BioMet Toolbox 2.0: genome-wide analysis of metabolism and omics data

Manuel Garcia-Albornoz1,†, Subazini Thankaswamy-Kosalai1,†, Avlant Nilsson1, Leif Väremo1, Intawat Nookaew1,2,* and Jens Nielsen1,*

1Department of Chemical and Biological Engineering, Chalmers University of Technology, Göteborg, Sweden and
2Comparative Genomics Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

Received December 20, 2013; Revised April 14, 2014; Accepted April 15, 2014

ABSTRACT
Analysis of large data sets using computational and mathematical tools have become a central part of biological sciences. Large amounts of data are being generated each year from different biological research fields leading to a constant development of software and algorithms aimed to deal with the increasing creation of information. The BioMet Toolbox 2.0 integrates a number of functionalities in a user-friendly environment enabling the user to work with biological data in a web interface. The unique and distinguishing feature of the BioMet Toolbox 2.0 is to provide a web user interface to tools for metabolic pathways and omics analysis developed under different platform-dependent environments enabling easy access to these computational tools.

INTRODUCTION
In the last few years, computer sciences and mathematics have contributed to the development of new strategies applied to biological research. Several free-access databases along with tools and software packages are available online aiming for the extraction of valuable information from raw data (1). In the field of biological sciences the use of computational tools has enabled a rapid expansion of new applications for data analysis and in silico simulations. New algorithms and software tools are being constantly developed helping to deal with the explosively expanding amount of data produced by science and industry (1,2).

Systems biology has been described as the holistic analysis of biological systems with the primary goal to understand the interactions between different components and their regulation. Genome-scale metabolic models (GEMs) and gene expression profiles are valuable sources of information for gene–gene interaction and phenotype predictions applied to systems biology (3,4).

The reconstruction and modification of GEMs and omics data analysis are tasks that require specialized software tools (4,5). Several tools for analysis, simulation, editing, running and visualization of GEMs and omics data have been developed and are already available; however, most of the new software and programs generated for bioinformatics applications require the installation of libraries or other additional software packages before their use. In addition to this, some programs require programming knowledge on a command-based platform, which can be bothersome for an inexperienced user. To overcome these difficulties, the upgraded version of the BioMet Toolbox (6) is, therefore, intended to provide a web user interface (WUI) to platform-dependent tools enabling the access by inexperienced users to these computational tools.

FEATURES
The main contribution of the BioMet Toolbox 2.0, is the online WUI for the previously developed RAVEN (4) and PIANO (3) tools. RAVEN is a software for GEM analysis and simulation developed in MATLAB. PIANO is a software developed in R for omics data analysis. The WUI to these platform-dependent tools provided by the BioMet Toolbox 2.0 enables the use of their functions under a user-friendly environment with no necessity of previous platform knowledge. The BioMet Toolbox v2.0 web site has been written in PHP, HTML and JavaScript. RAVEN can also be downloaded directly from the BioMet Toolbox 2.0 web site including installation guide and tutorial and PIANO can be downloaded from Bioconductor (7) through the provided link. In addition to the online WUI tools, the BioMet Toolbox 2.0 includes improvements to the user interface with a more logical layout, an expanded collection of high quality GEMs of different organisms (GEM repository) and a collection of legacy tools from the previous version of the BioMet Toolbox (Figure 1).
Figure 1. BioMet Toolbox v2.0 organization and functionalities.

The provided online WUI tools in BioMet Toolbox v2.0 offer two major groups of analysis which are: (i) GEM analysis and simulation and (ii) omics data analysis. The analysis and simulation of GEMs through the WUI tool (RAVEN powered) provide functionalities as: (i) GEM overview, (ii) GEM validation, (iii) Reporter metabolites, (iv) Flux balance analysis (FBA) and (v) GEM random sampling (Figure 2A). For the GEM overview, users are allowed to upload their GEM in specific Excel format or SBML format (8) together with omics data to check the entities of the uploaded GEM, such as number of genes, reactions and metabolites for each compartment, the reactions involving only products or only reactants and the metabolites that can be produced and consumed. In GEM validation the unbalanced reactions (Elemental balance), dead end reactions, dead end metabolites, connectivity and isolated networks can be evaluated and queried before going to the next step. For GEM analysis FBA simulations and integrative analysis of omics data using GEM as a scaffold, such as Reporter metabolite analysis (9,10), and the identification of transcripitional flux regulation using random sampling (11), are included.

Statistical values derived from omics data can be uploaded and analyzed through the WUI of omics analysis (PIANO powered), providing functionalities as: (i) Microarray quality check, (ii) Microarray differential expression analysis, (iii) Gene set analysis (GSA) and (iv) Consensus gene set analysis (Figure 2B). Since microarray data are widely used and shared in the research community, BioMet Toolbox 2.0 provides a standard microarray analysis work flow including quality assessment, normalization and differential expression analysis. Each analysis will generate result tables and appropriate plots which can be viewed directly in the WUI or be downloaded by the user. The Gene set analysis function collects a number of GSA methods into the same platform, making it easier to test different methods using the same settings, format and input. The input to this tool is a collection of gene sets and gene-level statistics. The gene sets can be, e.g. Gene Ontology terms (12) or any other terms, enabling the identification of statistically significant biological processes. The gene-level statistics can be, e.g. $P$-values and $t$-values from the Microarray differential expression analysis module or statistical values from RNA-seq data or other gene-centered omics data. The output of this tool is a network plot detailing the biological functions, and their connections, that are affected by differentially expressed genes along with a table in Excel format with the number of genes in each gene set, the gene set statistics and their $P$-values (normal and adjusted). The Consensus gene set analysis allows the user to combine results from different gene set analyses and is performed under a combination of different GSA methods in order to obtain a consensus heat map as an output.

Additionally to the WUI tools, the BioMet Toolbox 2.0 includes an expanded collection of GEMs (Models repository) including several high quality GEM reconstructions for different organisms. For fungi the available models are: *Saccharomyces cerevisiae*, *Pichia pastoris*, *Pichia stipitis*, *Aspergillus niger*, *Aspergillus oryzae* and *Aspergillus nidulans*. For bacteria the available models are: *Streptomyces coelicolor*, *Lactococcus lactis*, *Synechocystis sp. PCC6803* and *Amycolatopsis balhimycina*. All the models in the GEM repository are available in several formats. Submission of new GEMs is allowed and highly encouraged in order to expand the GEM database.

Several web sites are available for either GEM analysis or omics data analysis (13–19). Nevertheless, as outlined above, the BioMet Toolbox 2.0 offers an expanded selection of functions and tools all in one place enabling the user to combine the results of GEM analysis and omics data analysis.
Figure 2. Overview of the online tools workflow. Information flows through the model from left to right with each column showing the corresponding levels. (A) Online tool for GEM analysis and simulation. (B) Online tool for omics analysis.
Along with the WUI tools, some legacy tools (C13, BioOpt, Reporter Features and Reporter Subnetwork) from the previous version of the BioMet Toolbox (6) are still accessible from the BioMet Toolbox 2.0.

**SHOW CASE**

To illustrate the use of the BioMet Toolbox 2.0 some *in silico* simulations were performed by using the RAVEN powered online tools. These simulations were done by first uploading a yeast model (yeast 5.32) (20). The simulated media was a minimal chemostat medium (glucose-limited un...
under aerobic and anaerobic condition with a constrained up-
take rate for glucose specified from a condition of the previ-
ously reported in vivo values (21). Each model file (aerobic
and anaerobic) was then uploaded to the BioMet Toolbox
v2.0 for maximization of biomass production as an objec-
tive function. The GEM overview and GEM validation op-
tions were run in order to find any error in the model (Table
1A). These tools returned a summary of 897 genes, 2034 re-
actions and 1600 metabolites.

The shift from respiratory to fermentative metabolism
leading to ethanol production was further investigated us-
ging the GEM as a scaffold for integrative analysis of tran-
scriptome data from these growth conditions (22). The tran-
scriptome data were retrieved from the Array express num-
ber E-MEXP-3704. GEM random sampling was run, com-
paring the significance of change in flux between aerobic
and anaerobic conditions (Table 1B). This tool generates a
$N \times 3$ column matrix with the probabilities of a reaction:
(i) changing both in flux and expression in the same di-
rection, (ii) changing in expression but not in flux and (iii)
changing in flux but not in expression or changing in op-
posed directions in flux and expression. A comparison be-
tween simulated productions, uptakes and growth rates and
those obtained from previously published experimental in vivo
results for aerobic and anaerobic conditions (21) are shown
in Table 2A. The same model file was then uploaded
along with the $P$-values from differential gene expression
analysis under the two conditions to identify metabolite
hotspots. The top 10 ranking metabolites are presented in
Table 2B. Clearly, components associated with respiration
and adenosine triphosphate generation (including proton in
the mitochondria) are among the top-reporter metabolites
as the energy generation is completely shifted from respira-
tion to fermentation when changing from aerobic to anaer-
obic growth.

For the PIANO powered online tools a gene expression
data set from Saccharomyces cerevisiae was downloaded
from the Gene Expression Omnibus Database using access-
sion number GSE21988 containing the nutrient-dependent
regulation gene expression in S. cerevisiae. This data set
contains the gene expression profiles from S. cerevisiae
while growing in chemostat cultures on carbon or nitrogen
starvation using either glucose or ethanol as carbon source
(23). For this experiment, growth limitation was done by
either carbon or nitrogen. When carbon was limited, the
growth was tested on either glucose or ethanol (using am-
monium sulfate as the nitrogen source). When ammonium
sulfate was the limited factor, either glucose or ethanol
was used as the carbon source. Raw .CEL files were uploaded
to the online tool for omics analysis and the Microarray qual-
ity check was first performed, obtaining all the plots for raw
and normalized data (See example in Figure 3A). For the
Microarray differential expression analysis the compared
conditions were: Glucose versus Carbon limited, Glucose
versus Nitrogen limited, Ethanol versus Carbon limited and
Ethanol versus Nitrogen limited. The heatmap obtained
can be observed in Figure 3B. The heatmap shows the ex-
pression levels of the top significant differentially expressed
genes for the compared conditions. In this chart the dif-
fferences between the four conditions are clearly observed.
The results from gene set analysis were illustrated as a net-
work plot in Figure 3C detailing the biological functions en-
riched with significantly differentially expressed genes. Fur-
thermore the results from the consensus gene set analysis
are illustrated in Figure 3D as a heatmap and provides simi-
lar information as the network plot in the GSA, but by
showing the directionality of the gene set (up or down reg-
ulated) represents better detail for further biological inter-
pretation.

These examples show how the BioMet Toolbox v2.0 can
be used to obtain quantitative estimations of fluxes and
growth rates in good agreement with previously reported
in vivo results. At the same time, this tool provides a helpful
and robust way to perform analysis from omics data, which
can be used to identify new metabolic routes, gene targets
for genetic engineering and transcriptional changes occur-
rning in biological systems.

SUMMARY

The BioMet Toolbox v2.0 offers a selection of online soft-
ware tools for biological data analysis along with free ac-
cess to a collection of GEMs for their use in phenotype simu-
lations. Among the advantages of using the BioMet
Toolbox v2.0 are its web-based free availability and its user-
friendly and platform-independent online tools allowing for
omics data analysis and GEM analysis and simulation un-
der a WUI environment, suitable for both inexperienced
and advanced users. The online interface for the RAVEN
and PIANO powered tools represents an important ad-
advance in the field of system biology allowing the final user
to perform different features from an easy to use WUI en-
vironment avoiding any complicated software installation.
The BioMet Toolbox v2.0 offers the alternative option to
download the RAVEN and PIANO in order to perform
a wider range of functionalities under different command-
based platforms. Additionally, the BioMet Toolbox v2.0 of-
ers the possibility of constant growth in additional features
and updated functionalities.

AVAILABILITY

BioMet Toolbox v2.0 is freely available at www.biomet-
toolbox.org. Contact: biomet2 [at] sysbio.se.

FUNDING

European Research Council (INSYSBIO) [247013];
Chalmers Foundation; Knut and Alice Wallenberg Foun-
dation; Bioinformatics Infrastructure for Life Sciences
(BILS) and Samsung Electronics Inc. Source of open
access funding: Chalmers Library.

Conflict of interest statement. None declared.

REFERENCES

1. Garcia-Albornoz, M.A. and Nielsen,J. (2013) Application of
genome-scale metabolic models in metabolic engineering. Ind.
Biotechnol., 9, 203–214.
2. Osterlund,T., Nookaew,I. and Nielsen,J. (2012) Fifteen years of large
scale metabolic modeling of yeast: developments and impacts.
Biotechnol. Adv., 30, 979–988.
Table 1. (A) Part of the Script progress output for GEM overview showing the number of reactions, metabolites and genes for simulated minimal chemostat medium under aerobic condition. (B) Random sampling output for the GEM’s first top 20 reactions sorted by the highest probabilities of change in flux and expression in the same direction between two conditions (aerobic and anaerobic).

| GEM characteristics | Reaction name            | Change both in flux and expression in the same direction | Change in expression but not in flux | Change in flux but not in expression |
|---------------------|--------------------------|----------------------------------------------------------|-------------------------------------|-------------------------------------|
| id                  | 'ymn5_0'                 |                                                          | 0                                   | 0.3527                              |
| description         | 'Yeast metabolic network' |                                                          | 0.6473                              | 0                                   |
| annotation          | [1×1 struct]             | phosphoglycerate kinase triose-phosphate isomerase       | 0.6420                              | 0.0008                              |
| rxns                | [2034×1 cell]            | H+ diffusion                                             | 0.6164                              | 0.0071                              |
| mets                | [1600×1 cell]            | enolase                                                  | 0.6156                              | 0                                   |
| compNames           | [16×1 cell]              | glucose-6-phosphate isomerase                            | 0.5796                              | 0                                   |
| compOutside         | [16×1 cell]              | D-lactate/pyruvate antiport                               | 0.5664                              | 0                                   |
| rxnNames            | [2034×1 cell]            | D-lactate transport                                      | 0.5653                              | 0                                   |
| rxnComps            | [2034×1 double]          | ammonia transport                                        | 0.5526                              | 0.0886                              |
| grRules             | [2034×1 cell]            | oxygen exchange                                          | 0.5444                              | 0.0002                              |
| rxnGeneMat          | [2034×897 double]        | phosphoglycerate mutase glycine                          | 0.5210                              | 0                                   |
| subSystems          | [2034×1 cell]            | b-methyltransferase                                      | 0.5029                              | 0.0531                              |
| eccodes             | [2034×1 cell]            | fructose-bisphosphate aldolase                           | 0.4839                              | 0                                   |
| genes               | [897×1 cell]             | bicarbonate transport                                    | 0.4812                              | 0.0056                              |
| geneComps           | [897×1 double]           | bicarbonate formation                                   | 0.4711                              | 0.0056                              |
| metNames            | [1600×1 cell]            | glutamate dehydrogenase (NAD)                            | 0.4508                              | 0.1088                              |
| metComps            | [1600×1 double]          | pi-ribosyl imidazole. synths.                           | 0.4506                              | 0.2054                              |
| inchis              | [1600×1 cell]            | pi-ribosylglycinamidine synths.                         | 0.4421                              | 0.2054                              |
| metFormulas         | [1600×1 cell]            | ammonium exchange                                        | 0.4385                              | 0.0709                              |
| metMiriams          | [1600×1 cell]            | pantothenate transport                                   | 0.4385                              | 0.1894                              |
| unconstrained       | [1600×1 double]          | tyrosine transport                                       | 0.4360                              | 0.1665                              |

Table 2. (A) Comparison of previously reported in vivo fluxes and growth rates on aerobic and anaerobic conditions. Uptake and excretion fluxes are reported in mmol/g dry weight/h and growth rates in h - 1. Glucose was used as the limiting substrate. (B) Top 10 most significant reporter metabolites obtained from the reporter metabolites analysis for S. cerevisiae on minimal chemostat medium while changing from aerobic to anaerobic growth.

| A  | In vivo | BioMet Toolbox v2 |
|----|---------|-------------------|
|    |         | In silico simulation |
| **Aerobic** |         |                     |
| Carbon limited aerobic glucose uptake | 1.15 | 1                     |
| Carbon limited aerobic oxygen uptake | 2.7  | 2.74                  |
| Carbon limited aerobic CO₂ excretion | 2.8  | 2.78                  |
| Carbon limited aerobic ethanol production | 0   | 0                     |
| Carbon limited aerobic growth rate | 0.1  | 0.09                  |
| **Anaerobic** |         |                     |
| Carbon limited anaerobic glucose uptake | 2.3  | 2                     |
| Carbon limited anaerobic oxygen uptake | 0   | 0                     |
| Carbon limited anaerobic CO₂ excretion | 3.8  | 3.4                   |
| Carbon limited anaerobic ethanol production | 3   | 3.4                   |
| Carbon limited anaerobic growth rate | 0.03 | 0.04                  |

| B  |         |                     |
|----|---------|-------------------|
| Top 10 Reporter metabolites |         |                     |
| ferricytochrome c [mitochondrion] |         | ADP [mitochondrion] |
| ferrocytochrome c [mitochondrion] |         | H+ [mitochondrion]  |
| ubiquinone-6 [mitochondrion] |         | urea [cytoplasm]    |
| ubiquinol-6 [mitochondrion] |         | allantoate [cytoplasm] |
| phosphate [mitochondrion] |         | L-cysteine [extracellular] |
3. Väremo, L., Nielsen, J. and Nookaew, I. (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.*, 41, 4378–4391.

4. Agren, R., Liu, L., Shoaei, S., Vongsangnak, W., Nookaew, I. and Nielsen, J. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Comput. Biol.*, 9, e1002980.

5. Herrgard, M.J., Lee, B.S., Portnoy, V. and Palsson, B.O. (2006) Integrated analysis of regulatory and metabolic models reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res.*, 16, 627–635.

6. Cvijovic, M., Olivares-Hernandez, R., Agren, R., Dahr, N., Vongsangnak, W., Nookaew, I., Patil, K.R. and Nielsen, J. (2010) BioMet Toolbox: genome-wide analysis of metabolism. *Nucleic Acids Res.*, 38, W144–W149.

7. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, 5, R80.

8. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A. et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19, 524–531.

9. Patil, K.R. and Nielsen, J. (2008) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Nat. Acad. Sci. U. S. A.*, 102, 2685–2689.

10. Oliveira, A.P., Patil, K.R. and Nielsen, J. (2008) Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Syst. Biol.*, 2, 17.

11. Bordel, S., Agren, R. and Nielsen, J. (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.*, 6, e1000859.

12. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. et al. (2000) Gene ontology: tool for the unification of biology. *Gene Ontol. Consort. Nat. Genet.*, 25, 25–29.

13. Overbeck, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crâécy-Lagard, V., Diaz, N., Disz, T., Edwards, R. et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.*, 17, 5691–5702.

14. Maarleveld, T.R., Boele, J., Bruggeman, F.J. and Teusink, B. (2014) A data integration and visualization resource for the metabolic network of *Synechocystis sp.* PCC 6803. *Plant Physiol.*, 3, 1111–1121.

15. Cottret, L., Wildridge, D., Vinson, F., Barrett, M.P., Charles, H., Sagot, M.F. and Jourdan, F. (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.*, 38, W132–W137.

16. Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A. (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*, 13, 1805–1806.

17. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma’ayan, A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, 14, 128.

18. Glaab, E., Garibaldi, J.M. and Krasnogor, N. (2009) ArrayMining: a modular web-application for microarray analysis combining ensemble and consensus methods with cross-study normalization. *BMC Bioinformatics*, 10, 358.

19. Xia, X., McClelland, M. and Wang, Y. (2005) WebArray: an online platform for microarray data analysis. *BMC Bioinformatics*, 6, 306.

20. Heavner, B.D., Smallbone, K., Barker, B., Mendes, P. and Walker, L.P. (2012) Yeast 5—an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC Syst. Biol.*, 6, 55.

21. Nookaew, I., Jewett, M.C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J. and Bhumiratana, S. (2008) The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *Genome Res.*, 18, 1111–1121.

22. Jewett, M.C., Workman, C.T., Nookaew, I., Pizarro, F.A., Agosin, E., Barrett, M.P., Charles, H., Sagot, M.F. and Jourdan, F. (2010) Mappig condition-dependent metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.*, 38, W132–W137.

23. Jethava, V., Bhattacharyya, C., Dubhashi, D. and Vemuri, G.N. (2011) NETGEM: Network Embedded Temporal GENerative Model for gene expression data. *BMC Bioinformatics*, 12, 327.