Genome-scale *in silico atpE* gene knockout in *Escherichia coli* could drive novel biological discovery

Bashir Sajo Mienda

Bioinformatics Research Group (BIRG), Department of Biosciences and Health Sciences, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Skudai Johor Bahru, Malaysia

Abstract

One of the applications of *E. coli* genome-scale model is in the biological discovery of underground metabolic functions of partially characterized genes and/or enzymes. Here we report for the first time, a failed prediction of *atpE* gene knockout of no growth in the most recent *E. coli* reconstruction iJ01366 model, and a positive experimental growth on glucose, enabling a model-driven biological discovery of the underground metabolic function of this gene in *E. coli* metabolism. These findings unfolded what could be described as either scope gaps in the reconstruction or true biological gaps (knowledge gaps) on the missing *atpE* gene function in *E. coli* metabolism. This study informs other studies that the gaps could be pursued into the *E. coli* metabolism, leading to a model-driven discovery in the future. This can be achieved by using gap filling algorithms (such as GrowMatch, SMILEY and OMNI) in combination with ^{13}C labeling experiments and/or high throughput tools (such as phenotypic microarrays and robotic instruments) to update and uncover the missing *atpE* gene functions under different genetic and/or environmental conditions.

Keywords: *E. coli*, genome-scale model, ATP synthase, gene knockout prediction, knowledge gaps, biological discovery

1. Introduction

Genome-scale science has received remarkable attention in recent years because of the increasing development of Genome-scale Metabolic Models (GEMs). The assembled genome sequences and plethora of biochemical data in the form of an integrated biochemical reaction network (Reactome) of a microbial cell are called GEM (Monk and Palsson **2014**). These models have proven applications in metabolic engineering, model-driven discovery (Guzman *et al.* **2015**), prediction of cellular phenotype, studies of evolutionary process, and models of interspecies interactions (McCloskey *et al.* **2013**). Model-driven discovery application of GEM is of particular interest to genome-scale scientists, because it is used to explore a particular situation of models' failure of prediction.

Microbial model predictions are classified into four (4) computational prediction outcomes: (i) true positive, (ii) true negative, (iii) false positive and (iv) false negative predictions. The true positive prediction occurs when the model prediction of growth is consistent with the experimental measurement, and the true negative prediction occurs when the model prediction of no growth agrees (not agreed) with the experimental measurement of no growth (McCloskey *et al.* **2013**; Monk and Palsson **2014**). In contrast, predictions of growth and

Correspondence To: Bashir Sajo Mienda, Bioinformatics Research Group (BIRG), Department of Biosciences and Health Sciences, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Skudai Johor Bahru.Malaysia.

experimental measurement of no growth is known as a false positive outcome, while false negative prediction occurs when a GEM shows no growth outcome, but the experimental measurement results in growth (Monk and Palsson **2014**; O'Brien *et al.* **2015**).

Escherichia coli GEM has been used for biological discovery following a failure in prediction of gene essentiality, where multiple numbers of genes knockout reveal hidden reactions in central metabolism (Kenji *et al.* **2009**). Guided by GEM outcomes of false negative predictions, Kenji and colleagues (Kenji *et al.* **2009**) employed gap filling methods combined with systematic genes knockout in *E. coli* central carbon metabolism, that lead to the discovery of novel metabolic functions of the classic glycolytic enzymes phosphofructokinase and aldolase. In addition, the investigators performed metabolomics analysis that identified a new metabolite, sedoheptalose-1,7-bisphosphate, that had not been previously characterized (Kenji *et al.* **2009**). In a more recent study reported by Palsson and colleagues (Guzman *et al.* **2015**), they used a top-down model-driven approach and *in vivo* experimentation, where they examined three cases of genes that were inaccurately predicted as essential. These genes were evaluated separately and their isozyme's functions were discovered, reporting novel biological insight of their physiological role and underground metabolic functions in *E. coli* metabolism (Guzman *et al.* **2015**).

For the first time, the discrepancies between GEM predicted and observed growth states of *atpE* gene knockout in *E. coli* are being reported. In this study, we discovered the existence of what could be described as scope gaps and/or true biological gaps in the *E. coli* iJO1366 reconstruction. The current work also suggested gap-filling algorithms (such as GrowMatch, SMILEY and OMNI) that could help compute the possible reasons for the failure of this prediction, and possibly design targeted experiments that could correct inconsistencies in metabolic knowledge in the future.

2. Materials and methods

2.1. *In silico* analysis of gene knockout

E. coli genome-scale model iJO1366 (Orth *et al.* **2011**) was employed for the *in silico* evaluation of gene deletion by using the Minimization of Metabolic Adjustment (MOMA) algorithm (Segre *et al.* **2002**) with the OptFlux software platform-

(http://www.optflux.org) (Rocha *et al.* **2010**). *E. coli* iJO1366 model has been tested and proven to be of use in prediction of computations of growth rates and metabolite excretion rates from a range of substrates and genetic conditions (Feist *et al.* **2010**; Orth *et al.* **2011**). MOMA was described as a flux-based analysis technique that uses quadratic programming to search for the nearest point in the feasible solution space of the mutant model in relations to its wild-type optimal point feasible solution space (Segre *et al.* **2002**). The OptFlux software platform is an metabolic engineering (ME) platform that was implemented using the Java programming, which contains MOMA as a simulation algorithm. Flux Balance Analysis (FBA) was used for all phenotype simulations in this study. All the simulations of the mutant and the wild-type models were performed using the OptFlux software version 3.07.

Glucose was chosen as solitary carbon sources under aerobic conditions. The substrate uptake rates were constrained to a maximum of 18.5 mmol $gDW^{-1}h^{-1}$, whereas its corresponding oxygen uptake rate was set to 20 mmol $gDW^{-1}h^{-1}$, because the environmental condition is aerobic. These values were selected based on closely established experimental observations on aerobic and anaerobic growth in *E. coli* (Amit Varma **1993**; Varma and Palsson **1994**).

2.2. Bacteria and plasmid

E. coli JM109 (F¢ (traD36, proAB+ lacIq, D (lacZ) M15) endA1 recA1 hsdR17 (rk-, mk +) mcrA supE44 l- gyrA96 relA1 D (lacproAB) thi-1) was used for the maintenance of pKD4 template plasmid, pKD46 plasmid and the λ-Red helper plasmid. The plasmids were used strictly following the method described by Wanner and colleagues. The plasmid pKD4 was extracted from JM109 using the QIAprep Miniprep kit (QIAGEN) according to manufacturer's specifications.

2.3. Media chemicals and other reagents

E. coli cells used in this study were grown in LB medium containing 0.5% yeast extract (Difco), 0.5 NaCl and 1% Bacto tryptone (Difco) without or with antibiotics at the concentrations of 100 μg/ml ampicillin and 30 μg/ml of Kanamycin. L-arabinose, glucose were from Sigma Aldrich. KAPA HiFi Hotstart Ready Mix (2X) was obtained from KAPA BIOSYSTEMS. Agarose was purchased from Sigma Aldrich.

Table 1 Sources and characteristics of strains, plasmids and primers used in this study

E. coli strains	Relevant characteristics	Sources
JM109	Wild-type	Lab collection
Δ*atpE*/b3737	Δ*atpE* :: FRT-Kan-FRT	This study
Plasmids		
pKD4	bla FRT-kan-FRT	Datsenko and Wanner (**2000**)
pKD46	bla γ β exo (Red recombinase), temperature-conditional replicon	Datsenko and Wanner (**2000**)
Primers		
*atpE*_F	5'- CTACGCGACA GCGAACATCA CGTACAGACCCAGACCGTGTAGGCTGGAGCTGCTTC-3'	
*atpE*_R	5'- ATGGAAAACCTGAATATGGATCTGCTGTACATGGCTCATATGAATATCCTCCTTAG -3'	

2.4. PCR Primers

E. coli gene sequence was used to design forward and reverse primers with the pKD4 template plasmid sequence. The primers had a 36-nt 5' extension including the gene initiation codon (H1) and 20-nt sequence (p1) as described previously (Baba *et al.* **2006**; Datsenko and Wanner **2000**). See **Table 1** for details of the primers used in this study.

2.5. Generation of PCR fragments

PCR reactions were carried out in an Eppendorf thermo cycle using 25 µl reactions containing 12.5 µl of KAPA HiFi Hotstart Ready Mix (2X), 1µl of pKD4 template DNA, 1.0 µl of each primer. Reactions were performed for 30 cycles: 95ºC for 3 min, 98ºC for 20 sec, 55ºC for 15 sec, 72ºC for 1:30 sec, 72ºC for 60 sec and were cooled at 4ºC. PCR products were purified using the SV gel and PCR clean up system (Promega, USA), according to the manufacturer's protocol. PCR products were analyzed by 1% agarose gel-electrophoresis using 1X Tris-acetate buffer.

2.6. Electroporation and mutant selection

E. coli JM109 harboring the λ-Red helper plasmid pKD46 was grown in 100 ml of LB medium with ampicillin and 1mM L-Arabinose at 30ºC until an optical density at 600 reached 0.3. Competent cells for electroporation were prepared as described previously (Sharan *et al.* **2009**). Then 1 µl (400 ng) of the PCR fragment was mixed with 50 µl of competent cell in an ice-cold Eppendorf electroporation cuvette (0.2 cm). Electroporation was performed at 2.5KV with 2mF and 600Ω, followed by immediate addition of 1ml of SOC medium which contained (0.5% yeast extract (Difco), 2% Bacto tryptone (Difco), 2.5 mM KCl, 10 mM NaCl,
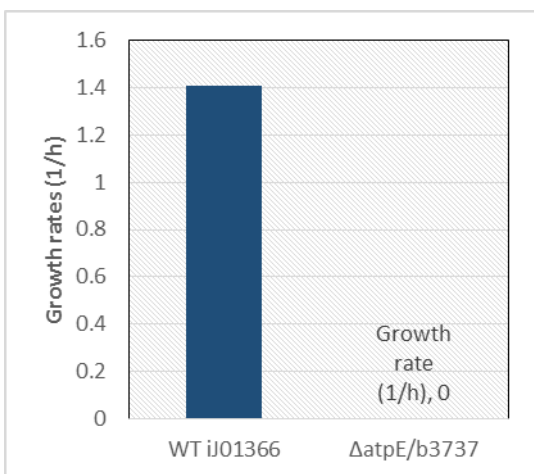
10 mM MgCl2, 10 mM MgSO4, 20 mM glucose) with 1 mM L-arabinose. The SOC medium mixed with the electroporated cells was incubated for 2 hours at 37ºC. Selection of kan^R transformant was followed immediately by spreading a one-tenth portion of it onto kanamycin agar plate as described by Baba and colleagues (Baba *et al.* **2006**). Growth tests of the *E. coli* mutant strain and the wild-type were performed in LB medium as described by Baba and colleagues (Baba *et al.* **2006**).

3. Results and discussions

E. coli genome scale metabolic models have been established to predict biological capabilities of genes knockout with notable application in systems metabolic engineering (Feist *et al.* **2010**; Guzman *et al.* **2015**; O'Brien *et al.* **2015**; Oberhardt *et al.* **2009**).However, sometimes models are often characterized by prediction failures, though it does not sound good if models fail to predict accurately. These predictive failures are perhaps of more interest to genome-scale scientists than their successes, because they represent opportunities for novel biological discoveries (Guzman *et al.* **2015**; Monk and Palsson **2014**; O'Brien *et al.* **2015**). In this study, we report for the first time how gene knockout was predicted to be essential in the model iJO1366 and the experimental results prove otherwise.

The results of our *in silico* prediction using *E. coli* iJO1366 wild type model under aerobic conditions indicated a true positive simulation outcome of growth-growth (GG) consistency (**Fig 1 and Fig 2A**). On the other hand, gene knockout result in mutant iJO1366 model under aerobic conditions indicated no growth (**Fig 1**), while in the actual experimental measurement, there is a growth (**Fig 2B**). This result indicated that there is a "no growth-growth inconsistency" (NGG) (Orth and Palsson **2010**). This type of prediction outcome is called false-negative prediction (Monk and
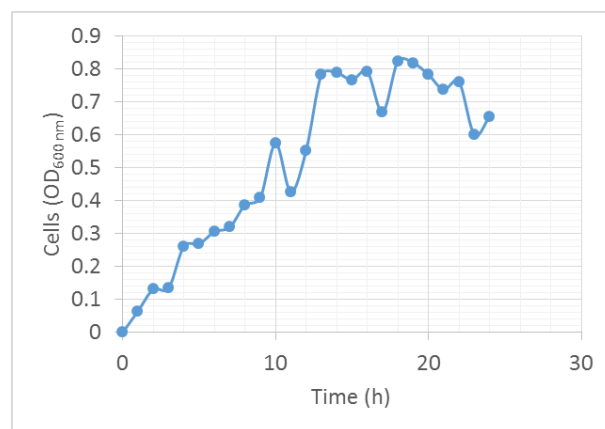
Palsson **2014**). This discrepancy indicated that the reconstructed reactome of the iJO1366 model is incomplete, representing either scope gaps in the reconstruction or true biological gaps (knowledge gaps). The scope gaps sometimes exist in a reconstruction, because large scale metabolic network models do not entail other systems such as signaling, transcription and translation, as such, metabolites produced during metabolism and enter these other systems could be left as gaps in the models, despite the fact that their biological functions are fully specified (Orth and Palsson **2010**). The true biological gaps on the other hand, represent cases of existing unknown biochemical reactions that produce or consume certain metabolites, which in turn, show a gap that could not be realized using a reconstruction (Orth and Palsson **2010**). These type of gaps represent our limited knowledge, and to fill them, new biological discoveries must be pursued, combining gap filling algorithms, such as GrowMatch (Kumar and Maranas **2009**), SMILEY (Reed *et al.* **2006**) and OMNI (Herrga *et al.* **2006**) with experimental verifications using ^{13}C labeling experiments and/or high throughput tools such as phenotypic microarrays and robotic instruments to screen cells at high rates (O'Brien *et al.* **2015**).
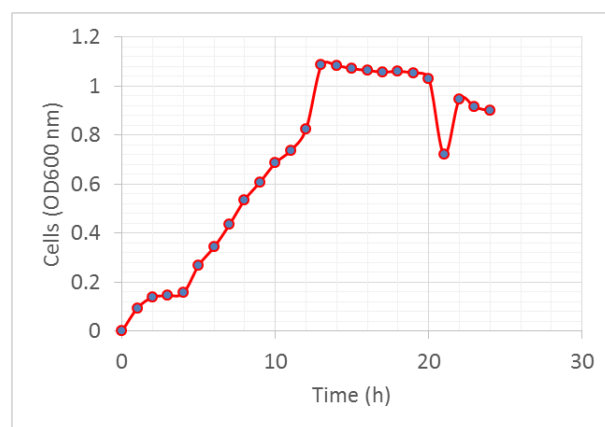

Fig 1 *In silico* growth rates of the wild-type (iJO1366 model) and the mutant Model (Δ*atpE*).

Genome-scale Metabolic Models (GEM) have been established to accurately predict metabolic engineering interventions (Feist *et al.* **2010**; McCloskey *et al.* **2013**; Oberhardt *et al.* **2009**; Orth *et al.* **2011**), although the accuracy of the prediction exercise of any GEM is predicated on an accurate reconstruction of the reactome (Monk & Palsson, **2014**). An incompletely reconstructed reactome leads to false negative predictions. Correct predictions usually align with the experimental results while inaccurate predictions do not

(O'Brien *et al.* **2015**). Prediction failures can be employed to systematically design strategies using a number of algorithms (GrowMatch, SMILEY, OMNI etc.) to build computer-aided hypotheses that could address these failures. The discrepancies seen in the current study with the gene knockout could be addressed by using computer-aided hypotheses building with gap-filling algorithm(s) and/or employed a previously established workflow methodology for model-driven discovery (Guzman *et al.* **2015**) to unfold the underground metabolic functions of the gene and its relevant isozyme. Reconciling the differences between GEM predicted results and observed experimental growth states has now been considered to be a proven strategy for novel biological discovery (O'Brien *et al.* **2015**).

Fig 2A Experimental growth profile of the wild-type *E. coli* strain.

Fig 2B Experimental growth profile of the *atpE* mutant strain.

4. Conclusion

Predictive failures in *E. coli* GEM are perhaps of more interest to genome-scale scientists, than success, as it paves way for biological discovery. The discrepancies between experimental observations and GEM predictions have been used to unfold gaps in our knowledge, where targeted experiments can be designed to correct inconsistencies in metabolic knowledge. In this study, we report for the first time, the discrepancies between GEM predicted and observed growth states of gene knockout in *E. coli*, where the existence of what could be described as scope gaps and/or true biological gaps in the *E. coli* iJO1366 reconstruction was discovered. The current work also suggested the use of gap-filling algorithms (such as GrowMatch, SMILEY and OMNI) that could help compute the possible reasons for the failure of this prediction, and possibly design targeted experiments that could correct inconsistencies in metabolic knowledge in the future.

5. Conflict of interest

There is no conflict in interest with authors during the work accomplishment.

6. References

Amit Varma BWB and Palsson BO (**1993**) Stoichiometric Interpretation of Escherichia coli Glucose Catabolism under Various Oxygenation Rates. Appl Environ Microbiol; 59(8): 2465-2473.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, *et al*. (**2006**) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol; 2:2006. 0008.

Datsenko KA and Wanner BL (**2000**) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA; 97(12), 6640-6645.

Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ and Palsson BO (**2010**) Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. Metab Eng; 12(3):173-186.

Guzman GI, Utrilla J, Nurk S, Brunk E, Monk JM, Ebrahim A, *et al.* (**2015**). Model-driven discovery of underground metabolic functions in Escherichia coli. Proc Natl Acad Sci USA; 112(3): 929-934.

Herrga MJ, Fong SS and Palsson BØ (**2006**) Identification of Genome-Scale Metabolic Network Models Using Experimentally Measured Flux Profiles. Plos Computational Biology; 2(7), 676-686.

Kenji N, Yoshihiro T, Nobuyoshi I, Tomoyoshi S, Miki H, Hisami W, *et al.* (**2009**) Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol Syst Biol; 5(306): 1-14.

Kumar VS and Maranas CD (**2009**). GrowMatch: An Automated Method for Reconciling /*In Vivo* Growth Predictions. Plos Computational Biology; 5(3):1-13.

McCloskey D, Palsson BO and Feist AM (**2013**) Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. Mol Syst Biol; 9: 661.

Monk J and Palsson BO (**2014**) Genetics. Predicting microbial growth. Science; 344(6191): 1448-1449.

O'Brien EJ, Monk J M and Palsson BO (**2015**) Using Genome-scale Models to Predict Biological Capabilities. Cell; 161(5): 971-987.

Oberhardt MA, Palsson BO and Papin JA (**2009**). Applications of genome-scale metabolic reconstructions. Mol Syst Biol; 5: 320.

Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM and Palsson BO (**2011**) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism--**2011**. Mol Syst Biol; 7: 535.

Orth JD and Palsson BO (**2010**) Systematizing the generation of missing metabolic knowledge. Biotechnol Bioeng; 107(3); 403-412.

Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, *et al.* (**2006**) Systems approach to refining genome annotation. Proc Natl Acad Sci USA; 103(46): 17480-17484.

Rocha I, Maia P, Evangelista P, *et al.* (**2010**). OptFlux: an open-source software platform for metabolic engineering. BMC Syst Biol; 4: 45.

Segre D, Vitkup D and Church GM (**2002**). Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA; 99(23): 15112-15117.

Sharan SK, Thomason LC, Kuznetsov SG and Court DL (**2009**) Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc; 4(2): 206-223.

Varma A and Palsson BO (**1994**) Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110. Appl Environ. Microbiol; 60: 3724 - 3731.

.