Project Milestone – Technical Brief Draft

To: President Avery D. Lion of NaturalCatalysts Inc.

From: Sohan Pramanik

RE: Technical Brief of Final Results

Date: April 22, 2020

Introduction

Our client, President Avery Lion of NaturalCatalysts, came to us to create an algorithm to analyze the kinetic enzyme test data provided to us. To answer his call for assistance, we will create a detailed description of our analysis of the dataset provided, including clear and easy-to-understand graphics that summarize the data. We must then provide an error analysis that characterizes the accuracy of our approach to determining performance characteristics of the system. Finally, we will craft a recommendation about what NaturalCatalysts, Inc. can honestly and ethically claim to its customers about the performance of the new enzymes.

Our algorithm calculates the Vmax, Km, and V0i with the possible smallest SSE for each of the given enzymes PGOX50, and A-E generating plots for the raw data and the Michaelis-Menten plot for V0i. We first find the range of seconds to use for the data, and then use Lineweaver Burk method to find all the calculated values accordingly.

The main critical decision we made was to change the best way to find the range of number of seconds to use. We incorporated a loop that will display how many seconds is best by finding and calculating a low SSE value. When the program calculates a low SSE value, this would allow the v0i to be more accurate, and then the Km and Vmax would follow as well. Another critical decision we made was the linearization of all Vo for each enzyme at each [S] utilizing the Lineweaver-Burk equation and calculating the Vmax and Km for each of the data sets using the equation. Using this model, a straight line is formed using the double reciprocal of the velocity and substrate concentration, since the data is linearized. Using the accurately found Km and Vmax values, we can use the Michaelis-Menten equation to find the theoretical velocity of each enzyme. Our last critical decision was we added a time range within the program for each enzyme. Since the data given is very large, we chose to use a time range for a couple of seconds in the beginning parts of the data. This accounts for noise as well since we are not using the entire data set and just a specific part.

Parameter Identification Procedure

The algorithm first loops around the PGOX50 enzyme first to automatically find the best range for number of seconds to use for calculations. It checks for the smallest SSE values for each of the ranges and picks the one with less error. Using this, it calculates the V0i by taking the linear regression of the initial points in the raw data. Then using the V0i set corresponding to the smallest SSE to calculate the theoretical velocity using the formula for Michaelis-Menten using the reference values given for Vmax and Km for PGXO50 enzyme. Then it finds the sum of SSE between the reference initial velocities and the reaction velocities. Next, it linearizes all the Vo for PGXO50 enzyme at each concentration [S] utilizing the Lineweaver-Burk equation and calculating the Vmax and Km for each of the data set using the equation. It repeats the step again to calculate the theoretical velocity using formula of Michaelis-Menten equation for the actual data and gets the sum of SSE values. Now since the PGOX50 parameters found, the remaining are the enzymes A-E which are followed in another loop using the same exact

steps as above to calculate the parameters. The last step is to check if the values of Km and Vmax are between the given ranges. Using all the calculated values it plots the Michaelis-Menten plots, and displays all the information accordingly, for all the enzymes.

Results

Enzyme	V_{max} (µM/s)	K_m (µM)	SSE (µM/s)²
NextGen-A	0.997	155.781	0.001482
NextGen-B	0.908	354.162	0.001519
NextGen-C	1.225	188.260	0.000913
NextGen-D	1.619	294.458	0.001557
NextGen-E	1.661	169.272	0.002147

Table 1. The Vmax, Km, and SSE value calculations from Algorithm for Enzymes A-E.

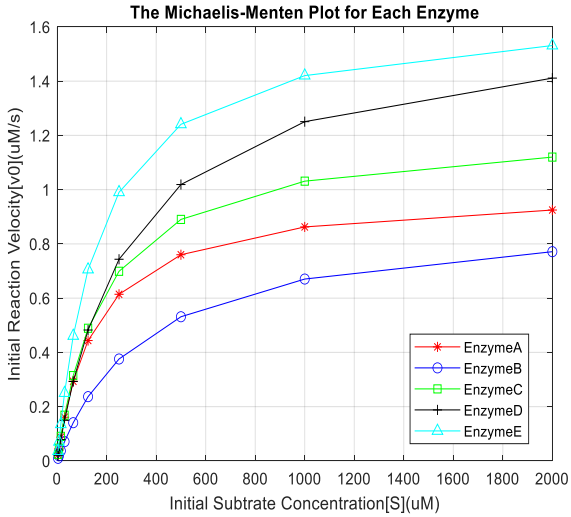

Figure 1. Michaelis-Menten Plot for Each Enzyme

Interpretation

The error in the process for the Vmax, is very low for all the enzymes given. According to our calculations the SSE values are very close to 0, for example the SSE value for Enzyme C is 0.00913. The SSE is the sum of the squares of residuals deviations predicted from actual empirical values of data. A low SSE indicates that the model has a smaller random error component, and that the fit will be more useful for prediction. Our algorithm is very concise and made it as accurate as possible by including a program to calculate and use a low SSE value for the number of seconds of each concentration. This disregards noise because we are using a specific amount of time instead of using the entire data, which shows the quality to be very reasonable. We also used the Lineweaver-Burk model to calculate the Vmax and Km values since it linearizes the data and gives a more accurate representation.

Generally, the enzyme products of NaturalCatalysts are reasonable due to their performance. Since there are many factors that affect enzyme prices, though using the data we collected, and the data we were given, enzymes that have a high reaction rate are more expensive than slower reaction rates. This is using the Michaelis Constant (Km) from the data given and displayed from our algorithm. Enzyme B is priced at $58.4115 because its Km value is 354.162 (µM). Then Enzyme A is priced at $420.7833 because its Km value is 155.781 (µM). Using the calculated general model trend, a lower Michaelis Constant result the price to be high, meaning the enzyme has higher performance and reaction rates. Enzymes A, C, and E are the best performances because it is Michaelis Constant is low, thus affecting the price to be high. Also, none of the prices were extrapolated using the price ranges. The manufacturing consistency is very stable and goes very well with the data given. The products follow a consistent linear trend that price decreases when Km increases which shows the x and y values are correlated.

References

Berg, J., Tymoczko, J., & Stryer, L. (2002). Appendix: Vmax and KM Can Be Determined by Double-Reciprocal Plots. Retrieved 17 April 2020, from https://www.ncbi.nlm.nih.gov/books/NBK22557/

Appendix: Figures and Tables

Metrics	Finding
Linearized Model equation	*y = -2.4042 * x + 7.8954*
SSE_{lin} [(log $)^2]	*0.1397*
SST_{lin} [(log $)^2]	*4.4451*
r^2_{lin}	*0.9686*
General Model equation	*y = 78594899.5504 * x ^ (-2.4042)*

Table 2. Regression metrics for price versus performance data.

Enzyme	Price ($)
NextGen-A	420.7833
NextGen-B	58.4115
NextGen-C	266.8872
NextGen-D	91.0470
NextGen-E	344.6161

Table 3. Price prediction for each enzyme using regression model.

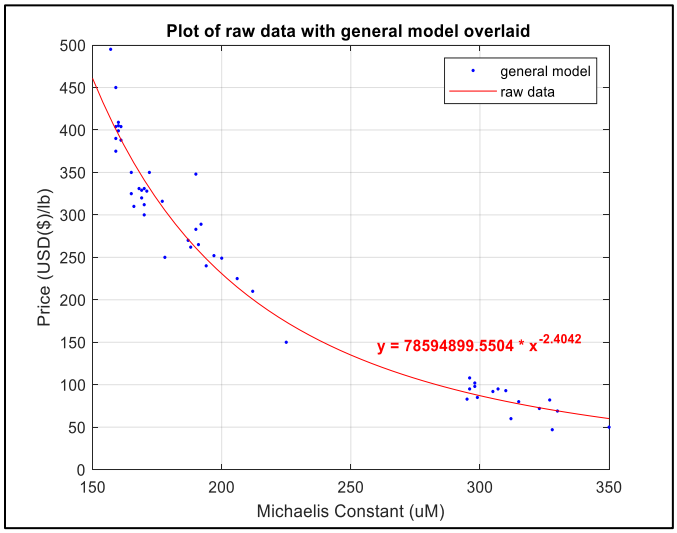

Figure 2. Plot of raw data with general model overlaid

Parameter (µM/min)	PGO-X50 Reference Values	Algorithm Calculations
v_{0_1}	0.028	0.0272
v_{0_2}	0.055	0.0533
v_{0_3}	0.11	0.1088
v_{0_4}	0.19	0.1837
v_{0_5}	0.338	0.3376
v_{0_6}	0.613	0.6132
v_{0_7}	0.917	0.9174
v_{0_8}	1.201	1.1757
v_{0_9}	1.282	1.2820
$v_{0_{10}}$	1.57	1.5704
V_{max}	1.61	1.614
K_m (µM)	214.28	218.771
SSE (µM/min)^2	0.0251	0.026

Table 4. PGOX50 Enzyme algorithm calculations of V0i, Vmax, Km, and SSE values

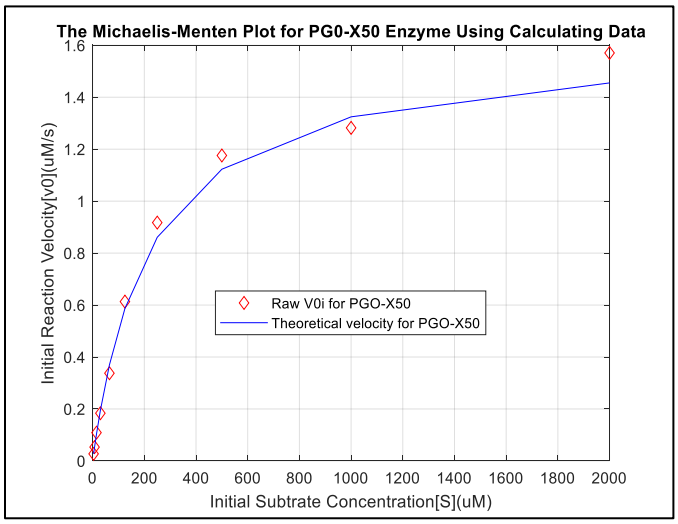

Figure 3. Michaelis Menten Plot of PGOX50 Enzyme