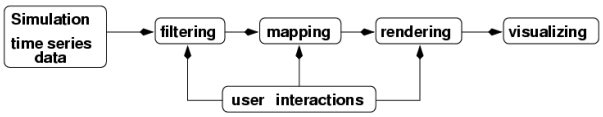

Figure 1: Visualisation pipeline.

2.1 Data formats

Like SimWiz it is able to read the network information from a SBML file (Systems Biology Markup Language [5]). SBML is a computer-readable file format to save biochemical reaction network information which describes all participating reactions and their reactants. This network structure can be visualised as a graph. In such a graph nodes represent reactants and edges reactions. The edge direction shows the direction of the according reaction. The graph layout of this network must be included in the SBML file and can be created with the $ReactionLayouter$ of the SimWiz package if necessary. There already exists a layout extension accepted by the SBML community by Gauges et al. [3] which will be part of SBML level 3 and enables SBML to save the coordinates and size of each reactant as well as the information about edges.

The numerical data representing time series should be saved in a simple text file (see figure 2). The first column represents the time and the other columns the reactants separated by tabulators or blanks. The first line consists of the reactant names or ids according to the SBML file. Each line in this table contains the concentration or particle numbers of each reactant for one time step. Many simulation programs are able to create such text files automatically.

```
time NADH O2 PER3+ PER2+ NAD.
0 0 0 843220 0 0
2 116025 83009 843220 0 0
4 231509 164609 843220 0 0
6 346657 244640 843220 0 0
```

Figure 2: Example of a text file with simulation results.

SimWiz3D reads that data formats and saves the network into a graph structure with nodes and edges and the concentrations in numerical arrays, one for each reactant. Thus, the tool needs a longer initial loading time but it allows a faster access for view updates because no further file accesses are needed.

2.2 3D view

The 3D view visualises the read layout information of the graph structure in a 2D plane (x,y) and uses the third axis (z) as time (z=0 first time step).

The time series are realized as tubes similar to e.g. the fund manager visualisation of Dwyer et. al. [1]. Each tube is composed of single conical frustums. The top radius of each conical frustum belongs to the time step n and the bottom radius to time step $n + 1$. The following conical frustum starts with time step $n+1$ and ends with $n+2$ and so on which creates a continuous appearance (see figure 3). Due to performance reasons time steps with the same concentration values are joined to one conical frustum.

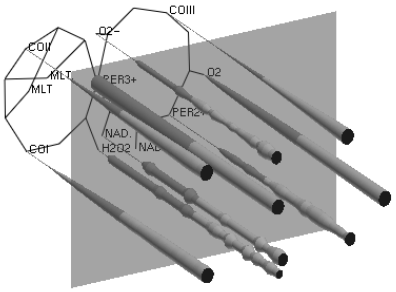

Figure 3: 3D view of biochemical simulation results and the cutting plane at a certain time step.

Similar to SimWiz and Dwyer et al. [2] the user can choose between two scaling types. The global scaling type takes the maximum concentration of all reactants as corresponding value for the maximum radius and the local scaling type the maximum concentration of each reactant. The maximum radius depends on the smallest distance between two nodes in the current network and corresponds to a fourth of this distance which prevents the view from overlapping of tubes. The minimum radius is by default 1 which corresponds to a thin line and the concentration value 0. Furthermore the user has the possibility to choose between two calculation methods. On the one hand the radius belongs to the square-root of the concentration which means it is proportional to the area of the cycle. On the other hand a logarithmic representation can be used which is helpful when very different value ranges exist. These options can be chosen in the button line on the top of the 3D view where also the projection can be switched between perspective and orthogonal (see figure 7). The zooming buttons together with rotation facilities allow the user to explore the 3D view in a comfortable way. After rotating the view when tubes overlap, the tool can expand the network to increase the distance of nodes, thereby reducing overlapping. By clicking two times on one tube this reactant and his reactions are highlighted in the 2D view and a small new window will show the following additional reactant information (see figure 4):

- name of this reactant

- selected time step by clicking
- statistical information (start value, minimum, maximum, mean value, variance)
- button to switch the current colour of this tube

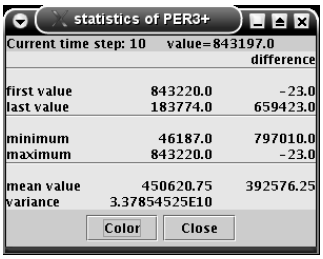

Figure 4: Additional statistical information of a reactant at a certain time step.

With the start and end time slider below the 3D view the user can reduce the number of shown time steps to focus on interesting time ranges or to reduce data which would also increase the update response time when long time series are used. The number of steps can also be reduced by setting a minimum concentration change value which filters steps out in which the concentration change is smaller than this value.

The last feature to mention similar to Dwyer et al. [2] a semi-transparent cutting plane is used to highlight a single time step. This time step and the corresponding concentration values are also shown in the 2D view which will be introduced in the next section.

The 3D visualisation is done by OpenGL which is a software interface to produce computer-generated images [6]. The time series data is mapped to the described conical frustums. These frustums are cached in display lists to reuse a precalculated version of them multiple times which increases the rendering time and the performance.

2.3 2D view

The 2D view is taken from SimWiz to obtain the topology of the network and to visualise the concentration of one single time step in the shapes of the nodes (see figure 5). These shapes are cycles whose radius is calculated accordingly to the above mentioned radius of conical frustums.

The user has different interaction possibilities to change both views simultaneously. First, by clicking on a cycle or by marking a group of cycles in this 2D view the 3D view focuses on these marked nodes and the other nodes can be removed or made translucent by choosing a transparency value between 0 and 100%. Single tubes of the 3D view can also be removed by clicking two times on the corresponding reactant in the 2D view. Second, with the time step option the user can change the shown time step by typing a new value into the text field or by moving the slider. This change evokes a new 2D view with the concentration values of the chosen time step and moves the semi-transparent plane to the according coordinates of this time step in the 3D view.

Another feature of this view offers an animation which changes the 2D view and moves the plane in the 3D view time step by time step automatically starting with the current time step. The speed is adjustable by the user. Furthermore when the user moves the mouse over an reactant in the 2D view the corresponding tube in the 3D view is highlighted.

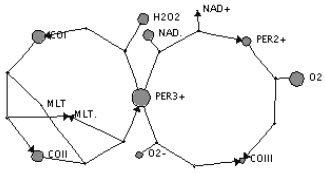

Figure 5: Example of a biochemical reaction network visualised in the 2D view. The concentration values at a certain time step are coded in the cycle size.

2.4 Settings

In the options menu the user has the possibility to make general settings to change colours and to influence the rendering process. First, since the tubes consist of single conical frustum, the user can change the minimum and maximum radius which is useful when only few time series are investigated. Second, the number of slices of conical frustum can be increased or decreased. This number influences the smooth appearance of a conical frustum which means the bigger the conical frustum is, the more slices are needed to smooth it out. But the user should have in mind that the more slices are used, the more time is consumed by the rendering process. Third, the colours of the tubes and the semi-transparent plane can be adapted. Finally, since the resolution of screens is limited, a maximum number of rendered time steps can be set. This number is used to resampling data sets with larger number of time steps. The default value is 2000 time steps and the following resampling methods can be chosen:

- nearest time step (dropping steps)
- maximum value of a group
- mean value of a group.

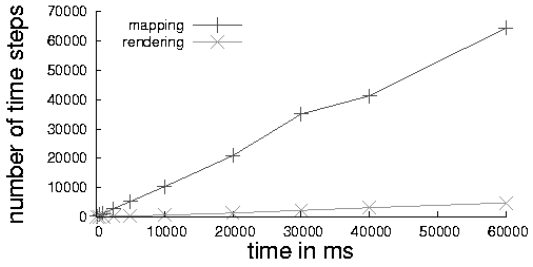

Figure 6: This chart shows how much the rendering process is faster than the mapping process for large time series data (hardware: AMD 1.6 GHz, Nvidia Geforce 2).

This resampling method can also be switched off to visualise all time steps or or a subset of the data.

The most time consuming process is the data mapping when the conical frustum are created. This process is always invoked when the time range, the number of steps shown or the underlying time data is changed. The rendering process which is started when the 3D view is rotated or zoomed using the display lists is much faster (see figure 6).

Conclusions and future work

SimWiz3D is a novel tool to visualise biochemical simulation results. It consists of a 3D view which visualises time series data received from biochemical network simulations and an additional 2D view to keep track of the network structure. A lot of user interactions are provided to adapt the views to user needs and ease the exploration of the data:

- navigating
  - zooming
  - rotating
- focussing:
  - single time step selection
  - highlighting interesting tubes (colour)
  - removing less interesting tubes
- filtering:
  - time range selection
  - minimum concentration change value
  - resampling
  - transparency

Especially the last two interactions (focussing and filtering) are overlapping because mostly the user filters data out to focus on a special issue. In contrast to Dwyer et al. [2] the user can follow each time series in a continuous manner and without disturbing arrows which makes the comparison of complete time series data easier. Furthermore SimWiz3D offers much more user interacting, focussing and filtering techniques which allows the user to adapt the tool to his demands and to explore the data in local or global views. It processes small as well as large data sets with high-dimensional data and a large number of time steps.

SimWiz3D is under further development to integrate correlation techniques and more advanced methods to avoid overlapping of tubes in the 3D view.

Acknowledgements

We would like to thank Jürgen Hesser, as well as Ralph Gauges, Ursula Kummer, Ursula Rost and Sven Sahle for helpful discussions. The project was funded by the Klaus Tschira Foundation (KTF).

References

[1] Tim Dwyer and Peter Eades. Visualising a Fund Manager Flow Graph with Columns and Worms. In *Proceedings of IEEE IV'02*, pages 147–158. IEEE Computer Society, 2002.

[2] Tim Dwyer, Hardy Rolletscheck, and Falk Schreiber. Representing experimental biological data in metabolic networks. In *Proceedings of the second conference on Asia-Pacific bioinformatics*, volume 29, pages 13 – 20, 2004.

[3] Ralph Gauges, Sven Sahle, Ursula Rost, and Katja Wegner. http://projects.villa-bosch.de/bcb/sbml/, 2005.

[4] Michael Himsolt. http://www.infosun.fmi.uni-passau.de/Graphlet/download/misc/GML.ps.gz, 2005.

[5] Michal Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, Ursula Kummer, N. Le Novere, L. M. Loew, D. Lucio, Pedro Mendes, Eric Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, Jrg Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup

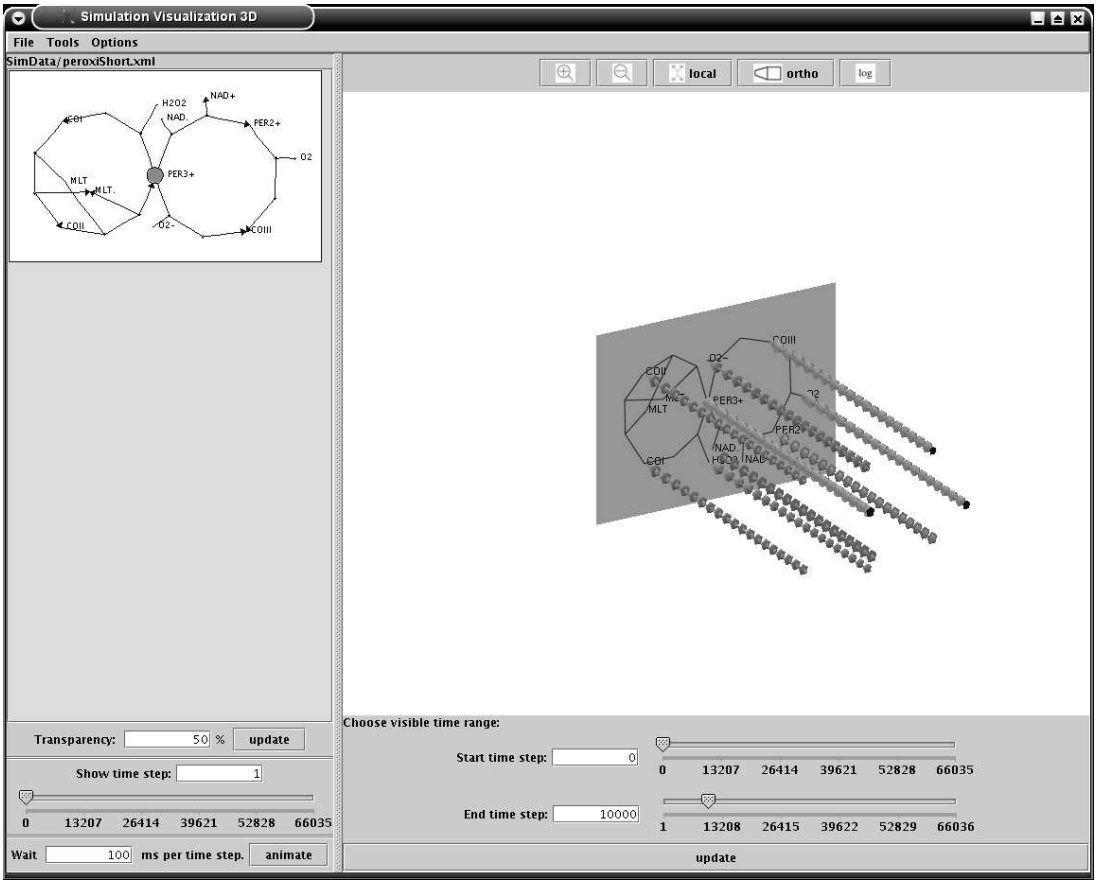

Figure 7: Screen shot of SimWiz3D. The 3D view shows oscillating time series data.

language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 1:524–531, 2003.

[6] OpenGL Architecture Review Board, Dave Shreiner, Mason Woo, Jackie Neider, and Tom Davis. *The OpenGL Programming Guide - The Redbook*. Addison-Wesley Pub Co, 4 edition, 2003.

[7] Ursula Rost and Ursula Kummer. Visualization of bicohemical network simulation with SimWiz. *IEE Systems Biology*, 1:184–189, 2004.

[8] SimWiz. http://atlas/bcb/software/, 2005.