Proceedings of the CELLmicrocosmos neXt workshop 2014

Bielefeld, 28.09.2014

Björn Sommer (Ed.)

Preface

A primary vision of Integrative Bioinformatics is the creation of a virtual cell. Cell Modeling and Visualization is an immense interdisciplinary task, bridging the gap between Biology, Chemistry, Physics, Design, and Informatics. For this purpose, in the context of the German Conference of Bioinformatics (28.09. - 01.10.2014), the first CELLmicrocosmos neXt workshop has taken place in Bielefeld University at the 28.09.2014. During this event, 10 years of the CELLmicrocosmos project were celebrated by presenting CELLmicrocosmos X and an additional hands-on workshop session.

To evaluate the potential of cell visualization and the prospects for future collaborations, seven talks were presented at the CELLmicrocosmos neXt workshop. In addition, two chapters were added here discussing some current and future work on the CELLmicrocosmos project. Each of the extended abstracts contains a maximum of two pages.

- Björn Sommer (Bielefeld University) started the workshop by showing "CELLmicrocosmos X", giving a short resume of 10 years development during an integrative cell modeling project (Page 5f).
- Niklas Biere et al. (Bielefeld University) presented a 3D stereoscopic animation of Chlamydomonas reinhardtii, following the principles of cell visualization as proposed by the CELLmicrocosmos project: "Towards a 3D Cell Model of Chlamydomonas reinhardtii" (Page 7f).
- Olga Popik et al. (SB RAS Novosibirsk) showed new achievements of our long-term Russian-German collaboration on the CELLmicrocosmos 4 project with the Institute of Cytology and Genetics in Novosibirsk. By using curated and text-ming-based database entries, pathway efficiencies were predicted: "Evaluation and Visualization of Pathway Efficiency based on Subcellular Protein Localizations" (Page 9f).
- Kirill Veselkov (Imperial London) discussed new findings in the area of Mass Spectrometry Imaging, targeting unmet medical needs of cancer network-driven diagnosis: "Top-down Systems Biology Approaches for -omics-based Tissue and Biofluid Analytics" (Page 11).
- Ahmet R. Öztürk (Erciyes University) was arguing that Data-Driven Documents (D3) are an indispensable technology for biological data visualization. By just using the browser, also complex relationships can be explored by the user: "Web-based Visualization with D3" (Page 12).
- Gökhan Kovanci (Bielefeld Unviersity) presented his master thesis, a combination of D3 and three.js which is able to visualize cytological localization scenarios within the browser: "Towards a web-based Viewer for the PathwayIntegration" (Page 14).
- Jens Krüger (Tübingen University), who is involved in the CELLmicrocosmos 2 project since several years, concluded the workshop. He presented how it is possible to initiate molecular dynamics simulation – which is usually a very complicated task – from within a browser using the Molecular Simulation Grid's portal technology: "Web-based MD Simulations using MoSGrid" (Page 16).
- Finally, this proceedings finishes with two additional chapters. First, "Towards Pathway Prediction and Subcellular Localization by using FraMeTex" based on the upcoming PhD thesis of Thorben Wallmeyer (Page 18). And second, "A novel Lecture Program in Bioinformatics: Interdisciplinary Cell Visualization and Modeling" by Björn Sommer and Ming Chen (Bielefeld Unviersity/Zheijang University), discussing first educative and scientific results of our Sino-German collaboration started in 2012 (Page 20).

Program Committee

Editor/Address

Björn Sommer

Bio-/Medical Informatics Department Bielefeld
Bielefeld University
Universitätsstr. 25
33615 Bielefeld
Germany

E-Mail : bjoern@CELLmicrocosmos.org

WWW : http://www.CELLmicrocosmos.org

Acknowledgments

Our thanks go to all workshop presenters and attendees for their participation, and their interesting and valuable contributions! In addition, many thanks go to our students, advisers and cooperation partners, who supported the project over the last decade:

http://team.CELLmicrocosmos.org

Additional thanks go especially to the Deutsche Forschungsgesellschaft (DFG) for supporting this work as part of the “Graduate College Bioinformatics (GK635)” from 2007-2009.

Workshop Website

http://neXt.CELLmicrocosmos.org

Citation

Sommer, Björn (Ed.): Proceedings of the CELLmicrocosmos neXt workshop, Bielefeld, 28.09.2014, DOI: 10.2390/biecoll-next2014-0.

Content

CELLmicrocosmos X

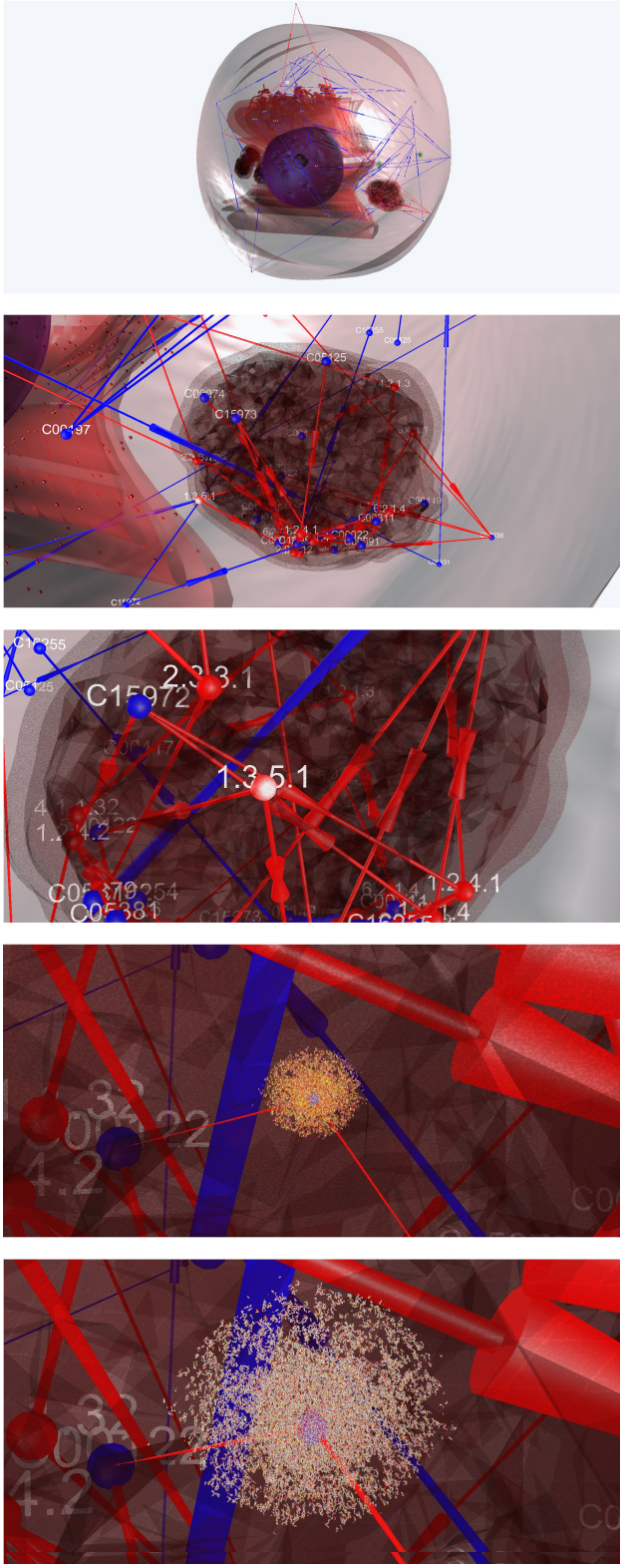
Björn Sommer*

Bio-/Medical Informatics Department, Bielefeld University/Center for Biotechnology
Universitätsstr. 25, 33615 Bielefeld, Germany

1 Introduction

The CELLmicrocosmos project represents a prototype of a static virtual cell environment combining different cytological levels. Ten years of development, starting from an initial cell animation, leading over a first educational interactive prototype, proceeding with a number of cell modeling projects. Today, approaches have been presented at the molecular, mesoscopic as well as functional level [1].

Integrative Bioinformatics combines different computer technology-related research areas to support life science research. Here, a major vision of bioinformatics is addressed: the creation of a virtual cell. The CELLmicrocosmos project combines the knowledge of more than ten databases with the objective to create a three-dimensional virtual cell environment integrating three cytological levels; the mesoscopic, the molecular and the functional level. Using the CellExplorer with the PathwayIntegration (CmPI)), virtual cell environments – directly or indirectly derived from microscopic data – are combined with protein-related networks by using localization information acquired from different well-known databases [2], [3]. In addition, these cell models can be extended by published data-based PDB membranes generated with the MembraneEditor (CmME) [4].



2 From Mesoscopic to Molecular Level

The right image shows the journey from the molecular to the mesoscopic level. Starting from the mesoscopic perspective, showing the complete cell model

*bjoern@CELLmicrocosmos.org

associated with two metabolic pathways from KEGG, the citrate cycle and the glycolysis, the user navigates towards the mitochondrion. The metabolic pathways were localized by using CmPI [3].

The mitochondrion model was associated with a texture of a mitochondrial PDB membrane model surrounding a protein which is part of the citrate cycle. Finally, the molecular structure is combined with the surface of the mitochondrion by using a novel technique we have recently introduced as Membrane Mapping [5].

3 Outlook

CmX is a small plug-in which can be used to combine the CellExplorer and the MembraneEditor to enable a smooth workflow, bridging the gap between the mesoscopic and molecular level. After the full release of the CellExplorer, CmX will also be freely available.

Now, the resulting projects can be used as a starting point for a collaborative project, providing a visual framework to combine and present different levels of cytological knowledge.

4 Acknowledgments

Instead of funding from large organizations, the driving force behind this project were students. Bachelor, master and diploma students were involved in the development of different modeling tools and 3D models. And during different student projects, even more students were involved in implementation, modeling and visualization work. Thanks go to them, and in addition to all colleagues supporting this project during the last decade:

http://team.CELLmicrocosmos.org

In addition, this work was funded in part by the “Graduate College Bioinformatics (GK635)” of Deutsche Forschungsgesellschaft (DFG).

5 References

[1] B. Sommer. CELLmicrocosmos-Integrative cell modeling at the molecular, mesoscopic and functional level. Doctorate Thesis, Bielefeld University, Bielefeld, Germany, 2012.

[2] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming, and B. Kormeier. CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. in *BIOSTEC BIOINFORMATICS 2010*, 2010:90–95.

[3] B. Sommer, B. Kormeier, P. S. Demenkov, P. Arrigo, K. Hippe, Ö. Ates, A. V. Kochetov, V. A. Ivanisenko, N. A. Kolchanov, and R. Hofestädt. Subcellular Localization Charts: A new visual methodology for the semi-automatic localization of protein-related data sets. Journal of Bioinformatics and Computational Biology, 11(1):1340005, 2013.

[4] B. Sommer, T. Dingersen, C. Gamroth, S. E. Schneider, S. Rubert, J. Krüger, and K. J. Dietz. CELLmicrocosmos 2.2 MembraneEditor: a modular interactive shape-based software approach to solve heterogeneous Membrane Packing Problems. Journal of Chemical Information and Modeling, 5(51):1165–1182, 2011.

[5] T. Waltemate, B. Sommer, and M. Botsch. Membrane Mapping: Combining Mesoscopic and Molecular Cell Visualization. Vienna, Austria, 2014:89–96.

Towards a 3D Cell Model of *Chlamydomonas reinhardtii*

Niklas Biere[1], Anja Doebbe[2], Nils Rothe[1], Thomas Huser[3], Olaf Kruse[2], Björn Sommer[1,*]

[1]Bio-/Medical Informatics Department, Bielefeld University/Center for Biotechnology
[2]Algae Biotechnology & Bioenergy, Bielefeld University/Center for Biotechnology
[3]Biomolecular Photonics Group, Bielefeld University
Universitätsstr. 25, 33615 Bielefeld, Germany

1 Introduction

Chlamydomonas reinhardtii is a green alga which is often used in Biotechnology as a model organism. This single-cell organism has a size of approx. 10 µm, contains a very large chloroplast relevant for the energy production which is partly used to move by using two flagella. An important biotechnological application is the production of biofuels.

2 Cell Modeling

A visualization approach of the interpretative abstraction level is used to create the C. reinhardtii model [1]. By combining the information of different publications, microscopic images and videos at multiple scales, this project aims to present a more precise three-dimensional structure of this cell type. The cell model is generated by using Blender [2]. In addition, 3D microscopy data was used as a base to model the *Endoplasmic reticulum* which was previously segmented with FIJI [3].

To visualize the energy-relevant pathways of C. reinhardtii, a simplified network was modeled and localized using the CELLmicrocosmos 4.2 PathwayIntegration [4]. This pathway was afterwards used in Blender to illustrate the intracellular processes.

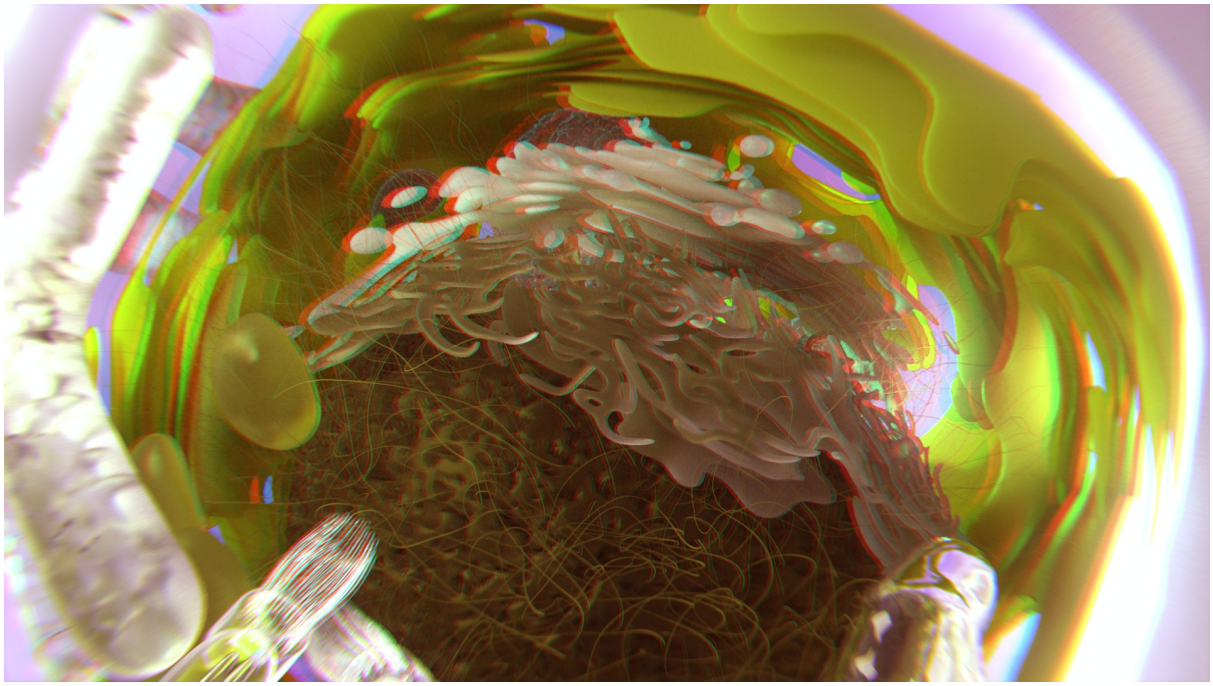


Figure 1: An anaglyph stereoscopic rendering (red/cyan) of Chlamydomonas reinhardtii

*bjoern@CELLmicrocosmos.org

3 Animation and Visualization

The colors of the different cell components are basically derived from the natural colors which are visible using a light microscope. For example, the color of the Chloroplast is green, caused by the Chlorophyl. There is also a number of publications and videos which are presenting the movement of the flagella. These sources have been taken into account during the modeling process [5].

4 Outlook

A first version of this video was presented at the conference “Prospects and challenges for the development of algal biotechnology” and the CELLmicrocosmos neXt workshop in September 2014. Future approaches may be the integration of the molecular level, e.g., by modeling the photo system complexes, by using the MembraneEditor [6].

Acknowledgments

This project is sponsored in part by the program “Bioinformatics of Signaling Networks” at the Center for Biotechnology of Bielefeld University, and by the Ministry of Innovation, Science, Research and Technology of North Rhine-Westphalia (MIWFT) as part of the research cooperation “MoRitS - Model-based Realization of intelligent Systems in Nano- and Biotechnologies” (grant no. 321 - 8.03.04.03 2012/02).

References

[1] B. Sommer. CELLmicrocosmos – Integrative cell modeling at the molecular, mesoscopic and functional level. Doctorate Thesis, Bielefeld University, Bielefeld, Germany, 2012.

[2] blender.org - Home of the Blender project - Free and Open 3D Creation Software. [Online]. Available: http://www.blender.org/. [Accessed: 22-Jul-2014].

[3] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, and B. Schmid. Fiji: an open-source platform for biological-image analysis. Nature Methods, 9(7):676–682, 2012.

[4] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming, and B. Kormeier. CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. in *BIOSTEC BIOINFORMATICS 2010*, 2010:90–95.

[5] V. F. Geyer, F. Jülicher, J. Howard, and B. M. Friedrich. Cell-body rocking is a dominant mechanism for flagellar synchronization in a swimming alga. Proceedings of the National Academy of Sciences, 110(45):18058–18063, 2013.

[6] B. Sommer, T. Dingersen, C. Gamroth, S. E. Schneider, S. Rubert, J. Krüger, and K. J. Dietz. CELLmicrocosmos 2.2 MembraneEditor: a modular interactive shape-based software approach to solve heterogeneous Membrane Packing Problems. Journal of Chemical Information and Modeling, 5(51):1165–1182, 2011.

Evaluation and Visualization of Pathway Efficiency based on Subcellular Protein Localizations

Olga V. Popik[1,*], Björn Sommer[2], Ralf Hofestädt[2], Vladimir A. Ivanisenko[1]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

[2] Bio-/Medical Informatics Department, Bielefeld University, Bielefeld, Germany

1 Introduction

Intracellular protein localization plays an important role in cell functioning as biological processes are usually distributed over different intracellular components. In some cell components processes might proceed with a high rate, and in others the rate is low.

For visualizing and analyzing subcellular localizations, we presented the subcellular localization charts in 2013 [1]. After a concrete localization scenario is found, it can be visualized in the context of a virtual cell environment. To acquire the heterogeneous localization data, two information sources where combined with the CELLmicrocosmos 4.2 PathwayIntegration, DAWIS-M.D. and ANDCell [2], [3].

Here, instead of analyzing only a subset of proteins or genes, a new approach is proposed to obtain a more global view of protein localizations. For this purpose, the localizations of all KEGG pathways were analyzed [4]. The results are combined in a localization matrix, showing cell component-internal reactions and those reactions occurring between two proteins localized in two different cell components.

2 Localization-based PPI frequency matrices

There are various databases containing data about PPI and intracellular localizations. ANDSystem integrates protein intracellular localization data and information on PPI, extracted from various databases such as UniProt, IntAct, KEGG, etc. as well as text-mining-based information derived from PubMed abstracts [5]. ANDSystem internally uses unique keys, mapping different synonyms to specific terms by using a mapping table.

By using data from ANDSystem we constructed a frequency matrix of PPI for all pair-wise combinations out of 14 intracellular localizations. In order to get a global view of subcellular reaction rates, the intra-localization rates and inter-localization rates are summarized in our approach. We propose that the resulting sum is an appropriate approach to estimate the overall pathway rate. The intracellular localization efficiency matrix $M_{i,j}$: is defined as

$M_{i,j} = \frac{K_{i,j}}{N_i * N_j}$, where

j, *i* are the indexes of the localizations, N_i, N_j are the number of proteins localized in *i* and *j* localizations, and $K_{i,j}$ is the number of interactions between proteins from localization *i* and *j*. The diagonal elements of each matrix show the localization-internal PPI frequency, and all remaining elements show the PPI frequency between different localizations. Initial analysis of the matrices revealed that the frequency of intra-localization PPI is higher than the one of inter-localization PPI.

*popik-olga@bionet.nsc.ru

3 Localization Visualization

The final matrix was used to generate a subcellular localization scenario by using the CELLmicrocosmos 4.2.1 PathwayIntegration [1], [6]. Figure 1 shows the result. The human KEGG pathways were evaluated using the matrix, ranked and compared with random pathways. The efficiency values can elucidate the problem on process flow differentiation, but to gain a more comprehensive picture, such data should be preferably visualized in 3D. The subcellular visualization is showing which regions are more preferable for highly efficient reactions. In the future, the proposed methods can be completed by data from other databases, such as BRENDA or BioModels, and extended with data concerning protein distribution over tissues.

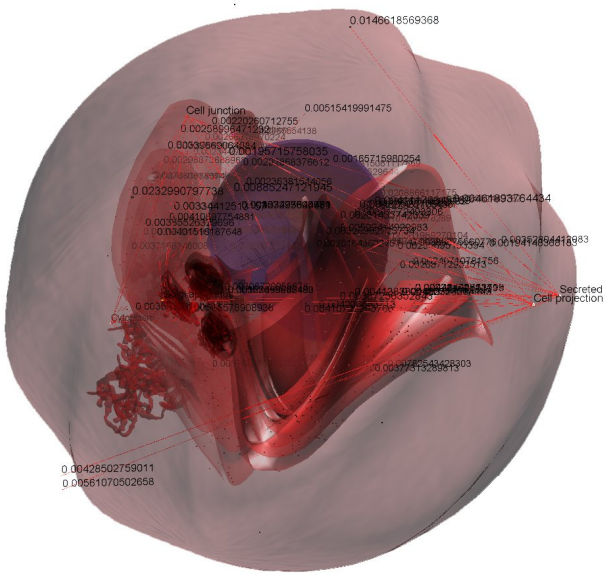


Figure 1: The matrix visualization in the context of an animal cell model using CELLmicrocosmos 4.2.1 PathwayIntegration, correlating the matrix values with the interrelated cell components

References

[1] B. Sommer, B. Kormeier, P. S. Demenkov, P. Arrigo, K. Hippe, Ö. Ates, A. V. Kochetov, V. A. Ivanisenko, N. A. Kolchanov, and R. Hofestädt. Subcellular Localization Charts: A new visual methodology for the semi-automatic localization of protein-related data sets. Journal of Bioinformatics and Computational Biology, 11(1):1340005, 2013.

[2] O. A. Podkolodnaya, E. E. Yarkova, P. S. Demenkov, O. S. Konovalova, V. A. Ivanisenko, and N. A. Kolchanov. Application of the ANDCell computer system to reconstruction and analysis of associative networks describing potential relationships between myopia and glaucoma. Russian Journal of Genetics: Applied Research, 1(1):21–28, 2011.

[3] K. Hippe, B. Kormeier, T. Töpel, S. Janowski, and R. Hofestädt. DAWIS-MD - a data warehouse system for metabolic data. GI Jahrestagung, 2:720–725, 2010.

[4] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research, 40(D1):D109–D114, 2012.

[5] P. S. Demenkov, T. V. Ivanisenko, N. A. Kolchanov, and V. A. Ivanisenko. ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. In silico biology, 11(3):149–161, 2012.

[6] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming, and B. Kormeier. CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. in *BIOSTEC BIOINFORMATICS 2010*, 2010:90–95.

Top-down systems biology approaches for "omics"-based tissue and biofluid analytics

Kirill A. Veselkov*

Faculty of Medicine, Department of Surgery & Cancer,
Imperial College London, London SW7 2AZ, United Kingdom

Top-down systems biology identifies molecular interaction networks in health and disease using system-wide data generated by high-throughput –omics technologies. The rapid advances in the post-genomic analytical technologies, most notably Mass Spectrometry (MS), has made possible to simultaneously analyze hundreds to thousands of metabolites in biological fluids and intact tissues at a relatively low cost. The ability to generate, model, and interpret such metabolic data in relation to benchmark patient clinical information, enabled by powerful bioinformatics platforms, makes personalized medicine a tangible proposition. The advancements in the field of metabolic phenotyping have focused primarily on instrument development and validation of the MS technologies.

However, despite the significant advances in these areas, the major barrier to its wider adoption and transition into main stream medicine and industry has been the challenge of interpreting and managing the huge data sets MS generates. The interpretation of metabolic data is dependent on the advances made by reconstruction of molecular networks via cellular bottom-up systems biology models.

This presentation outlines innovative bioinformatics solutions for emerging technologies of mass spectrometry imaging (MSI) and rapid evaporative ionization mass spectrometry (REIMS), targeting unmet medical needs of cancer network-driven diagnosis and rapid pathogen identification in sepsis. The translational bioinformatics platforms are presented that allow intuitive clinically-oriented interrogation of MS datasets in an integrated and streamlined fashion. These developments aim to overcome current methodological roadblocks for effective clinical translation of MS technologies since the existing workflows for data processing rely on a heterogeneous array of bioinformatic packages which are poorly optimized and only suitable for specialists in the field [1], [2].

References

[1] N. Abbassi-Ghadi, K. Veselkov, S. Kumar, J. Huang, E. Jones, N. Strittmatter, H. Kudo, R. Goldin, Z. Takáts, and G. B. Hanna. Discrimination of lymph node metastases using desorption electrospray ionisation-mass spectrometry imaging. Chemical Communications, 50(28):3661–3664, 2014.

[2] K. A. Veselkov, R. Mirnezami, N. Strittmatter, R. D. Goldin, J. Kinross, A. V. Speller, T. Abramov, E. A. Jones, A. Darzi, and E. Holmes. Chemo-informatic strategy for imaging mass spectrometry-based hyperspectral profiling of lipid signatures in colorectal cancer. Proceedings of the National Academy of Sciences, 111(3):1216–1221, 2014.

*kirill.veselkov04@imperial.ac.uk

Web-based Visualization with D3

Ahmet R. Öztürk*

Bioinformatics Group, Genome and Stem Cell Center, Erciyes University, Kayseri, Turkey

1 Current Visualization Approaches and D3

Visual representation of complex data reveals the motifs and patterns which cannot be easily understood by basic statistical approaches. However, in case of the analysis of real data, the phenomenon of interest usually involves many parameters and a researcher needs many different plots to visualize the diverse aspects of analysis results. In addition, current paradigm of visualization is based on displaying static images on a 2D surface. It might be a computer screen, a wall, a piece of paper, or any kind of material that provides a support for the image.

A solution for displaying complex data visualizations is to make visualization interactive through a computer or a smart device. This gives the user much more flexibility for looking at different aspects of the data with little effort. There are many standalone tools for specific visualization purposes such as Cytoscape and Gephi for network visualization whereas standalone software has some serious limitations for interaction flexibility and sharing the visualizations. Web technologies provide a great deal of opportunity for interactive data visualization and among many JavaScript library alternatives, D3 is well ahead [1]. D3 is a JavaScript library [2] that allows programmers to modify and transform any part of a web page and associate HTML elements with data. This results in interactive and dynamic graphs with a very wide range of layout options through a modern browser.

1.1 Visual consistency of D3

Another advantage of using D3 is object constancy. Usually, when you generate a plot with different elements, positions of each object are calculated from scratch using a built-in algorithm. This independent nature of calculations prevents continuity of the graph in response to a change of a parameter in underlaying data. This is a counterintuitive approach regarding our cognitive expectations. The visual continuity is achieved through animated transitions. There are examples showing how visual continuity helps the user to easily follow changes in response to a parameter change [3].

A suitable example for displaying the results of calculations of graph element positions without considering visual continuity is a BLOSUM network using specific BLOSUM matrices. Consider the network graphs below in Figure 1, with different BLOSUM matrices of 62% and 45% similarities, respectively.

On the graphs shown in Figure 1, when the matrix parameters change, positions of each node are calculated from the beginning. However, when you generate the initial network with D3 and enable the user to select the new matrix to be displayed, user can easily follow the changes in visualization as animations such as the introduction of new edges and smooth movement of nodes in a force-directed graph. Interactive version of this graph can be found at http://ahmetrasit.com/blosum/. User can move the mouse over an amino acid and get a detailed information about its molecular weight, frequency, codons, a brief information and chemical representation of the molecule at the bottom-right corner of the page. The user can also change BLOSUM identity value using the slider on the top-left, change substitution score threshold to change the network structure using the second slider on the top, change node color to represent different chemical properties

*ahmet.ozturk@metu.edu.tr

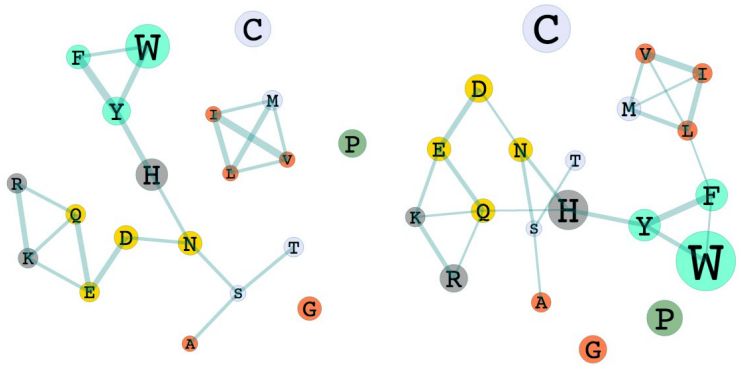


Figure 1: Network representations of BLOSUM62 (left) and BLOSUM45 (right) matrices.

of amino acids (horizontal bars on the center-right side is the color map of selected chemical properties), and lastly, change node size to represent different numerical properties of each amino acid according to the categories listed on the top-right side.

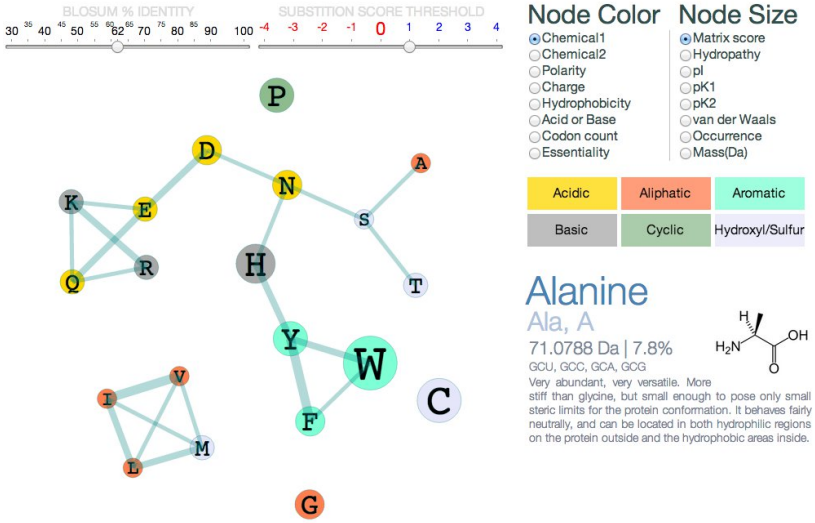


Figure 2: Screenshot of the Interactive BLOSUM Graph GUI available at http://ahmetrasit.com/blosum/

1.2 Content of the Workshop

In this talk, I will give some examples of how visualization helps us understand the story behind the data and I will demonstrate how to manipulate HTML elements using D3. Finally, I will introduce basic interactive graphs created with D3 and show how to manipulate a single graph layout with different data.

References

[1] D. Skau. Why D3.js is So Great for Data Visualization | Visually Blog. [Online]. Available: http://blog.visual.ly/why-d3-js-is-so-great-for-data-visualization. [Accessed: 17-Sep-2014].

[2] M. Bostock. D3.js - Data-Driven Documents. [Online]. Available: http://d3js.org. [Accessed: 13-Sep-2014].

[3] nsonnad. Object constancy with multiple sets of time-series. [Online]. Available: http://bl.ocks.org/nsonnad/4175202. [Accessed: 17-Sep-2014].

Towards a Web-based Visualization for the CELLmicrocosmos 4 PathwayIntegration

Gökhan Kovanci, Björn Sommer*

Bio-/Medical Informatics Department, Bielefeld University/Center for Biotechnology
Universitätsstr. 25, 33615 Bielefeld, Germany

1 Introduction

The CELLmicrocosmos 4.2 PathwayIntegration (CmPI) is a tool to localize and visualize protein and gene data in the context of a virtual cell environment [1]. Here, an alternative, easy-to-use web browser-based tool is presented (CmPIwv, “wv” for web viewer), providing only the basic functionality of CmPI.

First, the basic workflow of CmPI should be discussed. A cell model is constructed based on different available cell component models, or an existing cell model is loaded. Then, protein-related or gene-related data or networks are imported. Finally, information is required to combine the virtual cell with the appropriate data. For this purpose, subcellular localization entries from different biomedical databases is acquired. This data has to be analyzed, because for some proteins or genes many different potential localizations are available, for other ones there is no data available at all. The analysis of this data is supported by different visualization approaches, particularly the subcellular localization charts [2]. The result can be interactively visualized as a virtual cell model representing a specific localization scenario.

CmPI is implemented in Java and Java 3D 1.5.2 and has to be installed to a PC. Basically, there are three user groups for CmPI: "authors" who can create and edit a specific virtual localization scenario, "reviewers" who are interested in a specific application case but do not want to analyze the basic data, and the “viewers” who are interested in exploring the final model. CmPI can be used by authors after a short adaptation phase. It takes advantage of Java WebStart technology, but requires some time to finish the installation process. Moreover, the functionality and features of CmPI are far beyond the (re-)viewer's scope. The obvious idea would be to create a simplified version of CmPI by developing a Java applet which runs in the browser. But the nowadays very restricted access to applets in the browser – caused by security issues – plus the requirement to install Java 3D in the background onto the client computer makes it a very complicated task for the developer and the user. Therefore, an alternative approach is required optimized for the (re-)viewer groups.

2 CmPI Web Viewer Concept

The optimal tool would be an application, where only a click on a web link is required to load a complete virtual localization scenario into a browser window, compatible to different systems, such as PCs, tablets, mobile phones, etc. In this way it will be possible to send a link to the collaborators, or to publish a link online or in the context of a manuscript, which can then be explored by the (re-)viewers. For this purpose, a web tool is being developed based on a three tier architecture:

*bjoern@CELLmicrocosmos.org

[1] Three.js is a WebGL framework which provides a scene graph-based API and direct visualization in most browsers, such as Chrome®, Firefox, Internet Explorer® [3].

[2] D3 is a JavaScript-based visualization library which enables fast visualization of 2D graphics inside a browser [4].

[3] PHP is used for interchanging content with the server-based application and the local computer. It enables to visualize local files as well as to load cell environments into the application which are uploaded to an external webspace.

The layout of the web tool is based on the basic layout of CmPI. Two different subcellular localization charts are available: the localization overview chart which summarizes all localizations found for the complete pathway, and the protein localizations chart, which shows all available localization for each protein at a time [2]. Figure 1 shows a KEGG example based on a previously published application case [1], [5].

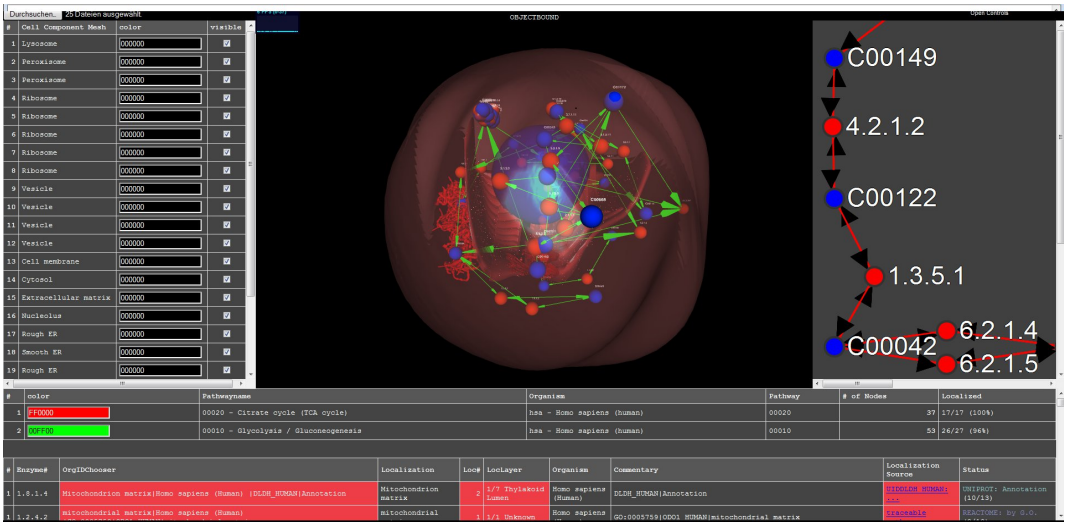


Figure 1: The CmPI WebViewer shows a virtual cell localization scenario correlated with the KEGG citrate cycle and the glycolysis as shown in the browser google Chrome.

3 References

[1] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming, and B. Kormeier. CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. in *BIOSTEC BIOINFORMATICS 2010*, 2010:90-95.

[2] B. Sommer, B. Kormeier, P. S. Demenkov, P. Arrigo, K. Hippe, Ö. Ates, A. V. Kochetov, V. A. Ivanisenko, N. A. Kolchanov, and R. Hofestädt. Subcellular Localization Charts: A new visual methodology for the semi-automatic localization of protein-related data sets. Journal of Bioinformatics and Computational Biology, 11(1):1340005, 2013.

[3] three.js - Javascript 3D library. [Online]. Available: http://threejs.org. [Accessed: 13-Sep-2014].

[4] D3.js - Data-Driven Documents. [Online]. Available: http://d3js.org. [Accessed: 13-Sep-2014].

[5] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, and T. Tokimatsu. KEGG for linking genomes to life and the environment. Nucleic Acids Research, 36(suppl 1):D480, 2008.

Web-based MD Simulations using MoSGrid

Jens Krüger*

Applied Bioinformatics Group, University of Tübingen
Sand 14, 72076 Tübingen, Germany

1 Introduction

Molecular simulations play an increasingly important role in the fields of chemistry, biology, physics and life sciences. Despite their importance most tool suites and applications are lacking a user-friendly interface and direct connection to high-performance or high-throughput compute resources (HPC/HTC). The molecular simulation grid (MosGrid; www.mosgrid.de) is a web-based science gateway bridging this gap and enabling easy access to molecular simulations [1].

2 MoSGrid

The gateway represents a multi-layered stack of different services needed to compose, submit, monitor and analyse simulations belonging to the domains of quantum chemistry, molecular dynamics or docking. The only component visible for users is the Liferay-based webpage, hiding the complexity beneath. In brief, MoSGrid relies on WSPGRADE/gUSE for workflow and job management, connected via the UNICORE middleware to HPC and HTC resources [2], using a federated storage solution based on XtreemFS. The authentication relies on personal X.509 certificates and trust delegation based on them. Currently access is possible for all researchers with an affiliation to a German research institution.

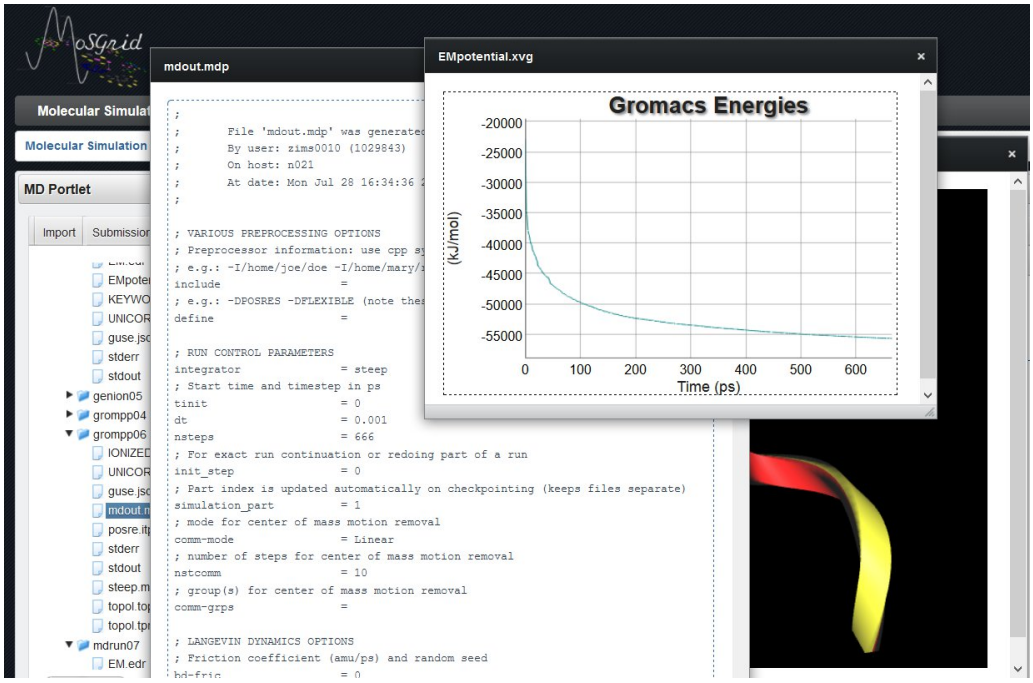


Figure 1: The MD portlet of the MoSGrid portal and some of its functionality are shown. In addition to monitoring and navigation capabilities, the visualization of molecules, numerical data and textual information is possible.

*krueger@informatik.uni-tuebingen.de

3 Molecular Dynamics

The focus of the presentation will be on molecular dynamics (MD) simulations, while MoSGrid is able to support a variety of other simulation types. MD simulations in particular with GROMACS, follow a distinct pattern: a molecular input structure is pre-processed using multiple small applications, a job description is compiled, the actual simulation is carried out, the raw data is processed and analyzed [3]. Such a simulation protocol can be conveniently represented as workflow. The presentation will highlight all essential steps for such a typical MD workflow, from the simulation system preparation, over equilibration steps to the analysis of a production run.

In order to ensure full metadata annotation and storage of provenance information consequently ensuring full reproducibility a molecular simulation markup language (MSML) template is associated to each workflow. Furthermore this documents serves as basis to dynamically create the graphical user interface (see Figure 1). Workflows for different kinds of MD simulations are provided aiming at different scientific use cases and different levels of user expertise.

4 Outlook

The maintenance and extension of the MoSGrid is ongoing work. Depending on the needs of the community modifications, additions and extensions are planned and implemented. In the past auxiliary applications such as a direct UNICORE connection for the Cm2 MembraneEditor or APL@Voro were developed [4]. Current endeavour focuses on the connection to PRACE resources via UNICORE and the implementation of an US version of MoSGrid, enabling convenient access to XSEDE resources.

Acknowledgments

The author would like to thank the BMBF (German Federal Ministry of Education and Research) for the opportunity to do research in the MoSGrid project (reference 01IG09006). The research leading to these results has partially been supported by the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 283481 (SCI-BUS). Special thanks are due to the university compute centers supporting MoSGrid.

References

[1] J. Krüger, R. Grunzke, S. Gesing, S. Breuers, A. Brinkmann, L. de la Garza, O. Kohlbacher, M. Kruse, W.E. Nagel, L. Packschies, R. Müller-Pfefferkorn, P. Schäfer, C. Schärfe, T. Steinke, T. Schlemmer, K. Warzecha, A. Zink and S. Herres-Pawlis: The MoSGrid Science Gateway – A Complete Solution for Molecular Simulations, *Journal of Chemical Theory and Computation*, 10 (6):2232–2245, 2014.

[2] J. Krüger, R. Grunzke, S. Herres-Pawlis, A. Hoffmann, L. de la Garza, O. Kohlbacher, W.E. Nagel and S. Gesing: Performance Studies on Distributed Virtual Screening, *BioMed Research International*, Volume 2014, Article ID 624024, 2014.

[3] S. Pronk, S. Páll1, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess and E. Lindahl: GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics*, 29 (7): 845-854, 2013.

[4] G. Lukat, J. Krüger and B. Sommer: APL@Voro: A Voronoi-Based Membrane Analysis Tool for GROMACS Trajectories, *Journal of Chemical Information and Modeling*, 53(11): 2908–2925, 2013.

Towards Pathway Prediction and Subcellular Localization by using FraMeTex

Thorben Wallmeyer*, Björn Sommer, Benjamin Kormeier

Bio-/Medical Informatics Department, Bielefeld University/Center for Biotechnology
Universitätsstr. 25, 33615 Bielefeld, Germany

1 Introduction

The visualization and localization of biochemical networks represent major tasks of bioinformatics. In the future, they could be used as additional criteria to find potential -omics interaction partners. Moreover, this visualization and localization can help scientists to find appropriate experiments or to create detailed problem-oriented pathways. Today, we can use database integration and text mining methods for the prediction of biological networks. However, we can use the same applications for the localization prediction of biological networks and their components [1]. Using this information, we are able to model and visualize biochemical networks in 2D or 3D. For this purpose, we use our database integration infrastructure BioDWH with the DAWIS-M.D., as well as ANDCell as the base for the subcellular localization of the CELLmicrocosmos 4.2 PathwayIntegration (CmPI) project [1–4]. Here, we discuss how this approach can be extended by using the new FraMeTex text mining framework.

2 Data integration

High-throughput methods generate, in short time, data concerning the whole genome of an organism. Hence, enormous, heterogeneous and versatile data is produced. The total number of databases, as well the data itself, is continuously increasing. Data distribution and heterogeneity causes big problems in biological data integration. Therefore, we introduced a bioinformatics data warehouse information system DAWIS-M.D. that can be accessed with the CmPI for tailor-made 3D cell and pathway visualization. DAWIS-M.D. is a platform-independent data warehouse approach for metabolic data that is based on the BioDWH data warehouse infrastructure. The information system combines a number of relevant biological databases. For the subcellular localization, BRENDA, Gene Ontology, Reactome, and UniProt are used. The content of the data warehouse is divided into 12 diverse domains, which can be queried via a web-bases graphical user interface that can be accessed with any common web-browser. The DAWIS-M.D. application provides search forms for the following domains: Compound, Disease, Drug, Transcription Factor, Enzyme, Gene, Glycan, Gene Ontology, Pathway, Protein, Reaction and Reaction Pair. Moreover, it is possible to identify relationships and interactions spanning multiple biological domains [3]. DAWIS-M.D. provides also a web-service to the CmPI. Networks can be localized and displayed in a cell environment, edited and extended by using the CmPI software application [1].

3 Text mining

The number of text mining algorithms that can be helpful in reconstructing biological networks from text data is unclear. Therefore, selecting and using the best text mining algorithm for a specific task is challenging. In addition, the heterogeneity of their interfaces and data structures for representing extracted facts is even complicating their

*thorben.wallmeyer@googlemail.com

application. We addressed this problem by developing a powerful framework. FraMeTex (Framework for Medical Text mining and knowledge engineering) offers a highly customizable analysis pipeline that is able to consult different text mining algorithms for knowledge extraction. Furthermore, it supports the adaption of text sources to be analyzed as well as the filtering of relevant data. Extracted knowledge like biological pathways or localizations can afterwards be persisted within a deductive knowledgebase. By using additional rules it is able to identify additional relations. The offered analysis pipeline of FraMeTex is divided into several logical modules that provide a convenient API (see figure). It allows a seamless and flexible integration of its functional modules within almost any application. Each module can be used standalone or chained with other modules to build a complex workflow. We developed a workflow that filters Medline abstracts by keywords and analyzes them with two different text mining algorithms. This enables us to extract reliable pathway and localization data from Medline. GENIAs Named Entity Recognition (NER) web service is responsible for tagging potential biomedical entities within previously selected Medline abstracts [5]. It distinguishes between proteins, cell lines, cell types, DNA and RNA terms. Additionally Enju performs a semantic parsing of the abstracts and returns Predicate Argument Structures (PAS). The analysis results of both algorithms are finally merged. PAS that deal with biomedical entities are most likely supposed to describe biological pathways or localizations.

4 Conclusion and Outlook

The FrameTex localization should be used to try to extend and verify the knowledge of CmPI. We will do this by using the described and a newly developed workflow. The additional workflow will use different text mining algorithms which allows an additional cross check of results. Especially the critical NER process may be performed better by other algorithms.

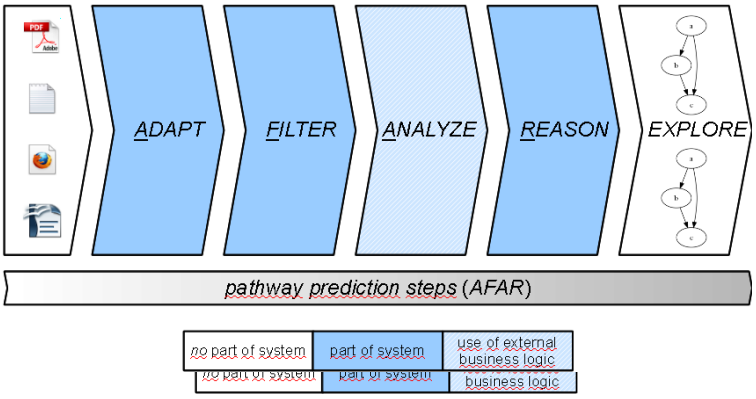


Figure 1: The FraMeTex pipeline

5 References

[1] B. Sommer, B. Kormeier, P. S. Demenkov, P. Arrigo, K. Hippe, Ö. Ates, A. V. Kochetov, V. A. Ivanisenko, N. A. Kolchanov, and R. Hofestädt. Subcellular Localization Charts: A new visual methodology for the semi-automatic localization of protein-related data sets. Journal of Bioinformatics and Computational Biology, 11(1):1340005, 2013.

[2] T. Töpel, B. Kormeier, A. Klassen, and R. Hofestädt. BioDWH: a data warehouse kit for life science data integration. Journal of Integrative Bioinformatics, 5(2):93, 2008.

[3] K. Hippe, B. Kormeier, T. Töpel, S. Janowski, and R. Hofestädt. DAWIS-MD-a data warehouse system for metabolic data. GI Jahrestagung, 2:720–725, 2010.

[4] O. A. Podkolodnaya, E. E. Yarkova, P. S. Demenkov, O. S. Konovalova, V. A. Ivanisenko, and N. A. Kolchanov. Application of the ANDCell computer system to reconstruction and analysis of associative networks describing potential relationships between myopia and glaucoma. Russian Journal of Genetics: Applied Research, 1(1):21–28, 2011.

[5] J.-D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii. GENIA corpus—a semantically annotated corpus for bio-textmining. Bioinformatics, 19(suppl 1):i180–i182, 2003.

A novel Lecture Program in Bioinformatics: Interdisciplinary Cell Visualization and Modeling

Björn Sommer[1,2*], Ming Chen[1,2]

[1] Bio-/Medical Informatics Department, Bielefeld University/Center for Biotechnology
Universitätsstr. 25, 33615 Bielefeld, Germany

[2] Bioinformatics Group, Zheijang University/College of Life Sciences,
Yuhagntang Road No. 866, 310058 Hangzhou, P.R. China

1 Introduction

The creation of a virtual cell scenario is a multidisciplinary vision. Cell modeling is a specific field which can be applied to a number of quite diverse approaches, starting from simulation-based ones, such as the Virtual Cell project – which is able to convert biological descriptions into a mathematical system of ordinary and/or partial differential equations – to visualization-based approaches, such as cellPACK – which is based on the idea to create visual filled cell representations [1], [2]. Integrative Bioinformatics is a field which tries to unite different information sources to solve biological questions. The CELLmicrocosmos project consists of a number of different projects trying to approach the vision of a virtual cell environment, such as [3], [4]. Starting from 2004, a remarkable number of students were involved in the aforementioned projects. In addition to our custom tools, also external tools, such as Blender, Fiji (IsJustImageJ), or GROMACS played an important role in our daily work [5–7].

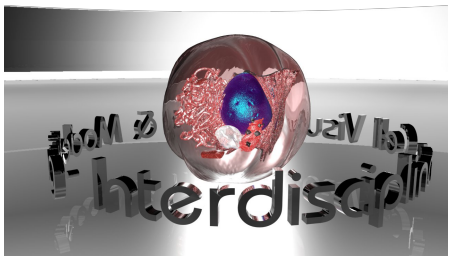


Figure 1: The logo of "Interdisciplinary Cell Visualization and Modeling". The cell model goes back to the initial Bachelor thesis from 2004 created with Autodesk 3ds max®.

2 The Lecture Program

In 2012, the idea emerged to develop a teaching module. First, these course should provide the students with background knowledge concerning cytology, microscopy, modeling and simulation, as well as bioinformatics. And second, practical lessons should be included. Finally, a module was initiated and developed in 2012 consisting of three different parts:

1. the lecture "Interdisciplinary Cell Visualization and Modeling", introducing the theoretical background,
2. different seminars, focusing and extending specific lecture topics (see below),
3. and the student project "CELLmicrocosmos Cell Modeling/Cell Visualization", where the practical work – programing, 3D stereoscopic modeling and visualization, database applications – was carried out.

*bjoern@CELLmicrocosmos.org

Three different seminars were offered. "(Multidisciplinary) Cell Visualization" was focusing on cytological visualization, e.g. in school books or digital edutainment. The seminar "Stereoscopic 3D Visualization" discussed the technical background and the difficulties of demanding biological stereoscopic visualizations, which resulted in a short film presentation in a cinema and an accompanying publication [8]. Also, some work presented at the CELLmicrocosmos neXt workshop is a result of this new teaching module. Moreover, the seminar "Genome-wide biological Network Modeling and Analysis" discussed aspects like graph analysis, Petri net simulation, and the visualization of biological networks.

Based on this Sino-German collaboration between the Bio-/Medical Informatics Department Bielefeld and Bioinformatics Group of Zheijang University, a bilateral exchange teaching and workshop program was started and there are a number of plans for our future collaborations. From 2012 to 2014, workshop events of this program took place in Bielefeld, Hangzhou, Harbin, and Tongliao. Moreover, a first book including some ideas of this collaboration was published in 2014 [9].

3 Acknowledgments

We would like to address our thanks for the kind support to the "International guest lectureship program of Bielefeld University" (2012 and 2014), as well as the "NSFC Research Fund for International Young Scientists" (No. 31450110068, 2014-2015).

4 References

[1] L. M. Loew and J. C. Schaff. The Virtual Cell: a software environment for computational cell biology. TRENDS in Biotechnology, 19(10):401–406, 2001.

[2] G. Johnson, D. Goodsell, L. Autin, S. Forli, M. Sanner, and A. Olson. 3D molecular models of whole HIV-1 virions generated with cellPACK. Faraday Discussions, 169:1–21, 2014.

[3] B. Sommer, T. Dingersen, C. Gamroth, S. E. Schneider, S. Rubert, J. Krüger, and K. J. Dietz. CELLmicrocosmos 2.2 MembraneEditor: a modular interactive shape-based software approach to solve heterogeneous Membrane Packing Problems. Journal of Chemical Information and Modeling, 5(51):1165–1182, 2011.

[4] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming, and B. Kormeier. CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. in *BIOSTEC BIOINFORMATICS 2010*, 2010:90–95.

[5] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, and B. Schmid. Fiji: an open-source platform for biological-image analysis. Nature Methods, 9(7):676–682, 2012.

[6] blender.org - Home of the Blender project - Free and Open 3D Creation Software. [Online]. Available: http://www.blender.org/. [Accessed: 22-Jul-2014].

[7] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput, 4(3):435–447, 2008.

[8] B. Sommer, C. Bender, T. Hoppe, C. Gamroth, and L. Jelonek. Stereoscopic cell visualization: from mesoscopic to molecular scale. Journal of Electronic Imaging, 23(1):011007–1 – 011007–10, 2014.

[9] M. Chen and R. Hofestädt. Approaches in Integrative Bioinformatics: Towards the Virtual Cell. Springer Publishing Company, Incorporated, 2014.