Project Milestone 4 – Technical Brief

To: <Avery D. Lion, President, NaturalCatalysts Inc.>

From: <Team 019-24: Gregory Szymchack, Nathan Thorson, Seena Pourzand, Sergio Monge>

RE: <Enzyme Reaction Technical Brief>

Date: <April 30, 2021>

Introduction

Natural Catalyst set our engineering team the problem of determining which of five given cleaning enzyme data sets is the best for mass production. The main deliverable for this project is a program that will return an array of initial velocities, the Menten constant, and the maximum velocity. The two main criteria for success were the analysis uses detailed data and the company can replicate and or use our data program. The main two constraints for this project were that the formulas provided by the company must be used so their engineers can replicate our methods and that the brief is no longer than two pages not including figures.

Our algorithm finds the slope between 2 adjacent points and repeats that process 55 times and averages it to get the v0 of a respective substrate concentration, doing that whole procedure 10 times in total. Those 10 v0s are utilized to make a Lineweaver Burke plot to find a linear regression model whose parameters we convert into a non-linear model, giving us Vmax and Km.

Our first decision to improve the accuracy of our program was to undo the hardcoding of the concentration values. The primary evidence in support of this was the possibility that we were presented with a dataset that has differing initial concentrations that would make the calculations, General model, and Michaelis-Menten Plot inaccurate. Our second decision was to reduce the number of data points used to determine the initial velocities [v(0)] of the substrate concentrations. The initial velocity of enzyme is the slope of a line that runs tangent to the first portion of the graph before the curve begins. The importance of this value is that it will later be used for the Michaelis-Menten plot (Berg). When the number of values used to find the tangent line is lessened, the initial velocity will change and not be affected by the curve (Berg). Our third decision to improve the accuracy was to use only ten standard concentration sets and not average the duplicates with the normal data sets. Our chief reason for this change is because of data consistency. When the models use a single test of data and the data has consistency, the model can use averaged data to make sure outlines are removed or dealt with (Fogg).

Parameter Identification Procedure

We take in parameters of a time column vector, a row vector for substrate concentrations, a matrix containing all of the enzyme tests for a given enzyme.

Then we enter a for loop that iterates 10 times, each iteration to represent the current substrate concentration. So, we create a column vector from the passed in enzyme data, the column being the current index of the for loop. For example, if it is the first iteration of the loop, the corresponding column vector for that iteration is all the rows of the provided enzyme data, but only the first column to represent the first concentration.

We smooth the current data in a column vector(called smoothed) using MATLABS movemean function.

We then enter a nested for loop that iterates from 1-56, incrementing by one, variable k being the index

Inside this loop, we find the slope between two adjacent points and add it to "avgSlopeSmoothed" vector at index k. Ex. we have the (k+1) value of the smoothed data – the k value of the smoothed data all divided by the (k+1) time value by the k time value. This process is repeated 54 more times.

We the take the average of these 55 values by using the mean function on avgSlopesmoothed to get one v0 value for the concentration. We store this average in originalV0Vec which will eventually hold 10 v0s.

We start the regression process by converting our v0s and substrate concentration to linearized data by taking their inverses. Ex. Xdata = 1 divided by the substrate concentrations & Ydata = 1 divided by v0s. Then to find the linear regression line parameters for this data, we use polyfit to find M and B. We then convert these linear model M & B values to the parameters values that we will use in our Michaelis Menten Model. Vmax equals 1 divided by linear B value. Km equals linear M values multiplied by Vmax.

The function then returns the v0s, Vmax and Km as outputs.

Results

Figures 2-6 each show the v0s (µM/s) for an enzyme (figure 2 for enzyme A, 3 for B, etc) due to the amount of substrate (µM/) they break down as well as the Michaelis-Menten model for that enzyme. Figure 1 shows Enzyme A's product concentration (uM) over time (seconds) for all substrate concentrations as well as their tangent lines for when time is 0 seconds. Finally, table 1 shows each enzyme's Km, vMax, and SSE.

Interpretation

In regards to our algorithms accuracy, all of our parameters fall within the accepted value range with the exception of one(NextGenE Km). However, error could still persist as our raw data was noisy and smoothing can only do so much when trying to determine v0(Figure 1). However, our algorithm has an average SSE of 0.00768 which is extremely close to zero and our Michaelis Menten Plots (Figures 2- 6) all show our determined v0s closely matching to the model line, showing that despite the noise, we were able to make accurately identify parameters. For the most part, NaturalCatalysts enzymes all have a relatively consistent shape and results. Looking at the performance of NaturalCatalysts Enzymes (Table1), we can categorize Enzymes A, C, & E as high effective compared to the rest of the data, they have relatively low KMs. Conversely, we can say Enzymes B and D are low effective as they have high Kms. Enzymes C, D & E are fast action as by their high vMaxes, while A and B are slow action. Enzyme E seems to be the best of the 5 with a relatively low Km and large Vmax, indicating fast action and high efficacy, making it the best for mass production

References

Berg, J. M. (1970, January 01). The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK22430/

Fogg, A. (2020, October 05). How to Ensure Data Consistency and Quality with Web Data Integration. Retrieved from https://www.import.io/post/how-to-ensure-data-consistency-and-quality/

Appendix: Figures and Tables

Table 1: NextGen Enzyme Parameters & SSE

Enzyme	M4 Algorithm		
	Enzyme Parameters		SSE $(\mu M/s)^2$
	V_{max} (µM/s)	K_m (µM)	
NextGen-A	0.9501	173.6767	0.0015
NextGen-B	0.8442	357.7180	0.0020
NextGen-C	1.2861	218.0345	0.0134
NextGen-D	1.4339	292.8052	0.0148
NextGen-E	1.5470	190.2243	0.0067

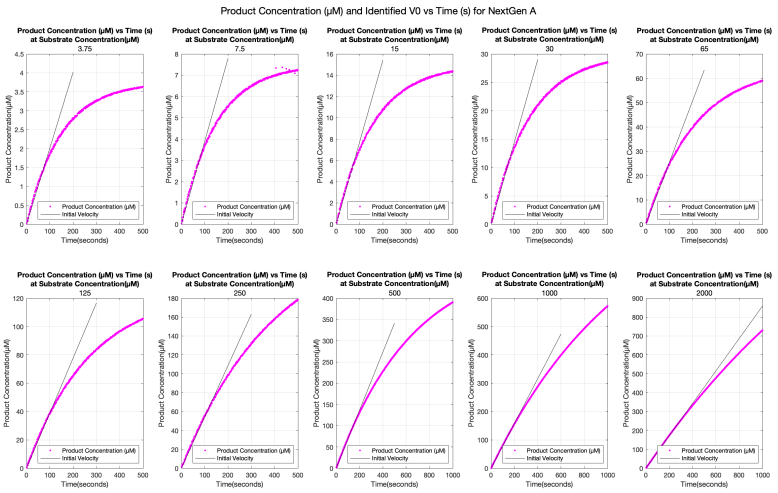

Figure 1: Enzyme A V0s & Product Concentrations

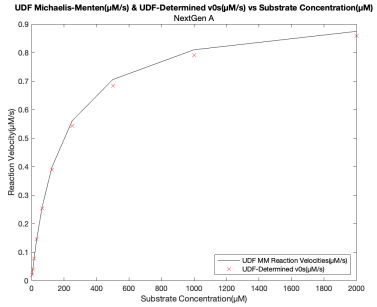

Figure 2: Enzyme A Michaelis-Menten Plot

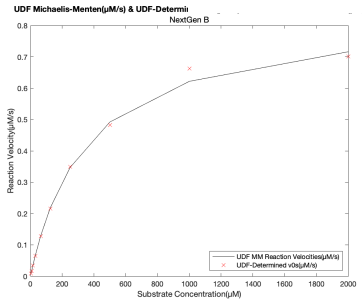

Figure 4: Enzyme B Michaelis-Menten Plot

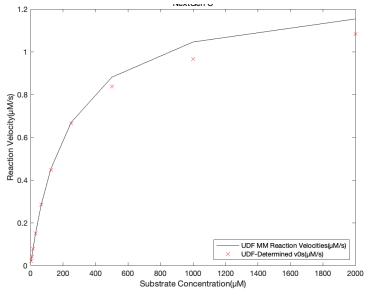

Figure 5: Enzyme C Michaelis-Menten Plot

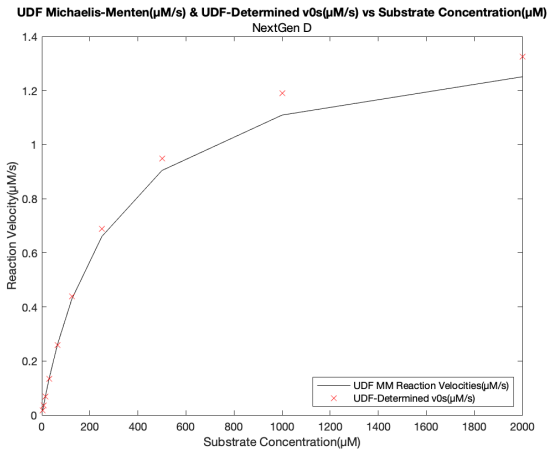

Figure 5: Enzyme D Michaelis-Menten Plot

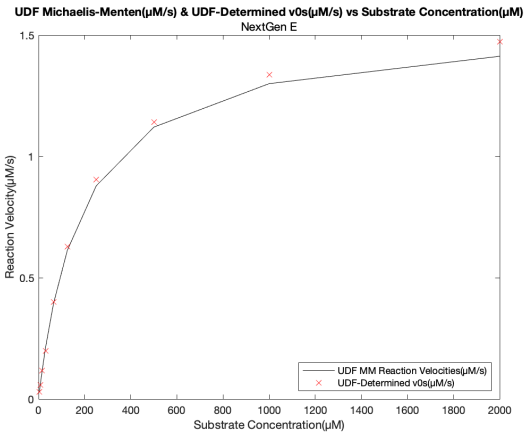

Figure 6: Enzyme E Michaelis-Menten Plot