doi 10.18699/bgrs2024-8.4-15

Metabolic modeling of the production of a recombinant protein by methanotrophs

Kulyashov M.A.[1]*, Richard H.[2], Kolmykov S.K.[1], Sokolova T.S.[1], Khlebodarova T.M.[1], Kalyuzhnaya M.G.[2], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *San Diego State University, San Diego, USA*

** kulyashov.ma@talantiuspeh.ru*

Key words: constrained-based modelling; methanotrophs; recombinant protein

Motivation and Aim: Methanotrophic bacteria are widely used in advanced biotechnology, beginning from the production of single cell protein (SCP), and ending with different cell metabolites: succinate, polyhydroxyalkanoates (PHAs), ectoine, methanol and others. Main advantage of their application in biotechnology is using a low-cost and available source – methane. One of these organisms widely studied among methanotrophs is *Methylotuvimicrobium alcaliphilum* 20Z^R, for which the production of recombinant proteins is the new biotechnology field of its application in a few last year's. Moreover, a genome-scale metabolic (GSM) model verified by a few omics experimental datasets was developed for the strain [1, 2]. The main goal of the study is an optimization of the recombinant protein production by *Methylotuvimicrobium alcaliphilum* 20Z^R using constraint-based modeling approaches.

Methods and Algorithms: The following proteins were used as target recombinant proteins: sfGFP protein, the nucleotide sequence of which was provided by Prof. Marina Kalyuzhnaya (San Diego California State University, USA); the amino acid sequence of the bovine β-casein protein (https://www.uniprot.org/uniprot/P02666), which was taken from the Uniprot database [3], was also used. Two marker proteins were also analyzed: spyCatcher and spyTAG which have 114 and 17 amino acids length, respectively. Their sequences as well as an estimation of ATP consumption for the production of recombinant protein were also provided based on the experiments in Kalyuzhnaya's lab. To run the algorithm, the growth medium on methane specified in the original model was used, with a methane consumption rate of 11.7 mmol*gDCW^{-1}*h^{-1} (gDCW – gram of dry biomass), with the exception of the context-specific model, in which a methane consumption rate of 2.85 mmol was used mmol*gDCW^{-1}*h^{-1} in accordance with experimental measurements.

Model modification and extension was made in the BioUML platform [4] using python libraries Biopython [5] and Cobrapy [6]. All original code and developed modules are available on the BioUML platform – https://shorturl.at/cmG47.

Co-optimization of recombinant protein production in an extended model was made using the Optflux tool [7] with WYELD evolutionary function, which gives an opportunity to fix the lower bound for the biomass equation, by taking in account the biomass fraction which is determined by the user.

Results: The first step was to write a program code in Python, which, based on the sequence of the recombinant protein, would create a pseudo-equation for its synthesis in a GSM model, and also add the set of reactions for its transport and exchange, taking

into account energy costs. Using the Biopython and Cobrapy libraries, the module was written. It allows a user to pass not only the amino acid sequence of a protein, but also the nucleotide sequence, which is converted into an amino acid, and then counts each type of amino acid, generates an identifier for it in the model, creates a pseudo-equation for the synthesis of a recombinant protein, taking into account the procedure for calculating stoichiometric coefficients and energy costs for the production of the corresponding protein. After generation of the pseudo-equation for protein synthesis, the transport reaction of its export from the cell is added to the model, taking into account ATP costs for its transport. In addition, this new BioUML module also enables to immediately co-optimize the production of recombinant protein using the evolutionary algorithm via the OptKnock algorithm implemented in the Mewpy library [8]. The original written code provides a feasibility to set various parameters for co-optimization, including the selection of environmental parameters of interest for simulation of the cell culture growth. The co-optimization can be conducted by means of 2 evolutionary functions, BPCY and WYIELD or via the use of both simultaneously. The written code implements options for running these algorithms for one target function, which expands the choice of algorithms to search for potential genetic modifications. A graphical interface of the code was also written that makes it easy to use for researchers without a programming background.

The previously constructed *i*IA409 GSM model [2] has been extended to integrate and account for the synthesis of recombinant proteins using the developed module. Furthermore, several models have been built in order to co-optimize and predict the production of different recombinant proteins in the combination with various marker proteins used simultaneously in the growth of the cell culture. Analysis of co-optimization results of these modified models depending on the used combination of both recombinant and marker proteins, enabled to identify potential genetic modifications facilitating the production of the corresponding protein simultaneously with the growth of the cell culture at various acceptable minimum values of the biomass obtained for wild type simulation.

Thus, three main directions of genetic modifications have been revealed both for the GFP protein and for the β-casein protein. The first set of modifications include modifications associated with a *direct change in the activity of TCA cycle reactions***: an increase in the activity** of citrate synthase, fumarate reductase, and malate dehydrogenase. The efficiency of these modifications can be experimentally verified. Another direction of modifications which is of interest, are reactions associated with the *redistribution of the carbon flux* in the cell due to the **activation** of the H_4MPT pathway, as well as a **decrease in the activity** of formate dehydrogenase, which will reduce the conversion of carbon into CO_2. Last one direction for modifications **is the increase of ATP production**: increase the activity of ATP synthase, increase the availability of H^+ in the cell, which would also lead to an increase of ATP production by the cell. However, this type of modifications is more challenging for experimental verification.

It is worth to note that the list of proposed modifications depends on the amino acid composition of the heterologous protein sequence. It indicates the developed program module with the implemented algorithm to search for potential modifications for various recombinant proteins takes into account the specificity of amino acids profile.

Conclusion: The Python-based module in the BioUML platform, which gives an opportunity to modify GSM model of methanotrophic organisms to account for synthesis of recombinant protein by the cell was developed The algorithm implemented in the

module considers the energy costs for protein synthesis and transport. Moreover, the module was applied to the original GSM model, *i*IA409 for *Methylotuvimicrobium alcaliphilum* 20Z^R, and allowed us to identify a set of potential genetic modifications contributing to the enhancement of recombinant protein production by the methanotroph taking into account the amino acids profile of the required protein.

Funding: The study was financially supported by the Russian Science Foundation (project № 23-24-00606, https://rscf.ru/en/project/23-24-00606/).

References

1. Akberdin I.R. et al. Methane utilization in Methylomicrobium alcaliphilum 20Z^R: a systems approach. *Sci Rep.* 2018;8(1):2512. doi 10.1038/s41598-018-20574-z
2. Akberdin I.R. et al. Rare earth elements alter redox balance in *Methylomicrobium alcaliphilum* 20Z^R. *Front Microbiol.* 2018;9:2735. doi 10.3389/fmicb.2018.02735
3. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 2023;51(D1):D523-D531. doi 10.1093/nar/gkac1052
4. Kolpakov F. et al. BioUML-towards a universal research platform. *Nucleic Acids Res*. 2022;50(W1):W124-W131. doi 10.1093/nar/gkac286
5. Cock P.J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-1423. doi 10.1093/bioinformatics/btp163
6. Ebrahim A., Lerman J.A., Palsson B.O., Hyduke D.R. COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst Biol*. 2013;7:74. doi 10.1186/1752-0509-7-74
7. Vilaça P., Maia P., Giesteira H., Rocha I., Rocha M. Analyzing and designing cell factories with OptFlux. *Methods Mol Biol*. 2018;1716:37-76. doi 10.1007/978-1-4939-7528-0_2
8. Pereira V., Cruz F., Rocha M. MEWpy: a computational strain optimization workbench in Python. *Bioinformatics*. 2021;37(16):2494-2496. doi 10.1093/bioinformatics/btab013