TOWARDS A TUTORING SYSTEM FOR BIOLOGY STUDENTS LEARNING MODELING AND SIMULATION

Géraldine Ruddeck[1]**, Alke Martens**[1]

[1]University of Rostock, Department of eLearning and Cognitive Systems
Albert-Einstein-Str. 21, 18055 Rostock, Germany

geraldine.ruddeck@uni-rostock.de(Géraldine Ruddeck)

Abstract

From the wet-lab to the dry-lab, Biologists have a big step to do – and little help. Conventional lectures can be completed by the use of a tutoring system to help Biology students with the learning of modeling and simulation of biochemical processes. In order to develop such a tutoring system, we first defined a modeling and simulation workflow aiming to minimize the sources of error for novice modelers and to give a structure to a tutoring system. We then identified teaching objectives through an expert survey. These teaching objective describe the most critical issues to be assimilated by the students to improve the modeling skills they need. With a user survey, we identified the mental model of the target group regarding biochemical processes and established the requirements on the structural and visual design of the tutoring system. Thus, we laid the foundation stone for the development of a tutoring system for Biology students learning modeling and simulation.

Keywords: systems biology, workflow, eLearning, mental model

Presenting Author's Biography

Géraldine Ruddeck is a Ph.D Student at the University of Rostock in the domain of eLearning and Cognitive Systems. She is part of the Research Training Group "Integrative Development of Modeling and SImulation Methods for RegeneratIve Systems" (dIEM oSiRiS), an interdisciplinary research group involving Biologists and Computer Scientists.

1 Introduction

Nowadays, in many research fields of Biology, experiments in the lab are no longer sufficient to make a step further into knowledge. How to predict the outcome of particularly complex processes like different treatments on a cancerous tumor? And how to find out which phenomena are involved in complex signaling pathways inside a cell, involving hundreds of proteins, most of them not measurable? For these examples and many more, computer experiments based on modeling and simulation are necessary. Scientists in the field of Biology or Chemistry have to learn the basics of modeling and simulation in order to work together with Modelers (e.g. Computer Scientists). But there is a big step to do from the wet-lab to the dry-lab – implicating acquisition of skills they are not used to or not even willing to learn.

Consulting Systems Biologists who teach modeling and simulation for Biologists, we found out that most Biology students properly fear differential equations and experience a real cognitive blockade when seeing one. This problem is addressed even in Biology journals like in [1]. Usual lectures are not sufficient to overcome such fundamental problems and an adaptive and long-term training directly on the target medium using a tutoring system would be a powerful complement in the curriculum of Biology students learning modeling and simulation of biochemical processes.

Strangely enough, there exists to the best of our knowledge no such tutoring system. On the one hand, plenty of systems, games and serious games, (e.g. Genomics Digital Lab [2] or Sims [3]), use simulation. Some of these systems have teaching aims, but the subject of the potential teaching is never the modeling itself. On the other hand, there are many modeling and simulation tools, but none with teaching aims related to the needs of Biologists. A survey at the Systems Biology working group of the University of Rostock (described in Section 3) showed that 88% of the Modelers use exclusively Matlab or some Matlab clones (e.g. Octave [4] and Scilab [5]) for their modeling and simulation works. The user interface of these tools is code oriented and requires skills in programming to handle them. For this reason, they are not very suitable for Biology students. For teaching purposes, the simpler syntax and quick examples of Sycomore are preferred. Standard in Systems Biology are also Copasi ([6]) and CellDesigner ([7]). These very specific tools offer a modeling and simulation environment without "apparent" programming code. The model editor JDesigner2 from the Systems Biology Workbench ([8]) even encomprises a very easy and user-friendly graphical user interface. Still, these are modeling and/or simulation tools and even if some have made a (more or less big) step towards user-friendliness, none has the intention of teaching or helping beginners towards modeling and simulation. Summarizing, there is a need of a tutoring system intended for Biology students learning modeling and simulation.

In this paper, we will first present a modeling and simulation workflow that is adapted to our purpose. There exists already all sorts of workflow, more or less detailed, focusing on some part or the other, but we need a common vocabulary when consulting experts, and the right granularity when designing contents. This makes the development of an appropriate workflow an important preliminary step for out intelligent tutoring system. In section 3, we then describe the expert survey we carried out in order to identify the key elements which Biology students should acquire in the training phase. However, an instructional curriculum design is far from enough for a successful tutoring system, and the representation of the content is essential to get and keep the learners attention. Therefore, we tried to identified and understand the mental model that Biology students have of biochemical processes. The last section describes the user survey carried out to this purpose and its results.

2 Workflow

One can find a plethora of workflows in the literature, as almost every book dealing with some sort of modeling or simulation presents one in its first few pages (see for instance [9, 10, 11, 12, 13]). Depending on the purpose of the author, some focus on the modeling part only, other on the simulation, fewer address the wet-lab experimenting. Klipp's "Systems Biology in Practice" ([9]) is a reference for most people from all horizons beginning with modeling in Systems Biology. The author presents in the general introduction a process of model development in 8 phases: 1. Formulation of the problem, 2. Verification of available information, 3. Selection of model structure, 4. Establishing a simple model, 5. Sensitivity analysis, 6. Experimental tests of the model predictions, 7. Stating the agreements and divergences between experimental and modeling results, 8. Iterative refinement of model. This workflow shows in the first three phases the importance of a good preparation rather than "ad hoc" modeling. The iteration is also a crucial part: Following the principle of parcimony, one should begin with a simple model and then add with each iteration as many details as needed – and no more. Law's "Simulation Modeling and Analysis" ([12]) is also a major reference in the simulation field and puts the modeling part as described above into the broader context of a "simulation study". The steps are: 1. Formulate problem and plan the study, 2. Collect data and define a model, 4. Construct a computer program and verify, 5. Make pilot runs, 7. Design experiments, 8. Make production runs, 9. Analyze output data, 10. Document, present, and use results. Step 3 and 6 are validation checks on the conceptual and programmed model. The wet-lab part is addressed very succinctly in step 2, put together with the development of the conceptual model.

In most descriptions, a model has to be developed for a specific simulation to obtain the answer to the original problem. The model isn't independent but designed to address the requirements of the specific simulation. For instance, Velten defines in [10] a "modeling and simulation scheme" in five phases: 1. Definitions, 2. Systems Analysis, 3. Modeling, 4. Simulation and 5. Validation. Which means that a verification and validation of the model is not intended, and that the final validation of the simulation results would be sufficient. This is incorrect in our point of view and after Law's workflow, as simulation experiments are to be made on a potentialy invalid model. Furthermore, this solid interlinking between model and simulation experiments exludes the possibility of reusing a model for different simulation experiments or, on the contrary, to test different generic models on one simulation experiment. Even if the global idea is maintained throughout all these workflow propositions, the inconsistent distinction between the modeling and the simulation part makes it difficult to use them as a foundation for our tutoring system.

Thus, we developed a workflow based on the experience gathered at our University and on a throughout literature review. This workflow distinguishes the three phases addressed in the previous workflows (see figure 1): the wet-lab experiment, the dry-lab modeling and the dry-lab simulation. On the one hand, these phases are strongly interconnected, each one beeing the necessary precursor of the next phase. On the other hand, the phases are strictly separated by the results "lying" inbetween, as these are the result of a complete phase walkthrough, inclusive feedback loops and iteration. The arrows are therefore not picturing a linear course but suggesting either feedback loops and iterations whithin one phase, or backwards steps from one phase to the previous one.

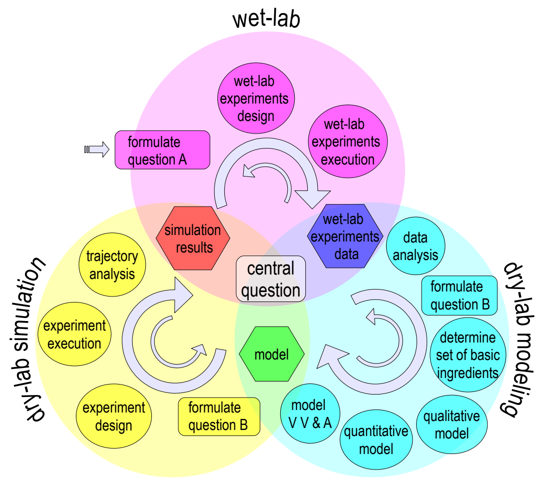

Fig. 1 Our workflow for modeling and simulation. The main phases (wet-lab, dry-lab modeling, dry-lab simulation) are separated by their concrete results. The arrows show not only the usual or ideal walkthrough direction, but also the omnipresent possibility of feedback loops, backwards steps and iterations.

As no experiment should ever been made without a clear purpose, a central question has to be formulated. Questions A, B and C are the refinements of this central question adapted to the current phase. For example, if one is interested in the propagation of the Yellow Fever, a possible central question could be: "Is a specific insecticide adapted to stop the propagation of the Yellow Fever?" Supposing that some variables are already known (e.g. incubation time or death rate of infected mosquitoes), a first wet-lab question (A1) could be "What influence has an insecticide on the mosquitoe population?". With these pieces of information, a model can be build following the question (B1) "Which influence has the insecticide on the human population?" This model can then be simulated to answer the question (C1) "Does the insecticide stop the propagation of the disease?" This first walkthrough may bring interesting results, but it considered very few parameters and didn't take into account e.g. the mutation of the virus or the side-effects of the insecticide on other animals or plants. This can be refined in further iterations or with feedback and back steps. Questions A, B and C are interconnected insofar as they derive from the central question, which assures the usability of the results for the respectively next phase.

This workflow has the advantage of giving a useful guidance structure for a tutoring system. We are now focusing on the modeling part and use therefore wet-lab results as basis. The modeling task ends when the model is ready to be used in simulation experiments. When a learner receives a modeling exercise, he can go through these steps (formulate question, define basic ingredients, make qualitative model, make quantitative model, make verification) and be sure to minimize the errors as he is following a best-practice workflow.

3 Teaching Objectives

As Biologists usually don't have to make models and dry-lab experiments all on their own but are collaborating with Systems Biologists or other Computer Scientists, there is no need to force Biology students into too much mathematical or modeling sophistications but better restrict to what is really useful in the first place for them and for their collaboration with Modelers. An expert survey was carried out involving 18 participants from the Systems Biology working group of the University of Rostock. All participants have experience in working with Biologists and two of them give lectures and exercise courses for modeling and simulation to both Biology and Computer Sciences students, in separate sessions. The survey took place as a semi-structured group interview, allowing high flexibility in the questioning as well as phases of free discussion between the attendees.

The expert survey confirmed the need for very basic skills and pointed out primary teaching aims for a tutoring system. It is interesting to point out that these most critical teaching aims can be classified into the categories 1 to 3 of Boom's taxonomy ([14]): knowledge, comprehension, application. The higher categories (analysis, synthesis, evaluation) were not addressed at all, both because knowledge and comprehension of modeling principles are quite sufficient to improve the collaboration and because Biology students don't have to become modeling experts.

The first requirement for Biology students is to understand the possibilities offered by dry-lab experiments but also their boundaries. The lack of motivation while learning modeling is, as the experts said, obvious. This is a general obstacle to any learning process. Motivation is a prerequisite condition to effective understanding and memorizing, and the first requirement by Bruns and Gajewski as cited in [15].

Another requirement was expressed concerning the general disposition of the Biology students towards Computer Sciences in general: the demystification of mathematics- or computer-based systems. Terms like ”fear”, ”afraid” and ”mental block” were often used by the attendees when describing the usual reaction of Biologists towards e.g. simple differential equations. This attitude obviously hinders the concerned people from even trying to understand the system or differential equations, and so give them no chance to improve their skills or change their opinion.

On a more concrete level, using a common vocabulary with the Modelers is a fundamental objective from the ”knowledge” category. Learning basic structures (as described e.g. in [16]) and beyond that, being able to use them and to identify them in a given model are a first step towards modeling. Usual lectures commonly see to that point to a certain extend but a tutoring system could offer some material to train more concretely and make students really familiar with these basics.

One step further, understanding the meaning of specific parameters in specific reactions and being able to read and interpret differential equations are the main goals of the ”comprehension” category. The comprehension issues require long-term training, because the learner doesn’t have just to learn something by heart but to really get a feeling for it. Complementary interactive units where the learner can experiment basic components by himself with immediate feedback are appropriate for such teaching goals.

The ”application” category contains two major points, both related to utilization of the already existing knowledge and communication within the community: the search and use of ontologies and the ability to recreate a model from a scientific publication. Understanding and using the standard exchange format SBML belongs naturally also to this category. Creating an own model from scratch is also an advanced skill and can be limited to rather simple models. These skills are more complex ones and therefore require more extensive exercises.

4 Biology Students’ Mental Model of Biological Processes

Knowing what to teach is not enough, the ”how” is determinant. As Biology is an empirical and explorative field, Biology students are used to problem solving approaches. Choosing the elements of relevance for a particular situation or determining the borders of the system to consider are also the common way of working in this field and familiar to the students. Therefore, the instructional strategy ”anchored instruction” is best suited to meet the previous knowledge of the target group and the pursued training objectives. The main idea there is to formulate exercises in form of problems with an appropriate complexity, to let the learner solve the problem and to propose similar exercises but with different contexts for a better knowledge transfer. This training form requires additionally modules for supporting self-directed learning, as the learner may need to look up some information while solving the problems.

In order to confirm this instructional choices and to determine the appropriated representation forms for the teaching and training content, a preliminary user survey was carried out. 30 students in Human Medicine and Medical Biotechnology at the University of Rostock, from the first to the last semesters, took part in the survey. The survey was a short questionary about representation forms of a system. We took the chemical synapse as example of a system, as it is sure to be known from all students and the answers wouldn’t depend on the previous knowledge of the person on the subject.

Description:

The survey was send as a pdf formular containing only text fields and radio buttons for the answers. It was important to make the answering easy, as we couldn’t assume the respondents to have computer skills above usual internet surfing. The survey contained only four pages and short questions, and was anounced to take less than five minutes to fill in. This way, we intended to get quick and spontaneous responses, minimizing the chances that the respondent interrupts during the survey, lowers concentration or gets lost in too much reflection. On the first page were three representations of a synapse:

a) A text: a short text giving basic information about the elements playing a role and describing the sequence of events taking place.

b) A drawing: a picture showing the pre- and postsynaptic cells containing the same elements as cited in the text. The elements are represented in a simplified illustration of the reality (e.g. vesicles are circles with little dots for the neurotransmitters) and annotated. The events and their sequence are partially represented by arrows.

c) A reaction map: a schematic graphic where the cells are represented as compartments, the elements as entities and the events as reactions or modification of reactions between entities. This representation is very abstract.

On the following two pages of the survey, eight simple statements were formulated in the first person perspective or impersonnaly, and to be rated for each of the three representations into ”applies completely”, ”applies generally”, ”doesn’t quite apply” and ”doesn’t apply at all”. This subdivision in four marks constains the respondant to take position: As no ”medium” mark is available, the respondant has to give either a ”good” or a ”bad” mark. The three representations were pictured in small size at the top of each page as a reminder. The statements were formulated in a positive form as to avoid double negation confusion and to permit a straitforward conversion of the answers into numerical marks ("applies completely" = 3, "doesn't apply at all" = 0). The statements were:

1. When I read the description, I understand at once how the process works.
2. When I work with this representation, I can retreive information quickly.
3. The representation contains all information needed to understand the process.
4. It is clear to me, what is happening when.
5. The succession of events is clearly to see in this representation.
6. My idea of a synapse is like this representation.
7. If I had to explain to a fellow student how a synapse works, I would use this representation.
8. I like this representation.

On the last page, information about the semester and subject of the student were gathered, as well as optional personal suggestions.

Results:

Converting the appreciations into marks 0 ("doesn't apply at all") to 3 ("applies completely"), the easiest result is the overall mean mark of each representation for each statement through all the respondants (see figure 2). The surface delimitated by each curve on the radarchart gives an idea of the general mark of each representation. If the textual and pictural representation cover quite similar surfaces, the map representation underperformes clearly. Indeed, the overall mean mark is 2.2 for a) and b), and only 1.2 for c).

Statement 2 and 7 address ease of use, statements 6 and 8 the personal liking of the respondant. For these four statements, the pictural representation scores best and the map representation is clearly disliked.

Statements 1, 4 and 5 address the sequence order. The textual representation there scores best, and the map representation scores in 4 and 5 comparatively to the other statements better. It is here interesting to note that statements 1, 4 and 5 address the same issue (sequence order), but statement 1 is very vague ("I understand how it works"), statement 4 is more precise, but still formulated subjectively ("It is clear to me, what is happening when.") and statement 5 is formulated objectively ("The sequence order is clearly expressed in the representation"). As statements 4 and 5 where on separate pages, this similitude was note clearly to spot for the respondant. From the intuitive understanding to the precise evaluation of the representations, drawing b) lost significanlty points (from 2.2 to 1.4) and model c) won significanlty points (from 1.2 to 1.8). The difference between the subjective and objective precise evaluation (statements 4 and 5) is much smaller but still interesting to notice: from 1.8 to 1.6 for b) and from 1.6 to 1.8 for c).

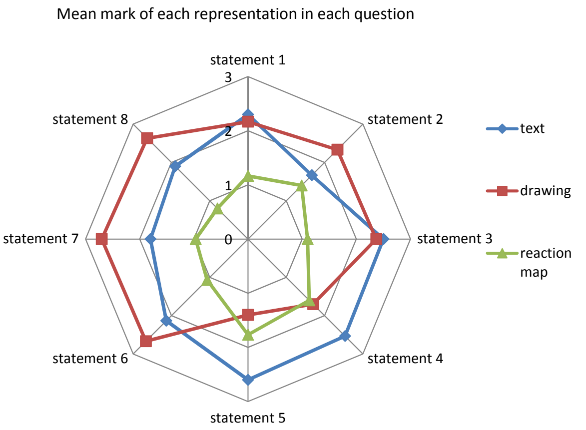

Fig. 2 Radarchart picturing the mean mark of the three representations for each statement.

A look at the repartition of the occurences of "good" and respectiveley "bad" marks, pictured on Figures 3 and 4, shows clearly the feeling of the students about the three representations. Apart from the lack of temporal information in statements 4 and 5, the drawing representation received less than 8% of "doesn't apply". On the contrary, the map representation scored almost two third of bad marks. The text representation scored globally good, although it is not very handy (statements 2 and 7) and obviously not the kind of representation that students have in mind when thinking of a synapse (statements 6 and 8).

Interpretation:

Correlating with the traditional representations in biology, it is no surprise to find out that the students are most at ease with text and reality-related drawings. Although the textual representation has some drawbacks in matter of easy and handy usage, its precision was praised and students are used to it. However, textual representations show important deficites in picturing more complex processes: Simultaneity and parallel sequences are difficult to render in a linear narration.

Despite its imprecisions, the drawing was closest to the mental model that students have of a synapse. The intuitive understanding of the representation (statements 1 and 2) as well as the personal evaluation (statements 6 to 8) attest of it. The student's affinity for the drawing made them even overlook the imprecision of the representation, as only 10% of the respondants recognized that the drawing doesn't contain all the needed information (statement 3). As some respondants noted, the best representation would be a mixture of the text and the drawing, which complement each other very well.

The third representation (model map) was for most of the students unintelligible. In the free comment part, some students quoted it as "too imprecise", "(very) confusing" or "illogical". One described it as "a simplification of picture b)", which is not really correct: the

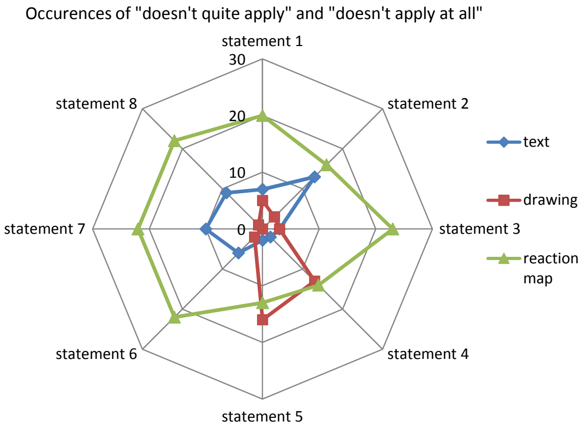

Fig. 3 Occurences of ”bad marks” to each statement and for each representation.

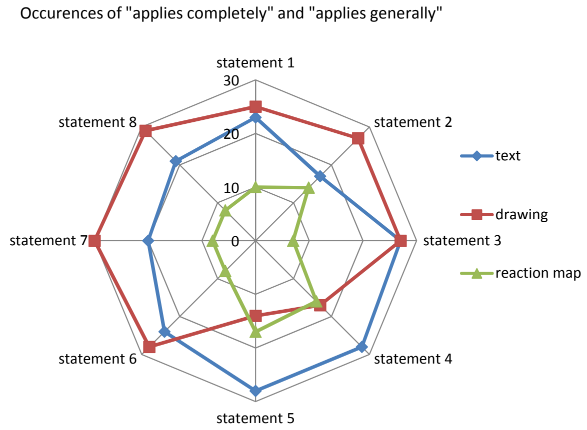

Fig. 4 Occurences of ”good marks” to each statement and for each representation.

visual elements are standardized in c) and don’t refere to their real appearance, but the process represented was exactly the same. The great misunderstanding of representation c), and even sometimes the unwillingness to understand it, are clearly to notice for most students. However, some students who understood it rated it consequently for almost all statements with highest marks.

From this survey, we can conclude that most students are not used to abstract, process-oriented model maps and prefere drawings with visual and spatial clues. Text descriptions, although unhandy and slower to apprehend, are mostly considered as a necessary complement to the drawing. From our work with Biologists, we know that most are very reluctant to mathematical issues and even more to programming code. We didn’t integrate this aspect into the survey in order to dissimulate our real goal and to obtain more accurate results on the other representation forms. It is now clear that a tutoring system for Biology students learning about modeling has to partially blind out and/or picture in an understandable way the most technical issues (e.g. program code, ontology entries), privilege graphical representations with visual and spatial clues, and the use of text (e.g. in descriptions, help, or as a notebook for the learner). The technical elements contained in the teaching objectives (e.g. basic differential equations, ontologie researches or SBML representation) have therefore to be introduced carefully and, when possible, explained in a way that is compatible with the mental model of the learners. Considering for example the understanding of basic structures of biological models, these insights into the mental model of our target group permit us to design appropriate contents in form of an ”interactive dictionary” in which the learner could see e.g. the mathematical description of a reaction, its graphical equivalent, a short textual explanation, a graphical representation of the evolultion of the entities’ concentration. A real understanding of how the reaction works could be then reached by giving the possibility to change the different parameters (e.g. initial values, rates, volumes) and to observe the impact on the outcome. Observing the changes in the outcome when modifying some parameters is a impressive experience to convince oneself that every part of the seemingly complex mathematical expression has a tangible sense and to understand the core of the process.

5 Conclusion

We presented here the foundation stone of a tutoring system for Biology students learning modeling and simulation. From our experience and the many workflows already existing for the modeling and simulation field, we developed an own workflow which can serve as a basis for the tutoring system. This workflow was validated by our experts in the same survey that enabled us to determine the most important teaching objectives to be achieved with the tutoring system. These teaching objectives address rather basic skills, but on a high level of acquisition. They can be put together into three major trends: communication with Modelers, understanding, recognition and use of basic structures of models, and scientific rigor.

The design of the tutoring system is also a key element to the success of the teaching. Thus, we identified requirements to take into account while developing tutoring systems for Biologists. Our current work with Biologists showed that our target group is used to work and interact on a visual basis and needs graphical and spatial clues to understand models. These are the reasons why the development of a graphical notation for Systems Biology is part of the current research ([17]). The preleminary survey we described confirmed this importance of an appropriate graphical representation and also pointed out the disposition to work with textual information. With these elements, it is now possible to design the teaching and training modules of the intelligent tutoring system and to integrate them into the modular architecture we parallely develop to promote content exchange, reuse and update.

6 References

[1] Steven Buckingham. Modelling in Biology - Don’t Be Shy of Maths. *Lab Times*, 04-2008.

[2] Spongelab. Genomics Digital Lab, 2009.

[3] Electronic Arts. Sims, 2000.

[4] James Rawlings and John Ekerdt. Octave, 2010.

[5] digiteo. SciLab, 2010.

[6] Virginia Bioinformatics Institute. Copasi, 2010.

[7] The Systems Biology Institute Tokyo. CellDesigner, 2008.

[8] Computational Systems Biology Sauro Lab. Systems Biology Workbench - Model Editor, 2010.

[9] E. Klipp. *Systems biology in practice: Concepts, implementation and application*. WILEY-VCH, Weinheim, 2005.

[10] Kai Velten. *Mathematical modeling and simulation: Introduction for scientists and engineers*. WILEY-VCH, Weinheim, 2009.

[11] Lawrence M Leemis and Stephen K Park. *Discrete-event simulation: A first course*. Pearson/Prentice Hall, Upper Saddle River, NJ, 2006.

[12] Averill M. Law and W. David Kelton. *Simulation modeling and analysis*. McGraw Hill series in industrial engineering and management science. McGraw Hill and McGraw-Hill, Boston, 3. ed. edition, 2000.

[13] Udo W. Pooch and James A. Wall. *Discrete event simulation: A practical approach*. CRC Press computer engineering series. CRC Press, Boca Raton, Fla., 1993.

[14] Lorin W Anderson. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, New York, complete ed., [nachdr.] edition, 2009.

[15] Michael Herczeg. *Einführung in die Medieninformatik*. Interaktive Medien. Oldenbourg, München, 2007.

[16] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, Chklovskii, D., and U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298, 22.10.2002.

[17] Nicolas Le Novère, Michael Hucka, and Herbert et al Sauro. The Systems Biology Graphical Notation. *Nature Biotechnology*, 06.10.2009.