Quantitative Biology 2017, 5(1): 99–104
DOI 10.1007/s40484-017-0098-1

RESEARCH ARTICLE

SynBioEcoli: a comprehensive metabolism network of engineered *E. coli* in three dimensional visualization

Weizhong Tu[2,†], Shaozhen Ding[1,†], Ling Wu[1], Zhe Deng[3], Hui Zhu[3], Xiaotong Xu[4], Chen Lin[4], Chaonan Ye[3], Minlu Han[3], Mengna Zhao[3], Juan Liu[4], Zixin Deng[3], Junni Chen[2], Dong-Qing Wei[5] and Qian-Nan Hu[1,*]

[1] Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
[2] Wuhan LifeSynther Cooperation Limited, Wuhan 430078, China
[3] Ministry of Education, Key Laboratory of Combinatorial Biosynthesis and Drug Discovery and School of Pharmaceutical Sciences, Wuhan University, Wuhan 430071, China
[4] State Key Laboratory of Software Engineering and School of Computer Sciences, Wuhan University , Wuhan 430072, China
[5] State Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University, Shanghai 200240, China
* Correspondence: hu_qn@tib.cas.cn

Received September 9, 2016; Revised December 19, 2016; Accepted January 15, 2017

***Background*: A comprehensive metabolism network of engineered *E. coli* is very important in systems biology and metabolomics studies. Many tools focus on two-dimensional space to display pathways in metabolic network. However, the usage of three-dimensional visualization may help to understand better the intricate topology of metabolic and regulatory networks.**

***Methods*: We manually curated large amount of experimental data (including pathways, reactions and metabolites) from literature related with different types of engineered *E. coli* and then utilized a novel technology of three dimensional visualization to develop a comprehensive metabolic network named SynBioEcoli.**

***Results*: SynBioEoli contains 740 biosynthetic pathways, 3,889 metabolic reactions, 2,255 chemical compounds manually curated from about 11,000 metabolism publications related with different types of engineered *E. coli*. Furthermore, SynBioEcoli integrates with various informatics techniques.**

***Conclusions*: SynBioEcoli could be regarded as a comprehensive knowledgebase of engineered *E. coli* and represents the next generation cellular metabolism network visualization technology. It could be accessed via web browsers (such as Google Chrome) supporting WebGL, at http://www.rxnfinder.org/synbioecoli/.**

Keywords: engineered *E. coli*; three dimensional metabolic network; biosynthetic ability

INTRODUCTION

Escherichia coli is one of the most important model organisms in biology and its metabolic GEM has aided the development of microbial systems biology [1]. To understand microbial biology at systems level, metabolic network reconstruction is a key technology to explore the structure and dynamics of cell system. Based on genome and literature data, several groups have reconstructed genome-scale metabolic network of *E. coli* [1–6]. Biological information contained in these metabolic networks always focuses on endogeny of *E. coli*, however, the larger amount of engineered information is not contained in these studies. On the other hand, many tools of visualizing information including pathways, reactions, compounds are based on traditional 2D. For instance, KEGG Atlas [7] is a graphical interface to the KEGG suite of databases, which contains a manually created global map for metabolism. Pathway Tools [8] applied in BioCyc, is a production-quality software environment for creating a type of model-organism database called

[†] These authors contributed equally to this work.

This article is dedicated to the Special Collection of Synthetic Biology, Aiming for Quantitative Control of Cellular Systems (Eds. Cheemeng Tan and Haiyan Liu).

Pathway/Genome Database and it is able to describe the genome and biochemical networks of organisms. Many other tools [9,10], such as Cytoscape, VisANT, Pathway Studio and Patik, have emerged for visually exploring biological networks. Due to the complexity of the metabolic network [10] and various types of information it contains, the traditional 2D representation of metabolism data can hardly be extended for hundreds pathways. Furthermore, metabolism data in traditional 2D visualization is a lack of compactness and information density. So, how to collect biological PRM data scattered in literature related with engineered *E. coli*, and visualize it in a global 3D overview is a big challenge. Several tools could be adapted to make the visualization of metabolism data more vivid. Arena3D [11] puts nodes into different layers to reveal interactions between node types. Since Arena3D computes separate layouts for each layer, edges between layers are often cluttered and difficult to follow. MetNetGE [12] utilizes a novel layout approach called the enhanced radial space-filling (ERSF) to give an overview of hierarchical pathway ontology and 3D tiered layouts, and its graphical user interface (GUI) is written with PyQt. However, there is a lack of experimental PRM data of engineered *E. coli* in MetaNetGE.

With great achievements that metabolic engineering and synthetic biology have received for the past 100 years, more and more biological knowledge has been discovered. At the same time, several groups have collected such information to establish biological reaction databases. For example, EcoCyc [13] combines information of metabolic process and genome of *E. coli.* KEGG [14] contains more than 9,000 biochemical reactions. BRENDA [15], an enzyme information system, collects information of enzymes, enzyme-ligands, reactions and pathways. BKM-react [16] is a non-redundant reaction database, which integrates with BREADA, KEGG and MetaCyc. However, most of these databases are not chassis-centered, and on the other hand, they did not curate comprehensively biosynthetic information for engineered *E. coli* from the original literature.

According to the above mentioned analysis, there is a lack of three dimensional online visualization web server that integrates with various informatics tools to represent a comprehensive metabolic network containing PRM data in engineered *E. coli*. Based on biological information in EcoCyc and large amount of experimental PRM data collected from science publications related with engineered *E. coli,* SynBioEcoli could make researchers have a global overview of biosynthetic ability of *E. coli* in three dimensional visualization.

RESULTS AND DISCUSSION

SynBioEcoli system

The SynBioEcoli system is constructed with several components, including data curation, quadratic partitioning, informatics tools and 3D web-based rendering

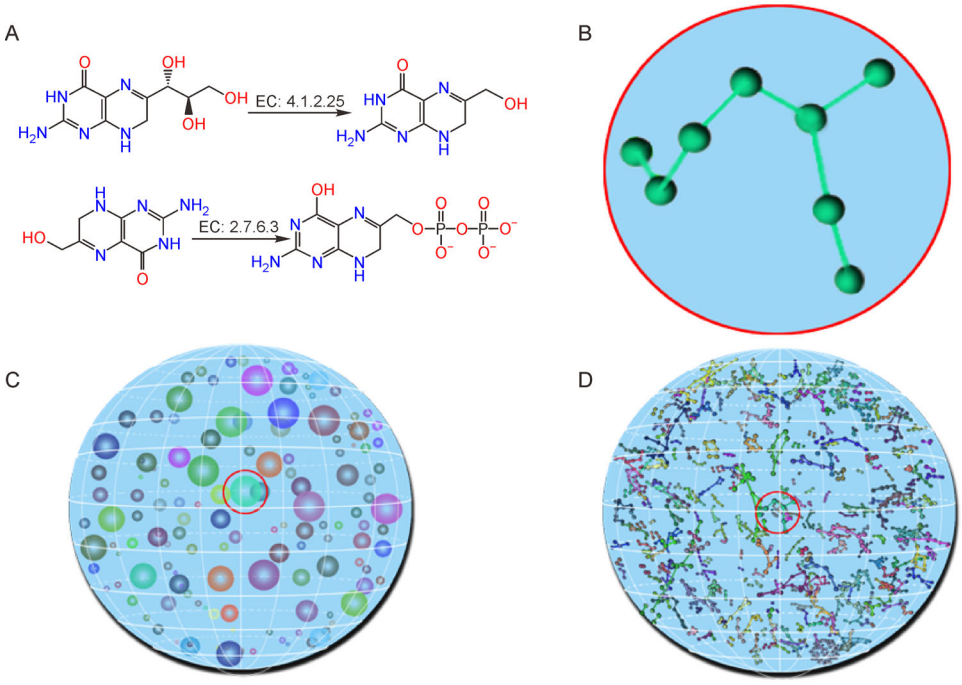

Figure 1. The building diagrams of SynBioEcoli system. (A) Metabolic reaction data curation. (B) One pathway data. (C) 3D space gridding. (D) 3D global overview after quadratic partitioning.

(shown in Figure 1). When comparing with traditional 2D visualization, the network in 3D could avoid overlapping (many nodes and edges are mixed together) that 2D network layout automatically generated (shown in Figure 2).

During metabolic network reconstruction, the solution of "reaction specificity" and "currency metabolites" is carried manually, and it is listed online at: http://www.rxnfinder.org/media/ecoli/data.html.

In the global graph of SynBioEcoli, the metabolites are treated as graph nodes, and the biochemical reactions among them are regarded as graph edges. SynBioEcoli allows researchers to retrieve the target item by names or (sub)structure smiles. Once the target item is determined, SynBioEcoli graph will reposition to make the target item shown in the center of the computer screen, researchers could click the colored object to get more information.

EcoCyc(biopax-level3.owl from its website) contains 338 biosynthetic pathways, 890 metabolic reactions, 964 chemical compounds. While SynBioEcoli contains 740 biosynthetic pathways, 3,889 biosynthetic reactions, and 2,255 chemical compounds, and it represents a more comprehensive knowledgebase to explore the biosynthetic ability of *E. coli*. Here, we provide some examples: lycopene and astaxanthin are value-added compounds. With the purpose of attaining some biosynthetic pathway information of them when utilize *E. coli* as cell factory, researchers could not retrieve any biosynthetic information in EcoCyc for the reason that they are non-native metabolite in *E. coli*. In PubMed, it is time-consuming for researchers to collect useful biosynthetic information including pathways, reactions, compounds and enzymes. While in SynBioEcoli, we collected such information manually. When users utilize lycopene as search terms, it will return five engineered pathways (Biosynthetic pathway of lycopene in *E. coli* from a foreign mevalonate pathway; Simplified diagram of glycolysis, TCA cycle, PPP, biosynthesis pathway of lycopene; lycopene biosynthesis 1; lycopene biosynthesis 2; lycopene biosynthesis 3), and when users utilize astaxanthin as search terms, it will return two engineered pathways (astaxanthin, astaxanthin biosynthetic pathway in astaxanthin-producing bacteria and the catalytic function of CrtZ and CrtW). What's more, many new clarity regarding biochemical reactions have been completely understood and successfully introduced to *E. coli*. For example, in the astaxanthin biosynthesis pathway, zeaxanthin was converted to adonixanthin, and then adonixanthin was converted to astaxanthin. Each reaction is supported by specified literature.

Substructure searching and sequence similarity searching

In SynBioEcoli, Chemoinformatics tools are implemented to search (sub)structure of metabolites and proteins/genes catalyzing enzymatic reactions. The following example in Figure 3A demonstrates a chemical substructure searching in SynBioEcoli.

After users input a query "O = C(O)c1ccccc1" (Benzoic acid), a list of compounds containing the substructure will be displayed in the panel list.

(i) When users click a compound name in the panel list, the structure image will be automatically loaded. Users could also move mouse over the structure picture to magnify for more details.

(ii) Users can click "≫" symbol to expand all pathways containing the specified compound in engineered *E. coli*. After users click a pathway, the SynBioEcoli view will be changed so that current compound will be repositioned to the center of the computer screen.

Network analysis

There are hundreds of metabolites (graph nodes) and reactions (graph edges) in SynBioEcoli according to thousands of literature related with engineered *E. coli*. Some network properties (Figure 3B), such as degree ranking and distributions of nodes, statistic of edges, the number of pathways, and network graph densities, could be automatically calculated in the network analysis function.

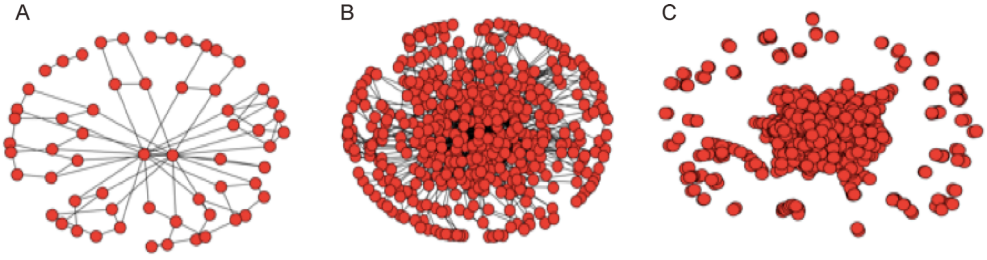

Figure 2. 2D Network/sub-network graph generated by NetWorkX. (A) Graph of sub-network containing 50 reactions in SynBioEcoli. (B) Graph of sub-network containing 300 reactions in SynBioEcoli. (C) Graph of whole network in SynBioEcoli containing 3,889 reactions.

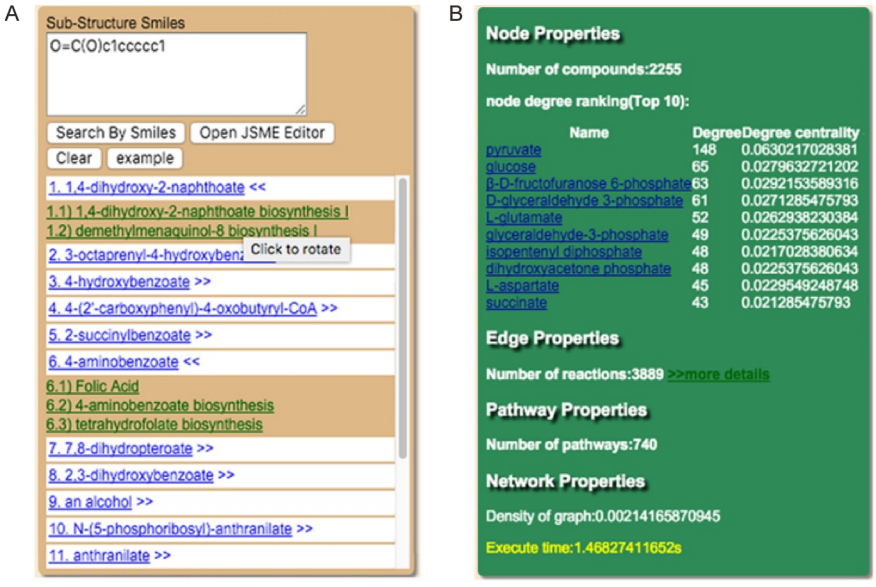

Figure 3. Bioinformatics tool examples implemented in the SynBioEcoli system. (A) Chemical substructure searching example. (B) Network property analysis.

CONCLUSIONS

SynBioEcoli utilizes novel three dimensional visualization technology to display metabolic network containing fruitful and experimental PRM data in engineered *E. coli*. With a focus on biosynthetic ability of *E. coli*, SynBioEcoli contains 740 pathways, 3,889 metabolic reactions, and 2,255 metabolites, and almost all of items are supported by specified literature. It could potentially provide a bridge between 2D metabolic network and 3D virtual reality of *E. coli* cellular metabolism, and it could be served as a comprehensive knowledgebase to explore biosynthetic ability of *E. coli* based on experimental PRM data in *E. coli* host.

METHODS

Data curation

Firstly, comprehensive PubMed biosynthetic publications were retrieved by using the search terms "(biosynthetic [Title/Abstract] OR biosynthesis [Title/Abstract] OR metabolic engineering [Title/Abstract]) AND (*Escherichia coli* [Title/Abstract] OR *E. coli* [Title/Abstract])", it returned about 11,000 publications related with engineered *E. coli* in various strain types. Most reactions and pathways are shown in diagram format in literature, so it is necessary to curate them manually. Our data curators downloaded related publications, read them and inputted the PRM data in our in-house website platform, and then several biological experts reviewed the data to ensure its correctness and completeness.

To increase reliability of data, each item in SynBioEcoli is supported by corresponding original literature. On the other hand, specific ID has been assigned to each item to avoid duplication. Three components (compounds, reactions as well as pathways) constitute the framework of SynBioEcoli, in which the EcoCyc data is also included.

Three dimensional graph visualization

A graph G = (V, E) is a set V (vertices) and E (edges), in which an edge joins a pair of vertices. A Fruchterman-Reingold force-directed graph drawing algorithm [17], which has been implemented in our previous study to visualize network pharmacology [18], was adopted again to generate the 3D coordinates of edges and nodes in this work. It is important to note that the algorithm was used twice (called as quadratic partitioning algorithm), firstly for gridding the 3D space for the pathway; and secondly for partitioning each pathway grid for chemical compounds contained in pathway. In order to make the edges have nearly equal length and avoid crossing, we utilized the algorithm above mentioned to assign forces among edges and nodes based on their relative positions, and then minimize their energy.

Substructure searching and sequence similarity searching

In the metabolic network, there are thousands of metabolites represented by nodes. In Chemoinformatics,

a chemical substructure is a subgraph of a molecule graph, and it is correspondingly labeled a manner reflecting the nature of the atoms and bonds in the original molecule. In this work, chemical substructure similar with molecular fragment searching in our previous studies [19,20], is used to retrieve related compounds and pathways.

In SynBioEcoli, similarity search methods [21–23] based on FASTA algorithm are used to retrieve specific enzyme or gene via sequence fragments of amino acid or base individually. The FASTA algorithm software package were downloaded from the FASTA team at Virginia University (http://faculty.virginia.edu/wrpearson/fasta/fasta36/).

Network analysis and web server

The network analysis function provides useful knowledge of SynBioEcoli, such as properties of node/edge/pathway/network. A python package, NetworkX (high-productivity software for complex networks) is used in the network analysis function. SynBioEcoli server system applies a Browser/Server framework under Linux environment, in which Ajax, Apache Http Server, CSS, Django, HTML5, JavaScript, Json, C++, Python, Online Molecular Editor as well as Network Graph Analysis Algorithms are included. What's more, WebGL technology and Three.js [24] Framework are used for interactively three-dimensional visualization in browser environment.

ABBREVIATIONS

PRM data,	data information involved pathways, reactions as well as metabolites
GEM,	genome-scale metabolism

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation of China (Nos. 31270101 and 31570092), the National High Technology Research and Development Program (No. 2012CB721000) and the Natural Science Foundation of Tianjin, China.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Weizhong Tu, Shaozhen Ding, Ling Wu, Zhe Deng, Hui Zhu, Xiaotong Xu, Chen Lin, Chaonan Ye, Minlu Han, Mengna Zhao, Juan Liu, Zixin Deng, Junni Chen, Dong-Qing Wei and Qian-Nan Hu declare they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. McCloskey, D., Palsson, B. O. and Feist, A. M. (2013) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. Mol. Syst. Biol., 9, 661–676
2. Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. and Palsson, B. O. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol., 3, 121–139
3. Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L. and Palsson, B. O. (2009) Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol., 7, 129–143
4. Feist, A. M. and Palsson, B. O. (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat. Biotechnol., 26, 659–667
5. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. and Barabási, A. L. (2000) The large-scale organization of metabolic networks. Nature, 407, 651–654
6. Ma, H. and Zeng, A. P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19, 270–277
7. Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S. and Kanehisa, M. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res., 36, W423–W426
8. Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., *et al.* (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Brief. Bioinform., 11, 40–79
9. Csermely, P., Korcsmaros, T., Kiss, H. J., London, G. and Nussinov, R. (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery. Pharmacol. Ther., 138, 333–408
10. Rojdestvenski, I. and Cottam, M. (2002) Visualizing metabolic networks in VRML. In Proceedings. Sixth International Conference on Information Visualisation, pp. 175–180
11. Pavlopoulos, G. A., O'Donoghue, S. I., Satagopam, V. P., Soldatos, T. G., Pafilis, E. and Schneider, R. (2008) Arena3D: visualization of biological networks in 3D. BMC Syst. Biol., 2, 104–111
12. Jia, M., Choi, S. Y., Reiners, D., Wurtele, E. S. and Dickerson, J. A. (2010) MetNetGE: interactive views of biological networks and ontologies. BMC Bioinformatics, 11, 469–485
13. Karp, P. D. and Riley, M. (1996) EcoCyc: an encyclopedia of *Escherichia coli* genes and metabolism. Nucleic Acids Res., 24, 32–39
14. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res., 42, D199–D205
15. Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., Sensen, C. W. and Schomburg, D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. Nucleic Acids Res., 43, D439–D446
16. Lang, M., Stelzer, M. and Schomburg, D. (2011) BKM-react, an integrated biochemical reaction database. BMC Biochem., 12, 42
17. Fruchterman, T. M. J. and Reingold, E. M. (1991) Graph drawing by force-directed placement. Softw. Pract. Exper., 21, 1129–1164
18. Hu, Q. N., Deng, Z., Tu, W., Yang, X., Meng, Z. B., Deng, Z. X. and Liu, J. (2014) VNP: interactive visual network pharmacology of diseases, targets, and drugs. CPT Pharmacometrics Syst. Pharmacol., 3, e105
19. Tu, W., Zhang, H., Liu, J. and Hu, Q.-N. (2015) BioSynther: a customized biosynthetic potential explorer. Bioinformatics, 32, 472–473
20. Hu, Q. N., Deng, Z., Hu, H., Cao, D. S. and Liang, Y. Z. (2011) RxnFinder: biochemical reaction search engines using molecular

structures, molecular fragments and reaction similarity. Bioinformatics, 27, 2465–2467
21. Smith, T. F. and Waterman, M. S. (1981) Identification of common molecular subsequences. J. Mol. Biol., 147, 195–197
22. Lipman, D. J. and Pearson, W. R. (1985) Rapid and sensitive protein similarity searches. Science, 227, 1435–1441
23. Pearson, W. R. and Lipman, D. J. (1988) Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA, 85, 2444–2448
24. Danchilla, B. (2012) Three.js Framework. In Beginning WebGL for HTML5, 173–203. Berkeley: Apress