Interactive Visualization and Automatic Analysis of Metabolic Networks – A Project Idea

Andreas Kerren
Institute of Computer Graphics and Algorithms
Vienna University of Technology
Favoritenstraße 9-11, A-1040 Vienna, Austria

Technical Report

September, 2003

Abstract

Visualization and analysis of metabolic networks is becoming increasingly important. Due to the tremendous research efforts in this area of Biochemistry, the level of detail in these networks is quickly increasing.
Both, existing systems and systems under development concentrate on the static visualization of metabolic networks based on classical graph layout methods or simple ad-hoc methods. These approaches cannot match the current needs of the users due to the following reasons. On the one hand, the visualization cannot adequately react to changes in the network and can only rarely take user constraints into account. On the other hand, the size and complexity of these networks requires methods for focusing on details without loosing the context.
In an interdisciplinary cooperation of three research groups with major competence in graph drawing, information visualization, and biochemistry, we wish to build a software tool for interactive visualization and analysis of metabolic networks that satisfies the needs of the dynamic nature of such networks and effectively supports the biochemical users within their daily analysis tasks. Various interaction possibilities as well as readable layouts will help the user to explore metabolic networks, to navigate within them, and to compare them.

1 General Information

The project idea described is highly interdisciplinary. We plan to establish a cooperation between three different research groups:

- Institute of Computer Graphics and Algorithms at the Vienna University of Technology, Austria (information visualization and automatic graph drawing).
- Institute of Computer Science at the University of Cologne, Germany (automatic graph drawing and discrete optimization).
- Institute of Biochemistry at the University of Cologne and Cologne University Bioinformatics Center (CUBIC), Germany (biochemical background and evaluation).

In the rest of this report, we give some information about the state of the art, some possible objectives, and a preliminary workplan, especially for the Institute of Computer Graphics and Algorithms in Vienna.

2 State of the Art

2.1 Visualization of Metabolic Networks

Due to the increasing amount of available data, computer science becomes increasingly important in biochemistry. Since evaluation by hand is impossible, the computer is used for investigating these data for several questions. Traditionally, the analysis is carried out abstractly, resulting in a certain parameter or number.

In recent years, the visual approach to analysis became ever more important. In this approach, the data are drawn in a way that allows the user to obtain information of these data quickly and reliably. The advantage of visualization over abstract analysis is its possibility of conveying much more information to the user at the same time. Interactive control over the drawing can increase significance and depth of information even further, because it allows the user to focus on the aspects of the drawing he/she is interested in.

For analyzing metabolic networks, visualization is an important tool, since a large amount of data is available in this area. In the KEGG-data base [43], 5711 metabolic reactions are recorded, involving 10743 chemical compounds.[1]

Any software for visualizing metabolic pathways and networks must meet several requirements. The graphical interface must show the user every information he asks for. Beyond the network itself, this may be the structural formula of a metabolite or a link to PDB-data base entries.

Furthermore, the drawings of the network must conform to some criteria that are examined systematically in the areas of Automatic Graph Drawing and Visualization. These criteria mainly concern the clearness of the drawing. The subject of research is how well-readable layouts can be characterized and how they can be generated automatically. The second question requires the development of algorithms for problems that are often hard to solve both theoretically and practically. Important techniques from Automatic Graph Drawing in the context of metabolic network visualization are the preservation of the mental map when applying small changes to the graph and the possibility of clustering nodes. Depending on the generated drawing, there are further important visualization approaches, e.g., navigation in the complete network, focusing on parts of the network, or gradual differentiability of nodes with less importance (side metabolites).

Uetz, Ideker and Schwikowski [54] underline the significance of visualizing biochemical networks and describe a number of requirements for future approaches: to name a few, they require the integration of additional information into the graph structure, a better layout, and the display of further information about existing nodes and compounds. They conclude with the following statement: "The challenge will then be to integrate the new data in such a way as to increase our understanding of the underlying biological processes, not obscure them in convoluted figures or excessive detail. Ultimately, these added layers of information will make the network even more powerful as a model on which simulations may be performed to predict experimental outcomes."

None of the existing visualization tools realizes these requirements. One reason is a lack of algorithmic know-how in Graph Drawing, another reason is a lack of understanding of useful visual representation. In the following, we give a short overview over the most important software tools for visualizing and editing metabolic networks.

The widespread drawings of the KEGG-data base [43], see Fig. 1, are produced by hand. They are connected via links, but real interaction is not available. Because of their manual generation, these maps are well readable and can thus serve as an example in terms of quality and user conventions.

The well-known and hand-drawn *Boehringer Biochemical Pathways poster* [40] also led to several visualization approaches. A cutting of the poster is shown in Fig. 2. It was digitized in *ExPASy* [41]; enzymes were linked with data base entries. Further interaction as well as any kind of influence on the drawing is not available.

The software *BioPath* [6, 33, 49] is a fully automatic visualization tool for metabolic pathways that also refers to this poster by taking over its drawing conventions and data. It is based on the layout tool *Graphlet* [30]. The drawings are produced by a Sugiyama-style hierarchical layout

[1] as of August 2003

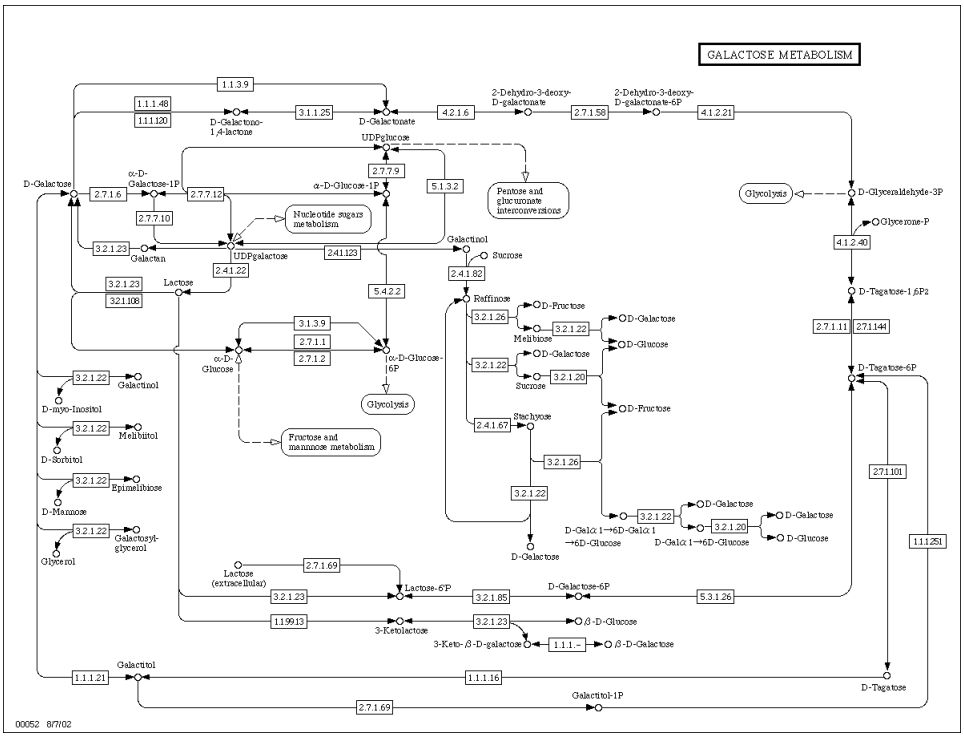

Figure 1: A drawing of the KEGG-data base [43].

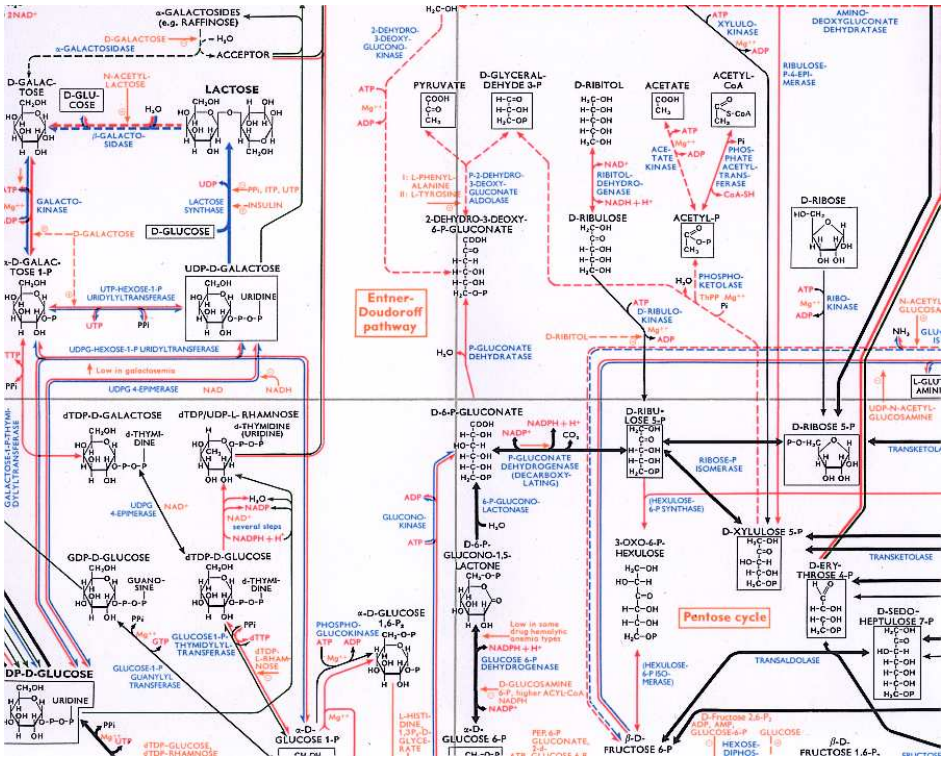

Figure 2: A cutting of the *Boehringer Biochemical Pathways poster* [40].

algorithm allowing restrictions on the horizontal and vertical order of nodes both for preserving the mental map and for drawing cyclic subnetworks in a special way. A number of requirements for visualization of biochemical networks are listed by Schreiber [49]. However, *BioPath* does not fully meet these requirements: there are no means of influencing the drawing interactively. Nodes cannot be grouped, as the algorithm does not support clustering. Furthermore, only small portions of the network are shown without displaying their context. Finally, only a few types of analysis are available.

Another method for drawing metabolic networks automatically was devised by Becker and Rojas [2]. For different kinds of subnetworks, different layout algorithms are used (hierarchical or force-directed). The drawings of subnetworks are then combined to a global drawing. This idea has a positive effect on the layouts, but the partition is only obtained heuristically. However, the good readability of the drawings is also due to the fact that neither side metabolites nor enzymes are displayed. No interaction is possible, and neither clustering nor preservation of mental maps is implemented.

Many visualization approaches produce drawings that are not readable at all, often because they use standard methods from Automatic Graph Drawing without adapting them to the special situation. One example is *DMD* [37]. The drawings of *PathViewer* [51] are also hardly readable; nevertheless, drawings of single pathways are very clear and there are different ways of choosing the information to be displayed.

Other tools produce better layouts but do not allow to group, focus on, fix, or even format parts of the graph, e.g., *PathFinder* [26]. This tool allows to identify paths between metabolites and to compare reaction paths with own data. Other types of analysis or interaction do not exist.

The *Pathway Tools* [34] stand out due to their ability to draw metabolic networks in different steps of abstraction, which is obtained by using symbols for certain subgraphs. This approach could lead to a better readability in general; in this case, however, this is prevented by the poor quality of the drawings.

Recently, a new approach was presented by Rojdestvenski. His *Metabolic Network Visualizer* [48] draws metabolic networks in three dimensions in order to avoid crossings between edges. Interaction is limited to standard techniques (zoom, translation, and rotation) for navigating in the network. Other disadvantages are the unusual visualization which will hardly be accepted by the user and the inadequate visualization of larger networks.

In summary, none of the existing tools fully meets the requirements for the visualization of metabolic networks. Most software products already fail when producing static drawings, either because they were developed by computer scientists using standard techniques from Automatic Graph Drawing without adapting them to the special situation or because they were developed by biologists or bioinformaticists who do not know this area. Additionally, most tools lack interaction and dynamic visualization techniques like Focus & Context. A visual representation of results from abstract analysis, experimental insight, or simulations (see e.g. [28]) is usually impossible.

2.2 Automatic Graph Drawing[2]

Some of the software tools described above originate from the area of Automatic Graph Drawing. However, they only use simple methods as the Sugiyama-style hierarchical layout or the spring embedder. As long as only a single path or a union of a small number of paths is to be drawn, the Sugiyama-style drawings are often very satisfactory, as in *BioPath* or *PathViewer*. For more complex subnetworks, the planarization method is preferable.

Planarization The idea of the planarization method is to make a graph planar in order to draw it by a layout algorithm for planar graphs. The usual planarization approach first searches for a planar subgraph that is as large as possible and computes a plane embedding of this subgraph. In the next step, the edges outside the subgraph are inserted into this embedding with a small

[2]This section was written by Christoph Buchheim, Institute of Computer Science at the University of Cologne, Germany.

number of edge crossings; these crossings are replaced by virtual nodes. The result is a plane graph. Since both crossing minimization and the detection of a maximum planar subgraph are NP-hard problems, planarization is often done heuristically.

The plane graph can be drawn by different methods now. Because of their very good quality, orthogonal drawings are a good choice. By Tamassia [53], an orthogonal representation with a minimum number of edge bends can be computed in polynomial time using network flow methods if the graph has maximum degree four. For extensions to general plane graphs, see [3, 18, 21].

By consistently assigning lengths to the edge segments, the orthogonal representation is transformed into a real drawing. Possible aims in this step are the minimization of the total edge length, the length of the longest edge, or the drawing area. All these minimization problems are NP-hard [44]. Heuristic compaction methods usually base on the idea of dissecting rectangles from the regions [53]. Another approach is presented in [11]. The algorithm of [19] is designed for graphs with different node sizes.

Clustering An important aim of our project is allowing clustering in the planarization method. An open problem in this context is c-planarity testing: a clustered graph is called c-planar if it admits a planar drawing that places all nodes of a cluster into a common region; edges are allowed to cross any region border at most once. It is unknown whether this problem can be solved in polynomial time. If the clustered graph is c-connected, i.e., if all subgraphs induced by clusters are connected, there are linear time algorithms for this problem [15, 20]. In practice, the problem is solved by adding virtual edges to the clusters making them c-connected [16, 27]. Now the resulting graph is tested for planarity. However, planarity may be destroyed by adding the "wrong" edges. If the graph is not c-planar, it can be planarized in a similar way as general graphs.

The c-planarity test also yields a c-planar embedding, being used to generate a c-planar drawing. Several algorithms for the latter problem exist, e.g., the algorithm of Feng [17] or an extension of the bend minimization algorithm of Tamassia. Before defining the auxiliary network, artificial nodes and edges representing the cluster borders are added to the graph. In the network flow algorithm, these edges can be forced to represent a rectangle. The resulting drawings have a good quality but require a relatively large area.

Mental Map The preservation of the "mental map" [42] is an important challenge in Automatic Graph Drawing. When adding single nodes and edges, the user usually wants to change the drawing as little as possible. However, most algorithms will create a completely new layout when applied to the new graph again. Hence one has to make sure explicitly that the general impression of the drawing is preserved, while at the same time keeping an eye on the general quality of the layout.

Existing approaches to this problem are diverse. Up to now, neither a generally accepted method for preserving the mental map nor a useful formalization of this problem have been presented. A first step towards this is [12], proposing certain metrics to measure the difference between the old and the new drawing.

The natural way to attack the problem is to intervene in the drawing algorithm and to enforce or prefer certain properties of the resulting drawing already at this stage. Differences from the previously generated drawing can be punished in a way that the optimization aim and the preservation of the mental map are tuned to each other. For the bend minimization algorithm of Tamassia, this approach was presented in [10]. A related approach is [9], where orthogonal drawings following a given sketch of the drawing are generated.

Common and Special Subgraphs The search for common subgraphs is a classical NP-hard problem [24]. In our project, however, we do not want to find exactly equal subgraphs but similar ones. In this weaker formulation, the problem was hardly examined up to now. One reason may be the increase of complexity (at least practically) induced by the fuzziness.

A closely related problem is the search for fuzzy symmetries. Few is known here, too. An approach based on symmetric subgraphs is devised in [13]. Here it is shown that the problem

of finding maximum symmetric subgraphs is NP-hard even for trees in general. Polynomial time algorithms are given for the special cases that the symmetry is a reflection, the graph is a tree or a plane graph, and the subgraph is node-induced.

In Automatic Graph Drawing, the search for special subgraphs is also important. This problem can be regarded as a special case of the search for common subgraphs; the first graph is fixed in this case. This problem arises when trying to draw certain substructures in a special way and thus having to identify these structures before. This unsolved problem is the theme of this year's graph drawing contest [5]. Again, the problem is complicated by introducing some tolerance.

2.3 Information Visualization

Biochemistry is a relatively new field of application for Information Visualization (InfoVis). One of the most important tasks is to find suitable visual metaphors: How should the abstract information be displayed, how should the data be arranged, and so on.

Metaphors In the case of biochemical networks, few work has been invested in developing metaphors, except for hand-drawn diagrams (see Fig. 1 and 2). Most of the tools described above use standard techniques without adapting them to the special requirements. Some first steps are done in recent publications (e.g., [49]), but they fail to allow global orientation in the network. A number of visual metaphors for the visualization of other networks exist (mostly of technical type, e.g., Internet connections and evolution of the WWW [7] or structure of web sites [14]). However, these metaphors can only partly be transferred to our special case, or they are limited to hierarchies.

Interaction and Navigation The same is true for classical interaction techniques as navigation, "Level of Detail (LOD)", or "Focus & Context (F&C)". In the tools described above, these methods are only available in a basic way, if at all. Navigation in the KEGG-data base is performed by opening new windows or diagrams. Geometric standard methods like "Perspective Walls" [39], "Table Lenses" [46], "Fisheye Views" [22], or "Hyperbolic Browser" [38] are not fully suitable. This is due to the fact that these methods have been developed without regarding other techniques. The software developed in our project will incorporate many other means of interaction (exploration, different types of analysis, clustering), which conflict with traditional concepts. The complexity of our application requires to develop completely new techniques that will be accepted by the user.

Further Aspects of Visualization The visual representation of different metabolites and reaction paths requires solutions that will be developed in our project. Further important problems are the display of information on the metabolites, the embedding of results of algorithms for abstract analysis into the visualization, or the visual comparison of two or more networks. Previous work on this subject was done by Rayson [47] in the case of visualizing geographic data on maps and by Keahey and Eick [35] for the visual path analysis of Internet traffic.

In summary, the seamless integration of data into a given network visualization is few examined up to now. Some attempts exist in so-called "battlespace" visualizations, in which often a lot of heterogeneous data have to be displayed [1].

3 Objectives

The purpose of our project is to develop a flexible tool for interactive visualization and analysis of metabolic networks. We target researchers from biology, in particular biochemistry and molecular biology. With our software, the user will be able to automatically draw and analyze data obtained from data bases or from his/her own research. He/she will be able to influence the type of visualization in many ways and to adapt it to his current problems and questions.

3.1 Visualization

Because of the huge size of the networks, it is important to focus on certain subnetworks being relevant for the current questions of the user. On one hand, less important parts of the network should be hidden or reduced to a single symbol, as done in the KEGG-data base [43], see Fig. 1. On the other hand, interesting regions should be highlighted and displayed in detail. In the area of information visualization, this concept is known as "Focus & Context (F&C)". For the special case of complex biochemical networks, there do not exist satisfying solutions yet.

The selection of relevant parts of the network will be much more flexible than in other tool. We will implement three different methods of selection:

- First of all, the software will support selection via the graphical user interface, allowing to hide, reduce, or highlight arbitrary subnetworks dynamically at any time.
- Second, it will provide a data base to store subgraphs that should be hidden, reduced, or highlighted in every drawing or in certain contexts.
- Third, we will incorporate algorithms for finding interesting subnetworks automatically, e.g., candidates for operons or subnetworks corresponding to orthologous genes.

A closely related technique is "Semantic Zooming", also known as changing the "Level of Detail (LOD)". The user should be able to arbitrarily choose the accuracy and detail of the structural or quantitative information given in the visualization. Here, one also has to consider additional information linked with metabolites or enzymes. The development of a useful navigation technique is crucial for our project.

Another feature of the software will be the ability of drawing two or more networks at the same time while finding similar regions automatically. These regions can be highlighted and visualized in a similar way, so that the user is able to identify them at first sight. The necessary algorithms for detecting similar subgraphs will be developed in the project. By this, it would also be possible to identify missing reaction paths (see [43]).

The software will also allow to add new metabolites or reactions to the network without destroying the current drawing. By this, the "mental map" [42] is preserved, sparing the user from confusion and the necessity of reorientation.

The placement of nodes and edges will base on the planarization method. This method usually produces high quality layouts, yet several modifications are necessary in order to meet the requirements of our application. First of all, we have to prevent functionally related nodes from being placed far from each other. This problem will be solved by allowing (manual or automatic) clustering. Another problem of the planarization method applied to metabolic networks is that some subgraphs are not drawn according to biochemical conventions, e.g., cycles. Substructures with a fixed or at least restricted drawing will be identified automatically and then drawn according to this restriction.

3.2 Methods of Analysis

For analyzing metabolic networks, our software system will not only contain visualization tools but also algorithms for abstract analysis, e.g., the computation of important parameters as the average path length [32] or the "Overall Closeness Centrality Index" [58] as well as the "Giant Strong Component (GSC)" [58]. The abstract results of such algorithms will be visualized, too. A smooth integration of both input and output of these algorithms into the visualization will strongly improve the usability of our software. As far as we know, this has not been realized by any other visualization tool yet.

Besides information about reaction paths obtained by experiments, the user may also want to estimate the influence of hypothetic reaction paths on the characteristics of the total reaction system. The formulation of such hypotheses will be possible interactively and the user will be able to check them via visualization.

Another practical requirement is the support of simulations of reaction processes, including an adequate visualization. An example for such a simulation is the knockout-experiment, where the influence of a missing enzyme is examined. Affected reaction paths and possible bypasses have to be displayed here.

4 Workplan

4.1 Joint Work

At the beginning, some work has to be done by all groups together:

- Determination of international standards and formats. On one hand, these are standards in the terminology of biological objects (e.g., enzymes) or formats for biochemical reactions as the *Systems Biology Markup Language (SBML)* [31]. On the other hand, these are standards for recording graphs, e.g., the *Graph Modeling Language (GML)* [25] or *GraphXML* [29]. Other examples are *GXL* [56] and *GraphML* [8], which were both presented on the *Graph Drawing 2001* symposium.
- Determination of interfaces and intermediate representations that will be used by the partners of the project during the development of the components.

In the course and at the end of the project we will also have to do some common work:

- "Formative Evaluation" of the system to be developed with the help of test persons: several evaluations (e.g., by usability tests) will be carried out during the complete implementation phase. By this, we can identify weak points in the approach or in the implementation at an early stage.
- "Summative Evaluation" of the developed system at the end of the project with the help of test persons, for obtaining a final rating of the software. This evaluation will rely on objective quantitative data and the use of statistic methods. For this, we will have to develop a concept, design questionnaires, execute the experiments with our new visualization software, and analyze the data obtained in this process. Our experience from evaluating learning software in the project GANIMAL [23, 36] with more than 100 students will be of great help here.
- Production of a documentation for using and extending the software as well as a tutorial at the end of the project.

4.2 Algorithms and Data Structures Group, Vienna

In order to be platform independent and to allow representation via WWW, we will use Java and Java3D for the visualization part.

Metaphors Visualization of biochemical networks is a relatively new field in information visualization. An important task in information visualization is the development of suitable metaphors which support the spatial as well as the temporal navigation needs of the user. They should facilitate the comprehension of the network structure by ensuring the identification of the user model with the model of the underlying visualization system. Concerning our project, the visualization must correspond with the natural expectations of the biochemists. With respect to this point, there are still a few open questions which go beyond our catalogue discussed in section 2.

Interaction The type of interaction that will be used in our visualization is called "direct manipulation" (see [45]) in the Human Computer Interaction (HCI) literature. Important properties of this interaction type are continuous representation of the relevant objects, the possibility of direct physical manipulation of objects (e.g., using the mouse), and the execution of incremental as

well as reversible operations in (soft) real time. The development of such interactive visualizations for users from the molecular biology community is one of the main goals of this project. This goal appears in almost every subsection of our work program.

In order to store the settings for the current user session, we need to develop a new extendable storage format describing the current state of the visualization. This format should not only contain obvious options such as the colours of the metabolits, but also properties of the visualization itself such as the clusters and the hidden parts of the network. To our knowledge, none of the tools for biochemical networks described in the literature has this ability.

Overview and Navigation Based on the literature mentioned in Section 2 and on recent papers from Information Visualization (e.g., [7, 52, 55, 57]) concerning Focus & Context (F&C) and navigation techniques in networks arising from other applications (e.g., the WWW, automata, ...), we need to develop a meaningful representation of the entire network. Our user needs to maintain an overview over the entire network while concentrating on a certain detail. Thereby, an intelligent navigation avoids a fragmented view of the subnetworks. One of the major aspects with navigation is the avoidance of disorientation, which arises if only details are emphasized. An important task for our project will be to avoid these navigation pitfalls.

One possibility could be to use a three-dimensional (3D) representation of the entire network. Our experience, however, shows that the user of such a visualization tends to be confused, if the 3D representation is not clearly arranged (like, e.g., a simple 3D Sugiyama layout). Open questions arising in the context of biochemical networks as well as in general information visualization, are: How can such arrangements look like, what will be accepted by the users, and how could the navigation look like in this case?

We suggest the following new idea in order to answer the above questions for our project: We will partition the 3D space into 2D levels (also, see Brandes and Corman [7]), each containing a plane drawing of a subnetwork. Relations between objects on different planes can be visualized via half-transparent edges, however, the number of these edges must be small. By intelligent navigation between the planes it is possible to get a complete overview of the entire network. Each subnetwork is visualized in 2D according to the quality and user conventions. These 2D visualizations will be generated using the new layout algorithms developed within our project (mainly by the Cologne group). The selection of the subnetworks for displaying on the planes is an important open problem which has to be examined in this project. It could be done by the user itself (interactively via the graphical user interface), by certain criteria (e.g., known reactions, certain cycles, ...) or automatically determined by algorithms (e.g., in [59]). Periodical consultations with our project partner, CUBIC, will guarantee that the visualization will be evaluated from a biochemist's point of view. For that purpose, we will develop simple 3D models in VRML with simple navigation (zoom, pan, rotate) right at the beginning of the project (rapid prototyping).

This visualization idea is also useful for integrating further visualization techniques: In order to focus on arbitrary parts of the network from the network overview (F&C), an arbitrary selected subpart of the network can thus be represented on one plane. Only adjacent planes are displayed and the rest of the entire network is hidden or is drawn transparent. The focussed plane can then be displayed in different ways, e.g., in top view.

Consequently, we need a possibility to select and to reselect parts of the network to hide them. This is desired by the biochemical users and needs complex solutions. Hiding network parts can form gaps in the network. We have to investigate how to represent the gaps in terms of place holders. A place holder could display the former content as an icon if the content is not too big. This is a property of the LOD-concept described in section 3.

There are further aspects that play an important role within this context. We will mention only the following: backtracking and to keep track on the history of all user actions, setting marks on the network, usage of audio and help functions.

Further Aspects of Visualization An open problem is the representation of additional quantitative information (enzyme number, DNA-sequence, protein structure, ...) at nodes (and edges)

of the displayed network. The visualization system should support the user-defined selection of the LOD of such information so that the user can concentrate on his/her own goals of his/her work. Default settings are important: A suggestion for this could be that the LOD in the focus should be as high as possible. But the visualization of these node attributes is too space-consuming in most cases to place them within the nodes. We have to find new ways because their simple presentation in a separate window is not sufficient in practice. Biochemists call for the integration of such additional information into the network visualization, see [54]. Further work to be undertaken, which affects the representation of the metabolites, is:

- Separate drawing of "side metabolites", i.e., metabolites of side reactions (e.g., ATP-consumption), in order to facilitate their differentiation from main metabolites and the clarity in context of the entire network. On the one hand, we have the possibility to change the graph layout itself (mainly by the Cologne group); on the other hand, we can modify the graphical representation of side metabolites without changing the graph layout (mainly by the Vienna group). Probably, a combined approach will be a promising solution for that problem.

- Integration of the very long names of metabolites into the visualization of the entire network without changing the graph layout. So, we can prevent big gaps in the network.

- Interactive selection of clusters (grouping of metabolic pathways) inside the visualization. This is an important point for biochemists. The selection should be as easy as possible because complex typed specifications are not accepted by the user.

Also, there is a number of working tasks that aim at the visual representation of the actual metabolic pathways:

- Obvious, and in other tools also realized, requirements are the drawing and highlighting of the metabolic pathways or the display of all reaction paths from a metabolit A to another metabolit B. We plan to develop two different views for such requirements: one within the entire visualization of the network and the other in a separate window with its own optimized graph layout.

- Implementation of an algorithm which supports the interactive insertion of new metabolites or pathways into the visualization of the entire network. Originally, these new objects can be defined by biochemical experiments or by the user as hypotheses. The system should distinguish both possibilities in the visualization and preserve the mental map of the previous network view. For this purpose, we nead to develop new algorithms in cooperation with our project partners.

- Automatic identification of circles and the development of a highlighting method to mark these circles in the entire network if they cannot be drawn as circles.

- Display of correlations between metabolites or pathways in the entire network.

A further important point is the development of a visualization concept for comparing two or more metabolic networks. The project group at Cologne will develop an algorithm that automatically computes identical or similiar areas of networks. Our working task is to visualize both (or all) networks and to highlight the identical or similiar parts in an adequate way, e.g., by putting them on top of each other in 3D space. But these parts can have a very complex structure so that we need to develop a dedicated visualization method. Its usability has also to be proved during the project by CUBIC.

All these aspects can occur in the visualization process at the same time. Therefore, we must be careful to not overload the visualization with too much visual information, e.g., by designing the visualizations so that the user can set his/her own priorities.

Visualization of Analysis Results The visual representation of analysis results and their integration into the visualization of the entire biochemical network have not been in the center of past research. One of the crucial points of our approach will be to "seamlessly" integrate such results into our visualization. Possible solutions range from simple coloring/grouping, e.g., of the GSC [58] of the network, to displaying quantitative results in terms of numbers, icons, etc. As before, we want to compare results of analyses and the visualization system should also support this. In this way, the user can discover correlations easily and fast. In addition, we need a seamless integration of the data input for these analyses. It depends on the kind of analysis, e.g., the user must select certain metabolites or pathways in some cases. In other cases, the input of simple parameters suffices.

In order to embed the specification of hypotheses into the visualization, we also have to develop new concepts. Like some approaches in the research of explorative educational systems, the visualization should stimulate the mental construction of knowledge. At educational systems, this construction facilitates the knowledge transfer that takes place self-controlled, self-organized and self-evaluated. The aim of the learning process is to shape the ability to solve problems. In case of the formulation of hypotheses within the visualization, the aim is to facilitate the discovery of novel correlations and the better understanding how certain pathways work. To reach this goal, we have to investigate how we can accomplish the exploration within metabolic networks. There are a lot of studies and experimental results in the area of educational software systems which confirm that explorative and interactive visualizations produce better learning results than static visualizations (see [4, 50]).

In this context the system should also support visual simulations of different biochemical reactions. In consultation with our project partner CUBIC, we have to detect the most interesting reactions to be simulated. It is possible that we will add dynamic elements to the simulations so that we get animations on biochemical networks.

During the project, other challenges will arise by formative evaluations and interviews with users. High-quality interactive visualization needs a big effort in design and implementation. In particular, the used visual metaphor as well as the quality of the implementation is of paramount importance because visualization systems must work reliably without any time lag and must be visually appealing in order to have a chance for practical success.

Acknowledgments

The author would like to thank Christoph Buchheim, University of Cologne, for writing Section 2.2, carefully proof-reading and his valuable comments on this report. Furthermore, I am thankful to Michael Jünger, University of Cologne, as well as Petra Mutzel, TU Vienna, for their support and proof-reading.

References

[1] K. Andrews, S. F. Roth, and P. C. Wong, editors. *Proceedings of the 7th IEEE Symposium on Information Visualization (InfoVis '01)*. IEEE Computer Society Press, 2001.

[2] M. Becker and I. Rojas. A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, 17(5):461–467, 2001.

[3] P. Bertolazzi, G. Di Battista, and W. Didimo. Computing orthogonal drawings with the minimum number of bends. In F. Dehne, A. Rau-Chaplin, J.-R. Sack, and R. Tamassia, editors, *5th International Workshop on Algorithms and Data Structures (WADS 1997)*, volume 1272 of *Lecture Notes in Computer Science*, pages 331–344. Springer-Verlag, 1997.

[4] A. Blumstengel. *Entwicklung hypermedialer Lernsysteme*. Wissenschaftlicher Verlag Berlin, WVB, 1998.

[5] F. J. Brandenburg. Graph Drawing Contest 2003: "Drawing Graphs within Graphs". `www.gd2003.org`, 2003.

[6] F. J. Brandenburg, M. Forster, A. Pick, M. Raitner, and F. Schreiber. BioPath – visualization of biochemical pathways. In *Proceedings of the German Conference on Bioinformatics (GCB '01)*, 2001.

[7] U. Brandes and S. R. Corman. Visual unrolling of network evolution and the analysis of dynamic discourse. In P.C. Wong and K. Andrews, editors, *Proceedings of the 8th IEEE Symposium on Information Visualization (InfoVis '02)*, pages 145–151. IEEE Computer Society Press, 2002.

[8] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and S. Marshall. GraphML progress report – structural layer proposal. In P. Mutzel, M. Jünger, and S. Leipert, editors, *Proceedings of the 9th International Symposium on Graph Drawing (GD 2001)*, volume 2265 of *Lecture Notes in Computer Science*, pages 501–512. Springer-Verlag, 2002.

[9] U. Brandes, M. Eiglsperger, M. Kaufmann, and D. Wagner. Sketch-driven orthogonal graph drawing. In S. G. Kobourov and M. T. Goodrich, editors, *Proceedings of the 10th International Symposium on Graph Drawing (GD 2002)*, volume 2528 of *Lecture Notes in Computer Science*, pages 1–11. Springer-Verlag, 2002.

[10] U. Brandes and D. Wagner. Dynamic grid embedding with few bends and changes. In K.-Y. Chwa and O. Ibarra, editors, *Proceedings of the 9th International Symposium on Algorithms and Computation (ISAAC 1998)*, volume 1533 of *Lecture Notes in Computer Science*, pages 89–98. Springer-Verlag, 1998.

[11] S. Bridgeman, G. Di Battista, W. Didimo, G. Liotta, R. Tamassia, and L. Vismara. Turn-regularity and optimal area drawings of orthogonal representations. *Computational Geometry*, 16(1):53–93, 2000.

[12] S. Bridgeman and R. Tamassia. Difference metrics for interactive orthogonal graph drawing algorithms. *Journal of Graph Algorithms and Applications*, 4(3):47–74, 2000.

[13] H.-L. Chen, H.-I. Lu, and H.-C. Yen. On maximum symmetric subgraphs. In J. Marks, editor, *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000)*, volume 1984 of *Lecture Notes in Computer Science*, pages 372–383. Springer-Verlag, 2001.

[14] E. H. Chi and S. K. Card. Sensemaking of evolving web sites using visualization spreadsheets. In G. Wills and D. Keim, editors, *Proceedings of the 5th IEEE Symposium on Information Visualization (InfoVis '99)*, pages 18–25. IEEE Computer Society Press, 1999.

[15] E. Dahlhaus. Linear time algorithm to recognize clustered planar graphs and its parallelization. In C. L. Lucchesi, editor, *Proceedings of the 3rd Latin American Symposium in Theoretical Informatics*, volume 1380 of *Lecture Notes in Computer Science*, pages 239–248. Springer-Verlag, 1998.

[16] G. Di Battista, W. Didimo, and A. Marcandalli. Planarization of clustered graphs. In P. Mutzel, M. Jünger, and S. Leipert, editors, *Proceedings of the 9th International Symposium on Graph Drawing (GD 2001)*, volume 2265 of *Lecture Notes in Computer Science*, pages 60–74. Springer-Verlag, 2002.

[17] P. Eades, Q.-W. Feng, and H. Nagamochi. Drawing clustered graphs on an orthogonal grid. *Journal of Graph Algorithms and Applications*, 3(4):3–29, 1999.

[18] M. Eiglsperger, U. Fößmeier, and M. Kaufmann. Orthogonal graph drawing with constraints. In *Proceedings of the 11th Annual Symposium on Discrete Algorithms (SODA 2000)*, pages 3–11. ACM Press, 2000.

[19] M. Eiglsperger and M. Kaufmann. Fast compaction for orthogonal drawings with vertices of prescribed size. In P. Mutzel, M. Jünger, and S. Leipert, editors, *Proceedings of the 9th International Symposium on Graph Drawing (GD 2001)*, volume 2265 of *Lecture Notes in Computer Science*, pages 124–138. Springer-Verlag, 2002.

[20] Q.-W. Feng, R.-F. Cohen, and P. Eades. Planarity for clustered graphs. In P. Spirakis, editor, *Proceedings of the 3rd Annual European Symposium on Algorithms (ESA 1995)*, volume 979 of *Lecture Notes in Computer Science*, pages 213–226. Springer-Verlag, 1995.

[21] U. Fößmeier and M. Kaufmann. Drawing high degree graphs with low bend numbers. In F. J. Brandenburg, editor, *Proceedings of the 3rd International Symposium on Graph Drawing (GD 1995)*, number 1027 in Lecture Notes in Computer Science, pages 254–266. Springer-Verlag, 1995.

[22] G. W. Furnas. Generalized fisheye views. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'86)*, pages 16–23. Addison-Wesley, 1986.

[23] GANIMAL – Projekt Homepage. *http://www.cs.uni-sb.de/GANIMAL*, 2002.

[24] M. Garey and D. Johnson. *Computers and Intractability – A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[25] GML – a Graph Modelling Language. `www.uni-passau.de/Graphlet/GML`, 1997.

[26] A. Goesmann, M. Haubrock, F. Meyer, J. Kalinowski, and R. Giegerich. PathFinder: reconstruction and dynamic visualization of metabolic pathways. *Bioinformatics*, 18:124–129, 2002.

[27] C. Gutwenger, M. Jünger, G. W. Klau, S. Leipert, P. Mutzel, and R. Weiskircher. AGD – a library of Algorithms for Graph Drawing. In P. Mutzel, M. Jünger, and S. Leipert, editors, *Proceedings of the 9th International Symposium on Graph Drawing (GD 2001)*, volume 2265 of *Lecture Notes in Computer Science*, pages 473–474. Springer-Verlag, 2002.

[28] M. D. Haunschild. Ein Werkzeug zur automatischen Modellgenerierung und verteilten Simulation metabolischer Netzwerke. Master's thesis, University of Siegen, Dept. of Electrical Engineering and Computer Science, 2001.

[29] I. Herman and M. S. Marshall. Graphxml - an XML-based graph description format. In *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000)*, pages 52–62, 2000.

[30] M. Himsolt. Graphlet: Design and implementation of a graph editor. *Software – Practice and Experience*, 30:1303–1324, 2000.

[31] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, and H. Kitano. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[32] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

[33] M. Jünger and P. Mutzel, editors. *Graph Drawing Software*. Mathematics and Visualization. Springer-Verlag, 2003.

[34] P. D. Karp, M. Ridley, S. M. Paley, A. Pellegrine-Toole, and M. Krummenacher. The Pathway Tools software. *Bioinformatics*, 18:225–232, 2002.

[35] T. A. Keahey and S. G. Eick. Visual path analysis. In P.C. Wong and K. Andrews, editors, *Proceedings of the 8th IEEE Symposium on Information Visualization (InfoVis '02)*, pages 165–168. IEEE Computer Society Press, 2002.

[36] A. Kerren. *Generierung interaktiver Animationen für den Übersetzerbau*. PhD thesis, Saarland University, Saarbrücken, July 2002. Shaker Verlag, Aachen, ISBN 3-8322-0899-2.

[37] R. Küffner, R. Zimmer, and T. Lengauer. Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics*, 16:825–836, 2000.

[38] J. Lamping and R. Rao. The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of Visual Languages and Computing*, 7(1):33–55, 1995.

[39] J. D. Mackinlay, G. G. Robertson, and S .K. Card. The perspective wall: detail and context smoothly integrated. In *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems*, pages 173–179. ACM, ACM Press, 1991.

[40] G. Michal. Biochemical Pathways (poster), 3rd edition, Boehringer Mannheim GmbH. *Biochemica*, 1993.

[41] G. Michal. *Biochemical Pathways (Biochemie-Atlas)*. Spektrum Akademischer Verlag, 1999.

[42] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout Adjustment and the Mental Map. *Journal of Visual Languages and Computing*, 6(2):183–210, 1995.

[43] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1):29–34, 1999.

[44] M. Patrignani. On the complexity of orthogonal compaction. *Computational Geometry*, 19(1):47–67, 2001.

[45] J. Preece, H. Sharp, D. Benyon, S. Holland, and T. Carey. *Human-Computer Interaction*. Addison-Wesley, Wokingham, Reading, Menlo Park, New York, 1994.

[46] R. Rao and S. K. Card. The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'94)*. ACM, 1994.

[47] J. K. Ray. Aggregate towers: Scale sensitive visualization and decluttering of geospatial data. In G. Wills and D. Keim, editors, *Proceedings of the 5th IEEE Symposium on Information Visualization (InfoVis '99)*, pages 92–99. IEEE Computer Society Press, 1999.

[48] I. Rojdestvenski. VRML metabolic network visualizer. *Computers in Biology and Medicine*, 33:169–182, 2003.

[49] F. Schreiber. High quality visualization of biochemical pathways in BioPath. *In Silico Biology*, 2, 2002.

[50] R. Schulmeister. Zukunftsperspektiven multimedialen Lernens. In K. H. Bichler and W. Mattauch, editors, *Multimediales Lernen in der medizinischen Ausbildung*. Springer, Heidelberg, 2000.

[51] M. Šīrava, T. Schäfer, M. Eiglsperger, M. Kaufmann, O. Kohlbacher, E. Bornberg-Bauer, and H.-P. Lenhof. BioMiner – modeling, analyzing and visualizing biochemical pathways and networks. *Bioinformatics*, 16:219–230, 2002.

[52] J. M. Six and I. G. Tollis. Effective graph visualization via node grouping. In K. Andrews, S. F. Roth, and P. C. Wong, editors, *Proceedings of the 7th IEEE Symposium on Information Visualization (InfoVis '01)*, pages 51–58. IEEE Computer Society Press, 2001.

[53] R. Tamassia. On embedding a graph in the grid with the minimum number of bends. *SIAM Journal on Computing*, 16(3):421–444, 1987.

[54] P. Uetz, T. Ideker, and B. Schwikowsi. Visualization and integration of protein-protein interactions. *Protein-Protein Interactions – A Molecular Cloning Manual*, pages 623–646, 2002.

[55] F. van Ham, H. van de Wetering, and J. J. van Wijk. Visualization of state transition graphs. In K. Andrews, S. F. Roth, and P. C. Wong, editors, *Proceedings of the 7th IEEE Symposium on Information Visualization (InfoVis '01)*, pages 59–66. IEEE Computer Society Press, 2001.

[56] A. Winter. Exchanging graphs with GXL. In P. Mutzel, M. Jünger, and S. Leipert, editors, *Proceedings of the 9th International Symposium on Graph Drawing (GD 2001)*, volume 2265 of *Lecture Notes in Computer Science*, pages 485–500. Springer-Verlag, 2002.

[57] Ka-Ping Yee, D. Fisher, R. Dhamija, and M. Hearst. Animated exploration of dynamic graphs with radial layout. In K. Andrews, S. F. Roth, and P. C. Wong, editors, *Proceedings of the 7th IEEE Symposium on Information Visualization (InfoVis '01)*, pages 43–50. IEEE Computer Society Press, 2001.

[58] A.-P. Zeng and H.-W. Ma. System structure analysis of metabolic networks constructed from genome data. Abstract, 2002.

[59] Y. Zheng, J. D. Szustakowski, L. Fortnow, R. J. Roberts, and S. Kasif. Computational identification of operons in microbial genomes. *Genome Research*, 12:1221–1230, 2002.