Development of text-mining solutions to facilitate lipid metabolism interpretation in Genome-Scale Metabolic Models

Adriano Silva[1], João Ribeiro[1], and Emanuel Cunha[1]

University of Minho

Abstract. The abstract should briefly summarize the contents of the paper in 15–250 words.

Keywords: First keyword · Second keyword · Another keyword.

1 Introduction

1.1 Context and motivation

In the past two decades, Systems Biology has emerged as a discipline capable of integrating molecular biological knowledge into an understanding at a system level, from a complete, precise, and efficient perspective. Biological systems represent a huge amount of data, with the need to be treated and contextualized where this discipline comes to the aid. Whether in the construction of stochiometric models or the reconstruction of genome scale metabolic models (GSM), with means to understand the genomic, biochemical, and physiological knowledge gathered [1, 2]. These approaches can guide to strain optimization and the production of a compound with industrial interest, such as biofuel produced by optimized yeasts [3].

Impulsed by the advances and cost-effectiveness in technologies that led to high-throughput biological data (Big Data), Systems Biology, and more precisely the reconstruction of GSM models, is gaining importance. The reconstruction of GSM models is taking advantage in the high quality data generated to create better simulations and predictions. In total, since the reconstruction of the first GSM model in 1999, 6239 GSM models where reconstructed until 2019 [4]. Nonetheless the pace in reconstruction of GSM models can't keep up with the growth of Big Data. The lack of integration of new data in GSM models is a problem inherent to this growth discrepancy.

Besides the usefullness of this models their reconstruction is limited due to the lack of biochemical and structural data incorporated. In fact, complex macromolecule's are often represented in its generic version not givin any biochemical and structural information. Particularly in the case of lipids, only a small chunk of GSM models reconstructed have defined lipids.

1.2 Objective

- integrar xyz bases de dados, numa base de dados base dados baseada em grafos
- mapear os modelos

2 State of art

2.1 GSM

explicar o que é GSM

2.2 Lipid computational representation

- Generic/structurally defined duality

2.3 Generic Representation in GSM models

- Lipid only with the backbone; - Name of the backbone; - Databases cross-references;

2.4 Structurally defined representation in GSM models

- Lipid with the whole structure - Name composed of backbone and side chains
- No databases cross-references

2.5 Lack in lipid annotation in GSM models

- Merge the two previous logics - Show that one is annotated and the other not. - Possível Resolução - integração dos sinonimos e abreviaturas das bases de dados com ETL (podes por uma frase para cada uma delas e um paragrafo sobre ETL) Anotação de modelos: - match direto - decomposição do nome em dois (backbone e side chain) - queries à base de dados

Sample Heading (Third Level) Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

Sample Heading (Fourth Level) The contribution should contain no more than four levels of headings. Table 1 gives a summary of all heading levels.
Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).

Table 1. Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Run-in Heading in Bold.** Text follows	10 point, bold
4th-level heading	*Lowest Level Heading.* Text follows	10 point, italic
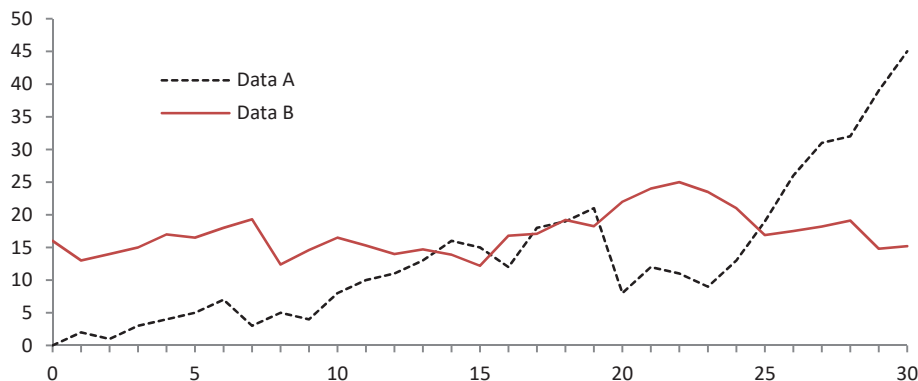
Fig. 1. A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

Theorem 1. *This is a sample theorem. The run-in heading is set in bold, while the following text appears in italics. Definitions, lemmas, propositions, and corollaries are styled the same way.*

Proof. Proofs, examples, and remarks have the initial word in italics, while the following text appears in normal font.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal

References

1. Y. Zou and M. D. Laubichler, "From systems to biology: A computational analysis of the research articles on systems biology from 1992 to 2013," *PLOS ONE*, vol. 13, p. e0200929, 7 2018.
2. I. Tavassoly, J. Goldfarb, and R. Iyengar, "Systems biology primer: The basic methods and approaches," *Essays in Biochemistry*, vol. 62, pp. 487–500, 10 2018.
3. H. W. Aung, S. A. Henry, and L. P. Walker, "Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism," *Industrial Biotechnology*, vol. 9, p. 215, 8 2013.

4

4. C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee, "Current status and applications of genome-scale metabolic models," *Genome Biology 2019 20:1*, vol. 20, pp. 1–18, 6 2019.