

Telling metabolic stories to explore metabolomics data – A case study on the Yeast response to cadmium exposure (Supplementary material)

Paulo Vieira Milreu^{1,*}, Cecilia Coimbra Klein^{1,2}, Ludovic Cottret³, Vicente Acuña^{1,4}, Etienne Birmelé^{1,5}, Michele Borassi⁶, Christophe Junot⁷, Alberto Marchetti-Spaccamela⁸, Andrea Marino⁹, Leen Stougie¹⁰, Fabien Jourdan¹¹, Pierluigi Crescenzi⁹, Vincent Lacroix^{1*}, Marie-France Sagot^{1,*}

¹ INRIA Grenoble Rhône-Alpes & Université de Lyon, F-69000 Lyon; Université Lyon 1; CNRS, UMR5558, LBBE, France

² Laboratório Nacional de Computação Científica (LNCC), Petrópolis, Brazil;

³ LISBP, UMR CNRS 5504 - INRA 792, Toulouse, France;

⁴ Mathomics (UMI 2807 CNRS) and Center for Genome Regulation (Fondap 15090007), University of Chile, Santiago;

⁵ Lab. Statistique et Génome, CNRS UMR8071 INRA1152, Université d'Évry, France;

⁶ Scuola Normale Superiore, 56126 Pisa, Italy;

⁷ Laboratoire d'Etude du Métabolisme des Médicaments, DSV/iBiTecS/SPI, CEA/Saclay, 91191 Gif-sur-Yvette, France;

⁸ La Sapienza University of Rome, Italy;

⁹ Università di Firenze, Dipartimento di Sistemi e Informatica, I-50134 Firenze, Italy;

¹⁰ VU University and CWI, Amsterdam, The Netherlands;

¹¹ INRA UMR1331 - Toxalim, Toulouse, France

* Corresponding authors: pvmilreu@gmail.com, vincent.lacroix@univ-lyon1.fr and marie-france.sagot@inria.fr.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXXX

SUPPLEMENTARY MATERIAL

1) Enumeration Algorithm

The algorithm to compute one story has as input a compressed network G and a total order π of the nodes and is illustrated in Figure 1. From π , we may easily compute what we call a **pitch**, which is defined as a story except for the maximality condition. Moving from a total order π to a pitch is done by keeping only arcs that are consistent with π and, after that, removing recursively any remaining white source or target. Completing a pitch into a story is done by adding paths between black nodes while avoiding cycles. The algorithm searches for extensions of the pitch following the order π of the nodes, moving to the next node when no new path may be added from the previous one. The resulting graph is a story. The enumeration was performed by examining all possible orderings of the nodes. More details on the mathematical modelling, the preprocessing step, the algorithms and their computational complexity are in [1].

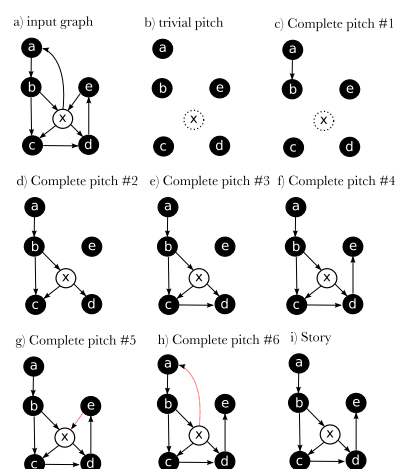


Fig. 1. a) The input graph with set of black nodes $B = \{a, b, c, d, e\}$ and white nodes $W = \{x\}$. b) The starting pitch, which is simply a graph $V = B$ and no arcs. c) The path $a \rightsquigarrow b$ is added to the pitch. d) Three paths starting from b are added to the pitch: $b \rightsquigarrow c$, $b \rightsquigarrow x \rightsquigarrow c$ and $b \rightsquigarrow x \rightsquigarrow d$. Notice that as x did not belong to the pitch at this point, the algorithm goes further and stops only when nodes in the pitch are found. e) The path $c \rightsquigarrow d$ is added to the pitch. f) The path $d \rightsquigarrow e$ is added to the pitch. g) The path $e \rightsquigarrow x$ is evaluated but cannot be added since e comes after x in the current partial order inferred from the pitch, and therefore such an addition creates at least one cycle, for instance $e \rightarrow x \rightarrow d \rightarrow e$. h) The path $x \rightsquigarrow a$ is evaluated but cannot be added since x comes after a in the current partial order inferred from the pitch, and therefore such an addition creates at least one cycle, for instance $x \rightarrow a \rightarrow b \rightarrow x$. i) There are no more nodes to traverse, the final object is a maximal pitch, i.e., a story.

2) Yeast metabolic network

We retrieved the reconstruction of the metabolic network of *Saccharomyces cerevisiae* s288c from MetExplore. This platform allows applying different filters to the network. Herein it is restricted to the small molecule metabolism, *i.e.* reactions involving one or more macromolecules such as proteins or nucleic acids are not represented. In addition, reactions involving pairs of cofactors were split into two reactions, such as the following transformation (or reverse): *compound A* + ATP → *compound B* + ADP + Pi will be represented as reaction 1: *compound A* ↔ *compound B*, and reaction 2: ATP → ADP + Pi. Ubiquitous compounds (*i.e.*, water, proton, carbon dioxide, phosphate, diphosphate, ammonia, hydrogen peroxide and oxygen) and cell compartments were as well removed from the network.

3) List of discriminating compounds

Table 1. List of discriminating compounds for the *S. cerevisiae* cell exposed to cadmium

Metabolite ID	intensity ratio	Present in the pathway
arginine	1.9	no
reduced glutathione	33.9	yes
O-acetylhomoserine (*)	0.5	yes
2-aminoadipate (*)	0.5	no
niacinamide (*)	4.8	no
pyridine-3-aldoxime (*)	4.8	no
pyrroline-hydroxy-carboxylate	0.7	no
methionine	0.3	yes
citrulline (*)	0.7	no
threonine	0.6	no
homoserine	0.6	no
glutamine	0.7	no
glutamate	0.8	yes
glutamylcysteine	192.2	yes
5-methylthioadenosine	11.0	no
serine	0.2	yes
glycine (*)	0.3	yes
cystathionine	50.5	yes
lysine	0.7	no
cysteinylglycine (*)	35.9	no
leucine/isoleucine	1.2	no
tyrosine	2.9	no
histidine	1.2	no
alanine	0.8	no

List of 24 metabolites from the yeast metabolic network whose concentration significantly varied under cadmium exposed. The intensity ratio column presents the ratio between the stress condition and the control. The 3rd column indicates whether the compound is present in the glutathione biosynthetic pathway (Fig. 1 of the main manuscript) or not. Metabolites identified with an (*) after their names were putative metabolites requiring more analysis for final identification.

4) Example of small metabolic story

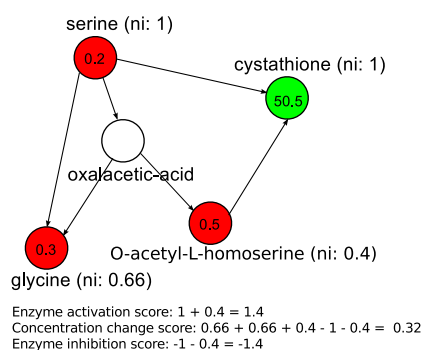


Fig. 2. Considering the story with 5 nodes presented in the figure, we may compute its score for the three different scoring schemes given in Table 2 of the main manuscript. The minimum concentration observed in the story for the red nodes is 0.2 and the maximum concentration observed for a green node is 50.5. Therefore, $ni(\text{serine}) = 0.2/0.2 = 1$, $ni(\text{cystathione}) = 50.5/50.5 = 1$, $ni(\text{glycine}) = 0.2/0.3 = 0.66$ and $ni(\text{O-acetyl-L-homoserine}) = 0.2/0.5 = 0.4$. Summing up the contribution of each arc as the product of the normalized intensity ratios of its extremities times the corresponding entrance in the score matrix, we obtain an enzyme activation score of 1.4, a concentration change score of 0.32 and an enzyme inhibition score of -1.4 . Notice that these scores cannot be compared between them, their role is to enable us to compare different stories.

5) Mapping of metabolites on the YeastCyc overview diagram

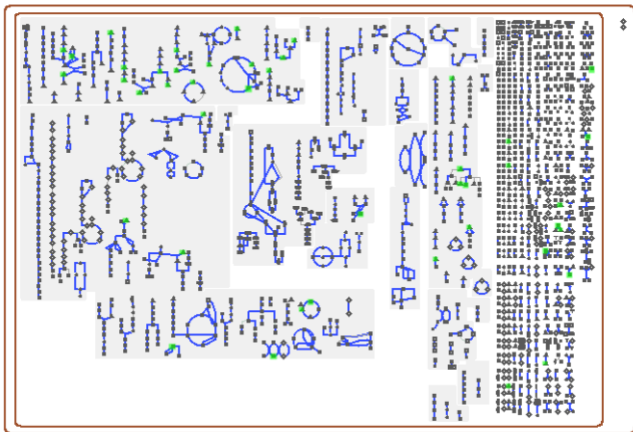


Fig. 3. Mapping of metabolites on the YeastCyc overview diagram. Notice that there are more highlighted metabolites than appear in the list. This is due to the fact that this view is pathway-oriented and metabolites are therefore duplicated. Moreover, there is no link between the pathways so the network connectivity is lost.

6) Glossary

Table 2. Main definitions used in the paper

Word	Definition
Arc	An arc of a graph $G(V, A)$ is an ordered pair $(u, v) \in A$, with $u, v \in V$. Such an arc is outgoing from u and incoming into v
Black nodes	Nodes corresponding to the discriminating compounds;
Directed graph (digraph)	A digraph is a pair (V, A) , where V is a set of nodes and A , the arc set, is a binary relation on V
Discriminating compounds	Compounds measured and whose concentration change is statistically significant
Green nodes	Nodes corresponding to the discriminating compounds whose concentration significantly increased
Metabolic story	Maximal directed acyclic subgraph that contains only black nodes as sources and targets
Red nodes	Nodes corresponding to the discriminating compounds whose concentration significantly decreased
Source	A node that has no incoming arc
Target	A node that has no outgoing arc
White nodes	Nodes corresponding to non-discriminating compounds, <i>i.e.</i> , compounds that were not measured or whose concentration did not significantly change

REFERENCES

- [1] V. Acuna, E. Birmelé, L. Cottret, P. Crescenzi, F. Jourdan, V. Lacroix, A. Marchetti-Spaccamela, A. Marino, P. V. Milreu, M.-F. Sagot, and L. Stougie. Telling stories: Enumerating maximal directed acyclic graphs with a constrained set of sources and targets. *Theor. Comput. Sci.*, 457:1–9, 2012.