MODELING AND DESIGN TOOLS FOR SYNTHETIC BIOLOGY

JURIJS MEITALOVS, NATALJA BULIPOPA, OLESJA KOVALONOKA

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

Abstract: *The aim of this paper is to review several computational tools for application in synthetic biology. Recently developed computational tools, for instance, for gap-filling, orphan reaction searching, metabolic engineering, pathway analysing, gene engineering, metabolic reconstruction and others, were reviewed and their functionality was described. Some tools have similar functionality and some tools can be used for different tasks. Tools for comparison were chosen from the most cited scientific literature.*

Keywords: synthetic biology, software tools, modeling and design tools.

Introduction

Synthetic biology brings together engineers and biologists to design and build novel biomolecular components, networks and pathways, and to use these constructs to rewire and reprogram organisms (Khalil and Collins, 2010). It is an extension of the biotechnologies that focuses on the design of new organisms not found in nature from a set of pre-characterized biological component (Gendrault et al., 2011).
Major application domains of synthetic biology apply to the medical field and environment, e.g. drugs synthesis, therapeutics, biosensing, biofuel production, pharmaceutical, biomaterials production and chemical production life understanding improvement (Khalil and Collins, 2010, Tepper and Shlomi, 2010, Gendrault et al., 2011)
Synthetic biology can be used to learn more about how natural living systems work. One way how to test our current understanding of biological system is to build an instance of the system in accordance with our current understanding of the system. Mathematical modeling plays an important and often indispensable role in synthetic biology because it serves as a crucial link between the concept and realization of a biological circuit. (Zheng and Sriram, 2010). Using models of biological systems instead of real systems and experiments is cheaper and does not require specific resources (Gendrault et al., 2011, Wiechert, 2002). Modeling is a tool to optimize the design of modified organism according to objective function. It is possible to use different computational tools for engineering approach in design and modeling (Macdonald et al., 2011).
In recent years, numerous computational tools have been developed, especially in the field of systems biology, and many of them have potential applications for synthetic biology (Marchisio et al., 2009). Several attempts have been made to describe synthetic biology software (Macdonald et al., 2011, Marchisio and Stelling, 2009, Orth and Palsson, 2010, Rubina and Stalidzans, 2010).
This article is devoted to the different groups of modeling software from the point of view of non-biologist. Below are shown some of these tools that can be used in different research processes, e.g. gene design tools, flux balance analysis, gap-filling, dynamic analysis. In different research fields of synthetic biology many processes are connected to each other, but each of them has its own purpose. In this paper are shown some of these processes.

Gene design tools

One of the first stages of metabolic network analysis is gene circuit design. Genetic regulatory circuits are the genes and gene products that are involved in the response to a signal or functional clusters of genes that impact each other's expression. These circuits can be modeled *in silico* to predict the dynamics of a genetic system. Understanding of genetic regulatory circuits is a key in the field of synthetic biology, where disparate genetic elements are combined to produce novel biological functions (Richardson et al., 2010).
GeneDesign is a set of web applications that provides public access to a nucleotide manipulation pipeline for synthetic biology. The server is public and freely accessible, and the source is available for download under the New BSD License. In the functionality of GeneDesign were added several algorithms and modules, were updated the restriction enzyme library, were added batch processing capabilities and several command line modules. GeneDesign was originally developed as an in-house project to automate the design of oligonucleotides for the construction of individual synthetic genes. The most popular modules are building block design, reverse translation and restriction site addition (Richardson et al., 2010).
GenoCAD is an application to design protein expression vectors, artificial gene networks and other genetic constructs composed of multiple functional blocks called genetic parts. GenoCAD automatically derives the construct sequence from its comprehensive libraries of genetic parts. GenoCAD includes design rules describing how parts should be combined in genetic constructs. These rules are used to build a wizard that guides users through the process of designing complex genetic constructs and artificial gene networks. GenoCAD facilitates the design of artificial DNA sequences in different ways. GenoCAD includes a flexible system to manage libraries of public and user-defined genetic parts and it relies on formal design strategies to guide both novice

and experienced users in the design of structurally valid constructs for various biological applications (Czar et al., 2009).

ProMoT was originally designed for the computer-aided modeling of chemical processes as well as for the implementation of libraries that contain reusable modeling. ProMoT enables the use of object-oriented modeling techniques including encapsulation, aggregation, and inheritance. In ProMoT, dynamic models are built by aggregating structural and behavioral modeling entities, which in ProMoT are organized as an object-oriented class hierarchy with multiple inheritances. ProMoT supports models based on differential algebraic equation systems and on a logical modeling formalism. Modeling in ProMoT is subdivided into several steps which are associated to different components. ProMoT provides a special modeling language as well as a graphical user interface (GUI) for interactive modeling (Mirschel et al., 2009).

TinkerCell is a synthetic biology computer-aided design **(**CAD) tool for visually constructing and analyzing biological networks. The visual interface alone does not have any analysis capabilities. The analyses are performed by third-party libraries and modules that interface with TinkerCell. Numerous functions are available in TinkerCell for analyzing a constructed model, including deterministic and stochastic simulation algorithms, metabolic control analysis, structural analysis, flux-balance analysis, one-dimensional and two-dimensional steady state analysis, and centrality measurements. This software supports SBOL standard (Chandran et al., 2009).

Gene Designer is a stand-alone software for fast and easy design of synthetic DNA segments. Software allows adding, editing and combining genetic elements such as promoters, open reading frames and tags through an intuitive drag-and-drop graphic interface and a hierarchical DNA/Protein object map. Using advanced optimization algorithms, open reading frames within the DNA construct can readily be codon optimized for protein expression in any host organism. GeneDesigner also includes features such as real-time sliding calculator of oligonucleotide annealing temperatures, sequencing primer generator, tools for avoidance or inclusion of restriction sites, and options to maximize or minimize sequence identity to a reference (Villalobos et al., 2006).

Databases

All necessary information for genetic modeling, microorganism reconstruction and optimization can be taken from specific databases. Below is given description of a few databases that are useful for a metabolic reconstruction.

Kyoto Encyclopedia of Genes and Genomes (KEGG): This is a bioinformatics database containing information on genes, proteins, reactions and pathways. The organisms section is divided into eukaryotes and prokaryotes, encompasses many organisms for which gene and DNA information can be searched by typing in the enzyme of choice. This resource can be extremely useful when building the association between metabolism enzymes, reactions and genes (Ogata et al., 1999).

BioCyc and MetaCyc: BioCyc is a collection of 1690 pathway/genome databases (Nov 2011). Each database in the BioCyc collection describes the genome and metabolic pathways of a single organism. For example, Additionally, MetaCyc, an encyclopedia of experimentally defined metabolic pathways and enzymes, contains 1790 pathways from more than 2216 different organisms and 8800 metabolic reactions in Nov 2011 (Caspi et al., 2008).

metaTIGER: is a collection of metabolic profiles and phylogenomic information on a taxonomically diverse range of eukaryotes. Phylogenomic information is provided by 2,257 large phylogenetic trees which can be interactively explored. High-throughput tree analysis can also be carried out to identify trees of interest, e.g. trees containing horizontal gene transfers. metaTIGER also provides novel facilities for viewing and comparing the metabolic profiles (Whitaker et al., 2009).

BRENDA: A comprehensive enzyme database, BRENDA, allows you to search for an enzyme by name or EC number. You can also search for an organism and find all the relevant enzyme information. Moreover, when an enzyme search is carried out, BRENDA provides a list of all organisms containing the particular enzyme of interest (Schomburg et al., 2002).

Model SEED: This is an online resource for the analysis, comparison, and reconstruction of genome-scale metabolic models. Users can submit genome sequences to the RAST annotation system, and the resulting annotation can be automatically piped into the Model SEED to produce a draft metabolic model. The Model SEED automatically constructs a network of metabolic reactions, gene-protein-reaction associations for each reaction, and a biomass composition reaction for each genome to produce a model of microbial metabolism that can be simulated using Flux Balance Analysis (Henry et al., 2010).

BioBrick standard biological parts are DNA sequences of defined structure and function; they share a common interface and are designed to be composed and incorporated into living cells to construct new biological systems. BioBrick parts represent an effort to introduce the engineering principles of abstraction and standardization into synthetic biology (Shetty et al., 2008).

Gap-filling tools

Even the most complete models are not perfect, however, they all contain gaps or missing information. There are two types of missing information in metabolic network reconstructions. The first type is gaps in a network, places where a reaction that consumes or produces a metabolite is missing, creating a dead-end. The second type is orphan reactions. These are reactions that are known to exist, but it is not yet known which gene or genes code for their enzymes. Both types of gaps are the result of our incomplete knowledge of the metabolism of an organism (Orth et al., 2010). Gap-filling methods can thus be used to improve metabolic network models while simultaneously leading to discovery of new metabolic gene functions.

MetaRoute is a computational method to search for relevant routes between a source and a product in metabolic networks. Its speed allows the method to be used in a web interface for interactive navigation through genome scale networks and local network visualization. The underlying approach is based on graph-theory and the basic algorithm uses atom mapping rules and path weighting schemes to search for relevant paths in a directed graph representing the metabolic network. Precalculated atom mapping rules are incorporated into both the graph representation and the path finding algorithm to ensure calculation and to enforce biochemical constraints. To this end, the weighted path search is constrained to only trace feasible paths transferring atoms from the user-specified source to the product metabolite. The underlying data stems from the Biological Network Database (BNDB) and is based on data from KEGG (Blum et al., 2008).

BNICE - algorithm developed to identify all biochemical reactions that could realistically link two metabolites. In this algorithm, molecules are represented as bond- electron matrices (BEMs) and reactions are represented by matrix addition. It is possible to model the activity of an enzyme catalyzed reaction occurring with alternative substrates by adding the same reaction matrix to different BEMs, resulting in different product BEMs from each one. BNICE iteratively searches for all possible pathways of chemical transformations from one metabolite to another, offering gap-filling hypotheses (Hatzimanikatis et al., 2005).

GapFind and **GapFill** were developed in attempt to minimize the total number of gaps in a metabolic network model. GapFind is a mixed-integer linear programming (MILP) algorithm that can identify every gap in a network by identifying blocked metabolites, that is, those that cannot be produced or consumed at steady-state under any conditions. Although it may be easy to identify some gaps by inspection of network or pathway maps, certain unusual model structures such as cycles can lead to non-intuitive sets of blocked metabolites. GapFill is another MILP algorithm, and its objective is to minimize the total number of gaps by reversing the directionality of existing reactions, adding new reactions, adding transport reactions for blocked metabolites, or adding intracellular transport reactions between compartments for multi-compartment models (Kumar et al., 2009).

SMILEY -with a constraint-based metabolic network model, flux balance analysis can be used to simulate the growth of microbial organisms on different substrates, including many different carbon and nitrogen sources. When an *in vivo* experiment shows that an organism can grow on a certain substrate but the model predicts that it cannot, the model is likely missing one or more of the reactions required to consume the substrate. SMILEY was created to predict which reactions are likely missing from a network when the model predicts no growth but experiment indicates growth. SMILEY uses MILP to attempt to identify a flux distribution that leads to growth on the substrate of interest while minimizing the total number of reactions added from a universal database of reactions. It thus predicts which reactions should be added to the model parsimoniously, adding the smallest number of reactions necessary to reconcile *in silico* and *in vivo* growth predictions (Reed et al., 2006).

GrowMatch - this algorithm uses experimentally determined gene essentiality data to identify incorrect model predictions. GrowMatch attempts to correct two types of disagreements differently, when the model predicts growth but an experiment shows that the organism cannot grow with this gene knockout and when the model predicts no growth but an experiment shows that growth is possible. GrowMatch then uses the three single compartment methods from GapFill (reversing reaction directions, adding reactions from a database, or adding transport reactions) to attempt to correct the model (Kumar et al., 2009).

OMNI - is another MILP-based algorithm that compares *in silico* predictions to experimental measurements in an attempt to improve a constraint-based model. In this case, OMNI uses measured metabolic flux data. OMNI compares experimentally measured fluxes to fluxes predicted by flux balance analysis, and then attempts to minimize the total difference between measured and predicted fluxes by adding or removing reactions while maintaining a predicted growth rate above a defined minimum. Because it can remove reactions as well as add them, it can be used to remove reactions corresponding to poorly annotated genes if they do not match the experimental data (Herrgård et al., 2006).

PathoLogic is a program for automatically constructing metabolic networks from annotated genomes. It uses EC numbers, Gene Ontology terms, or annotated gene names to map reactions to genes, and then assembles the reactions into pathways by comparing the reactions to the reference database MetaCyc and adding any missing reactions. After performing these steps, however, many of the reactions in the new pathways may be orphans. Pathologic includes a PHFiller program that attempts to identify the genes associated with these reactions (Karp et al., 2002).

ADOMETA is a bioinformatics framework that combines analysis of gene co-expression, phylogenetic profiles, chromosomal clustering and protein fusions has been developed. The genes in an organism with unknown functions and genes with known functions as well, can be compared to the genes in the local metabolic network surrounding an orphan reaction using different types of functional association evidence. Genes that are close to each other in a metabolic network are more likely to have similar co-expression profiles, phylogenetic profiles, and so forth. In ADOMETA, the genes in an organism are compared to the surrounding genes using different methods, and the different association scores are combined. The products of the highest scoring genes are the most likely to catalyze the orphan reaction (Kharchenko et al., 2006).

Flux balance analysis (FBA) tools

FBA is a mathematical method for analyzing metabolism and distribution of fluxes in large metabolic networks. Flux-balance analysis based on linear optimization is widely used to compute metabolic fluxes in large metabolic networks and gains increasingly importance in network structural analysis. Thus, a computational tool flexible enough to realize a wide variety of FBA algorithms and able to handle batch series of flux-balance optimizations is of great benefit (Hoppe et al., 2011). The first FBA applications relied on the steady-state assumption and biomass maximization only. Several software solutions for FBA are currently available.

COBRA is a constraint-based reconstruction and analysis toolbox, a software package running in the Matlab environment, which allows for quantitative prediction of cellular behavior using a constraint-based approach. Specifically, this software allows predictive computations of both steady-state and dynamic optimal growth behavior, the effects of gene deletions, comprehensive robustness analyses, sampling the range of possible cellular metabolic states and the determination of network modules. Functions enabling these calculations are included in the toolbox, allowing user to input a genome-scale metabolic model distributed in SBML format and perform these calculations with just a few lines of code. The results are predictions of cellular behavior that have been verified as accurate in a growing body of research (Becker et al., 2007).

OptFlux covers an even larger range of flux optimization methods accessible through a graphical user interface. It is an easy-to-use solution well suited for biotechnologists with lesser interest in the algorithmic details (I. Rocha et al., 2010).

FASIMU offers the first available implementation of thermodynamic feasibility as a quickly computable MILP, is flexible in the choice of objective functions and constraints, and allows for batch processing of heterogeneous computations and automatic evaluation of the solutions and visualization of the computed fluxes (Hoppe et al., 2011).

FluxAnalyzer is a package for MATLAB and facilitates integrated pathway and flux analysis for metabolic networks within a graphical user interface. Arbitrary metabolic network models can be composed by instances of four types of network elements (Klamt et al., 2003).

FAME is a web-based modeling tool that combines the tasks of creating, editing, running, and analyzing/visualizing stoichiometric models into a single program. Analysis results can be automatically superimposed on familiar KEGG-like maps. FAME allows users to either upload their own preexisting model, or to build a new model based on the information in KEGG. When building from scratch, it is possible to select a subset of pathways from KEGG, foregoing the inclusion of unnecessary reactions that may be present in existing genome-scale reconstructions (Boele et al., 2012).

Optimization of pathways

Sometimes it is necessary to modify metabolic network to achieve better result by additions and deletions of reaction in the network. It can be used also for flows optimization in the pathways.

OptORF is a bi-level optimization problem which identifies the optimal metabolic and regulatory gene deletions as well as gene over expressions that maximize biochemical production at the maximum cellular growth under transcriptional regulatory constraints. The inner problem of OptORF, which is a linear programming problem, maximizes growth under the given gene deletions and regulatory states that are determined by the constraints of the outer problem. OptORF is formulated as a single level MILP by replacing the inner maximization problem with its optimality conditions as constraints (Kim et al., 2010).

OptForce procedure identifies all possible engineering interventions by classifying reactions in the metabolic model depending upon whether their flux values must increase, decrease or become equal to zero to meet a pre-specified overproduction target. This leads to the identification of a sufficient and non-redundant set of fluxes that must change to meet a pre-specified overproduction target. Starting with this set we subsequently extract a minimal set of fluxes that must actively be forced through genetic manipulations to ensure that all fluxes in the network are consistent with the overproduction objective (Ranganathan et al., 2010).

OptKnock computational framework is introduced for suggesting gene deletion strategies leading to the over-production of chemicals or biochemical. This is accomplished by ensuring that a drain towards growth resources must be accompanied, due to stoichiometry, by the production of a desired product. Computational results for gene deletions are in good agreement with mutant strains published in the literature. While some of the suggested deletion strategies are straightforward and involve eliminating competing reaction pathways, many

others suggest complex and nonintuitive mechanisms of compensating for the removed functionalities (Burgard et al., 2003).

OptStrain computational framework is aimed at guiding pathway modifications, through reaction additions and deletions, of microbial networks for the overproduction of targeted compounds. These compounds may range from electrons or hydrogen in biofuel cell and environmental applications to complex drug precursor molecules. OptStrain provides a useful tool to aid microbial strain design and, more importantly, it establishes an integrated framework to accommodate future modeling developments (Pharkya et al., 2004).

Tools for metabolic network dynamic modeling

Having identified the key components of the system, reacting entities and topology, one needs to design the biochemical structure of the model. Genome-scale metabolic network reconstructions are built from all of the known metabolic reactions and genes in a target organism. Reconstructions are useful because they can be converted into constraint-based models, allowing practical calculations and simulations to be performed.

Celldesigner is a graphical editor for biological networks, into which existing models in SBML format can be imported and modified. Notes can also be added and mathematical equations entered for the different interactions. Substrates, products and modifiers can be added either directly by drawing and setting the appropriate connections on the canvas or by entering them in the corresponding lists. Information about the graphical layout is stored in a Celldesigner - specific format in the annotation elements of the newly created SBML file. While Celldesigner is mainly an editor, it has some built-in deterministic integration capability features, based on third party solvers. Using SBML as its native format Celldesigner generates models that can be used by the whole family of SBML aware tools (Funahashi et al., 2003).

COPASI is a software application for simulation and analysis of biochemical networks and their dynamics, creating and solving mathematical models of biological processes such as metabolic networks, cell-signaling pathways, regulatory networks, infectious diseases, and many others. COPASI includes features to define models of biological processes, simulate and analyze these models, generate analysis reports, and import/export models in SBML format. COPASI carries out several analyses of the network and its dynamics and has extensive support for parameter estimation and optimization. COPASI provides means to visualize data in customizable plots, histograms and animations of network diagrams (Hoops et al., 2006).

AMIGO is a toolbox which facilitates parametric identification by means of advanced numerical techniques which cover the full iterative identification procedure putting especial emphasis on robust methods for parameter estimation and practical identifiability analyses, plus flexible capabilities for optimal experimental design. The toolbox offers several dynamic simulation tasks to solve system dynamics under given values of model unknowns and given experimental schemes, run sensitivity analysis and rank of parameters, parameter estimation allows for multi-experiment fitting of local and global unknowns, the toolbox can solve the optimal sequential-parallel experimental design problem as a truly general optimal control problem (Balsa-Canto and Banga, 2011).

Conclusion

There are many and varied modeling and design tools for synthetic biology. That is why authors have shown only a fraction of all possible software tools and possible application areas. Many of these tools are connected and can be used to solve problems together as COPASI and CellDesigner for example. Some of the tools have similar functionality, e.g., Cobra and FAME offers similar computational procedures, so each user can select the most appropriate software for solving synthetic biology problems.

Some of all tools, described above, are web based applications, some – desktop applications that are not available for all users. Some tools are freely available for academic purpose, but there are also tools which work under other software, like Matlab, which needs license. There are many different databases that are useful for researcher in the field of synthetic biology. Each database is developed with a specific goal and collects specific information. Reconstruction models of microorganisms are not perfect and may contain errors. There can be used a gap filling software which can improve metabolic network models. For each type of problem it is necessary to choose correct software tool to achieve better results. It depends on many aspects of research field and possible employment of these tools.

Acknowledgments

The author is acknowledging the financial support received from the State Education Development Agency of Latvia and the European Social Fund within the scientific research project "Establishment of Latvian interdisciplinary interuniversity scientific group of systems biology" (ref.no. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128).

References

Balsa-Canto, E., Banga, J.R., 2011. AMIGO, a toolbox for advanced model identification in systems biology using global optimization. *Bioinformatics (Oxford, England)*, 27(16), pp.2311-3.

Becker, S., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., & Herrgard, M. J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2(3), pp.727-38.

Blum, T., Kohlbacher, O., 2008. MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization. *Bioinformatics (Oxford, England)*, 24(18), pp.2108-9.

Boele, J., Olivier, B.G. & Teusink, B., 2012. FAME, the Flux Analysis and Modeling Environment. *BMC systems biology*, 6(1), p.8.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, 84(6), pp.647-57.

Caspi, R., Foerster, H., Fulcher, C. a, Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., et al., 2008. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic acids research*, 36, pp.D623-31.

Chandran, D., Bergmann, F.T., Sauro, H.M., 2009. TinkerCell: modular CAD tool for synthetic biology. *Journal of biological engineering*, 3, p.19.

Czar, M.J., Cai, Y., Peccoud, J., 2009. Writing DNA with GenoCAD. *Nucleic acids research*, 37(Web Server issue), pp.W40-7.

Funahashi, A., Morohashi, M., Kitano, H., Tanimura, N., 2003. CellDesigner : a process diagram editor for gene-regulatory and. *Biosilico*, 1(5), pp.159-162.

Gendrault, Y., Madec, M., Lallement, C., Pecheux, F., & Haiech, J., 2011. Synthetic biology methodology and model refinement based on microelectronic modeling tools and languages. *Biotechnology journal*, pp.796-806.

Hatzimanikatis, V., Li, C., Ionita, J., Henry, C. S., Jankowski, M. D., Broadbelt, L. J., 2005. Exploring the diversity of complex metabolic networks. *Bioinformatics (Oxford, England)*, 21(8), pp.1603-9.

Henry, C. S., DeJongh, M., Best, A., Frybarger, P. M., Linsay, B., Stevens, R. L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology*, 28(9), pp.977-82.

Herrgård, M.J., Fong, S.S., Palsson, B.Ø., 2006. Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS computational biology*, 2(7), p.e72.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., et al., 2006. COPASI-a COmplex PAthway SImulator. *Bioinformatics (Oxford, England)*, 22(24), pp.3067-74.

Hoppe, A., Hoffmann, S., Gerasch, A., Gille, C., & Holzhütter, H.-G., 2011. FASIMU: flexible software for flux-balance computation series in large metabolic networks. *BMC bioinformatics*, 12(1), p.28.

Karp, P.D., Paley, S., Romero, P., 2002. The Pathway Tools software. *Bioinformatics (Oxford, England)*, 18, pp.S225-32.

Khalil, A.S., Collins, J.J., 2010. Synthetic Biology : Applications Come of Age., Nat Rev Genet, 11(5), pp.367-379.

Kharchenko, P., Chen, L., Freund, Y., Vitkup, D., & Church, G. M., 2006. Identifying metabolic enzymes with multiple types of association evidence. *BMC bioinformatics*, 7, p.177.

Kim, J., Reed, J.L., 2010. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC systems biology*, 4, p.53.

Klamt, S., Stelling, J., Ginkel, M., & Gilles, E. D., 2003. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, 19(2), p.261.

Kumar, V.S., Maranas, C.D., 2009. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS computational biology*, 5(3), p.e1000308.

Macdonald, J. T., Barnes, C., Kitney, R. I., Freemont, S., Stan, G.-bart V., 2011. Integrative Biology Computational design approaches and tools for synthetic biologyw. *Society*, pp.97-108.

Marchisio, M., Stelling, J., 2009. Computational design tools for synthetic biology. *Current opinion in biotechnology*, 20(4), pp.479-85.

Mirschel, S., Steinmetz, K., Rempel, M., Ginkel, M., Gilles, E. D., 2009. P RO M O T : modular modeling for systems biology. *Bioinformatics*, 25(5), pp.687-689.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*, 27(1), pp.29-34.

Orth, J.D., Palsson, B.Ø., 2010. Systematizing the generation of missing metabolic knowledge. *Biotechnology and bioengineering*, 107(3), pp.403-12.

Pharkya, P., Burgard, A.P.. Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. *Genome research*, 14(11), pp.2367-76. A

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS computational biology*, 6(4), p.e1000744.

Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., et al., 2006. Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), pp.17480-4.

Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D., Bader, J. S., 2010. GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic acids research*, 38(8), pp.2603-6.

Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., et al., 2010. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC systems biology*, 4, p.45.

Rubina, T., Stalidzans, E., 2010. Software tools for structure analysis of biochemical networks, *In Proceedings of Applied Information and Communication Technologies 2010*, Jelgava, Latvia, pp. 33-49.

Schomburg, I., Chang, A., Schomburg, D., 2002. BRENDA, enzyme data and metabolic information. *Nucleic acids research*, 30(1), pp.47-9.

Shetty, R.P., Endy, D., Knight, T.F., 2008. Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering*, 2, p.5.

Tepper, N., Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics (Oxford, England)*, 26(4), pp.536-43.

Villalobos, A., Ness, J. E., Gustafsson, C., Minshull, J., Govindarajan, S., 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC bioinformatics*, 7, p.285.

Whitaker, J. W., Letunic, I., McConkey, G. a, & Westhead, D. R., 2009. metaTIGER: a metabolic evolution resource. *Nucleic acids research*, 37(Database issue), pp.D531-8.

Wiechert, W., 2002. Modeling and simulation: tools for metabolic engineering. *Journal of biotechnology*, 94(1), pp.37-63.

Zheng, Y., Sriram, G., 2010. Mathematical modeling: bridging the gap between concept and realization in synthetic biology. *Journal of biomedicine & biotechnology*, p.541609.