Supporting Multi-Level Models in Systems Biology by Visual Methods

S. Biermann, A.M. Uhrmacher, H. Schumann
Department of Computer Science
University of Rostock

Abstract

Multi-level models describe a system on different organization levels explicating the structure of a system, in terms of its components and the interaction between those. In Systems Biology, multi-level models result in a hierarchical structure, whose different layers might comprise thousands of model components. To make full use of the advantage of multi-level models, i.e. the explicit representation of the model's structure, visualization techniques are required that support overview and detail inspections in an interactive manner. Based on a model, which describes the tryptophan synthase as a multi-level model in JAMES, we show how different information visualization techniques, i.e. overview and detail techniques, Rings, interactive foldings of substructures and sinks for information hiding, can be combined to support the analysis of even highly unbalanced model structures.

1 Introduction

Bioinformatics is concerned with conceptualizing biology in terms of molecules (in the sense of physical-chemistry) and then applying informatics techniques (derived from disciplines such as applied mathematics, computer science, and statistics) to understand and organize the information associated with these molecules(Luscombe et al. 2001). Systems Biology can be interpreted as a sub-branch of Bioinformatics which is aimed at improving our insights into the dynamics of cellular systems. In this context modeling and simulation methods are crucial (Wolkenhauer et al. 2003).

During all phases of modeling and simulation, classical visualization techniques are employed (Nocke et al. 2003), i.e. during designing the model, experimenting with the model, and analyzing the results of the simulation. As do the different phases of modeling and simulation in general, the different visualization techniques serve an easier understanding of the system and its behavior(Slavik et al. 2003). Among the different phases, the modeling process is considered crucial in gaining an insight into the system. Its effect is considered the larger, the less knowledge about the system is available, as is also argued in explorative modeling approaches (Davis 2000).

To support the modeling of complex systems, many formalisms, languages and tools allow to hierarchically compose models. They typically integrate the different traditional views in modeling systems, i.e. as functional models, as networks of interactions, and as hierarchical composition of models. Thereby, composition and interaction determines the overall structure of a model (Zeigler 1996). The more complex the structure of a model becomes, in terms of numbers, heterogeneity of components and interaction patterns, the more new visualization techniques are required that address the problems of these "large scale models". Under the term "Information Visualization" visualization methods have been developed to provide a "compact graphical presentation and user interface for rapidly manipulating large numbers of items (10^2 – 10^6), ... Effective information visualizations enable users to make discoveries, decisions or explanations about patterns (correlations, clusters, gaps, outliers, ...), groups of items or individual items." (Shneiderman 2001).

The paper focuses on supporting the development and analysis of multi-level models based on applying and adapting advanced visualization techniques. The paper is organized as follows. First the concept of multi-level models and its role in Systems Biology is explained. Afterward a short overview on applying visualization techniques for modeling purposes in Systems Biology is presented. Thereafter, the multi-level model of the tryptophan synthase in JAMES is shortly described. In the following sections some of the employed visualization techniques are explained in more detail.

Proceedings 18th European Simulation Multiconference
Graham Horton (c) SCS Europe, 2004
ISBN 3-936150-35-4 (book) / ISBN 3-936150-36-2 (CD)

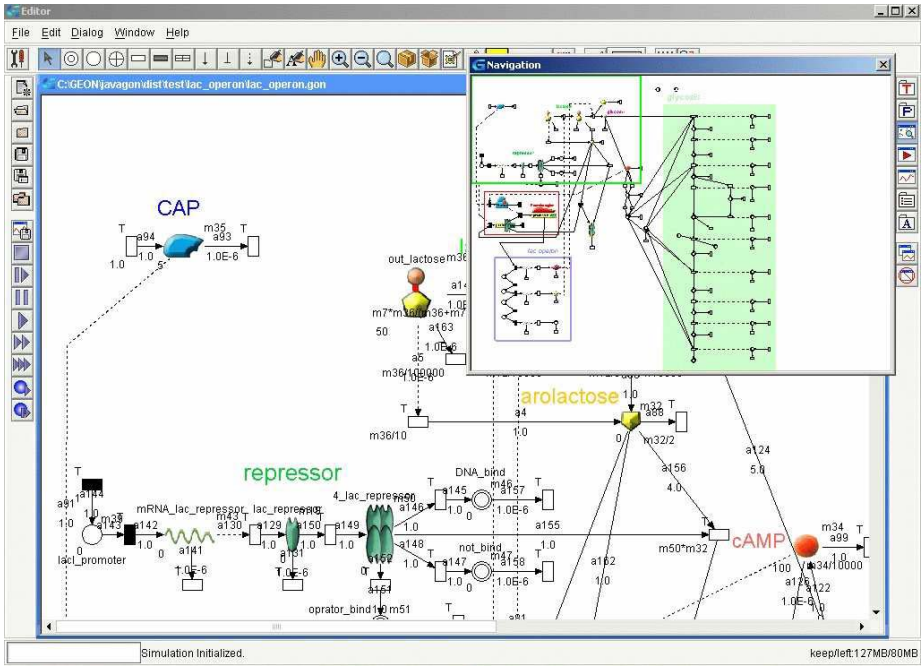

Figure 1: The interface of Genomic Object Net (Nagasaki et al. 2003)

2 Multi-Level Models in Systems Biology

When the term simulation is used in the context of systems biology most often it refers to continuous systems modeling and simulation (de Jong 2000). A series of simulation tools for continuous systems modeling and simulation in general and systems biology applications in particular exist, e.g. GEPASI (Mendes 1993), PROMOT/DIVA (Ginkel et al. 2003). Continuous models reflect nicely what is measured in cellular biology. Small samples of cell cultures are analyzed by extracting the DNA, enzymes or metabolites and quantifying the concentration of the respective species over time. This type of model emphasizes a continuous, deterministic, macro perception on cellular systems.

However, other approaches interpret cellular systems as being composed of vast amounts of entities, each of which has a state and an individual behavior pattern. Their activities are triggered by discrete events, like the arrival and release of interacting species, or by the time flow, like the time required for intra-molecular rearrangements. This perception suggests using other modeling approaches (Strohman 2000) like a discrete event modeling approach, which has been applied to the dynamics of cellular systems already more than 20 years ago, e.g. by Bernard Zeigler (Zeigler 1981). One of the classical approaches in Systems Biology, which has been developed by Gillespie (Gillespie 1976), is based on discrete event simulation as well, even though it is best known for its stochastic modeling and simulation. It has lead to several extensions and refinements during the 90s. Among them are some that turn from the macro perspective of the analyzed system to a micro perspective where individual entities are described.

Individual-based models consist of multiple homogeneous entities, which do not interact directly but via the macro level. The macro-level model contains information about the number of individuals and other information about groups of individuals. The macro level is not restrained to simply aggregate the information of the micro level. It might have variables and a behavior of its own. In the later case not only upward but also downward causation of biological systems can easily be modeled (Campbell 1974). Individual-based models can naturally be employed to describe phenomena in Systems Biology, e.g. on the macro-level the cytoplasma or bulk solution keeps track of the concentrations and changes of concentrations and the micro-level comprises individual species like DNA, enzymes, metabolites etc, with their individual states and behavior pattern. If more than two levels of organizations are considered whose entities are not restricted to indirect interaction via the macro model, we arrive at the more general concept of multi-level models (Uhrmacher and Swartout 2003).

3 Visualization Techniques in Systems Biology

In systems biology visualization techniques are used throughout all phases of modeling and simula-

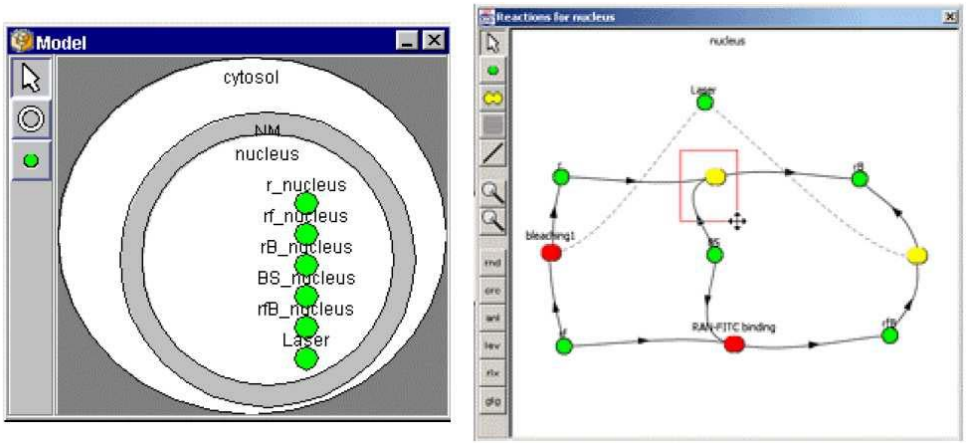

Figure 2: Model design with Virtual Cell (Schaff et al. 1997)

tion, i.e. to support the model design, the monitoring during simulation and the analysis of the achieved results (Allen et al. 2003). In the following we will concentrate on the first phase: designing or, to be more specific, representing the model and its structure.

Some modeling formalism lend themselves directly for a translation into graphical presentations. In this category formalism for discrete event systems modeling, e.g. Petri Nets or State Charts, and continuous systems modeling, e.g. Block Diagrams or Bond Graphs, belong equally. They are used by simulation systems to support the modeler in designing new models and are often provided as an alternative to non-graphical modeling languages. It is important to note that simulation systems support different types of users. Getting acquainted with an abstract modeling formalism holds little appeal to many users. If simulation systems are aimed at specific application areas, e.g. network simulation, or manufacturing simulation, the visual model can be based on metaphors and components libraries specific to this application area. Formal modeling formalisms and languages, on the one hand and pre-defined components, on the other hand, obviously address the need of different types of users.

To accommodate both types of users, two different avenues in supporting modeling via graphical interfaces are followed in System Biology.

- One is to enrich a general graphical modeling formalism with symbols from the application area. E.g. Genomic Object Net (Nagasaki et al. 2003) combines the graphical formal representation of Hybrid Petri Nets with icons, that represent e.g. DNA, mRNA, or processes like transcription, to visualize the semantics of certain places and transitions (figure 1). These approaches inherit from the adopted general modeling formalism, that they are expressive, semantically unambiguous, extendable, and supported by simulation engines.

- The second approach is to base the graphical modeling on metaphors and notations commonly used in the application area and thus, to obviate the need for defining the model in differential equations or other formal abstract languages. E.g. interaction diagrams are widely used in Biology and Chemistry. However, they lack the required expressiveness, semantic unambiguity, and consequently direct support by simulation engines. Their derivatives, like the Kohen diagram, have inherited these problems (Kitano 2003). To be executable by simulation engines, they have to be extended and refined. In Virtual Cell (Schaff et al. 1997), a user defines a model of a cell by defining compartments and substrates, relating substrates to compartments (figure 2, left hand side), and interrelating substrates, the former is based on an interaction diagram variant (figure 2, right hand side).

One problem all modeling and simulation systems share: to display "large scale models", so that users can explore details without losing orientation in the model.

For this purpose Genomic Object Net already provides a technique from information visualization referred to as overview and detail. As can be seen in Figure 1, the user interface is divided into two windows. Whereas the larger window allows the user to explore his range of interest, the smaller window serves as a coarse-grained overview of the entire model. Thus, also a larger model can locally be explored in detail without losing the global orientation. However there are many more information visualization techniques, that can be used to reduce complexity and dimension of model visualizations in systems biology.

Generally, when faced with complex models, the visualization seeking mantra introduced by Shneiderman: overview first, zoom and filter, then details-on-demand, should be employed to provide different views with different granularity (Shneiderman 2001).

4 The Challenge of Multi-Level Model Visualization

Multi-level models integrate the different traditional views in modeling systems, i.e. as functional models, as networks of interactions, and as hierarchical composition of models.

At the lowest layer we find functional models of individuals. To define the interaction between models, models are equipped with ports. In modeling and simulation the distinction between system and its environment is crucial, so model components can be grouped to form more complex models, so called composite or coupled models that interact with their environment via their input and output ports. Most modeling formalism assume a strong composition, i.e. one model component belongs only to one coupled model. To support a successive hierarchical construction, the property of being closed under composition is crucial (Zeigler 1996).

Thus, as structuring elements we have the network perspective, who interacts with whom, and the composition. Both we also find in continuous macro models. A multi-level model contains sub-models that describe single enzymes, and sub-models that describe enzyme populations. So unlike hierarchically structured continuous models which typically comprises a medium number of heterogeneously structured sub-models, multi-level models are faced with representing different populations. Each of its sub-models might comprise 1000s of homogeneously structured individuals. The overall structure of the model might become rather unbalanced, as at the higher level of organization a couple of heterogeneously structured models (same as in the continuous realm) interact, however, one of which might contain several 1000s sub-models.

Thus, one perspective on the model will hardly suffice to visualize the model structure in a compact manner and to enable users to manipulate rapidly the model structure. Therefore, different perspectives shall be supported. At least the functional, the network, and the composition perspective should be distinguished.

5 Modeling the Tryptophan Synthase in James

Based on the multi-level model of the tryptophan synthase we will illustrate the visualization techniques and their use in supporting multi-level models in Systems Biology. The visualization techniques are implemented as part of a Graphical User Interface which is currently under development for the simulation system JAMES (A Java based Agent Modeling Environment for Simulation) (Uhrmacher et al. 2000).

JAMES (A Java based Agent Modeling Environment for Simulation) (Uhrmacher et al. 2000) has been developed for simulating multi-agent systems. Systems are interpreted as communities of interacting autonomous entities, with the ability to adapt their behavior, interaction, and composition patterns. JAMES is based on the formalism DYNDEVS (Uhrmacher 2001) which adds reflection to DEVS (Zeigler et al. 2000) capturing the notion of self aware and self manipulating entities. The model design supports a hierarchical, compositional construction of models. It distinguishes between atomic and coupled models. Atomic models are equipped with input and output ports by which they communicate with their environment. Their behavior is defined by transition functions, an output function, and a time advance function which determines how long a state persists "per se". DEVS models can be interpreted as time triggered automata and thus graphically represented as STATECHARTS, for a detailed discussion see e.g. (Schulz et al. 2000). A coupled model is described by a set of component models, which may be atomic or coupled, and by the couplings that exists among the components and between the components and its own input and output ports. Thus, the structure of coupled models forms a compound directed graph (see section 6).

The application example, i.e., the tryptophan synthase, catalyzes the final two reactions of the biosynthesis of tryptophan. In bacteria it exists as a nearly linear $(\alpha\beta)_2$ complex. Each α subsystem catalyses the cleavage of indole 3-glycerol phosphate (IGP) to produce indole and D-glyceraldehyde-3-phosphate, and each β subsystem produces tryptophan from L-serine and the channeled indole. The α and the β subsystem are connected by a largely hydrophobic tunnel. The α subsystem transforms IGP to indole and D-glyceraldehyde 3-phosphate (GAP). The first is forwarded to the β subsystem whereas the latter is released into the bulk solution. Its functionality is hampered by glycerol 3-phospate (G3P).

In JAMES, the entire tryptophan synthase model is described as a multi-level model (Degenring et al. pear). The macro level contains models that describe the state and dynamics of the different populations of the bulk solution. The macro models responsible for the indole, the serine, the IGP, and the G3P interact with the "micro model" responsible for the synthase. The former keep track of the amounts of substrates, products and enzymes and defines the behavior at the level of concentrations and collision probabil-

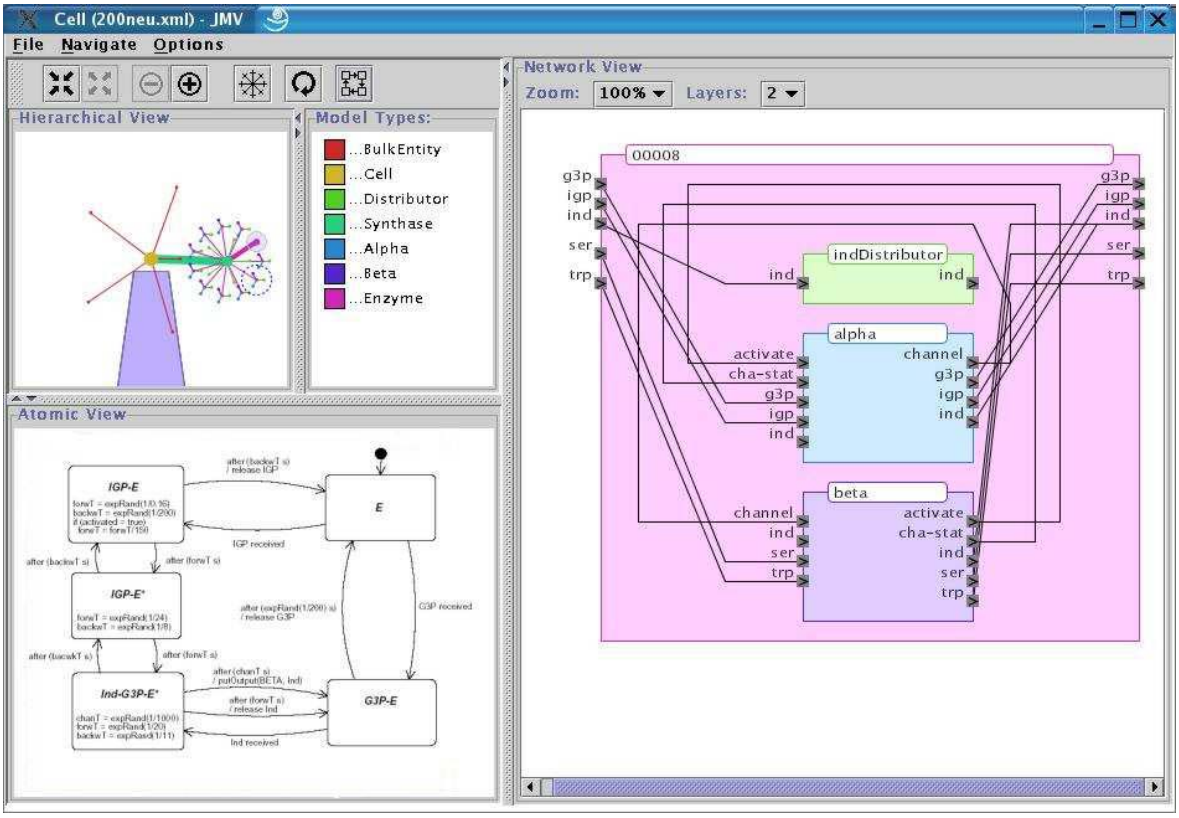

Figure 3: Screen shot showing the three different perspectives

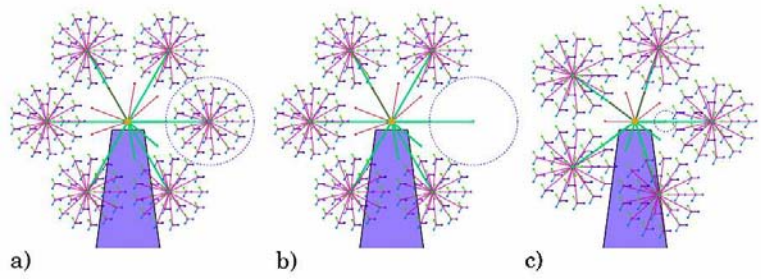

Figure 4: Interactive Folding.

ity. The micro model synthase contains thousands of models each of which describes a single enzyme, in upper left corner. As we are interested in the role, the channel plays in the tryptophan synthase, we define the enzyme model to consist of two different sub-units, i.e. alpha and beta, which communicate via the channel, see figure 3 on the right hand side.

The behavior of each sub-unit is modeled as discrete transitions from one state to the other. State changes might be triggered by the arrival of metabolites or by the flow of time, see figure 3 in the lower, left corner. The overall model structure is highly unbalanced, one of the children has more than 200 children, see figure 3, in the upper left corner. In the following we will discuss some of the realized features.

6 Employed Visualization Techniques

The visualization has to be divided into different windows referring to the different perspectives of the model. The three different perspectives (see section 4) are combined to visualize models developed with JAMES: the atomic view, the network view and the hierarchical view (figure 3). This reflects the traditional distinction between functional, network, and hierarchical modeling and reduces the complexity and dimensions of each individual visualization. Thereby, the hierarchical structure of the model acts as the mental map, as it guides the user up and down

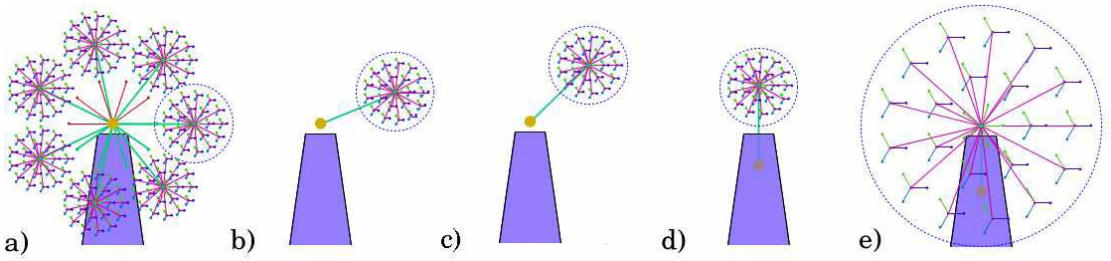

Figure 6: Zooming in based on sinks.

the different organizational levels. Only the compositional structure is shown and the user can interactively select an area of interest. The selected area is shown in a more detailed second window: the network view which shows the interaction structure of the selected model. If an atomic model is selected, its internal structure will be displayed in the functional view.

Since a model in JAMES can comprise a plethora of models at different compositional layers, we decided to use a radial techniques to present the hierarchical structure. Those techniques allow to display the entire hierarchical structure even with huge numbers of nodes. Among those techniques, the visualization technique RINGS (Ringed Interactive-Navigation Graph System) (Teoh and Ma 2002) exploits more efficiently the limited screen space than other radial techniques (figure 5). In addition, its ringed circular layout provides good support for the mental map of the user. In RINGS a parent node is placed in the center of all its child nodes. The child nodes are placed in concentric rings around the parent. Thereby nodes with many children are placed on outer rings, those with fewer ones on inner rings. For a full description of the layout algorithm of RINGS, see (Teoh and Ma 2002).

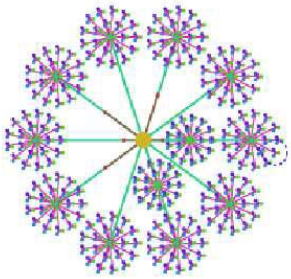

Figure 5: RINGS: Ringed Interactive-Navigation Graph System

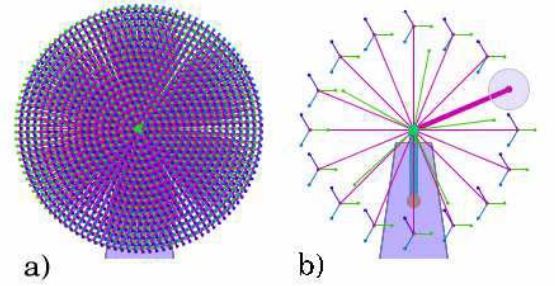

Figure 7: The functioning of a children sink

Nodes with many children occupy a lot of space. To more easily explore the structure of the model, appropriate information hiding techniques have to be integrated. Sub-structures can be interactively folded, which leaves more space for the remaining sub-trees. A folded sub-tree is shown as a blue circle within the window (Figure 4).

Another visualization concept are sinks where an explicit container collects the currently not visible entities. Two different types of sinks have been implemented.

- To zoom in, the sub-tree of interest can be selected. Thereby, all other sub-trees fade away into the sink, the root of the hierarchy moves visually into the sink while the root of the selected sub-tree moves gradually into the center. Now the subtree occupies the entire window. The sink is transparent and shows the predecessors of the sub-tree (Figure 6). The process is of course reversible.

- Using sinks, the number of displayed children can be reduced, as well. Here, the sink is moved interactively over the children that shall not be displayed (Figure 7). Different types of sinks can be used for "removing" children of different types.

The above techniques have been used in the context of exploring the hierarchical structure of the model and are features of the window HIERARCHY VIEW.

The window NETWORK VIEW utilizes the technique developed by Sander (Sander 1996). His approach is based on compound structured graphs. Whereas it clearly emphasizes the interaction structure between entities, the composition structure still remains visible. This facilitates the orientation of the user. However, as the composition hierarchy is

shown in a separate window and the system might comprise different composition layers with a plethora of components, it is important to be able to hide information on demand. Therefore, the depth of shown composition levels can be determined interactively.

If in the NETWORK VIEW an atomic model is selected, it is shown in the third window, the ATOMIC VIEW. Here the common Statecharts by Harel (Harel 1987) are used, to represent the functional level of the model. What is currently not visualized is the ability of a model to access its own structure, i.e. to change the composition, interaction, and behavior pattern of model.

7 Conclusion

The current work has concentrated on visualizing the structure of an existing model. The visualization of a multi-level model has to consider different perspectives, the atomic view, the network or interaction perspective and the hierarchical or compositional perspective. Whereas the atomic view could be realized by classical visualization techniques, for the network and hierarchical perspective more advanced visualization techniques have been employed due to the high number of models and interactions. The focus of the work has been on exploring the hierarchical structure of a model. A special problem for applying existing visualization techniques has been the typical, rather unbalanced structure of multi-level models in Systems Biology. To support the user in the analysis of, and orientation in the model structure, the presentation technique RINGS has been equipped with interaction techniques like folding, and installing sinks. The work presented is only a first step. Specific interaction features for the network perspective has still to be added, same as reflecting interaction facilities across the different views.

The current visualization is based on existing models. The next step will be to analyze the developed concepts and to adapt them to help fostering new multi-level models in Systems Biology. The visualization of the model structure has still to be connected to the running simulation, to animate the development of the structure. This is not trivial, due the complexity of the models and due to the possibly changing model structures in JAMES. Thus, components and interactions might appear and disappear, and components might even move through the model structure.

Multi-level models promise a more flexible approach toward the understanding of cellular systems. However, they also provide new challenges for modeling, simulation, and visualization techniques, alike – which we have just started to address.

References

Allen, N., C. Shaffer, N. Ramakrishnan, M. Vass, and L. Watson (2003, Dec.). Improving the development process for eukaryotic cell cycle models with a modeling support environment. *SIMULATION: Transactions of the Society for Modeling and Simulation, 79*(12), 674–688.

Campbell, D. (1974). Downward causation in hierarchically organized biological systems in studies in the philosophy of biology. In F. Mayala and T. Dobzhansky (Eds.), *Studies in the Philosophy of Biology: Reductionism and Related Problems*, pp. 179–186. New York: Macmillan Press.

Davis, P. (2000). Exploratory analysis enabled by multiresolution, mulitperspective modeling. In *Proceedings of Winter Simulation'2003*, pp. 293–302. SCS.

de Jong, H. (2000). Modeling and simulation of genetic regulatory systems: A literature review. Technical Report 4032, INRIA.

Degenring, D., M. Röhl, and A. Uhrmacher (to appear). Discrete event, multi-level simulation of metabolite channeling. *BioSystems*.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics 22*, 403–434.

Ginkel, M., K. A., T. Nutsch, R. Rehner, and E. Gilles (2003). Modular modeling of cellular systems with promot/diva. *Bioinformatics 19*(9), 1169–1176.

Harel, D. (1987, June). Statecharts: A visual formalism for complex systems. *Science of Computer Programming 8*(3), 231–274.

Kitano, H. (2003). A graphical notation for biochemical networks. *Biosilico 1*(5), 169–176.

Luscombe, N., D. Greenbaum, and M. Gerstein (2001). What is bioinformatics? a proposed definition and overview of the field. *Methods Inf Med 40*, 346–358.

Mendes, P. (1993). GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Computer Applications in the Biosciences 9*(5), 563–571.

Nagasaki, M., A. Doi, H. Matsuno, and S. Miyano (2003). Genomic object net: a platform for modeling and simulating biopathways. *Applied Bioinformatics*.

Nocke, T., U. Boehm, H. Schumann, and M. Flechsig (2003). Information visualization supporting modelling and evaluation tasks for

climate models. In *Proceedings of Winter Simulation'2003*, New Orleans.

Sander, G. (1996). Layout of compound directed graphs. Technical Report A/03/96, Universität des Saarlandes, FB 14 Informatik.

Schaff, J., C. Fink, B. Slepchenko, J. Carson, and L. Loew (1997). A general computational framework for modeling cellular structure and function. *Biophys. J.*, 1135–1146.

Schulz, S., T. Ewing, and J. Rozenblit (2000, March). Discrete event system specification (devs) and statemate statecharts equivalence for embedded systems modeling. In *Proceedings of the 2000 IEEE Conference and Workshop on Engineering of Computer-Based Systems*, Edinburgh, Scotland.

Shneiderman, B. (2001). Supporting creativity with advanced information abbundant user interfaces. In *Human-Centred Computing, On-Line Communities, and Virtual Environments*, London, UK. Springer.

Slavik, P., M. Gayer, F. Hrdlicka, and O. Kubelka (2003). Problems of visualization of technological processes. In *Proceedings of Winter Simulation'2003*, New Orleans.

Strohman, R. (2000). Complexity - organization becomes cause in the matter. *Nature Biotechnoloy 18*, 575–576.

Teoh, S. and K.-L. Ma (2002). RINGS: A technique for visualizing large hierarchies. In *Graph Drawing*, London, pp. 268–275. Springer.

Uhrmacher, A. (2001). Dynamic Structures in Modeling and Simulation - a Reflective Approach. *ACM Transactions on Modeling and Simulation 11*(2), 206–232.

Uhrmacher, A. and W. Swartout (2003). Agent-oriented simulation. In M. Obaidat and G. Papadimitriou (Eds.), *Applied System Simulation*. Kluwer Academic Press.

Uhrmacher, A. M., P. Tyschler, and D. Tyschler (2000). Modeling Mobile Agents. *Future Generation Computer System 17*, 107–118.

Wolkenhauer, O., H. Kitano, and K.-H. Cho (2003). An introduction to systems biology. *IEEE Control Systems Magazine 23*(4), 38–48.

Zeigler, B. (1996). A Framework for Modeling and Simulation. In D. Cloud and L. Rainey (Eds.), *Applied Modeling and Simulation: An Integrated Approach to Development and Operation*, Chapter 3. US Air Force.

Zeigler, B., H. Praehofer, and T. Kim (2000). *Theory of Modeling and Simulation* (2nd ed.). London: Academic Press.

Zeigler, B. P. (1981). Simplification in biochemical systems. In L. Segel (Ed.), *Mathematical Models in Molecular and Cellular Biology*. Cambridge University Press.

ACKNOWLEDGMENTS

The research on the simulation system JAMES is supported by the DFG (German Research Foundation).

AUTHOR BIOGRAPHIES

SUSANNE BIERMANN holds a MSc in Computer Science from the University of Rostock. She is currently participating at a postgraduate studies program at the Imedia Academy in Providence / USA. She is particularly interested in data visualization, information architecture, information technology, design and information visualization.

ADELINDE M. UHRMACHER is an Associate Professor at the Department of Computer Science at the University of Rostock and head of the Modeling and Simulation Group. Her research interests include the development of modeling and simulation methods, particularly for agent-oriented modeling and simulation, and the applications of these methods. Current projects centre on developing a component-based simulation framework to support the design and testing of software agents and on exploring the role of multi-level modeling and simulation in Systems Biology. She is co-editor-in-chief of the Simulation - Transactions of the SCS.

Heidrun Schumann is heading the Computer Graphics Research Group at the Institute of Computer Science at the University of Rostock. Her research profile covers Information Visualization and Visual Data Mining, Mobile Interfaces, and Rendering as well as Image Display. She was heading the cross- institutional research group "Movi-Visualization of Multimedia Information on Mobile Computer Systems," supported by the German Research Society - DFG. Current research projects, supported by research institutions or industry focus e.g. on the development of scalable frameworks for information visualization, on the visualization of climate data and human health data as well as on problems of user defined image display and transmission.