Scholastic Medical Sciences

Research Article

Volume 3 Issue 2- 2025

# *Ab Initio* Whole Cell Kinetic Models of *Escherichia coli* BL21 (ebeTBSW25) and MG1655 (ecoMAL25)

**Travina BS Wong[1,2], Minh Anh Le[1,2], Madhunisha Arivazhagan[1,2], Ashmitha Senthilkumar[1,2], Keng Yao Yeo[1,2], Tanisha Saisudhanbabu[1,2], Victor R Lukianto[1,2], Maurice HT Ling[1,2,3]**

[1]*School of Life Sciences, Management Development Institute of Singapore, Singapore*

[2]*Department of Applied Sciences, Northumbria University, United Kingdom*

[3]*HOHY PTE LTD, Singapore*

**Corresponding Author:** Maurice HT Ling, School of Life Sciences, Management Development Institute of Singapore, Singapore.

**Received:** February 28, 2025; **Published:** March 16, 2025

## Abstract

*Escherichia coli* is a well-studied organism and has been metabolic engineering for wide variety of chemical production. Mathematical models are useful tools in this endeavour as they can guide engineering approaches. Such models can be classified primarily into genome-scale models (GSMs) or kinetic models (KMs). Over the years, several GSMs have been developed but only 1 KM (ecoJC20) has been developed for *E. coli*. KM ecoJC20 is only about 19-24% the number of reactions compared to notable *E. coli* GSMs, as it is an adaptation of a generic cell KM of the same size. This study aims to construct larger whole cell kinetic models of *E. coli* MG1655 and BL21 using previously described *ab initio* approach via annotated genomes. The result is 2 simulatable whole cell kinetic models of *E. coli* – one for BL21 strain, ebeTBSW25, which consists of 1624 metabolites and 2340 enzymatic reactions; and another for MG1655 strain, ecoMAL25, which consists of 1801 metabolites and 3012 enzymatic reactions. The number of reactions and metabolites in both ebeTBSW25 and ecoMAL25 are substantially larger than ecoJC20 and within the range of existing GSMs.

## Introduction

*Escherichia coli* is likely the most studied organism [1], and undoubtedly the workhorse of metabolic engineering for chemical production [2]. It has been engineered to produce many different metabolites; such as, 1,3-butanediol [3], L-homoserine [4], 2-phenylethylacetate [5], Vitamin B12 [6], benzoic acid [7], 5-aminolevulinic acid [8], acetins [9], and short-chain primary amines [10]. *E. coli* MG1655 [11] and BL21 [12] were often used due to their robustness, and high biomass yield.

Predicting metabolic perturbations using mathematical models are useful in metabolic engineering [13] as it can guide engineering approaches [14, 15]. Genome-scale models (GSMs, also known as constraint-based models) and kinetic models (KMs) are the two main modelling approaches [16, 17]. GSMs are steady-state stoichiometric models which lacks enzymatic regulation [18] and difficult to add genes (transgenes) into the system as its original purpose is to evaluate changes in native gene expression on its metabolism [19, 20]. On the other hand, kinetic models (KMs) can have regulation and is much easier to add transgenes. Furthermore, KMs can predict both rates and yield of metabolites [21] while GSMs are primarily for rates. At the same time, large-scale KMs; such as, whole cell KMs; offer more intricate details [22] and have important applications [23, 24] compared to smaller-scale KMs; such as, KMs of pathways. There are a number of GSMs for *E. coli*. Notable ones include iAF1260 (based on Accession NC_000913.3; containing 1668 metabolites, 1261 genes, and 2382 reactions) [25], iJO1366 (based on Accession NC_000913.3; containing 1805 metabolites, 1367 genes, and 2583 reactions) [26], iDK1463 (based on Accession CP007799; containing 1313 metabolites, 1463 genes, and 2984 reactions) [27], iML1515 (based on Accession NC_000913.3; containing 1877 metabolites, 1516 genes, and 2712 reactions) [28], iHM1533 (based on Accession CP007799; containing 2069 metabolites, 1487 genes, and 2867 reactions) [29].

A whole cell kinetic model of *E. coli*, ecoJC20, has been reported [30] by adapting a generic whole cell kinetic model, UniKin1 [31]. However, UniKin1 consists of only 566 reactions, 306 metabolites, and 310 enzymes; which is substantially fewer than include iAF1260 [25], iJO1366 [26], iDK1463 [27], iML1515 [28], or iHM1533 [29].

Recent studies demonstrated that whole cell kinetic models may also be constructed by identifying reactions from annotated genomes via enzyme names [32, 33]. Hence, this study aims to construct larger whole cell kinetic models of *E. coli* MG1655 and BL21 via annotated genomes. We present whole cell kinetic models of *E. coli* BL21, ebeTBSW25, comprising of 1624 metabolites, and 2340 enzymatic reactions; and *E. coli* MG1655, ecoMAL25, comprising of 1801 metabolites, and 3012 enzymatic reactions; thereby, bypassing the structural limitations of UniKin1 [31].

## Methods

The reactomes of *E. coli* MG1655 and BL21 were identified from its genomes, Accession number NC_000913.3 [34] and NZ_CP053601.1 , respectively; via identification of enzymatic genes using the process previously described [32, 33, 35]. The end result was a list of enzymes, a list of substrates and products of each enzymatic reactions, and a list of metabolites deduced from the substrates and products. The production of each enzyme was modelled as the production of mRNA and peptide as previously described [33]. The model was constructed in accordance to AdvanceSyn Model Specification [16], and tested for simulatability.

Initial concentrations of all mRNA and enzymes were set to 0 mM. Initial concentrations of all metabolites were set to 1 mM except the following which were set to 1000 mM – For *E. coli* MG1655's model (ecoMAL25): (i) C00001 (Water), (ii) C00002 (ATP), (iii) C00003 (NAD+), (iv) C00004 (NADH), (v) C00005 (NADPH), (vi) C00006 (NADP+), (vii) C00007 (Oxygen), (viii) C00011 (Carbon Dioxide), (ix) C00014 (Ammonia), (x) C00025 (L-Glutamate), (xi) C00031 (D-Glucose), (xii) C00041 (L-Alanine), (xiii) C00047 (L-Lysine), (xiv) C00049 (L-Aspartate), (xv) C00062 (L-Arginine), (xvi) C00064 (L-Glutamine), (xvii) C00065 (L-Serine), (xviii) C00073 (L-Methionine), (xix) C00077 (L-Ornithine), (xx) C00078 (L-Tryptophan), (xxi) C00079 (L-Phenylalanine), (xxii) C00082 (L-Tyrosine), (xxiii) C00097 (L-Cysteine), (xxiv) C00123 (L-Leucine), (xxv) C00135 (L-Histidine), (xxvi) C00148 (L-Proline), (xxvii) C00152 (L-Asparagine), (xxviii) C00183 (L-Valine), (xxix) C00188 (L-Threonine), and (xxx) C00221 (beta-D-Glucose). For *E. coli* BL21's model (ebeTBSW25): (i) C00001 (Water), (ii) C00002 (ATP), (iii) C00003 (NAD), (iv) C00004 (NADH), (v) C00005 (NADPH), (vi) C00006 (NADP), (vii) C00007 (Oxygen), (viii) C00008 (ADP), (ix) C00009 (Phosphate), (x) C00010 (CoA), (xi) C00011 (Carbon Dioxide), (xii) C00014 (Ammonia), (xiii) C00022 (Pyruvate), (xiv) C00025 (L-Glutamate), (xv) C00031 (Glucose), (xvi) C00037 (Glycine), (xvii) C00041 (L-Alanine), (xviii) C00047 (L-Lysine), (xix) C00049 (L-Aspartate), (xx) C00062 (L-Arginine), (xxi) C00064 (L-Glutamine), (xxii) C00065 (L-Serine), (xxiii) C00073 (L-Methionine), (xxiv) C00077 (L-Ornithine), (xxv) C00079 (L-Phenylalanine), (xxvi) C00082 (L-Tyrosine), (xxvii) C00097 (L-Cysteine), (xxviii) C00123 (L-Leucine), (xxix) C00135 (L-Histidine), (xxx) C00148 (L-Proline), (xxxi) C00152 (L-Asparagine), (xxxii) C00183 (L-Valine), (xxxiii) C00188 (L-Threonine), (xxxiv) C00407 (L-Isoleucine), and (xxxv) G10008 (N-Acetyl-D-glucosamine. The model was simulated using the fourth-order Runge-Kutta method [36, 37] from time zero to 3600 seconds with timestep of 0.1 second, and the concentrations of metabolites were bounded between 0 millimolar and 1000 millimolar. The simulation results were sampled every 2 seconds.

## Results and Discussion

753 enzymes, corresponding to 2340 enzymatic reactions and 1624 metabolites, were identified from the annotated genome of *E. coli* BL21 via KEGG [38]. Similarly, 961 enzymes, corresponding to 3012 enzymatic reactions and 1801 metabolites, were identified from the annotated genome of *E. coli* ML1655 via KEGG [38]. These reactions were developed into simulatable models based on AdvanceSyn Model Specification [16]. The number of reactions and metabolites identified for both strains are substantially higher than that of ecoJC20 [30], and within the range of existing GSMs (Table 1), suggesting that our resulting models are considered large enough to be whole cell models.

The resulting models, denoted as ecoMAL25 and ebeTBSW25, were simulated using AdvanceSyn Toolkit [16]. Our simulation results (Figure 1) suggests that the models are free from syntax error as the presence of simulation results suggests that the constructed models can be simulated. Hence, we present simulatable whole cell KMs of *E. coli* MG1655 and BL21, which can be a base template for incorporating other cellular and growth processes [39-41], or as a system to examine cellular resource allocations [24, 42-44].

| Model Type | Model | Number of Metabolites | Number of Reactions |
|---|---|---|---|
| Genome-Scale Model (GSM) | iAF1260 [25] | 1668 | 2382 |
| | iJO1366 [26] | 1805 | 2583 |
| | iDK1463 [27] | 1313 | 2984 |
| | iML1515 [28] | 1877 | 2712 |
| | iHM1533 [29] | 2069 | 2867 |
| | ecoJC20 [30] | 306 | 566 |
| Kinetic Model (KM) | ebeTBSW25 (This study) | 1624 | 2340 |
| | ecoMAL25 (This study) | 1801 | 3012 |

**Table 1:** Comparison of Models.

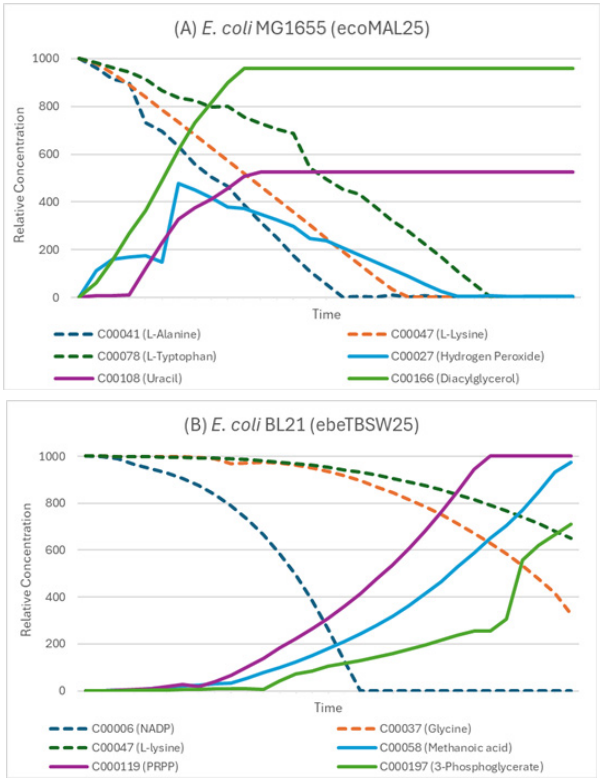


**Figure 1:** Selection of Simulation Results. Panel A shows a selection of results for ecoMAL25 while Panel B shows a selection of results for ebeTBSW25.

## Conclusion

In this study, we present an *ab initio* whole cell kinetic model of *Escherichia coli* BL21, ebeTBSW25, comprising of 1624 metabolites, 753 enzymes with corresponding transcriptions and translations, and 2340 enzymatic reactions; and *Escherichia coli* MG1655, ecoMAL25, comprising of 1801 metabolites, 961 enzymes with corresponding transcriptions and translations, and 3012 enzymatic reactions.

## Supplementary Materials

Reaction descriptions and model can be download from https://bit.ly/ebeTBSW25, and https://bit.ly/ecoMAL25; for *Escherichia coli* BL21 and MG1655 respectively.

## Note

Travina BS Wong, and Minh Anh Le are joint first authors.

## Conflict of Interest

The authors declare no conflict of interest.

## References

1. Ruiz N, Silhavy TJ (2022) "How Escherichia coli Became the Flagship Bacterium of Molecular Biology". *Journal of Bacteriology* 204(9): e0023022.
2. Valle A, Bolívar J (2021) "Escherichia coli, The Workhorse Cell Factory for the Production of Chemicals. Microbial Cell Factories Engineering for Production of Biomolecules". *Elsevier*, pp 115–137.
3. Islam T, et al;(2023) "Metabolic Engineering of Escherichia coli for Biological Production of 1, 3-Butanediol". *Bioresource Technology* 376: pp128911.
4. Sun Y, et al; (2024) "Metabolic Engineering of Escherichia coli for the Production of l-Homoserine". *Chem & Bio Engineering* 1(3): pp223–230.
5. Guo D, et al; (2017) "Metabolic Engineering of Escherichia coli for Production of 2-Phenylethylacetate from L-Phenylalanine". *Microbiology Open* 6(4): e00486.
6. Fang H, et al; (2018) "Metabolic Engineering of Escherichia coli for De Novo Biosynthesis of Vitamin B12". *Nature Communications* 9(1): pp4917.
7. Luo ZW, Lee SY (2020) "Metabolic Engineering of Escherichia coli for the Production of Benzoic Acid from Glucose". *Metabolic Engineering* 62: pp298–311.
8. Yang Y, et al; (2024) "Metabolic Engineering of Escherichia coli for the Production of 5-Aminolevulinic Acid Based on Combined Metabolic Pathway Modification and Reporter-Guided Mutant Selection (RGMS)". *Biotechnology for Biofuels and Bioproducts* 17(1):82.
9. Zada B, et al; (2020) "Metabolic Engineering of Escherichia coli for Production of Non-Natural Acetins from Glycerol". *Green Chemistry* 22(22): pp7788–7802.
10. Kim DI, et al; (2021) "Microbial Production of Multiple Short-Chain Primary Amines via Retrobiosynthesis". *Nature Communications* 12(1):173.
11. Wang J, et al; (2020) "Converting Escherichia coli MG1655 into a Chemical Overproducer Through Inactivating Defense System Against Exogenous DNA". *Synthetic and Systems Biotechnology* 5(4): pp333–342.
12. Waegeman H, et al; (2013) "Increasing Recombinant Protein Production in Escherichia coli K12 Through Metabolic Engineering". *New Biotechnology* 30(2):pp255–261.
13. Khanijou JK, et al; (2022) "Metabolomics and Modelling Approaches for Systems Metabolic Engineering. *Metabolic Engineering Communications* 15: ppe00209.
14. Gudmundsson S, Nogales J (2021) "Recent Advances in Model-Assisted Metabolic Engineering". *Current Opinion in Systems Biology* 28:pp100392.
15. Strutz J, et al; (2019) "Metabolic Kinetic Modeling Provides Insight into Complex Biological Questions, but Hurdles Remain". *Current Opinion in Biotechnology* 59: pp24–30.
16. Ling MH (2020) "AdvanceSyn Toolkit: An Open Source Suite for Model Development and Analysis in Biological Engineering". *MOJ Proteomics & Bioinformatics* 9(4): pp83–86.
17. Richelle A, David B, et al; (2020) "Towards a Widespread Adoption of Metabolic Modeling Tools in Biopharmaceutical Industry: A Process Systems Biology Engineering Perspective". *npj Systems Biology and Applications* 6(1):6.
18. Srinivasan S, et al; (2015) "Constructing Kinetic Models of Metabolism at Genome-Scales: A Review". *Biotechnology Journal* 10(9): pp1345–1359.
19. Gu C, et al; (2019) "Current Status and Applications of Genome-Scale Metabolic Models". *Genome Biology* 20(1):121.
20. O'Brien EJ, et al; (2015) "Using Genome-scale Models to Predict Biological Capabilities". Cell 161(5):971–987.
21. Prabhu S, et al; (2025) "Derivative-Free Domain-Informed Data-Driven Discovery of Sparse Kinetic Models". Industrial & Engineering Chemistry Research 64(5):2601–2615.
22. Lucido A, et al; (2025) "Multiscale Mathematical Modeling in Systems Biology: A Framework to Boost Plant Synthetic Biology". *Plants* 14(3): pp470.
23. Carrera J, Covert MW (2015) "Why Build Whole-Cell Models?" *Trends in Cell Biology* 25(12): pp719–722.
24. Thornburg ZR, et al;(2022) "Fundamental Behaviors Emerge From Simulations of a Living Minimal Cell". *Cell* 185(2): pp345-360. e28.
25. Feist AM, et al; (2007) "A Genome-Scale Metabolic Reconstruction for Escherichia coli K-12 MG1655 that Accounts for 1260 ORFs and Thermodynamic Information". *Molecular Systems Biology* 3:121.
26. Orth JD, et al; (2011) "A Comprehensive Genome-Scale Reconstruction of Escherichia coli Metabolism — 2011". *Molecular Systems Biology* 7(1):535.
27. Kim D, et al; (2021) "Development of a Genome-Scale Metabolic Model and Phenome Analysis of the Probiotic Escherichia coli Strain Nissle 1917". *International Journal of Molecular Sciences* 22(4):2122.

28. Monk JM, et al; (2017) "iML1515, A Knowledgebase that Computes Escherichia coli Traits". *Nature Biotechnology* 35(10): pp904–908.

29. van 't Hof M, et al; (2022) "High-Quality Genome-Scale Metabolic Network Reconstruction of Probiotic Bacterium Escherichia coli Nissle 1917". *BMC bioinformatics* 23(1):566.

30. Cho JL, Ling MH (2021) "Adaptation of Whole Cell Kinetic Model Template, UniKin1, to Escherichia coli Whole Cell Kinetic Model, ecoJC20". *EC Microbiology* 17(2): pp254–260.

31. Murthy MV, Balan D, et al; (2020) "UniKin1: A Universal, Non-Species-Specific Whole Cell Kinetic Model". *Acta Scientific Microbiology* 3(10): pp04–08.

32. Kwan ZJ, et al; (2024) "Ab Initio Whole Cell Kinetic Model of Stutzerimonas balearica DSM 6083 (pbmKZJ23)". *Acta Scientific Microbiology* 7(2): pp28–31.

33. Arivazhagan M, et al; (2025) Ab Initio Whole Cell Kinetic Model of Bifidobacterium bifidum BGN4 (bbfMA24). *Acta Scientific Nutritional Health* 9(1):42–45.

34. FR, Plunkett G, et al; (1997) "The complete genome sequence of Escherichia coli K-12". *Science* 277(5331):1453–1462".

35. Senthilkumar A, et al; (2025) "Ab Initio Whole Cell Kinetic Model of Lactobacillus acidophilus NCFM (lacAS24)". *Journal of Clinical Immunology & Microbiology* 6(1): pp1–5.

36. Yong B (2019) The Comparison of Fourth Order Runge-Kutta and Homotopy Analysis Method for Solving Three Basic Epidemic Models". *Journal of Physics: Conference Series 1317:012020*.

37. Ling MH (2016) "COPADS IV: Fixed Time-Step ODE Solvers for a System of Equations Implemented as a Set of Python Functions". *Advances in Computer Science: an International Journal* 5(3): pp5–11.

38. Okuda S, Yamada T, et al; (2008) "KEGG Atlas mapping for global analysis of metabolic pathways. Nucleic Acids Research 36(Web Server issue)":W423–W426.

39. Ahn-Horst TA, et al; (2022) "An Expanded Whole-Cell Model of E. coli Links Cellular Physiology with Mechanisms of Growth Rate Control". *npj Systems Biology and Applications* 8(1):30.

40. Chagas M da S, et al; (2023) "Boolean Model of the Gene Regulatory Network of Pseudomonas aeruginosa CCBH4851". *Frontiers in Microbiology* 14:1274740.

41. Hao T, Song Z, et al; (2024) "Reconstruction of Metabolic-Protein Interaction Integrated Network of Eriocheir sinensis and Analysis of Ecdysone Synthesis". Genes 15(4):410.

42. Bianchi DM, et al; (2022) "Toward the Complete Functional Characterization of a Minimal Bacterial Proteome". *The Journal of Physical Chemistry* B 126(36): pp6820–6834.

43. Sun G, DeFelice MM, et al; (2024) "Cross-Evaluation of E. coli's Operon Structures via a Whole-Cell Model Suggests Alternative Cellular Benefits for Low- Versus High-Expressing Operons". Cell Systems 15(3):227-245.e7.

44. Choi H, Covert MW (2023) "Whole-cell modeling of E. coli confirms that in vitro tRNA aminoacylation measurements are insufficient to support cell growth and predicts a positive feedback mechanism regulating arginine biosynthesis". *Nucleic Acids Research* 51(12):5911–5930.