PROCESS OF GENETIC MODIFICATION OF MICROORGANISMS FROM THE POINT OF VIEW OF THE SOFTWARE ENGINEERING

JURIJS MEITALOVS

Biosystems Group, Department of Computer Systems, Latvia University of Agriculture, Latvia
jurijs.meitalovs@gmail.com

Abstract: *Genetically modification of organisms using approaches of systems and synthetic biology becomes very popular in recent years. Implementing genes from one organism to another or creating synthetic genes is useful in many fields to create products that are necessary for people . Also it is possible to use genes or devices to get new functions of organisms which not occur in the nature. Also this science uses approaches of engineering to create modified organisms. In this paper is shown modified organisms development process from the point of view of a software engineer. Also it focuses on the usage of modeling techniques in this process creation of a prototype of system.*

Keywords: synthetic biology, systems biology, genetically modified organisms, modeling software tools

Introduction

Genetically modified organisms engineering has received considerable attention in recent years for enabling the production of chemicals and novel functionality achievement. A genetically modified organism (GMO) is referred to as a living organism whose genome has been modified by the introduction of a gene able to express an additional protein that confers new characteristics (Mannelli et al., 2003).

Industrial genetically modified microorganisms, such as *Escherichia coli*, have been extensively used in major industrial processes including food, chemicals, pharmaceuticals and environment (Kizer et al., 2008; Lee et al., 2008; Shams Yazdani and Gonzalez, 2008; Trinh and Srienc, 2009). The microbial metabolism transforms feeding substrates into a wide variety of cellular products, many of which are materials for industrial and commercial purposes. Standard industrial operations have been developed to cultivate microbial cells in a controlled growth environment (Wang and Hatzimanikatis, 2006, W.-C. Lee and Huang, 2000)

There are many steps in creating genetically modified organisms. The main steps that be distributed to the following steps: design, fabrication, integration and testing. First step - target of modification should be defined – getting of new functions, new product production or utilization of some substrate. Different possible options of organisms where such a target can be implemented and possible sources of functional parts should be examined. That can be one of the most time-consuming processes and the result of it will affect the next steps.

This process can be viewed from the point of view of software engineering. We can assume that a cell is a computer, genome is an operating system and additional gene is a program (Andrianantoandro et al., 2006) . In this case it is possible to use software engineering approach to program new function or to implement genetic modification into organism. Reprogramming a cell involves the creation of synthetic biological components by adding, removing, or changing genes and proteins. That means that software engineering methods, e.g. specification, design, modeling, implementation, validation, can be used during genetically modified organisms production (Andrianantoandro et al., 2006). For example, a waterfall model can be used for this purpose (See Fig.1). The waterfall model is a sequential design process, often used in software development processes, in which progress is as flowing steadily downwards through the phases of development process (Wescon, 1970). After each step it is possible to return to previous step and to make correction to improve results. Also in the end of the process in is necessary to compare the achieved result with specification. Modeling tasks can be performed and we can assume that modeling is prototyping of a real organism.

As a result, design of synthetic biological systems has become an iterative process of modeling, construction, and experimental testing that continues until a system achieves the desired behavior. There can be found a possible problem – a mistake in one of the previous steps cannot be found at once and can cause possible troubles in the next steps. To reduce the possibility of such problem it is possible to use a technical journal where it is necessary to make documentations about all the changes in the development process.

The aim of this paper is to the show process of implementation of genetically modification into microorganism from the point of view of a computer scientist.

Modeling tasks in GMO production

It is important to predict behavior of systems using modeling in the phase when they are under development and therefore cannot be examined experimentally (Stepina and Stalidzans, 2010). The process begins with the abstract design and specification of devices, modules, or organisms, and is often guided by mathematical models where parameters of organisms and devices are described. However, such initial attempts rarely yield fully functional implementations due to incomplete biological information (Andrianantoandro et al., 2006). It is possible to create the model in many ways, e.g. at genome level or at metabolic network level. There are

different software tools for it. For genome level modeling such tools as Tinkercell, Genocad, Genedesigner and others can be used. That can help understand how a modified gene should work, is it possible to make such modification and are the correct genes chosen.

Various computational methods are based on constraint-based modeling, which allows to predict the effect of genetic manipulations on cellular metabolism considering a genome- scale metabolic network (Tepper and Shlomi, 2010). For constraint-based metabolic network modeling can be used tools like Cobra, Copasi, Metatool and many others. They can be used on different levels of modeling - stoichiometry, dynamic modeling, and elementary mode analysis. This is the level of metabolic reconstruction of the whole organism or a part of it. Using these methods it is possible to understand how organisms will work after modification, will metabolic network supply all necessary compounds and how metabolic kinetics works and how much possible metabolic pathways are presented. Each of these methods employs an optimization problem that searches for a specific genetic manipulation that would lead to a possible metabolic flux distribution, which maximizes the production rate of a chemical of interest (Tepper and Shlomi, 2010).

After genome level modeling is complete, next step is inserting genes or chemical reactions that represent these genes into metabolic network model (Fig. 1). Metabolic network model or metabolic reconstruction model shows all possible chemical reaction in the organism or in the part of organism. If this model is developed in sufficient level and satisfies defined criteria, there can be implemented new functionality. If the model shows that there are some gaps in the network, different tools can be used to predict possible compounds and reactions that can be implemented into model, e.g. GapFill, OptKnock. That allows simplifying the model - deleting unnecessary reactions, removing deadens and cycles in the network, getting better flow through chemical network. In the end that brings a list of reactions which can be removed and which can be inserted into organism and a list of corresponding genes.
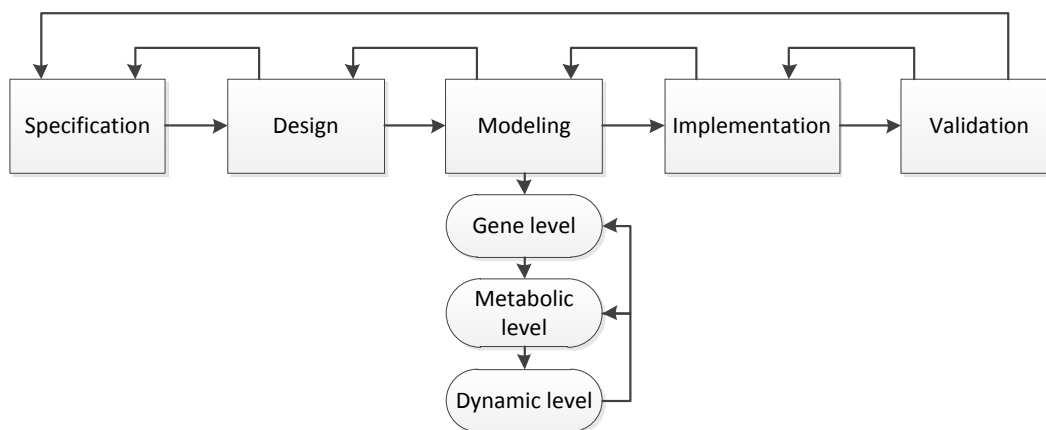
Figure 1. Iterations of GMO development process

When all necessary modifications in metabolic network are completed it is possible to make dynamic analysis of kinetic parameters of the organism. During dynamic analysis parameters estimation and optimization are made. It can show how modified organisms can behave in time scale and return data of compounds concentrations. Also it is possible to achieve necessary parameters of reactions speed and enzymes concentrations. After each modeling step it is possible to return to the previous step to improve the model.

When all parameters are obtained it is possible to make biological experiment in the lab based on the achieved data. After the modification all modified organisms are separated from unmodified. That can be done using additional markers. After that it is possible to make an analysis of microorganisms – do the planned function according specification or not. That gives back experimental data that can be used in optimization. All data can be used to make modeling process and modification process until the best result will be achieved. It depends on quality of the model and accuracy of biological modification.

Next step is validation or testing of the organism to compare it with a specification. If there are some differences in the specification and the real organisms that mean that there are some errors in others steps of the process. It is possible to make experiments with modified organism in flask at first to understand how organisms behave under specific conditions and then make cultivation in a bioreactor to test industrial production. Results between experiments in flask and bioreactor may vary because of different parameters of the environment. It is a good practice to make a design of the experiment before starting a cultivation (Tye, 2004). That can help cover all possible parameters and reduce the count of physical experiments. Also during cultivation processes the model of the process can help check how microorganisms behave under given conditions. Such model can take into account all parameters of the process and recognize the differences between the planned processes and a real one. If physical process matches with data from the model it can be assumed that the modification of the organism is made successfully.

During each start of modeling process it is necessary to mark start parameters and in the end to mark end parameters to get information about changes. Also the result of model should be marked. In case when many models are running in the same time, e.g. for optimization or parameter estimation processes, parameters of each running model should be marked. This approach can help track changes in each model and the obtained results. Data exchange format should be taken into account between different software because at? this step gaps in data can occur, even if one data standard, e.g. SBML, is used. During a physical experiment even when it is happening according to the protocol, each parameter of the process should be noted so it would be possible to repeat this experiment in the future.

It is also possible to implement cost calculation for modification into the model. If the quality of the final result and costs of the modification process is satisfying, real experiment can be started. If in the beginning of the process prediction is that it cannot return investments, it is desirable to think about making this kind of research project. If the goal of the project is to get profit but the model of all processes shows that it is not possible, it is needed to examine all the aspects of the project and make some changes in it or decline it.

Conclusions

Genetically modified organisms production is an interactive process where can be used many approaches from software engineering. Steps of software development can be performed to create new functionality of microorganisms in the similar way how new software extends functionality of the computer. Also computer modeling can be performing to create the model or prototype of organisms to see how it should work.

Computational modeling in metabolic engineering has traditionally been used to guide experimental attempts by predict the effect of genetic modifications on metabolism. Modeling involves the prediction of genetic manipulations that would lead to optimized microbial strains, maximizing the production rate for chemicals of interest (Tepper and Shlomi, 2010). From an engineering perspective, mathematical modeling is one of the most successful scientific tools available for this task. Starting with the modeling process for a given complex system the aim of the model should be specified. Rational design based on mathematical models improves system behavior during new system development (Andrianantoandro et al., 2006). The development of large models must be supported by tools that help keeping track of all the structural assumptions and enzyme kinetic terms in the model (Wiechert, 2002).

Each change in data during modeling process should be carefully documented. Also process of data transfer from one step to another should be documented. All, even small, changes in the process should be taken into account, e.g. which parameters are changed in the model during each launch and when it was made, what are the results of these changes. That can help to analyze results of changes and understand which changes are excess and minimize their count doing only needed of them. That can reduce time consuming and reduce costs of the process. The journal of changes should be accessible for all participants of the process- modeler, genetics, and biologists. Also that can help find errors in the process and avoid them.

Acknowledgments

References

Andrianantoandro, E., Basu, S., Karig, D.K., Weiss, R., 2006. Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology,* 2, 2006.0028.

Kizer, L., Pitera, D.J., Pfleger, B.F., Keasling, J.D., 2008. Application of functional genomics to pathway optimization for increased isoprenoid production. *Applied and environmental microbiology,* 74, pp. 3229-41.

Lee, S.K., Chou, H., Ham, T.S., Lee, T.S., Keasling, J.D., 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current opinion in biotechnology,* 19, pp. 556-63.

Lee, W.-C., Huang, C.-T., 2000. Modeling of ethanol fermentation using Zymomonas mobilis ATCC 10988 grown on the media containing glucose and fructose. *Biochemical Engineering Journal,* 4, pp. 217-227.

Mannelli, I., Minunni, M., Tombelli, S., Mascini, M., 2003. Quartz crystal microbalance (QCM) affinity biosensor for genetically modified organisms (GMOs) detection. *Biosensors & bioelectronics,* 18, pp. 129-40.

Shams Yazdani, S., Gonzalez, R., 2008. Engineering Escherichia coli for the efficient conversion of glycerol to ethanol and co-products. *Metabolic engineering,* 10, pp. 340-51.

Stepina, D., Stalidzans, E. 2010. Applications of modelling in synthetic biology, *In Proceedings of Applied Information and Communication Technologies 2010*, Jelgava, Latvia, pp. 10-17.

Tepper, N., Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics (Oxford, England),* 26, pp. 536-43.

Trinh, C.T., Srienc, F., 2009. Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol. *Applied and environmental microbiology* ,75, pp. 6696-705.

Tye, H., 2004. Application of statistical "design of experiments" methods in drug discovery. *Drug discovery today,* 9, pp. 485-91.

Wang, L., Hatzimanikatis, V., 2006. Metabolic engineering under uncertainty--II: analysis of yeast metabolism. *Metabolic engineering,* 8, pp. 142-59.

Wescon, I., 1970. MANAGING THE DEVELOPMENT OF LARGE SOFTWARE SYSTEMS Dr. Winston W. Rovce. *Electronics,* pp. 1-9.

Wiechert, W., 2002. Modeling and simulation: tools for metabolic engineering. *Journal of biotechnology,* 94, pp. 37-63.