UniKin1: A Universal, Non-Species-Specific Whole Cell Kinetic Model

Madhurya V Murthy[1,2], Dakshahini Balan[1,2], Nur Jannah Kamarudin[1,2], Victor CC Wang[1,2], Xue Ting Tan[1,2], Avettra Ramesh[1,2], Shermaine SM Chew[1,2], Nikita V Yablochkin[1,2], Karthiga Mathivanan[1,2] and Maurice HT Ling[1,2,3]*

[1]*Department of Applied Sciences, Northumbria University, United Kingdom*
[2]*School of Life Sciences, Management Development Institute of Singapore, Singapore*
[3]*HOHY PTE LTD, Singapore*

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

Abstract

Mathematical models of metabolism can be a useful tool for metabolic engineering. Genome-scale models (GSMs) and kinetic models (KMs) are the two main types of models. GSMs provide steady-state fluxes while KMs provide time-course profile of metabolites, which has more advantage in identifying metabolic bottlenecks. However, KMs require greater degree of accuracy for parameters than GSMs resulting in fewer large-scale KMs than GSMs. Recently, large-scale KMs have been developed but are not based on standard enzymatic rate equations resulting in difficulty in interpreting results in terms of enzyme kinetics. Here, we construct a universal, non-species-specific KM of core metabolism, based on Michaelis-Menten Equation, from glucose to the 20 amino acids and 5 nucleotides based on reactions listed in Kyoto Encyclopaedia of Genes and Genomes (KEGG). Non-species specificity is achieved by using the same Michaelis-Menten constant (Km), turnover number (Vmax), and concentration for each metabolite and enzyme for each equation. This forms a base model for developing species-specific whole cell KMs. The resulting model consists of 566 reactions, 306 metabolites, and 310 enzymes, involving in 1284 metabolite productions, and 1249 metabolite usages. Sensitivity analysis shows that 85% of the metabolite concentration changes with the change of one enzyme kinetic parameter. This forms a base model for developing species-specific whole cell KMs.

Keywords: Kyoto Encyclopaedia of Genes and Genomes (KEGG); Turnover Number; Kinetic Models

Introduction

The term "fourth industrial revolution (FIR)" is coined by Klaus Schwab [1] as the fusion between information and communication technologies and human capacities, with the potential to influence the world by increasing efficiency in different sectors by implementing advanced technologies; such as, artificial intelligence, big data analytics, automation, and robotics. New innovations in metabolic engineering and synthetic biology has shaped the biotechnological platform in the new generation; thus, asserting its place in FIR [2].

Metabolic engineering aims to optimize metabolic process by genetic engineering and pathway optimization for optimal production of one or more desired products in the most sustainable order [3]. Identification of pathway bottlenecks by tracking individual substrates can be aided by mathematical modelling and simulations [4,5]. Mathematical models of a metabolic pathway play an important role in the simulation of cellular events *in silico* experiments under different genetic and environmental conditions. There are mainly two types of models – kinetic models and constraint-based models [6].

Constraint based models of metabolism, commonly known as genome-scale models (GSMs), are based on reaction stoichiometries and reversibilities, offering every possible solutions by imposing constraints [5,7] while providing steady-state production rates of metabolites [8]. Yang., *et al.* [9] used a GSM of *Escherichia coli* to study the production of aromatic polyester after cloning isocaprenoyl-CoA:2-hydroxyisocaproate CoA-transferase from *Clostridium difficile* and an evolved polyhydroxyalkanoate synthase. Kinetic models (KMs) generally require rate equations in the form of ordinary differential equations (ODE) that defines the rate of change of concentrations of the substrates involved [10], which offers a transient dynamic approach as it provides specific solutions in time for steady-state fluxes from the initial concentration of the substrates [5]. Hence, KMs are considered unconstrained [6] and can provide time-course profile of metabolites rather than steady-state, which is more advantageous than GSMs [5]. However, KMs require greater degree of accuracy for parameters than GSMs [5]; hence, more demanding and as a result, there are fewer large-scale KMs than GSMs. Khodayari., *et al.* [11] presented a KM of *E. coli* core metabolism, which had been used to explore succinate overproduction [12]. The original model [11] has also been expanded by the same group [13]. However, the model [13] is not based on standard enzyme rate equations; such as Michaelis-Menten equation [14] or Haldane Equation [15]. Hence, it makes interpretation of results difficult as the correspondence between the model parameters and enzyme kinetics is not obvious.

In this study, we construct a universal, non-species-specific KM of core metabolism from glucose to the 20 amino acids and 5 nucleotides based on reactions listed in Kyoto Encyclopaedia of Genes and Genomes (KEGG) database [16]. Non-species specificity is achieved by using the same Michaelis-Menten constant (Km), turnover number (Vmax), and concentration for each metabolite and enzyme for each equation. This forms a base model for developing species-specific whole cell KMs.

Methods

Model construction

Reference metabolic reactions linking glucose to the synthesis of 20 amino acids and 5 nucleotides were identified from 12 reaction maps (map00220, map00230, map00240, map00250, map00260, map00270, map00300, map00330, map00360, map00350, map00290, and map00340) in Kyoto Encyclopaedia of Genes and Genomes (KEGG) database [16]. The concentration of each metabolite, which are being modelled mathematically as an ODE [17], is in the form of $\frac{d[metabolite]}{dt} = \sum_{i=1}^{N} production_i - \sum_{i=1}^{N} usage_i$, where production represents a formation or synthesis of the metabolite, and usage represents a usage of the metabolite to form another metabolite. As production and usage terms are in pairs, they can be modelled as a Michaelis-Menten expression, $\frac{k_{cat}[enzyme](\prod_{i=1}^{N}[substrate_i])}{K_m + (\prod_{i=1}^{N}[substrate_i])}$, where k_{cat} is the turnover number (per second) of the enzyme, K_m is the Michaelis-Menten constant, [enzyme] and [substrate] are the concentrations (in molar) of the enzyme and substrates respectively and N represents the number of molecules. The concentrations of metabolites and enzymes were set at 1 micromolar and 100 millimolar, respectively; while the k_{cat} and K_m were set at 10 per second and 1 millimolar, respectively. The model is written as an AdvanceSyn Model Specification (https://bit.ly/ADSToolkit).

Baseline simulation

Model simulation was performed using AdvanceSyn Toolkit (https://bit.ly/ADSToolkit). The model was simulated using the fourth-order Runge-Kutta method [18,19] from time zero to 21600 seconds with timestep of one second, and the concentrations of metabolites were bounded between 0 millimolar and 1 millimolar. The simulation results were sampled every 60 seconds.

Sensitivity analysis

Sensitivity analysis was performed using AdvanceSyn Toolkit (https://bit.ly/ADSToolkit) using One-Factor-at-a-Time method [20,21] was used to evaluate the effects of varying enzyme concentrations, Michaelis-Menten constant, and turnover number (as listed in the Variables section of the model) by comparing the baseline simulation results with that after increasing the factor by 100 times. The model was simulated using the fourth-order Runge-Kutta method [18,19] from time zero to 21600 seconds with timestep of one second, and the concentrations of metabolites were bounded between 0 millimolar and 1 millimolar. A total of 930 variations of the parameters (310 variations each of concentration, K_m, and k_{cat}) were simulated and compared against that of baseline simulation. Mean squared error (MSE) and Pearson's coefficient of determination (R-square) was calculated for each parameter variation using the 2 sets of metabolite concentrations at 21600 seconds. Hence, MSE is directly proportional to the effect of the parameter to the

entire metabolite network.

Results and Discussion

Model consists of 306 metabolites in 566 reactions

The resulting model consists of 566 reactions, 306 metabolites, and 310 enzymes, involving in 1284 metabolite productions, and 1249 metabolite usages. This amount to 1.63% of the 18,720 metabolites (accessed on July 09, 2020) found in KEGG [16]. From the perspective of reactions, the number of substrates is correlated to the number of products (R-square = 0.337, F-statistic = 286, p-value = 2.5E-52). More than 62% of the reactions have 2 substrates and 2 products – 67 (11.8%), 355 (62.7%), 108 (19.1%), and 36 (6.4%) reactions with 1, 2, 3, and 4 production terms, respectively; whereas 49 (8.6%), 351 (62.0%), 131 (23/1%), and 35 (6.2%) reactions with 1, 2, 3, and 4 usage terms, respectively. From the perspective of metabolites, the number of reactions producing and/ or using a metabolite is correlated (Figure 1; R-square = 0.649, F-statistic = 563, p-value = 3.6E-71).

85% of the metabolite concentration changes with one parameter variation

Sensitivity analysis shows near perfect correlation (Figure 2; R-square > 0.996, F > 91131, p-value < 1E-240) between the mean squares errors (MSE) and Pearson's coefficient of determination (R-square). R-square measures the degree of correlation between the metabolic profiles of parameter variation and baseline while MSE measures the differences between the metabolic profiles of parameter variation and baseline. As this study involves comparison between the concentrations of the original reaction and the concentrations in the simulated reactions; difference measures, such as MSE, is a more reliable measure [22] than correlation measures, such as R-square, for model selection [23]. Hence, MSE used for further analysis and shows that the average concentration varia-
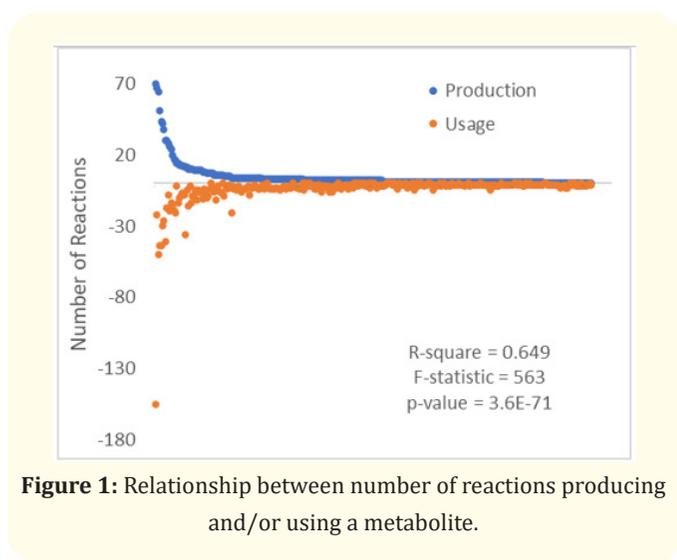

Figure 1: Relationship between number of reactions producing and/or using a metabolite.

Bins (by MSE)	Enzyme Parameters		
	Turnover Number (k_{cat})	Concentration	Michaelis-Menten Constant (K_m)
>1.0E-09	303	292	278
>1.0E-08	298	282	273
>1.0E-07	203	166	145
>1.2E-07	129	74	68
>1.4E-07	55	23	20
>1.6E-07	35	8	2
>1.8E-07	15	0	1
>2.0E-07	4	0	0

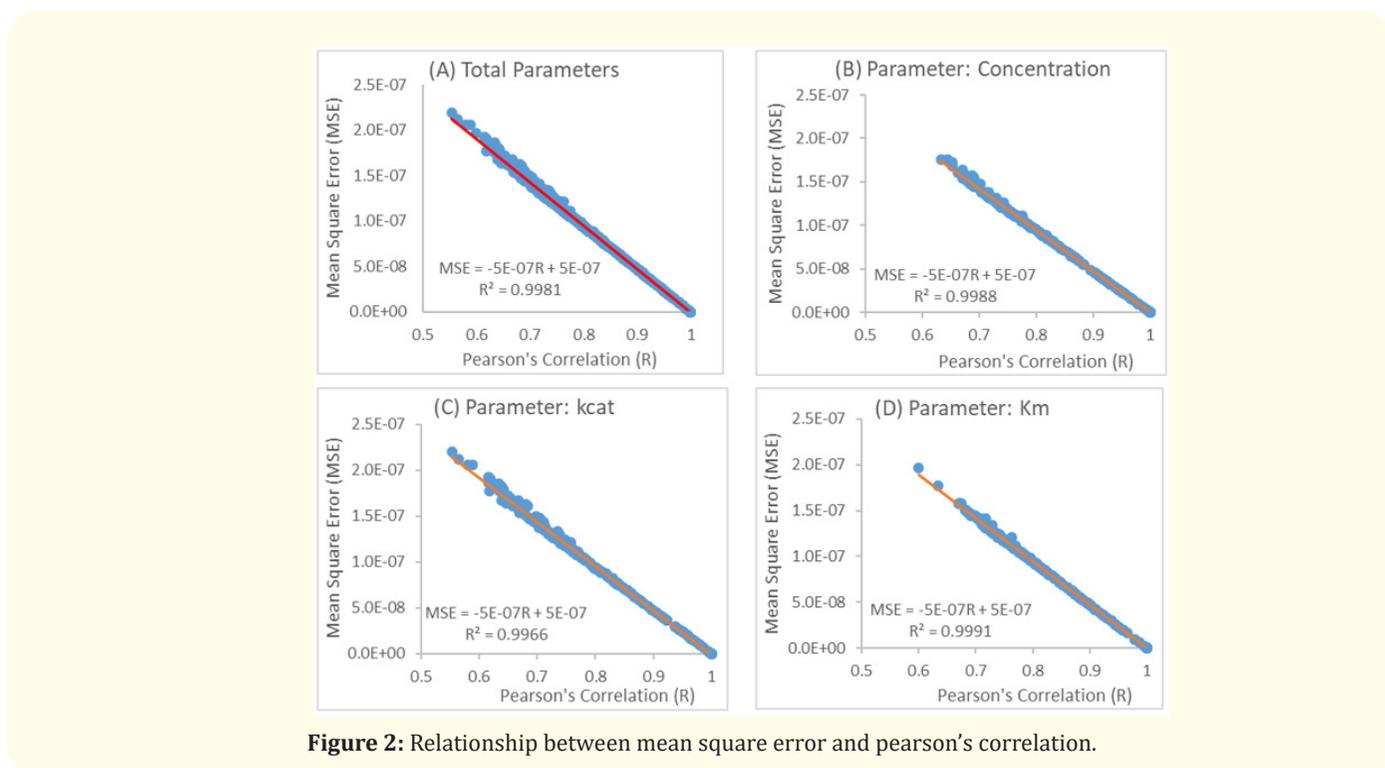**Table 1:** Distributions of MSE by parameters.

Figure 2: Relationship between mean square error and pearson's correlation.

tion in metabolites as a result of one change in enzyme parameter; namely, concentration, k_{cat}, or K_m; is 305 micromolar with the median at 324 micromolar as the unit for MSE is squared-micromolar.

Our results show that MSE can be affected by the three enzyme parameters (Table 1). However, comparing the experimental work needed to affect each of the parameters; changes in concentration requires gene overexpression [24] which is often easier and more routine than changes in k_{cat} and K_m, which often requires screen of mutant libraries or directed evolution [25]. Hence, changes in enzyme concentration is generally considered before changes in k_{cat} and K_m.

Further analysis suggests that 15% of the metabolites do not change in concentration at the end of simulation time regardless of changes in any enzyme parameters while 45% and 40% of the metabolites increases or decreases in concentration at the end of simulation time with the change of one enzyme parameter. From the number of affected metabolites, the median number of enzyme parameter changes that can increase or decrease metabolite concentration at the end of simulation time are 35 and 52.5 respectively. Importantly, 5 metabolite concentrations (KEGG compound IDs C00302, C01250, C05125, C06195, and C09306) can be decreased by only one parameter change while 15 metabolite concentrations (KEGG compound IDs C00019, C00188, C00212, C00460, C00499, C00606, C02350, C02718, C04272, C05147, C05381, C05519, C12248, C15972, and C19706) can be increased by only one parameter change. On the extreme end, the concentration of O-Acetyl-L-serine (KEGG compound ID C00979) can be reduced by 669 parameter changes while the concentration of carbon dioxide (KEGG compound ID C00011) can be increased by 673 parameter changes. The goal of metabolic engineering is to optimize cellular processes for the production of a metabolite of interest [3] and these results illustrate the potential use of KMs as an exploratory tool.

Conclusion

This study presents a universal, non-species-specific kinetic model of core metabolism, based on Michaelis-Menten Equation; from glucose to the 20 amino acids and 5 nucleotides; consisting of 566 reactions, 306 metabolites, and 310 enzymes, involving in 1284 metabolite productions, and 1249 metabolite usages. This forms a base model for developing species-specific whole cell kinetic models.

Data Availability

The data files for this study can be downloaded at https://bit.ly/UniKin1, which is a zip file containing the following files: (i) Reactions.xlsx containing information of the reactions with its associated metabolites and enzymes, (ii) Fluxes.xlsx contains analysis of the productions and usages of each metabolite, (iii) UniKin1.modelspec is the model in AdvanceSyn Model Specification, (iv) UniKin1.py is the executable simulation model, (v) Simulation.xlsx contains results of baseline simulation, and (vi) Sensitivity_Analy-

Bibliography

1. Shwab K. "The Fourth Industrial Revolution: What It Means, How to Respond". *Foreign Affair* 12 (2015): 2015-7.

2. Ramzi AB. "Metabolic Engineering and Synthetic Biology". *Advances in Experimental Medicine and Biology* 1102 (2018): 81-95.

3. García-Granados R., *et al*. "Metabolic Engineering and Synthetic Biology: Synergies, Future, and Challenges". *Frontiers in Bioengineering and Biotechnology* 7 (2019): 36.

4. Keasling JD. "Manufacturing Molecules Through Metabolic Engineering". *Science* 330.6009 (2010): 1355-1358.

5. Kim OD., *et al*. "A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering". *Frontiers in Microbiology* 9 (2018): 1690.

6. Machado D., *et al*. "Exploring the Gap Between Dynamic and Constraint-Based Models of Metabolism". *Metabolic Engineering* 14.2 (2012): 112-119.

7. Reed JL. "Genome-Scale Metabolic Modeling and Its Application to Microbial Communities". In: The Chemistry of Microbiomes. National Academies Press (2017).

8. Fernandez-de-Cossio-Diaz J., *et al*. "Characterizing Steady States of Genome-Scale Metabolic Networks in Continuous Cell Cultures". Nielsen J, editor. *PLOS Computational Biology* 13.11 (2017): e1005835.

9. Yang JE., *et al*. "One-Step Fermentative Production of Aromatic Polyesters from Glucose by Metabolically Engineered Escherichia coli Strains". *Nature Communications* 9.1 (2018): 79.

10. Wang RS. "Ordinary Differential Equation (ODE), Model". In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. Encyclopedia of Systems Biology. New York, NY: Springer New York (2013): 1606-1608.

11. Khodayari A., *et al*. "A Kinetic Model of Escherichia coli Core Metabolism Satisfying Multiple Sets of Mutant Flux Data". *Metabolic Engineering* 25 (2014): 50-62.

12. Khodayari A., *et al*. "Succinate Overproduction: A Case Study of Computational Strain Design Using a Comprehensive Escherichia coli Kinetic Model". *Frontiers in Bioengineering and Biotechnology* 2 (2014): 76.

13. Khodayari A and Maranas CD. "A Genome-Scale Escherichia coli Kinetic Metabolic Model k-ecoli457 Satisfying Flux Data for Multiple Mutant Strains". *Nature Communications* 7 (2016): 13806.

14. Johnson KA and Goody RS. "The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper". *Biochemistry* 50.39 (2011): 8264-8269.

15. Sonnad JR and Goudar CT. "Solution of the Haldane Equation for Substrate Inhibition Enzyme Kinetics Using the Decomposition Method". *Mathematical and Computer Modelling* 40.5-6 (2004): 573-82.

16. Kanehisa M., *et al*. "The KEGG databases at Genome Net". *Nucleic Acids Research* 30.1 (2002): 42-46.

17. Chen WW., *et al*. "Classic and Contemporary Approaches to Modeling Biochemical Reactions". *Genes and Development* 24.17 (2010): 1861-1875.

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

19. Ling MH. "COPADS IV: Fixed Time-Step ODE Solvers for a System of Equations Implemented as a Set of Python Functions". *Advanced Computer Science* 5.3 (2016): 5-11.

20. Abou-Taleb KA and Galal GF. "A comparative study between one-factor-at-a-time and minimum runs resolution-IV methods for enhancing the production of polysaccharide by Stenotrophomonas daejeonensis and Pseudomonas geniculate". *Annals of Agricultural Sciences* 63.2 (2018): 173-180.

21. Razavi S and Gupta HV. "What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models: A Critical Look at Sensitivity Analysis". *Water Resources Research* 51.5 (2015): 3070-3092.

22. Willmott CJ. "Some Comments on the Evaluation of Model Performance". *Bulletin of the American Meteorological Society* 63.11 (1982): 1309-1313.

23. Waldmann P. "On the Use of the Pearson Correlation Coefficient for Model Evaluation in Genome-Wide Prediction". *Frontiers in Genetics* 10 (2019): 899.

24. Prelich G. "Gene Overexpression: Uses, Mechanisms, and Interpretation". *Genetics* 190.3 (2012): 841-854.

25. Ko Y-S., *et al*. "Tools and Strategies of Systems Metabolic Engineering for the Development of Microbial Cell Factories for Chemical Production". *Chemical Society Reviews* 49.14 (2020): 4615-4636.