

# Bridging the scales: semantic integration of quantitative SBML in graphical multi-cellular models and simulations with EPISIM and COPASI

Thomas Sütterlin<sup>1,2</sup>, Christoph Kolb<sup>2</sup>, Hartmut Dickhaus<sup>2,3</sup>, Dirk Jäger<sup>1,2</sup> and Niels Grabe<sup>1,2,\*</sup>

<sup>1</sup>Department of Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg,

<sup>2</sup>Hamamatsu Tissue Imaging and Analysis Center (TIGA), BIOQUANT, Heidelberg University and <sup>3</sup>Department of Medical Informatics, Institute of Medical Biometry and Informatics, University Hospital Heidelberg, 69120 Heidelberg, Germany

Associate Editor: Mario Albrecht

## ABSTRACT

**Motivation:** Biological reality can *in silico* only be comprehensively represented in multi-scaled models. To this end, cell behavioural models addressing the multi-cellular level have to be semantically linked with mechanistic molecular models. These requirements have to be met by flexible software workflows solving the issues of different time scales, inter-model variable referencing and flexible sub-model embedding.

**Results:** We developed a novel software workflow (EPISIM) for the semantic integration of Systems Biology Markup Language (SBML)-based quantitative models in multi-scaled tissue models and simulations. This workflow allows to import and access SBML-based models. SBML model species, reactions and parameters are semantically integrated in cell behavioural models (CBM) represented by graphical process diagrams. By this, cellular states like proliferation and differentiation can be flexibly linked to gene-regulatory or biochemical reaction networks. For a multi-scale agent-based tissue simulation executable code is automatically generated where different time scales of imported SBML models and CBM have been mapped. We demonstrate the capabilities of the novel software workflow by integrating Tyson's cell cycle model in our model of human epidermal tissue homeostasis. Finally, we show the semantic interplay of the different biological scales during tissue simulation.

**Availability:** The EPISIM platform is available as binary executables for Windows, Linux and Mac OS X at <http://www.tiga.uni-hd.de>. Supplementary data are available at <http://www.tiga.uni-hd.de/supplements/SemSBMLIntegration.html>.

**Contact:** niels.grabe@bioquant.uni-heidelberg.de

Received on August 8, 2012; revised on October 10, 2012; accepted on November 4, 2012

## 1 INTRODUCTION

Many systems biological models have been developed in the past years. The vast majority of these models focus exclusively on the sub-cellular or the cellular level, whereas only a minority are really multi-scaled (Andasari *et al.*, 2012; Hoehme *et al.*, 2010; Osborne *et al.*, 2010; Owen *et al.*, 2009; Munaron and Scianna, 2012). However, semantic linking of systems biological scales with models on sub-cellular, cellular and tissue level is a

prerequisite for a truly comprehensive *in silico* representation of biological reality (Scianna and Preziosi, 2012). Although intra-cellular molecular processes are frequently represented by ordinary differential equation (ODE) systems, on the multi-cellular level, more abstract descriptions of cell behaviour like process diagrams are used (Amirkroll *et al.*, 2008; Sütterlin *et al.*, 2009). Semantic linking connects both modelling approaches by flexibly mapping time scales, cross-referencing of model variables and by this linking of cellular states like proliferation and differentiation to molecular processes. Thereby, the mapping of time-scales enables model integration and re-use without the necessity of reparameterization.

Databases like BioModels provide a huge variety of mainly sub-cellular models for download (Le Novère *et al.*, 2006). Systems Biology Markup Language (SBML) is the most commonly used standard for this kind of quantitative models representing biochemical reaction or gene-regulatory networks. This standard is supported by >200 different software packages (Hucka *et al.*, 2003; Hucka and Le Novère, 2010) like CellDesigner (Funahashi *et al.*, 2003) and COPASI (complex pathway simulator) (Hoops *et al.*, 2006). However, almost none of these packages directly support building multi-scaled, multi-cellular tissue models and simulations, respectively. COPASI has previously been used in a hard-coded tissue simulation integrating TGF- $\beta$ 1 (transforming growth factor) signalling in epidermal wound healing (Sun *et al.*, 2009). A software workflow allowing flexible semantic integration of SBML models has to be established to make use of the valuable source of SBML-based models in a multi-scaled tissue modelling context. Such a workflow has to hide the technical complexity of realizing multi-scaled models and particularly model simulations to address a rather biologically than computationally skilled community. The software CompuCell3D (Andasari *et al.*, 2012; Swat *et al.*, 2009) allows to directly reference and simulate SBML-based models within a multi-cellular tissue simulation. However, CompuCell3D models are manually built with extensible mark-up language (XML)-files and Python scripts requiring dedicated knowledge of these computer science-related concepts. To our knowledge, there is no intuitive and user-friendly software workflow available to semantically link and integrate models on different scales to a multi-cellular tissue model without any kind of coding.

\*To whom correspondence should be addressed.

We previously introduced the EPISIM platform as a graphical cell behavioural modelling and tissue simulation workflow (Sütterlin *et al.*, 2009). Here, we extend our work towards a platform allowing direct semantic integration of SBML-based sub-cellular and graphical cell behavioural models. (i) We implemented an SBML model importer for our graphical modelling system (GMS). (ii) The GMS was moreover extended by a graphical editor for SBML model species, parameters and reactions. Any SBML model item can be directly referenced and manipulated in a graphical cell behavioural model (CBM). This allows building a conceptual semantic model link between the biological scales. (iii) To realize the according technical link, we built a code generator for automatic generation of a model connector software component (MCC). MCC connects a cell's multi-agent-based CBM simulation with an imported SBML model's time course simulation. MCC handles multi-directional dataflow and mapping of the different model time scales. (iv) COPASI was integrated in our simulation environment (SE) for time course simulation of imported quantitative SBML-based models.

We demonstrate the capabilities of the new software workflow by importing and semantically integrating Tyson's cell cycle model (Tyson, 1991) in our model of human epidermal tissue homeostasis (Grabe and Neuber, 2005, 2007). We directly link cellular states like proliferation and differentiation to the behaviour of Tyson's model. Finally, we show the semantic interplay between the sub-cellular and the cell behavioural model in the tissue simulation.

## 2 METHODS

GMS is a set of Java-based plug-ins for the open source software platform Eclipse. The GMS is subdivided in four independent components: (i) Variable-Sheet Editor; (ii) Graphical Model Editor (GME); (iii) Function Library; and (iv) SBML Model Editor (SBME). Components (i)–(iii) have been described earlier (Sütterlin *et al.*, 2009). The new SBME component (iv) was designed using the Unified Modelling Language (UML) (Supplements S1). The resulting model represents an Ecore meta-model of those parts of an imported SBML model that should be made accessible by the SBME. The Eclipse Modelling Framework (EMF) is used to generate large parts of the SBME's source code based on the aforementioned Ecore meta-model (Gamma and Beck, 2004; Moore *et al.*, 2004).

SBML model files to be imported are parsed using the library JSBML [Fig. 1a (1)] offering purely Java-based data structures to read, write and manipulate SBML files (Dräger *et al.*, 2011). An imported SBML model's species, parameters and reactions are stored in a separate file in the XML-Metadata Interchange (XMI) format being an open standard of the Object Management Group. This standard is the SBME's required input format. The original SBML file remains unchanged. Graphical CBMs are translated into executable code by our EPISIM code generator (Sütterlin *et al.*, 2009). This code is stored together with the SBML files in a model archive.

Such a model archive can be loaded and executed by our multi-agent-based graphical SE. SE is built on top of the simulation framework MASON (Multi Agent Simulation Of Neighborhoods) (Luke, 2005). We integrated COPASI as a plug-in for our SE (Fig. 1a). COPASI is used for deterministic time course simulations (LSODA solver) of imported SBML models (Hoops *et al.*, 2006). The multi-scale tissue simulations for determining the computation time per simulation step with and without optimization were run on an Intel(R) Core(TM) i7-2600 processor (3.4 GHz, 8 GB random access memory).

## 3 SOFTWARE WORKFLOW

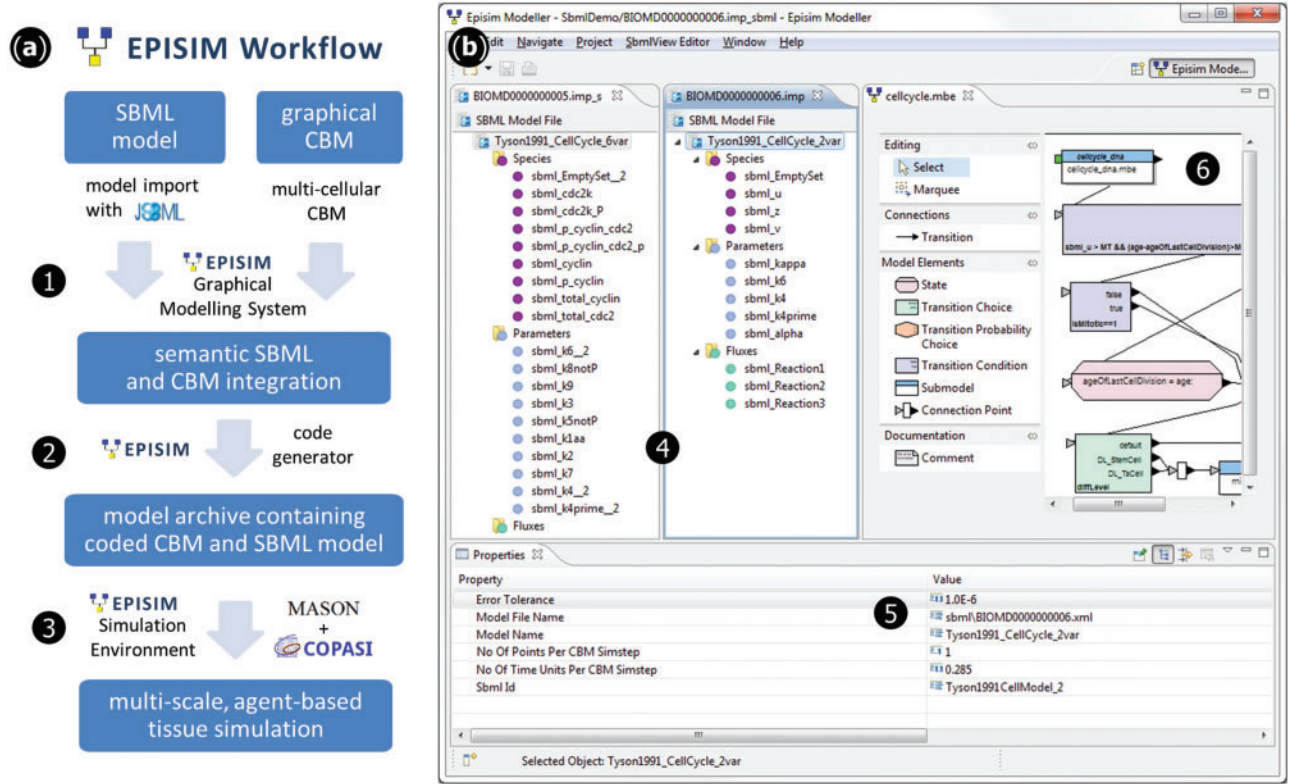
The newly developed software workflow (Fig. 1a) allows graphical modelling and simulation of cellular behaviour in tissues on multiple scales. This software workflow consists of two ready to use software tools: (i) the GMS called EPISIM Modeller and (ii) the SE called EPISIM Simulator. Quantitative SBML models mostly representing biochemical reaction or gene-regulatory networks can be imported by the GMS (Section 3.1). Using the SBME and GME, an imported SBML model's species, parameters and reactions can be directly referenced in a CBM and by that semantically integrated into cellular decisions and a cell's behaviour, respectively [Fig. 1b (4), (6)]. CBMs are automatically translated into executable code with the EPISIM code generator. Moreover, a MCC linking the multi-agent-based tissue simulation with the SBML model time course simulation is generated (Section 3.2). The code is stored together with the imported SBML files in a model archive, which in turn is loaded by the SE. The resulting multi-scale agent-based tissue simulation is subdivided in three steps: (i) SBML model time course simulation; (ii) CBM simulation; and (iii) biomechanical model simulation. In this work, we focus on steps (i) and (ii) (Section 3.2).

### 3.1 SBML model import

The GMS allows importing an unlimited number of quantitative SBML-based models into a CBM project. The import process has five consecutive steps: (i) parsing of the SBML file; (ii) extraction of all species, parameters and reactions; (iii) conversion of local parameters to global parameters; (iv) elimination of namespace conflicts; and (v) storage of the extracted elements in a separate file (Supplementary Movie M1). Only information contained in the SBME's data model (Supplements S1) is extracted from an SBML file. Reactions are named fluxes in the SBME [Fig. 1b (4)], as only a reaction's flux can be referenced in a CBM. A reaction's local parameters are converted into global ones, as each parameter should be unambiguously accessible within a CBM. If local parameters exist, concerned reactions in the imported SBML file are automatically changed. Otherwise, the original SBML file remains unchanged. The SBML files are copied into the model archive during translation of the graphical CBMs into executable code. Each name of a species, parameter or reaction must be unique, as otherwise an unambiguous reference in a CBM would not be possible. This general requirement can lead to namespace conflicts when importing multiple SBML-based models. Those conflicts are eliminated by adding a numbered suffix to duplicate names. Additionally, the prefix 'sbml\_' is added to each name. The original name is kept as a separate property visible in the SBME. Namespace conflicts arising during the conversion of local into global parameters are eliminated accordingly.

### 3.2 Model simulation and time scale mapping

COPASI was integrated in the SE for multi-agent-based tissue simulations. The SE loads a model archive compiled by the GMS, and each cell holds a reference to an instance of the coded CBM. The model archive, moreover, contains imported SBML model files. These files are loaded by COPASI providing one instance of each model for time course simulation. A tissue



**Fig. 1. (a)** EPISIM multi-scale modelling and simulation software workflow. (1) Graphical CBMs are created with EPISIM GMS, which is able to import SBML-based models. The imported model's species, reactions and parameters can be directly referenced, and thus semantically integrated in a CBM. (2) CBMs are translated into executable code by the EPISIM code generator and stored in a model archive together with the imported SBML files. (3) The model archive is loaded by the SE built using the framework MASON and the software COPASI. MASON offers the infrastructure for a multi-agent tissue simulation based on a CBM. COPASI loads the imported SBML models and runs an individual time course simulation for each SBML model and each cell in the simulated tissue. **(b)** EPISIM Modeller (GMS), (4) SBME, (5) an SBML model's properties and time course simulation parameters and (6) GME for CBM specification

simulation step can be subdivided in two consecutive sub-steps for each cell: (i) time course simulation of each imported SBML model and (ii) simulation of one CBM time step. Thereby, the CBM simulation uses the results computed in sub-step (i).

The semantic link between simulation sub-step (i) and (ii) requires a mapping of the model time scales. As depicted in Figure 2, the clock frequency of the multi-cellular tissue simulation causes the assumption of an atomic time step for a CBM here denoted  $\Delta t_{\text{cbm}}$ . The SBML model's simulation time  $\Delta t_{\text{sbml}}$  denotes the model's time-unit (e.g. hours). An SBML model's mapped time  $t_{\text{mapped}}$  then corresponds to the time interval simulated by COPASI for each tissue simulation step  $\Delta t_{\text{cbm}}$ . To calculate  $t_{\text{mapped}}$ , the modeller has to determine semantically equivalent time periods  $t_{\text{sem\_cbm}}$  in the CBM and  $t_{\text{sem\_sbml}}$  in the SBML model. Such periods can be the cell cycle time in the CBM and the according time in the semantically linked SBML-based cell cycle model (Section 4.1).

$$t_{\text{mapped}} = \frac{\Delta t_{\text{cbm}} \times t_{\text{sem\_sbml}}}{\Delta t_{\text{sbml}} \times t_{\text{sem\_cbm}}} \quad (1)$$

COPASI numerically calculates  $n_{\text{points}}$  time points subdividing  $t_{\text{mapped}}$  in the time course simulation. The interval between two

time points in an SBML model's time course simulation  $\Delta t_{\text{point}}$  can be calculated according to equation 2.

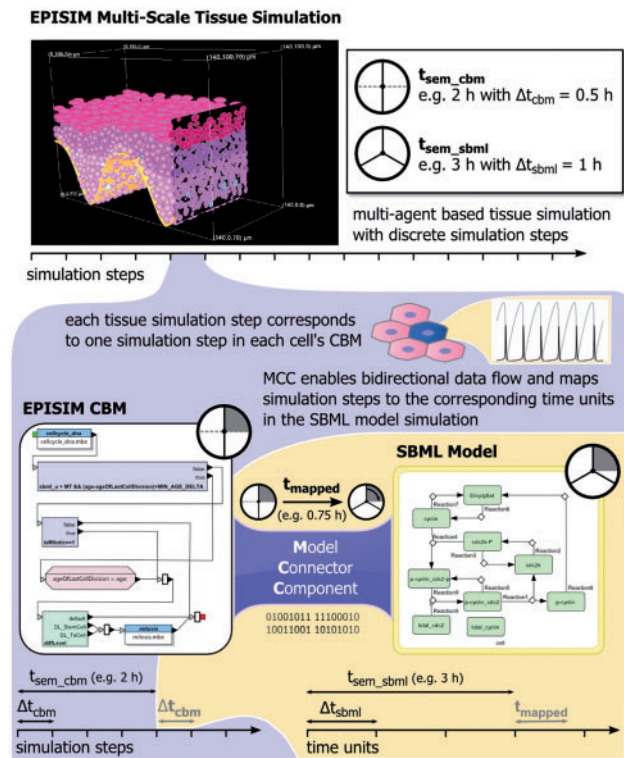
$$\Delta t_{\text{point}} = \frac{t_{\text{mapped}}}{n_{\text{points}}} \quad (2)$$

The value of  $\Delta t_{\text{point}}$  has to be sufficiently small for an accurate numerical time course simulation of a particular SBML-based model. Too small values lead to an unnecessary increase of computation time without improving the simulation accuracy. An optimal value for  $n_{\text{points}}$  has to be determined depending on the kinetics of SBML model. The GMS allows individual setting of  $t_{\text{mapped}}$  (*No Of Time Units Per Simstep*) and  $n_{\text{points}}$  (*No Of Points Per CBM Simstep*) for each imported SBML model [Fig. 1b (5)]. These values are passed on to COPASI.

A MCC is automatically generated by the GMS to link the multi-agent-based tissue simulation of CBMs with the time course simulation of imported SBML models (Fig. 2). MCC is the mediator between both kinds of models, and it enables the semantic interplay during the simulation:

- **Time mapping:** mapping of tissue simulation steps to the time scales of the individual imported SBML models as described earlier in the text.





**Fig. 2.** Each tissue simulation step corresponds to one simulation step in each cell's CBM instance. Moreover, one simulation step represents the atomic model time  $\Delta t_{\text{cbm}}$ . This atomic time has to be mapped to the semantically equivalent time  $t_{\text{mapped}}$  in an SBML model's time course simulation. An SBML model's basic time interval resulting from the chosen time unit (e.g. hours) is denoted by  $\Delta t_{\text{sbml}}$ . The time  $t_{\text{mapped}}$  is calculated based on the two time intervals  $t_{\text{sem\_cbm}}$  and  $t_{\text{sem\_sbml}}$ , representing semantically equivalent periods from a biological perspective. In the given example, these time periods as well as the atomic times differ. Hence, 0.5h in the tissue simulation is mapped to 0.75h in the SBML model simulation. The time mapping as well as bi-directional data flow is facilitated by an automatically generated MCC

- *Administration of SBML model simulation states:* each imported SBML model is only loaded once by COPASI. A single SBML model instance is used for all cells. A data structure stores an SBML model's time course simulation state after each simulation step. The MCC administers all time course simulation states of all imported SBML models for all cells.
- *Initialization of time course simulations:* if the tissue simulation contains  $m$  cells at simulation step  $n$ , we need  $m$  individual time course simulations of the same SBML model. MCC reinitializes COPASI after each simulation step. The SBML model state of tissue simulation step  $n-1$  is the initial time course simulation state of simulation step  $n$ .
- *Enable access to time course simulation results:* In a graphical CBM, each species' concentration, each parameter's value and each reaction's flux can be referenced and modified. This enables the semantic integration of both modelling levels. Technically, this bi-directional dataflow is realized by the automatically generated MCC. The MCC serves as

a data pipe between CBM instances and the according SBML model simulation states.

- *Mapping of original names to unique names:* as described in Section 3.1, the name of a species, parameter or reaction is converted to a unique name when an SBML model is imported. The SBML model file used for time course simulation by COPASI remains unchanged. MCC maps the original names to the unique names and vice versa.

Wrapping COPASI with the MCC also allows the user to store the SBML simulation results together with the CBM states in a simulation snapshot file at chosen time points. A tissue simulation can be fully resumed by loading a snapshot serving as simulation starting point (Supplementary Movie M7). Finally, the MCC can switch an SBML model's time course simulation for a particular cell on or off for saving computation time. Depending on a cell's 'type' or 'differentiation stage', the simulation of a certain SBML model might semantically not be relevant. The GMS's function library offers the command *SbmlSim* ('sbmlId', mode) to control an SBML model's simulation mode.

### 3.3 Workflow generality

The introduced EPISIM software workflow provides a ready to use, out-of-the-box software solution in contrast to dedicated software development frameworks like Chaste (Pitt-Francis et al., 2009). EPISIM targets a community without extensive computational skills. Hence, we designed and implemented easy to use computer programs that hide the technical complexity of realizing multi-scale models and simulations. Out of the box computer programs providing a graphical user interface can hardly compete with programming frameworks in terms of flexibility and generality. Nevertheless, the EPISIM workflow or parts of it can be either transferred or re-used within other SEs. The GMS can be linked to a totally different SE like CompuCell3D, as the graphical CBMs are not directly translated into Java code. In an intermediate step, a programming language independent, XML-based model representation is generated. This representation in turn is used to generate Java code using Extensible Stylesheet Language (XSL) transformation. By exchanging the XSL transformation style sheet, translation into any programming language can be realized. As the generated Java code does not include components of the multi-agent framework MASON, one can re-use and transfer the introduced MCC-based link to COPASI in another Java-based multi-agent SE with reasonable time and effort. Finally, the Java-based EPISIM SE intends to integrate novel biomechanical models besides those that are already available. This requires the implementation of the chosen method in Java using a given interface. The connector, used to link a graphical CBM to the biomechanical model, can be automatically generated.

## 4 RESULTS

To demonstrate the capabilities of our newly developed multi-scale tissue modelling and simulation workflow, we semantically integrated Tyson's cell cycle model (Tyson, 1991) in our graphical CBM of epidermal homeostasis (Grabe and Neuber,

2005, 2007; Sütterlin *et al.*, 2009). To this end, we imported both the 2 variable and the 6 variable version of Tyson's model with our GMS [Fig. 1b (4)]. Both versions of the SBML model files were obtained from BioModels database (Le Novère *et al.*, 2006). The 2 variable model version was semantically integrated in our graphical CBM. We, therefore, completely remodelled our epidermal CBM's cell cycle sub-model (Section 4.1) and parts of our cell differentiation sub-model (Section 4.2). The 6 variable model version was imported to show the GMS's automatic elimination of namespace conflicts (Section 3.1) and the MCC's capabilities (Section 3.2) to handle multiple imported SBML models. Finally, we ran a multi-scale agent-based tissue simulation with the SE to demonstrate the semantic interplay of SBML-based sub-cellular models and graphical CBM's in a multi-cellular context (Section 4.3).

The whole multi-scale workflow is illustrated by a series of supplementary movies (Supplementary Movie M1–M7). The graphical modelling language is described in Supplements S2. The software tools and all modelling projects or simulation data used in the movies can be downloaded from the supplements website.

#### 4.1 Cell cycle model

We remodelled the cell cycle sub-model of our CBM of epidermal homeostasis. After importing Tyson's 2 variable cell cycle model with the GMS, we semantically integrated the model's readout (Fig. 3 and Supplements M2). Tyson's mathematical model of *cdc2* and cyclin interaction is able to operate in the three modes (i) steady state with high maturation promoting factor (MPF) activity; (ii) spontaneous oscillator; or (iii) excitable switch. We associated mode (ii) with cell division cycles in stem cells and transit amplifying (TA) cells. Mode (i) is used in no longer proliferating but differentiated spinosum and

granulosum cells (Section 4.2). The SBML model species *sbml\_u* corresponds to concentration of active MPF relative to *cdc2*. Cells divide at a significantly high active MPF concentration modelled by exceeding mitotic threshold (MT) [Fig. 3 (1)]. We defined a minimum period *MIN\_AGE\_DELTA* between two cell divisions, avoiding multiple cell divisions in consecutive simulation steps [Fig. 3 (2)]. The whole condition represents a semantic link between a sub-cellular model and a cell behavioural decision. A proliferating cell goes into cell cycle arrest in case of too high cell density in its microenvironment. In such a case of high external pressure, cell division is delayed and *isMitotic* is set to true (T) in the mitosis sub-model. This triggers cell division in one of the following simulation steps [Fig. 3 (3)].

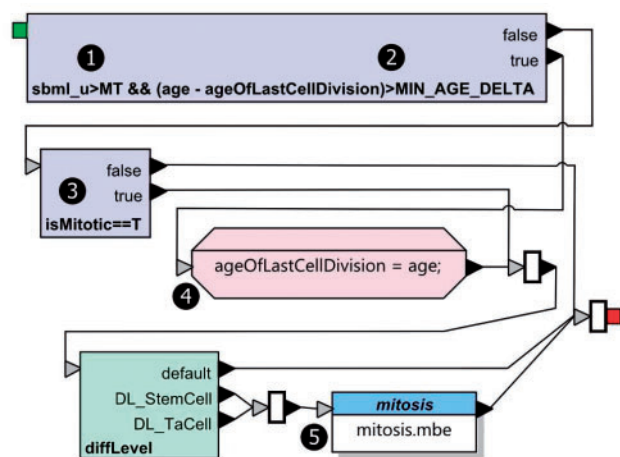
With default parameterization from BioModels database, Tyson's model yields an active MPF concentration peak every 35 time units interpreted as hours in our case ( $\Delta t_{\text{sbml}} = 1$  h). Threshold MT is reached every 34.2 h. In our epidermis model, a proliferating cell divides every 60 h. The CBM's atomic time interval is  $\Delta t_{\text{cbm}} = 0.5$  h. With  $t_{\text{sem\_sbml}} = 34.2$  h and  $t_{\text{sem\_cbm}} = 60$  h, the mapped time  $t_{\text{mapped}}$  is 0.285 h for this SBML model. We set  $n_{\text{points}} = 1$ , as resulting  $\Delta t_{\text{point}}$  is 0.285 h. This is sufficiently small for an accurate time course simulation of Tyson's model.

#### 4.2 Cell differentiation model

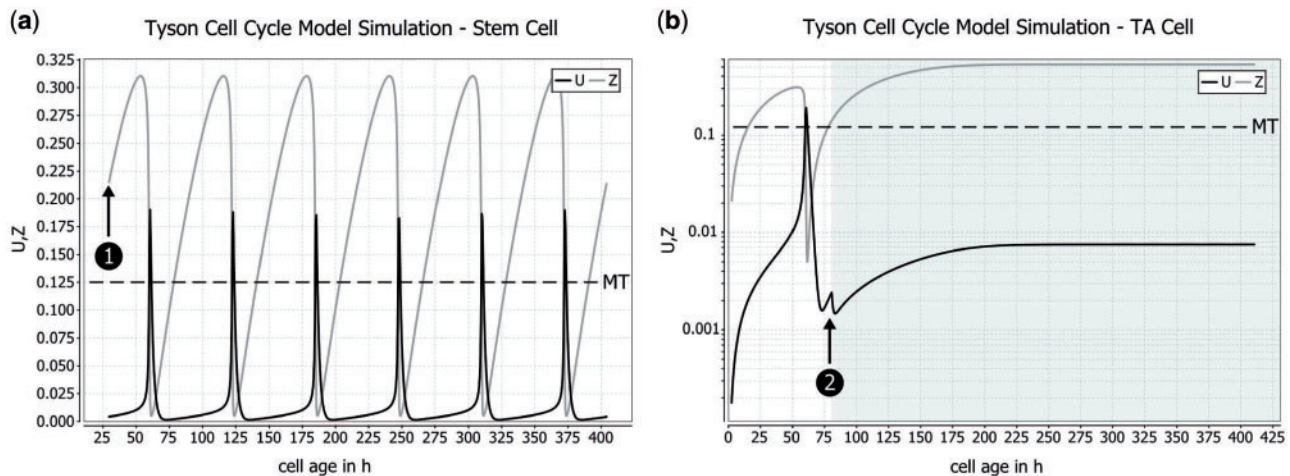
In our model of human epidermis (Grabe and Neuber, 2005), we distinguish proliferating (stem cells, TA cells) and differentiating (spinosum and granulosum cells) keratinocytes. The cell differentiation sub-model (Supplements S4) links sub-cellular model behaviour to cell states. TA cells differentiate to early spinosum cells depending on their age. TA cell differentiation to early spinosum cells corresponds to a cell state transition from proliferation to differentiation. Tyson's cell cycle model can operate in three modes. The model's parameter *k6* (named *sbml\_k6* after import) can be used to change between these modes. By setting *sbml\_k6* = 2, the model behaviour changes from oscillation to steady state. Hence, proliferation stops, as cell division is triggered by high active MPF concentrations, and threshold MT is never reached.

#### 4.3 Multi-scale tissue simulation

To demonstrate the semantic interplay of the imported sub-cellular SBML model and the CBM, we compiled a model archive with the GMS. The SE runs a multi-scale agent-based tissue simulation based on the loaded model archive (Supplementary Movies M5–M7). A CBM is per default linked with 2D biomechanical model (Grabe and Neuber, 2005). Supplementary Movie M6 shows how a graphical CBM can be connected to the 3D version of this biomechanical model. Depending on the linked biomechanical model, the SE automatically starts a 2D or 3D tissue simulation. SE allows generating charts for real-time monitoring of single cells or the tissue as a whole. Figure 4a shows the time course simulation results of Tyson's cell cycle model within a stem cell. Stem cells are initialized with a random age avoiding synchronized cell behaviour. SE initializes imported SBML models according to this random initial age [Fig. 4a (1)]. We used  $t_{\text{mapped}} = 0.285$  for the imported cell cycle model and observed active MPF (U) peaks every 60 h in stem cells.



**Fig. 3.** Semantic integration of Tyson's cell cycle model in a graphical CBM. The condition for cell division holds if the active MPF (*sbml\_u*) concentration exceeds threshold *MT* (1) and the minimum time *MIN\_AGE\_DELTA* between two cell divisions has elapsed (2). A cell also divides if it went into cell cycle arrest in one of the previous simulation steps (3). After saving the cell's age at cell division (4), the mitosis submodel (5) responsible for cell division is called



**Fig. 4.** Semantic interplay between sub-cellular and cellular model behaviour during tissue simulation. **(a)** Oscillation between high and low active MPF concentrations ( $U$ ) within a monitored stem cell. Stem cells are initialized with a random age (1) before simulation. Cell division is triggered when mitotic threshold  $MT$  is reached. **(b)** The monitored TA cell differentiates to an early spinosum cell after one cell division (2). This corresponds on the sub-cellular level to a transition of Tyson's cell cycle model from oscillating to steady-state behaviour

This sub-cellular model behaviour triggers cell division by exceeding threshold  $MT$ . The transition from cell proliferation to cell differentiation is semantically linked to a change from oscillating to steady-state model behaviour on the sub-cellular level [Fig. 4b (2)]. After one cell division, a TA cell differentiates to an early spinosum cell, which no longer proliferates because of the steady state of Tyson's cell cycle model [Fig. 4b (2)].

We optimized the computation time per tissue simulation step using the command *SbmlSim*. The regular computation time per tissue simulation step is almost eight times higher than the optimized computation time in a 2D tissue simulation and more than six times higher in a 3D tissue simulation. Independent of optimization, the time per tissue simulation step increases linearly with an increasing cell number. Further details regarding computation time optimization and the parallelization of simulations using the SE on a computer cluster are included in Supplements S3.

## 5 DISCUSSION

We developed and introduced the first entirely graphical multi-scale cell behavioural modelling and simulation software workflow that semantically links the sub-cellular and cellular level in tissue simulations. Semantic linking comprises flexible mapping of time scales, enabling of bi-directional model variable access and by this linking of general cellular states like proliferation and differentiation to molecular processes and vice versa.

Cellular behaviour is graphically modelled with process diagrams. An unlimited number of quantitative SBML-based models can be imported with the GMS. This enables a semantic integration of sub-cellular in cell behavioural models. We integrated COPASI in the SE for calculating deterministic time course simulations of the imported SBML models for each individual cell. An automatically generated MCC handles the data flow between sub-cellular and cell behavioural models during tissue simulation. Time course simulation states can be stored together with the overall tissue simulation state in a simulation

snapshot. Based on such snapshot, a tissue simulation can be fully resumed. Moreover, this enables archiving, re-using and exchanging of tissue simulation results among users of the SE. Simulation snapshots allow rerunning a published tissue simulation that goes far beyond the information conveyed by a simulation movie. However, the lack of a commonly accepted standard for representing cell-based models as well as the according simulation results hinders the exchange of both, models and results, between different existing modelling and simulation environments. Such a standard is in our view the prerequisite for realizing re-usable and ready to run multi-scale models in a fashion that is comparable with sub-cellular models using SBML.

To demonstrate the applicability of our software workflow, we imported the SBML version of Tyson's well known 2 variable cell cycle model with the GMS. In our model of epidermal homeostasis, we linked the imported sub-cellular model with the cellular states proliferation and differentiation. Cell division in proliferating keratinocytes is controlled by the active MPF concentration modelled on the sub-cellular level. In turn, the behaviour of Tyson's model is coupled to a keratinocyte's differentiation stage on the cellular level. By this, we were able to demonstrate a bi-directional semantic link between the sub-cellular and cellular modelling and simulation level. Additionally, we demonstrated the automatic elimination of namespace conflicts by importing the 6 variable version of Tyson's.

For both model versions, the SBML model time units could be successfully mapped to the required CBM time steps. This shows that with the developed workflow, SBML models can be re-used in a context requiring a different time scale without re-parameterizing the SBML model. One might argue that such a mapping of semantically equivalent periods should not be necessary. But actually, the necessity of directly manipulating parameters of an ODE-based model would contradict our initial requirement of re-usability of SBML models. Moreover, we see time mapping as a valuable technique for model synchronization and simpler multi-scale model integration. Finally, we demonstrated that the SE is able to administer time course simulations



of multiple imported quantitative SBML models. The SE successfully handled >9000 individual time course simulations per tissue simulation step in the 3D simulation. The computation time per simulation step increases linearly for increasing cell numbers independent of optimization by the modeller. Moreover, this indicates the power of the EPISIM code generator and the approach of automatically translating graphical CBMs into highly efficient and optimized executable code (Sütterlin *et al.*, 2009).

Our software workflow allows graphically building multi-scaled models without any coding. This is a unique approach compared with other available systems (Andasari *et al.*, 2012; Pitt-Francis *et al.*, 2009). The available model elements (Supplements S2) of the graphical modelling language allow deterministic, stochastic or hybrid modelling. Even continuous models can be manually integrated (Supplements S3). All control structures of a programming language like branching based on conditions or values are available. Loops are automatically induced when neighbouring cells are referenced. Graphical CBMs are linked to a hard coded biomechanical model provided by the SE. This modularization into biomechanical and cell behavioural models is not necessarily a limitation. Moreover, it allows re-using one and the same CBM in 2D and 3D as demonstrated. The integration of new biomechanical models in the SE is explicitly intended. Furthermore, the GMS could be adopted and connected to other simulation environments like CompuCell3D with manageable workload. The GMS as well as the SE hides to a large extent the technical complexity of modelling and simulating cellular behaviour in a tissue context. Furthermore, the separation of technical complexity and systems biological model complexity is in our view a key requirement for building highly complex multi-scale models. This is the main lesson learned from the field of software engineering where model driven architectures are analogous. This complexity separation fosters the creation of more complex truly multi-scaled models by not computer but rather biologically skilled scientists.

Public databases like BioModels provide complete ready-to-run SBML models (Hucka and Le Novère, 2010). With our novel software workflow, we established a link to this large and up to now unexploited potential for building multi-scale models in a multi-cellular context. To come up with more realistic and consequently more complex biological models, it is inevitable to rely on an existing, ready to use model base, instead of starting from scratch solely based on literature. However, this requires that increasingly sub-cellular models are developed with a focus on the functional, behavioural change of cells. Such models could be models of cell decision for proliferation, differentiation, migration or apoptosis. This currently in literature underrepresented aspect is in our view an essential prerequisite for building higher complexity models. In this context, our multi-scale workflow motivates the integration of existing models or the creation of novel ones. In consequence, we expect this to be the route to more holistic models of larger biological systems.

**Funding:** BMBF FORSYS-PARTNER (grant number: 0315263), BMBF MEDSYS (grant number: 0315401 B) and BMBF Virtual Liver Network (grant number: 0315761).

**Conflict of Interest:** none declared.

## REFERENCES

- Amirkroll, H. *et al.* (2008) GemCell: a generic platform for modeling multi-cellular biological systems. *Theor. Comput. Sci.*, **391**, 276–290.
- Andasari, V. *et al.* (2012) Integrating intracellular dynamics using CompuCell3D and Bionetsolver: applications to multiscale modelling of cancer cell growth and invasion. *PLoS One*, **7**, e33726.
- Dräger, A. *et al.* (2011) JSBML: a flexible Java library for working with SBML. *Bioinformatics*, **27**, 2167–2168.
- Funahashi, A. *et al.* (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, **1**, 159–162.
- Gamma, E. and Beck, K. (2004) *Contributing to Eclipse: Principles, Patterns, and Plug-ins*. Munich, Addison-Wesley, Boston.
- Grabe, N. and Neuber, K. (2005) A multicellular systems biology model predicts epidermal morphology, kinetics and Ca<sup>2+</sup> flow. *Bioinformatics*, **21**, 3541–3547.
- Grabe, N. and Neuber, K. (2007) Simulating psoriasis by altering transit amplifying cells. *Bioinformatics*, **23**, 1309–1312.
- Hoehme, S. *et al.* (2010) Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. *Proc. Natl Acad. Sci. USA*, **107**, 10371–10376.
- Hoops, S. *et al.* (2006) COPASI—a complex pathway simulator. *Bioinformatics*, **22**, 3067–3074.
- Hucka, M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
- Hucka, M. and Le Novère, N. (2010) Software that goes with the flow in systems biology. *BMC Biol.*, **8**, 140.
- Luke, S. (2005) MASON: a multiagent simulation environment. *Simulation*, **81**, 517–527.
- Moore, B. *et al.* (2004) *Eclipse Development Using the Graphical Editing Framework and the Eclipse Modeling Framework*. 1st edn. IBM Redbooks, Research Triangle Park, Durham, NC, USA.
- Munaron, L. and Scianna, M. (2012) Multilevel complexity of calcium signaling: modeling angiogenesis. *World J. Biol. Chem.*, **3**, 121–126.
- Le Novère, N. *et al.* (2006) BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, **34**, D689–D691.
- Osborne, J. M. *et al.* (2010) A hybrid approach to multi-scale modelling of cancer. *Phil. Trans. R. Soc.*, **368**, 5013–5028.
- Owen, M. R. *et al.* (2009) Angiogenesis and vascular remodelling in normal and cancerous tissues. *J. Math. Biol.*, **58**, 689–721.
- Pitt-Francis, J. *et al.* (2009) Chaste: a test-driven approach to software development for biological modelling. *Comput. Phys. Commun.*, **180**, 2452–2471.
- Scianna, M. and Preziosi, L. (2012) Multiscale developments of the cellular potts model. *Multiscale Model. Simul.*, **10**, 342–382.
- Sun, T. *et al.* (2009) Exploring hypotheses of the actions of TGF- $\beta$ 1 in epidermal wound healing using a 3D computational multiscale model of the human epidermis. *PLoS One*, **4**, e8515.
- Swat, M. H. *et al.* (2009) Multicell simulations of development and disease using the CompuCell3D simulation environment. *Methods Mol. Biol.*, **500**, 361–428.
- Sütterlin, T. *et al.* (2009) Modeling multi-cellular behavior in epidermal tissue homeostasis via finite state machines in multi-agent systems. *Bioinformatics*, **25**, 2057–2063.
- Tyson, J. J. (1991) Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc. Natl Acad. Sci. USA*, **88**, 7328–7332.