Scholastic Medical Sciences

Research Article

Volume 3 Issue 1 January 2026

Ab Initio Whole Cell Kinetic Model of *Thermus aquaticus* Y51MC23 (taqTT26)

Tristan Zhi Xian Tay[1,2], Magaa Lakshmi Dhinakaran[1,2], Sohnnakshee Murugesu[1,2], Aguilar Normi Luisa Cinco[1,2], Pandiyan Srinithiksha[1,2], Maurice Han Tong Ling[2,3,4]

[1]*School of Health & Life Sciences, Teesside University, UK*

[2]*Management Development Institute of Singapore, Singapore*

[3]*Newcastle Australia Institute of Higher Education, University of Newcastle, Australia*

[4]*HOHY PTE LTD, Singapore*

Corresponding Author: Maurice Han Tong Ling, Newcastle Australia Institute of Higher Education, University of Newcastle, Australia.

Received: December 18, 2025; **Published:** December 24, 2025

Abstract

Thermus aquaticus is a thermophilic bacterium best known for its thermostable DNA polymerase used in polymerase chain reaction (PCR). This intrinsic thermostability of its proteins has also driven their use and consideration in other high-temperature engineering applications in molecular biotechnologies. Metabolic modelling using genome-scale or constraint-based models (GSMs) and kinetic models (KMs) can guide potential routes for investigations by providing predictive, *in silico* analyses. While GSMs are more often used, KMs enable a more comprehensive, dynamic visualization of the metabolic networks of the modelled organisms, and more naturally incorporate *in silico* gene modifications. Although a published GSM of *T. thermophilus* exists, there has been no GSMs or KMs of *T. aquaticus* to-date. Hence, a KM of *T. aquaticus* Y51MC23 was constructed in this study using *ab initio* approach by identifying enzymes from its published genome and identifying the corresponding reaction from KEGG (Kyoto Encyclopedia of Genes and Genomes). The resulting kinetic model, taqTT26; comprising of 1059 metabolites, 431 enzymes with corresponding transcriptions and translations, and 1522 enzymatic reactions; and can be a base template for future refinement.

Keywords: Whole-cell model, Kinetic model, Differential equations, AdvanceSyn Toolkit

Introduction

Thermus aquaticus is a thermophilic, yellow-pigmented, non-sporulating, non-motile, Gram-negative bacillus; first isolated from a thermal spring in Yellowstone national park USA in 1969; and often forming long filaments and aggregate into “rotund bodies” resembling larger spheroplasts [1, 2]; with an optimum temperature at 70 °C for aerobic respiration. Although characterized as obligate aerobes, many strains of *T. aquaticus* and most *Thermus* species possess the ability to undergo anaerobic respiration via nitrate reductase in the presence of nitrate, using it as the terminal electron acceptor [3, 4]; except *T. aquaticus* Y51MC23, in which the genes encoding nitrate reductase and its other essential enzymes are absent. Instead, Y51MC23 can undergo fermentation alongside aerobic respiration. Hence, its growth leans closer towards that of facultative anaerobes than its type strain YT-1, being able to grow well in both aerobic and anaerobic conditions [3]. The most well-known feature of *T. aquaticus* is the thermostable DNA polymerase, which used in polymerase chain reaction (PCR) [5]. Other enzymes and proteins from *T. aquaticus* have also been utilized in molecular biology techniques such as Taq ligase in DNA assembly technology [6], and Taq single-stranded DNA-binding (TaqSSB) protein in multiplex PCR to reduce primer-dimer formation [7]. Due to its thermostability, *T. aquaticus* has been considered for engineering for high-temperature applications in the industry [8].

In metabolic engineering, modelling has long served as a compass for identifying promising intervention strategies [9, 10]. The field typically relies on two main modelling paradigms [11, 12]: genome-scale or constraint-based models (GSMs) and kinetic models (KMs). Although GSMs are widely used, they are constrained by their emphasis on reaction rates rather than yields. KMs, by contrast, not only simulate rates but also predict metabolite yields [13], offering a richer picture of pathway behaviour. They also accommodate *in silico* gene knock-ins more naturally than GSMs [14]. Because of these advantages, KMs are increasingly recognised as valuable tools for evaluating alternative engineering options before experimentation. This has prompted renewed interest in building and expanding kinetic model resources [15, 16].

Although a GSM of *Thermus thermopilus* has been published [17, 18], there has been no GSMs nor KMs of *T. aquaticus* to-date. Hence, this study aims to construct a KM of *T. aquaticus* Y51MC23 using *ab initio* approach by identifying enzymes from its published genome [3], and identifying the corresponding reaction from KEGG [19]. The resulting model is named as taqTT26 using the nomenclature proposed by Cho and Ling [20], which consists of 1059 metabolites, 431 enzymes with corresponding transcriptions and translations, and 1522 enzymatic reactions.

Materials and Methods

Identification of Reactome: The genome of *Thermus aquaticus* Y51MC23 (NCBI RefSeq assembly GCF_001399775.1; NCBI GenBank Accession NZ_CP010822.1) was used as source to identify enzymatic genes using the process previously described [14, 21, 22]. Briefly, each enzymatic gene was identified as a presence of complete Enzyme Commission (EC) number in the GenBank record and mapped into reaction IDs via KEGG (Kyoto Encyclopedia of Genes and Genomes) Ligand Database for Enzyme Nomenclature [19]. For example, EC 1.1.1.23 (https://www.genome.jp/entry/1.1.1.23) catalyses reactions R01158, R01163, and R03012; where the substrates and products of each rection can be identified.

Model Development: The model was developed using methodology described in Sim et al. [23]. Based on BioNumbers data, an *Escherichia coli* cell contains roughly 3000 RNA polymerase molecules (BioNumbers 106199) [24], and only about one quarter of these are actively transcribing at any given time (BioNumbers 111676) [25]. With an elongation rate of 22 ribonucleotides per second (BioNumbers 104109) [26] and an average nucleotide mass of 339.5 Da, the cell generates close to 5600 kDa of RNA per second, equivalent to 9.3e-18 grams per second. Considering a cell volume of 7e-16 litres [27] and 4225 protein-coding genes (BioNumbers 105443) [28], the synthesis rate is approximately 2.92 micromolar per gene per second. Given an average mRNA lifespan of 107.56 seconds (BioNumbers 107666) [29] (0.93% decay per second), the mRNA dynamics can be represented as d[mRNA]/dt = 0.00292 – 0.0093[mRNA]. For translation, median protein output is roughly 1000 peptides per transcript per hour (BioNumbers 106382) [30], or 0.278 peptides per second, while protein degradation is about 1% per hour (2.78e-6 per second) (BioNumbers 109924) [31]. Thus, peptide concentration is d[peptide]/dt = 0.278[mRNA] – 0.00000278[peptide]. All metabolites were modelled using ODEs [21, 32] with kcat = 13.7 per second and Km = 1 mM, consistent with Bar-Even et al. [33], formatted per AdvanceSyn Model Specification [34].

Model Simulation. The constructed model was tested for simulatability using AdvanceSyn Toolkit [34]. 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: (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) C00037 (Glycine), (XIII) C00041 (L-Alanine), (XIV) C00045 (Amino acid), (XV) C00047 (L-Lysine), (XVI) C00049 (L-Aspartate), (XVII) C00064 (L-Glutamine), (XVIII) C00065 (L-Serine), (XIX) C00073 (L-Methionine), (XX) C00078 (L-Tryptophan), (XXI) C00079 (L-Phenylalanine), (XXII) C00082 (L-Tyrosine), (XXIII) C00097 (L-Cysteine), (XXIV) C00099 (beta-Alanine), (XXV) C00123 (L-Leucine), (XXVI) C00133 (D-Alanine), (XXVII) C00135 (L-Histidine), (XXVIII) C00148 (L-Proline), (XXIX) C00151 (L-Amino acid), (XXX) C00152 (L-Asparagine), (XXXI) C00155 (L-Homocysteine), (XXXII) C00162 (Fatty acid), (XXXIII) C00178 (Thymine), (XXXIV) C00183 (L-Valine), (XXXV) C00188 (L-Threonine), (XXXVI) C00221 (beta-D-Glucose), (XXXVII) C00380 (Cytosine), (XXXVIII) C00407 (L-Isoleucine), (XXXIX) C00638 (Long-chain fatty acid), (XL) C01000 (Ferrocytochrome c2), (XLI) C05167 (alpha-Amino acid). The model was simulated using the fourth-order Runge-Kutta method [35, 36] 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

The annotated genome of *T. aquaticus* Y51MC23 consists of 2560 genes as of the latest annotation dated 16 December 2024, including 2442 protein coding sequences. 431 unique EC numbers consisting of 1522 enzymatic reactions involving 1059 metabolites were identified and developed into a model based on AdvanceSyn Model Specification [34]. The resulting model taqTT26 is larger than the GSM for *T. thermophilus*, iTT548 [18], which consists of 796 reactions and 635 metabolites. In addition, 862 ODEs acting as placeholder for enzyme transcriptions and translations were added.

Based on the criteria of successful simulation as proposed by previous studies as evidence of a syntactically correct model [14, 22, 37–41], we simulated taqTT26 model using AdvanceSyn Toolkit [34] and verified that it compiles and runs successfully, as illustrated by the simulation profiles in Figure 1. Although the concentration of phosphoric acid shows signs of oscillation which suggests that the model has feedback mechanisms, this behaviour should be viewed as a modelling artifact caused by enforcing median kinetic constants from a large enzyme dataset [42]. These uniform parameters flatten biological variation and shift relative flux magnitudes. In essence, what we provide here is a working whole-cell kinetic framework for *T. aquaticus* Y51MC23, designed intentionally as a base template for future refinement, such as adding organism-specific enzyme kinetics [43–45], incorporation of additional regulatory layers or as a system to examine cellular resource allocations [46–49].

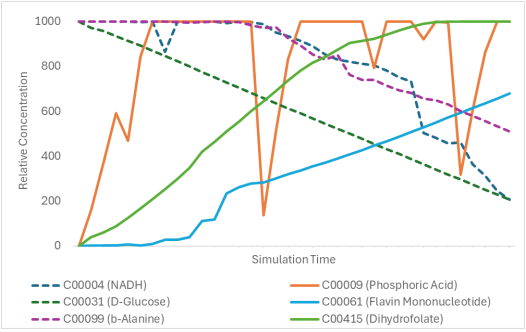

Figure 1: Selection of Simulation Results.

Conclusion

In this study, we present an *ab initio* whole cell kinetic model of *T. aquaticus* Y51MC23. The resulting kinetic model, taqTT26; comprising of 1059 metabolites, 431 enzymes with corresponding transcriptions and translations, and 1522 enzymatic reactions.

Supplementary Materials

Reaction descriptions and model can be download from https://bit.ly/taqTT26.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors wish to thank the institute, Management Development Institute of Singapore, for its support towards this work. The cost of publication fees was borne by the authors.

References

1. Brock TD, Freeze H et al: (1969) "Thermus aquaticus gen. n. and sp. n., a Nonsporulating Extreme Thermophile". *Journal of Bacteriology* 98(1): pp289–297.
2. Brock TD, Edwards MR et al: (1970) "Fine Structure of Thermus aquaticus, an Extreme Thermophile". *Journal of Bacteriology* 104(1): pp509–517.
3. Brumm PJ, Monsma S, et al; (2015) "Complete Genome Sequence of Thermus aquaticus Y51MC23. *PLOS ONE* 10(10): e0138674.
4. Munster MJ, Munster AP et al; (1986) "Isolation and Preliminary Taxonomic Studies of Thermus Strains Isolated from Yellowstone National Park, USA". *Microbiology* 132(6): pp1677–1683.
5. Chien A, Edgar DB, Trela JM et al: (1976) "Deoxyribonucleic Acid Polymerase from the Extreme Thermophile Thermus aquaticus". *Journal of Bacteriology* 127(3): pp1550–1557.
6. Lei C, Li S-Y, Liu J-K et al; (2017) "The CCTL (Cpf1-Assisted Cutting and Taq DNA Ligase-Assisted Ligation) Method for Efficient Editing of Large DNA Constructs In Vitro". *Nucleic Acids Research*: gkx018.
7. Olszewski M, Rębała K, Szczerkowska Z, Kur J et al: (2005) "Application of SSB-Like Protein from Thermus aquaticus in Multiplex PCR of Human Y-STR Markers Identification". *Molecular and Cellular Probes* 19(3): pp203–205.
8. Loder AJ, Zeldes BM, Conway JM, et al; (2017) "Extreme Thermophiles as Metabolic Engineering Platforms: Strategies and Current Perspective". *Industrial Biotechnology, eds Wittmann C, Liao JC (Wiley),* 1st Ed., pp 505–580.
9. Khanijou JK, Kulyk H, et al; (2022) "Metabolomics and Modelling Approaches for Systems Metabolic Engineering". *Metabolic Engineering Communications* 15: ppe00209.
10. Gudmundsson S, Nogales J et al: (2021) "Recent Advances in Model-Assisted Metabolic Engineering". *Current Opinion in Systems Biology* 28: pp100392.
11. 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): pp6.
12. Lee YQ, Choi Y-M, et al; (2025) "Genome-scale metabolic model-guided systematic framework for designing customized live biotherapeutic products". *NPJ systems biology and applications* 11(1): pp73.
13. Prabhu S, Kosir N, Kothare MV, Rangarajan S. et al;(2025) "Derivative-Free Domain-Informed Data-Driven Discovery of Sparse Kinetic Models". *Industrial & Engineering Chemistry Research* 64(5): pp2601–2615.
14. Yeo KY, Arivazhagan M, et al; (2025) "Ab Initio Whole Cell Kinetic Model of Yarrowia lipolytica CLIB122 (yliYKY24)". *Medicon Medical Sciences* 8(4): pp01–06.
15. Foster CJ, Wang L, et al; (2021) "Building Kinetic Models for Metabolic Engineering". *Current Opinion in Biotechnology* 67: pp35–41.
16. Lázaro J, Wongprommoon A, Júlvez J, et al; (2025) "Enhancing genome-scale metabolic models with kinetic data: resolving growth and citramalate production trade-offs in Escherichia coli". *Bioinformatics Advances* 5(1): ppvbaf166.
17. Dahal S, Poudel S, Thompson RA et al; (2016) "Genome-Scale Modeling of Thermophilic Microorganisms. Network Biology, Advances in Biochemical Engineering/Biotechnology". *ed Nookaew I (Springer International Publishing, Cham)*, Vol. 160, pp 103–119.
18. Lee N-R, Lakshmanan M, et al; (2014) "Genome-Scale Metabolic Network Reconstruction and In Silico Flux Analysis of the Thermophilic Bacterium Thermus thermophilus HB27". *Microbial Cell Factories* 13: pp61.
19. Okuda S, Yamada T, et al; (2008) "KEGG Atlas mapping for global analysis of metabolic pathways". *Nucleic Acids Research 36(Web Server issue)*: ppW423–W426.
20. Cho JL, Ling MH et al:(2021) "Adaptation of Whole Cell Kinetic Model Template, UniKin1, to Escherichia coli Whole Cell Kinetic Model, ecoJC20". *EC Microbiology* 17(2): pp254–260.
21. Kwan ZJ, Teo W, Lum AK, Ng SM, Ling MH et al; (2024) "Ab Initio Whole Cell Kinetic Model of Stutzerimonas balearica DSM 6083 (pbmKZJ23)". Acta Scientific Microbiology 7(2): pp28–31.
22. Maiyappan S, Sim SS, et al; (2025) "Four Ab Initio Whole Cell Kinetic Models of Bacillus subtilis 168 (bsuLL25) 6051-HGW (bshSM25), N33 (bsuN33SS25), FUA2231 (bsuGR25)". *Journal of Clinical Immunology & Microbiology* 6(2): pp1–6.
23. Sim BJH, Tan NTF, Ling MHT et al; (2025) "Multilevel Metabolic Modelling Using Ordinary Differential Equations. Encyclopedia of Bioinformatics and Computational Biology (Second Edition), eds Ranganathan S, Cannataro M, Khan AM "(*Elsevier, Oxford*), pp 491–498.
24. Müller-Hill B (1996) "The lac Operon: A Short History of a Genetic Paradigm" *Berlin, Germany*.
25. Churchward G, Bremer H, Young R et al; (1982) "Transcription in Bacteria at Different DNA Concentrations". *Journal of Bacteriology* 150(2): pp572–581.
26. Gray WJ, Midgley JE et al;(1971) "The Control of Ribonucleic Acid Synthesis in Bacteria. The Synthesis and Stability of Ribonucleic Acid in Rifampicin-Inhibited Cultures of Escherichia coli". *The Biochemical Journal* 122(2): pp161–169.

27. Kubitschek HE et al: (1990) "Cell Volume Increase in Escherichia coli After Shifts to Richer Media". *Journal of Bacteriology* 172(1): pp94–101.

28. Hu P, Janga SC, et al; (2009) "Global Functional Atlas of Escherichia coli Encompassing Previously Uncharacterized Proteins". *PLoS biology* 7(4): ppe96.

29. So L-H, Ghosh A, Zong C, et al; (2011) "General Properties of Transcriptional Time Series in Escherichia coli". *Nature Genetics* 43(6): pp554–560.

30. Schwanhäusser B, Busse D, et al; (2013) "Corrigendum: Global Quantification of Mammalian Gene Expression Control". *Nature* 495(7439): pp126–127.

31. Maurizi MR (1992) "Proteases and Protein Degradation in Escherichia coli. Experientia" 48(2): pp178–201.

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

33. Bar-Even A, Noor E, et al; (2011) "The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters". *Biochemistry* 50(21): pp4402–4410.

34. Ling MH et al; (2020) "AdvanceSyn Toolkit: An Open Source Suite for Model Development and Analysis in Biological Engineering". *MOJ Proteomics & Bioinformatics* 9(4): pp83–86.

35. Yong B et al; (2019) "The Comparison of Fourth Order Runge-Kutta and Homotopy Analysis Method for Solving Three Basic Epidemic Models". *Journal of Physics: Conference Series* 1317: pp012020.

36. Ling MH et al; (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.

37. Saisudhanbabu T, Yeo KY, et al; (2025) "Ab Initio Whole Cell Kinetic Model of Limosilactobacillus fermentum EFEL6800 (lfeTS24)". *EC Clinical and Medical Case Reports* 8(4): pp01–04.

38. Arivazhagan M, Senthilkumar A, et al; (2025) "Ab Initio Whole Cell Kinetic Model of Bifidobacterium bifidum BGN4 (bbfMA24)". Acta Scientific Nutritional Health 9(1): pp42–45.

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

40. Wong TB, Le MA, et al; (2025) "Ab Initio Whole Cell Kinetic Models of Escherichia coli BL21 (ebeTBSW25) and MG1655 (ecoMAL25)". Scholastic Medical Sciences 3(2): pp01–04.

41. Ambel WB, Tan LP, Toh D, et al; (2025) "UniKin2 – A Universal, Pan-Reactome Kinetic Model". *International Journal of Research in Medical and Clinical Science* 3(2): pp77–80.

42. Bar-Even A, Milo R, Noor E, et al; (2015) "The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness". *Biochemistry* 54(32): pp4969–4977.

43. Ahn-Horst TA, Mille LS, Sun G, 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): pp30.

44. Chagas M da S, et al; (2023) "Boolean Model of the Gene Regulatory Network of Pseudomonas aeruginosa CCBH4851". *Frontiers in Microbiology* 14: pp1274740.

45. 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): pp410.

46. Thornburg ZR, Bianchi DM, et al; (2022) "Fundamental Behaviors Emerge From Simulations of a Living Minimal Cell". *Cell* 185(2): pp345-360.e28.

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

48. 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): pp227-245.e7.

49. Choi H, Covert MW et al; (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): pp5911–5930.