Supplementary material

Index

Supplementary text

Biomass composition

Medium composition

Scripts

Tarefas	2023								2024			
	mai	jun	jul	ago	set	out	nov	dez	jan	fev	mar	abr
Tarefa 1: Avaliação das vias metabólicas associadas ao metano												
Tarefa 2: Otimização in silico da fixação de metano												

Figure 1. Work plan timeline

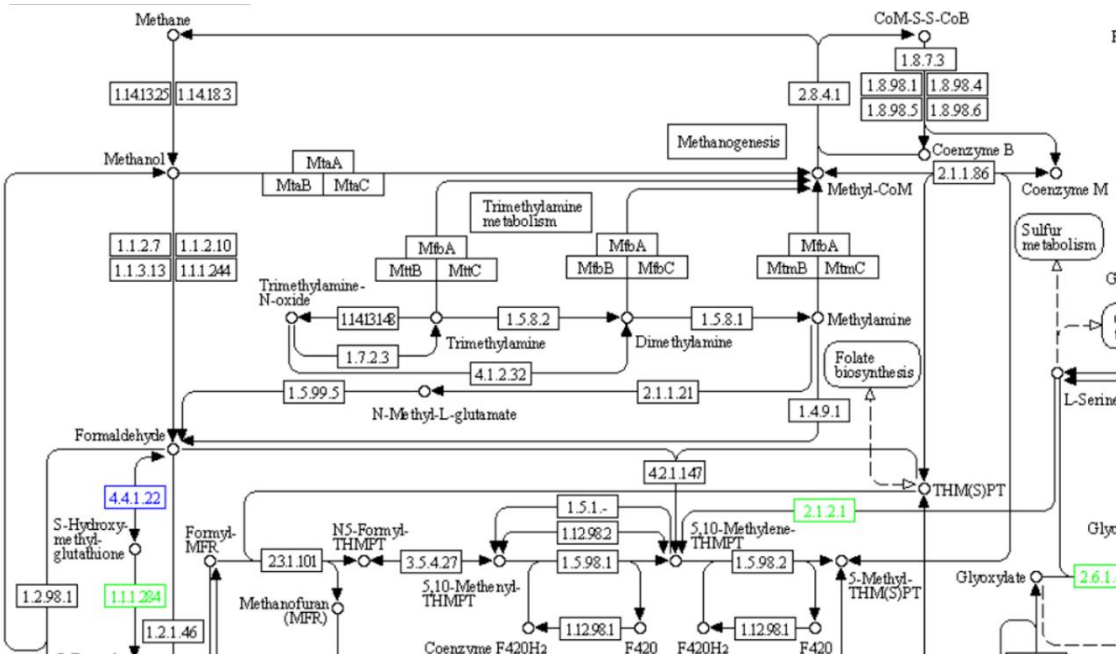

Figure 2. Segment of the methane metabolism of *C. vulgaris*, highlighting the absence of reactions and enzymes associated with it colored by *merlin*. Image extracted from KEGG.

The colors provived by *merlin* are interpreted as follows:

Green: The enzyme and reaction are present in the model;

Dark Blue: The reaction is present in the pathway, but the correspondent enzyme is missing; however, another enzyme present in another pathway might be associated with the reaction;

Cyan Blue: The enzyme is present in the model, but the reaction is connected to a dead-end;

Red: The enzyme is present in the network, but the reaction is a blocked and associated with a dead-end metabolite;

Colorless (Black): The enzyme is not present in the network (gap)

Table 1. Tools available for assisting the reconstruction of GSM models and generic biological databases

	Name	Web address	Reference
Databases	NCBI	https://www.ncbi.nlm.nih.gov/	[1]
	GOLD	https://gold.jgi.doe.gov/	[2]
	BRENDA	http://www.brenda-enzymes.org/	[3]
	KEGG	http://www.kegg.jp/	[4]
	KBase	https://www.kbase.us/	[5]
	BIGG models	http://bigg.ucsd.edu/	[6]
	BioCyc	https://biocyc.org/	[7]
	TCDB	https://www.tcdb.org/	[8]
	CCAP	https://www.ccap.ac.uk/	
Tools			
	BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi	[9]
	LocTree3	https://rostlab.org/services/loctree3/	[10]
	DeepLoc	https://services.healthtech.dtu.dk/services/DeepLoc-2.0/	[11]
	COBRA Toolbox	https://opencobra.github.io/cobratoolbox/latest/	[12]
	MEWpy	https://mewpy.readthedocs.io/en/latest/	[13]
Aiding software	Merlin	http://www.merlin-sybio.org/	[14]
	AureMe		[15]
	ModelSeed	https://modelseed.org/	[16]
	Raven Toolbox	https://github.com/SysBioChalmers/RAVEN/wiki	[17]
	CarveMe	https://carveme.readthedocs.io/en/latest/	[18]
	Pathway Tools	http://bioinformatics.ai.sri.com/ptools/ptools-features.shtml	[19]

Table 2. Biomass composition

Macromolecule	mass % (g MM/gDW)					Mean	Normalized
Protein	0,52	0,54	0,37	0,51	0,29	0,400	*0,457*
DNA	-	-	-	-	0,001	0,001	*0,001*
RNA	-	-	-	-	0,028	0,0028	*0,032*
Lipid	0,09	0,018	0,26	0,12	0,11	0,158	*0,180*
Carbohydrates	0,09	0,015	0,3	0,12	0,53	0,240	*0,274*
Pigments	-	-	-	-	0,04	0,040	*0,045*
Cofactors	-	-	-	-	0,01	0,010	*0,011*
Total						0,08795	*1*
Reference	[20]	[21]	[22]	[23]	[24]		

Table 3. RNA composition [25]

Metabolite	Chemical Formula	mmolmonomer/gDW	mmolmonomer/ge-RNA
UTP	$C_9H_{11}N_2O_{15}P_3$	0,058615616	1,841158
GTP	$C_{10}H_{12}N_5O_{14}P_3$	0,104205539	3,27317
CTP	$C_9H_{12}N_3O_{14}P_3$	0,104205539	3,27317
ATP	$C_{10}H_{12}N_5O_{13}P_3$	0,058615616	1,841158

Table 4. Carbohydrates composition [26].

Metabolite	**MW - dTDP or UDP or H2O (g/mol)**	**g M/ gDW (%)**	**g M/ gDW**	**mmol M / gDW**	**g M/ gDW (normalized)**	**mmol M / g MM**
UDP-L-arabinose	186,18	0,72	0,0072	0,038672253	0,010464834	0,14122732
UDP-D-Xylose	436,29	0,57	0,0057	0,013064705	0,00828466	0,047711036
UDP-alpha-D-galactose	145,13	4,36	0,0436	0,300420313	0,063370385	1,097105876
GDP-mannose	144,1211	0,68	0,0068	0,047182543	0,009883454	0,172306076
UDP-glucose	147,1418	9,91	0,0991	0,673499984	0,144036815	2,45955669
Starch	48660,2	2,6	0,026	0,000534318	0,037789679	0,001951276
Total	49848,1929	18,84	0,1884	1,073374115	0,273829828	3,919858274

Table 5. Fatty-acid composition [27].

Metabolite	Molar fraction	MW (g/mol)	MW of Each Fatty Acid in the final Fatty Acid ($g/mol_{Fatty\ Acid}$)
Myristic acid	0,01	227,37	2,2737
Pentadecanoic acid	0,004	241,4	0,9656
Palmitic acid	0,164	255,41	41,88724
Palmitoleic acid	0,015	253,41	3,80115
Margaric acid	0,012	269,45	3,2334
Oleic acid	0,042	281,47	11,82174
5-Octadecenoic acid	0,003	281,5	0,8445
7,10-Hexadecadienoic acid	0,204	251,4	51,2856
11-Octadecenoic acid	0,02	281,46	5,6292
Linoleic acid	0,351	279,45	98,08695
Stearic acid	0,028	283,48	7,93744
Nonadecylic acid	0,002	297,5	0,595
Linolenic acid	0,052	277,43	14,42636
Arachidic acid	0,057	311,53	17,75721
9-Eicosenoic acid	0,004	310,5	1,242
Eicosadienoic acid	0,013	308,5	4,0105
Homolinolenic acid	0,004	306,48	1,22592
Heneicosylic acid	0,002	325,56	0,65112
Arachidonic acid	0,006	303,47	1,82082
Behenic acid	0,009	339,58	3,05622
Total	1,002		

Supplementary text

Biomass composition

The overall composition of biomass is given on Table 2. The determination of the composition of DNA and protein is automatically provived by merlin through the tool e-BiomassX [28]. Composition of RNA, carbodydrates and fatty acids is stated on Tables 3, 4 and 5, respectively. Future work will envolve gathering experimental data for lipids and cofactors of *Chlorella vulgaris.*

Medium composition

The common medium composition for *C. vulgaris* is: water, oxygen, orthophosphate, CO_2, sulfate, H^+, photon, nitrate, magnesium cation and Fe^{2+}. The missing compounds were added as drains to the model which then automatically generated transport reactions.

Scripts

The research group involved in this project has developed scripts to facilitate te gap-filling process. In these scripts, the EC number of a missing reaction is inputted in the python file, which then retrieves BLAST results containing genes associated with corresponding e-values. (search_ec_in_genome.py)

Additionally, a jupyter notebook for simulating the flux on pathways was created to determine whether each pathway was functional and producing the required metabolites. (pathway_fluxes.ipynb)

References

1. E. W. Sayers, J. Beck, E. E. Bolton, D. Bourexis, J. R. Brister, K. Canese, D. C. Comeau, K. Funk, S. Kim, W. Klimke, et al., “Database resources of the national center for biotechnology information,” Nucleic acids research, vol. 49, no. D1, p. D10, 2021.
2. K. Liolios, I.-M. A. Chen, K. Mavromatis, N. Tavernarakis, P. Hugenholtz, V. M. Markowitz, and N. C. Kyrpides, “The genomes on line database (gold) in 2009: status of genomic and metagenomic projects and their associated metadata,” Nucleic acids research, vol. 38, no. suppl_1, pp. D346–D354, 2010.
3. . Chang, L. Jeske, S. Ulbrich, J. Hofmann, J. Koblitz, I. Schomburg, M. Neumann-Schaal, D. Jahn, and D. Schomburg, “Brenda, the elixir core data resource in 2021: new developments and updates,” Nucleic acids research, vol. 49, no. D1, pp. D498–D508, 2021.
4. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res 28 (2000) 27–30
5. A. P. Arkin, R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, et al., “Kbase: the united states department of energy systems biology knowledgebase,” Nature biotechnology, vol. 36, no. 7, pp. 566–569, 2018
6. Z.A. King, J. Lu, A. Dr¨ ager, et al., BiGG models: a platform for integrating, standardizing and sharing genome-scale models, Nucleic Acids Res. 44 (2016) D515–D522.
7. R. Caspi et al, “The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases,” Nucleic Acids Research, 2014.
8. M.H. Saier, V.S. Reddy, G. Moreno-Hagelsieb, et al., The Transporter Classification Database (TCDB): 2021 update, Nucleic Acids Res. 49 (2021) D461–D467
9. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” Journal of molecular biology, vol. 215, no. 3, pp. 403–410, 1990.
10. T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansorge, K. Balasz, et al., “Loctree3 prediction of localization,” Nucleic acids research, vol. 42, no. W1, pp. W350–W355, 2014
11. J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, and O. Winther, “Deeploc: prediction of protein subcellular localization using deep learning,” Bioinformatics, vol. 33, no. 21, pp. 3387–3395, 2017.
12. N. Vlassis, M.P. Pacheco, T. Sauter, Fast reconstruction of compact contexto specific metabolic network models, PLoS Comput. Biol. (2014)
13. V. Pereira, F. Cruz, M. Rocha, MEWpy: a computational strain optimization workbench in Python, Bioinformatics (2021)
14. O. Dias, M. Rocha, E. C. Ferreira, and I. Rocha, “Reconstructing genome-scale metabolic models with merlin,” Nucleic acids research, vol. 43, no. 8, pp. 3899–3910, 2015.
15. M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron, N. Guillaudeux, et al., “Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models,” PLoS computational biology, vol. 14, no. 5, p. e1006146, 2018.
16. S. M. Seaver, F. Liu, Q. Zhang, J. Jeffryes, J. P. Faria, J. N. Edirisinghe, M. Mundy, N. Chia, E. Noor, M. E. Beber, et al., “The modelseed biochemistry database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes,” Nucleic acids research, vol. 49, no. D1, pp. D575–D588, 2021.
17. H. Wang et al., ‘RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor’, PLoS Comput Biol, vol. 14, no. 10, Oct. 2018
18. P. D. Karp, P. E. Midford, R. Billington, A. Kothari, M. Krummenacker, M. Latendresse, W. K. Ong, P. Subhraveti, R. Caspi, C. Fulcher, et al., “Pathway tools version 23.0 update: software for pathway/genome informatics and systems biology,” Briefings in bioinformatics, vol. 22, no. 1, pp. 109–126, 2021.

19. D. Machado, S. Andrejev, M. Tramontano, and K. R. Patil, “Fast automated reconstruction of genome-scale metabolic models for microbial species and communities,” Nucleic acids research, vol. 46, no. 15, pp. 7542–7553, 2018
20. S. Weber, P. M. Grande, L. M. Blank, and H. Klose, “Insights into cell wall disintegration of chlorella vulgaris,” PLoS One, vol. 17, no. 1, p. e0262500, 2022.
21. Y. Panahi, A. Y. Khosroushahi, A. Sahebkar, and H. R. Heidari, “Impact of cultivation condition and media content onchlorella vulgaris composition,” Advanced pharmaceutical bulletin, vol. 9, no. 2, p. 182, 2019.
22. Y. Liang, N. Sarkany, and Y. Cui, “Biomass and lipid productivities of chlorella vulgaris under autotrophic, heterotrophic and mixotrophic growth conditions,” Biotechnology letters, vol. 31, pp. 1043–1049, 2009.
23. J. A. Coronado-Reyes, J. A. Salazar-Torres, B. Juarez-Campos, and J. C. Gonzalez-Hernandez, “Chlorella vulgaris, a microalgae important to be used in biotechnology: a review,” Food Science and Technology, vol. 42, 2020.
24. M. Muthuraj, B. Palabhanvi, S. Misra, V. Kumar, K. Sivalingavasu, and D. Das, “Flux balance analysis of chlorella sp. fc2 iitg under photoautotrophic and heterotrophic growth conditions,” Photosynthesis research, vol. 118, pp. 167–179, 2013.
25. C. Zuñiga, C.-T. Li, T. Huelsman, J. Levering, D. C. Zielinski, B. O. McConnell, C. P. Long, E. P. Knoshaug, M. T. Guarnieri, M. R. Antoniewicz, et al., “Genome-scale metabolic model for the green alga chlorella vulgaris utex 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions,” Plant physiology, vol. 172, no. 1, pp. 589–602, 2016.
26. B. O. McConnell and M. R. Antoniewicz, “Measuring the composition and stable-isotope labeling of algal biomass carbohydrates via gas chromatography/mass spectrometry,” Analytical chemistry, vol. 88, no. 9, pp. 4624–4628, 2016
27. H. A. Pantami, M. S. Ahamad Bustamam, S. Y. Lee, I. S. Ismail, S. M. Mohd Faudzi, M. Nakakuni, and K. Shaari, “Comprehensive gcms and lc-ms/ms metabolite profiling of chlorella vulgaris,” Marine drugs, vol. 18, no. 7, p. 367, 2020
28. S. T. Santos, “Development Of Computational Methods For The Determination Of Biomass Composition And Evaluation Of Its Impact In Genome-Scale Models Predictions,” Masters thesis, Universidade do Minho, Braga, 2013