Developing and setting up a platform and protocols for microscopy image driven-based modeling in context of Systems Biology

Szymon Stoma

1 Introduction

During the last decade systems biology – the systematic study of such functional networks as cellular metabolism, signaling and gene expression – matured into a distinct, interdisciplinary field of research. Systems biology can strongly benefit from applying methods of Computer Science. This interdisciplinary approach may facilitate the understanding of various biological systems [1].

As a part of this systematic approach, it is often required to integrate and standardize the data coming from different biological protocols (e.g. micro-array, 2D-gels, mass spectrometry). In this respect Computer Science appears to be especially effective and efficient. An application of standard Computer Science routines allows for systematic integration of the biological data into models through the development of (i) standards such as Systems Biology Markup Language (SBML) [2] (ii) the open-access signaling databases [3] and iii) software such as Copasi or CellDesigner [4, 5] allowing to systematically integrate these data into models.

Recently, the progress in single cell microscopy such as fluorescent proteins (FP), used as reporters of bio-molecular interactions [6], together with new image acquisition techniques (e.g. multichannel confocal microscopy [7] and two-photon microscopy [8]) led to a significant increment of the role of the imaging methods in molecular biology. These techniques open up many new areas of research - the popularization of a spatial context - to name explicitly one of the most important among them.

Currently, a common technique is to deal with images in a qualitative way - images are attached to the publications and their properties are literally listed in the articles. A step forward from this methodology is to quantify the information within these images (e.g. the measurements of the light intensities coming from FP expressions in different parts of the cell) in a spatially resolved way. This approach has two major advantages: i) it enables us to broaden the usage of imagining techniques in Systems Biology, which require quantified data ii) it increases the quality of the information and allows for further automatic processing (e.g. analysis and modeling). However, this crucial step requires definition of a new methodology, formulation of new approaches and algorithms, their experimental evaluation and - last but not least - development of specifically dedicated software based on the proposed methods

2 Proposed project details

2.1 Motivation and goals of the project

The goal of the proposed project is to develop and set up a software platform and protocols for microscopy image driven-based simulations which will allow for applying more systematic and quantitative approaches in biology.

The software platform will be based on the preliminary work of the Project Leader performed together with Theoretical Biophysics Department on Humboldt University in Berlin (HU), an open-source framework called Spatio-Temporal Simulation Environment (STSE) (http://stse-software.org/). The development of such a software appeared to be necessary in the context of the research on the *Arabidopsis* meristem (within 3 year Marie Curie Action Grant, Sy-Stem) and studies on *Aquaglyceroporin* turnover (within 2 year Marie Curie Action Grant, Aqua(glycero)porin) in which the Project Leader was actively involved.

The significant difference is that the proposed project has its main interest in **algorithms, methodology and software development**. It is contrary to the previous projects, where the software platform and protocols were developed as a side-effect of the study on biological

object. However, it is important to underline, that **studying the biological object is also scheduled in the final phase** of the project. This approach enables **setting up a detailed protocol of image acquisition, processing and analysis prior to its application to real data** (with scientific significance), which is usually contrary in biology-centered projects. This approach allows to create quality software, which can be used in future projects.

2.2 Realization of the proposed project

The host institution of the project is West Pomeranian University of Technology, Szczecin (WPUT). The implementation of the developed algorithms and methods will be supported by student projects at Faculty of Computer Science and Information Systems, WPUT. WPUT will provide the key resources required for project realization: i) expertise in image processing, filtering and recognition ii) Master Students writing their theses related to the project iii) individual student projects under the supervision of Project Leader and WPUT members which will be used to develop and integrate the scientific output into STSE framework. It is important to note that WPUT starting from 10/2010 is starting Bioinformatics studies, which may eventually provide additional support and resources for the proposed project, since the project scope will be coherent with the one of the directions of WPUT development.

The partner for this project is Theoretical Biophysics Departament on Humboldt University in Berlin (HU), led by prof. Edda Klipp. The role of the partner would be to provide the expertise in Systems Biology applyied to yeast, which is internationally recognized team specialization. The members of this team will be actively cooperating with WPUT, sharing their biological expertise and they everyday routines with the members of WPUT team. Project Leader will be holding a position also at HU in the group of Edda Klipp, which will facilitate the cooperation.

2.3 Proposed project outline, resources and outputs

The key principle of the project is to develop and set up a software platform and protocols for microscopy image-driven simulations. The main concept of such an approach is presented in the supporting text. To achieve this goal particular steps are scheduled (the visual project schedule is on the last page of this document):

1. **Developing and integrating the STSE software within the openalea framework and guidelines** (http://openalea.gforge.inria.fr). This step assures the quality of the software, its installation and documentation which is crucial in the project distribution and persistence phase. The web page of the project will be activated, together with automatic documentation and development protocols and scenarios.

   Experience and Partners: In this part Project Leader is planning to use his experience as a former openalea developer. Developing individual components and a web page will be supported by students projects realized at WPUT and supervised by Project Leader.

   Output: STSE will be available as a module of Openalea software. From this early stage on, the project web page will contain actual project status, required jobs and directions of development.

2. **Creating expertise on yeast image processing, analysis and recognition**. Yeast will be used as a model system that allows embedding of newly gained information into pre-existing knowledge. The software is planned to be used also with different model systems (e.g. human cells), however it is important to start the study with a system which is i) easy to deal with ii) the teams which will be running the project have experience with. These conditions are filled by Yeast, therefore we select them in the early stage of the project. It is then needed to provide an semi-automation of image processing. Thus in this step it is planned to perform a scientific research on specification of image acquisition, noise filtering, image segmentation and shape recognition adjusted to the microscopic yeast images. Based on the state of the art analysis and experiments, it is planned to develop specific image processing algorithms.

   Experience and Partners: It will be crucial to use the experience of WPUT at this place which will be actively represented by Dr Dariusz Frejlichowski, Dr Mariusz Borawski, Dr Edward Półrolniczak, who have former experience in image processing, filtering and recognition. To provide a biolog-

ical context for such a study images from yeast will be used. The yeast images will be acquired at HU in the laboratory of prof. Edda Klipp. The implementation of the developed algorithms will be supported by student projects at WPUT.

Output: Publication about image segmentation/processing in the images with yeasts. Algorithms prototypes running and tested.

3. **Developing of semi-automatic image processing protocols in the context of STSE**. This point will be used to integrate the algorithms developed in the previous step into STSE framework. The algorithms would have to be integrated with data acquisition protocols and compatible with common biological routines.

   Experience and Partners: The majority of the work in this sub-project will be performed by the Master Student. The Student will be guided by the WPUT team, Project Leader and by an external expert, Dr Matteo Barberis (HU, Berlin), sharing his expertise on yeast imagining techniques.

   Output: Publication about STSE software in Bioinformatics. GUI tools required for the protocol integrated into STSE software. Defense of Master Thesis.

4. **Integrating the simulation framework into STSE software**. During this subproject the image processing protocols from previous step will be extended by providing the framework to perform a time-course simulation in between the digitized steps. The goal of this is study is explained on the example in a separate document attached to this application.

   Experience and Partners: Here the previous experience of Project Leader in the area of spatial simulations together with prof. Edda's Klipp experience on modeling of signaling pathways kinetics will be crucial. The implementation of the routines and algorithms will be actively supported by WPUT students.

   Output: STSE modules allowing for spatial simulations in between digitized steps.

5. **Integrating Systems Biology Markup Language (SBML) into the STSE platform**. The goal of this sub-project is to provide a link between currently used non-spatial modeling software and the STSE platform. It is crucial since it allows to exchange the data in the Systems Biology community.

   Experience and Partners: The majority of the work in this sub-project will be performed by the Master Student. The Student will be guided by the WPUT team, Project Leader and by external expert, Dr Wolfram Liebermeister (HU, Berlin), sharing his expertise on XML/SBML used as descriptors of kinetics in biochemical systems.

   Output: Exporters/Importers of models to modified SBML integrated into STSE software. Defense of Master Thesis.

6. **Spatio-temporal analysis of the distribution of selected biologically significant compound in yeast**. The target of this study is to apply the developed tools in the project with significant biological results. The exact topic will be determined at later stage, since it will depend on the tool capabilities after four first periods of the proposed project. During these periods the STSE team will attract interested groups, especially from Berlin.

   Experience and Partners: In this subproject a good collaboration with prof. Edda Klipp and the former experience of Project Leader will be crucial mainly because of i) contacts and recognition in the yeast signaling pathways community ii) former extraordinary experience in running such a multidisciplinary study allowing to wisely choose partners and scope of the research.

   Output: Collaboration on biological paper. Accomplished example of scientific-significant results obtained with STSE.

2.4 Sub-tasks realized by students

2.4.1 Master students

An important part of the project are student master theses. The theses are closely related with the work of Project Leader, which will actively support and supervise the students. It is then crucial to provide students the international standards to work. In general the Project Leader allocates the budget for the following activities for each student in the project:

1. monthly scholarship (granted by default by FNP),

2. mobile computer (granted for the time of student's presence in the project),

3. two weeks time spend in the host institution of the external expert for the given thesis (if possible in two one-week sessions),

4. possibility to present his/her work in the international conference.

These conditions are provided to allow the students for: i) maximum efficiency and focus on the subjects (1 and 2) ii) international standard career development opportunities for the students (3 and 4) iii) diffusion and popularization of the project (4).

The theses are scheduled to be performed during the proposed projects (they cover the areas described in the project details). The following thesis is foreseen in the first and second period:

- "Integrating Systems Biology Markup Language (SBML) into the STSE platform"
  - external expert: Dr. Wolfram Liebermeister (Humboldt University(TBP), Berlin, Germany)

Thesis foreseen for the third and fourth period (it profits from the completion of the first and second phase thesis):

- "Creating a semi-automatic image processing work flows on in the context of STSE platform"
  - external expert: Dr. Matteo Barberis (Humboldt University(TBP), Berlin, Germany)

2.4.2 Students projects

Except of Master students officially involved in the proposed project it is planned to integrate the work of students from the individual projects. The individual projects will be officially announced in the WPUT and web page together with their bounty (allocated in the budget), which will be paid from the budget of the project, giving the WPUT students the possibility to i) gain the experience by working on a real project, with commercial development procedures ii) assure the quality of developed code by contract (this approach was successfully tested e.g. at University of Wroclaw while obtaining official 25k$ grant from Microsoft Research during the realization of the project Nemerle, 2004). The topics of the sub-projects will be mainly focused on the implementation of various parts of the software and will be based on the expertise and guidelines of the scientists actively supporting STSE development. The topic list is included as separate document.

3 Persistence of the results

STSE is an open-source project. That means that its source code is officially available without any restrictions and it must remain in the public domain. The project sources are kept in the http://sourceforge.net which is guarantying the persistence of code even if the supporting institutions will withdraw their support.

4 Strategic importance of the project

Currently, to the knowledge of Project Leader, there is no platform allowing for spatial simulations based on microscopy images. In the same time the importance of imagining techniques is constantly increasing, due to the progress in the acquisition equipment, as well as in fluorescent markers. **Successful development of such tools would allow for becoming an important partner for experimental laboratories all around the world, possibly leading to world-level science.** Last but not least, the topic of the proposed project is coherent with the current directions of development suggested by Polish Ministry of Science and Higher Education.

References

[1] Bornholdt, S. Systems biology: Less is more in modeling large genetic networks. *Science* **310**, 449–451 (2005). URL http://dx.doi.org/10.1126/science.1119959.

[2] Hucka, M. *et al.* The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003). URL http://dx.doi.org/10.1093/bioinformatics/btg015.

[3] Klipp, E. *Systems biology : a textbook* (Wiley-VCH, 2009), 1 edn. URL http://www.worldcat.org/isbn/3527318747.

[4] Funahashi, A., Matsuoka, Y., Jouraku, A., Kitano, H. & Kikuchi, N. Celldesigner: a modeling tool for biochemical networks. In *WSC '06: Proceedings of the 38th conference on Winter simulation*, 1707–1712 (Winter Simulation Conference, 2006). URL http://portal.acm.org/citation.cfm?id=1218112.1218422.

[5] Hoops, S. *et al.* Copasi–a complex pathway simulator. *Bioinformatics (Oxford, England)* **22**, 3067–3074 (2006). URL http://dx.doi.org/10.1093/bioinformatics/btl485.

[6] Ward, T. H. & Lippincott-Schwartz, J. The uses of green fluorescent protein in mammalian cells. *Methods of biochemical analysis* **47**, 305–337 (2006).

[7] White, J. G., Amos, W. B. & Fordham, M. An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy. *J. Cell Biol.* **105**, 41–48 (1987).

[8] Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).

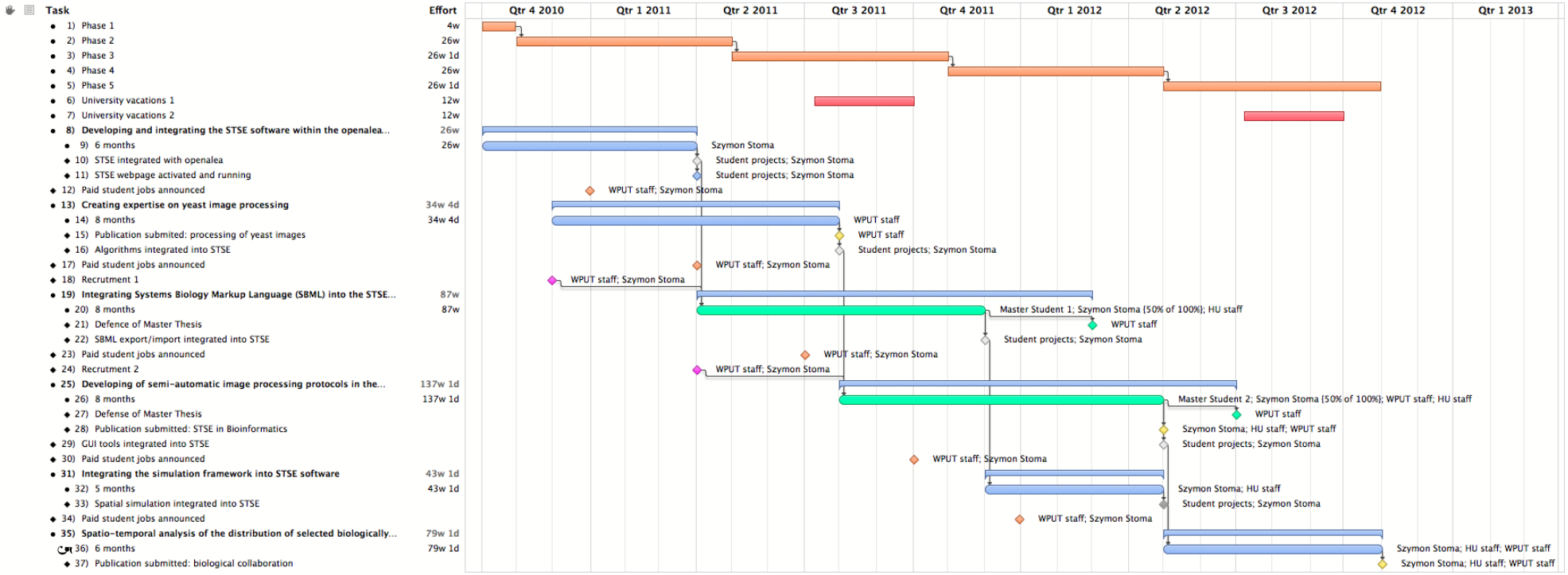

Figure 1: Gantt plan of the proposed project.