Project Milestone 4 – Technical Brief

To: President Avery D. Lion of NaturalCatalysts Inc.

From: Team 14: Jun Shern Lim, Kevin Crowley, Tyler Maslak, and Shravan Ranganathan

RE: NaturalCatalysts Enzyme Analysis Technical Brief

Date: 30th April 2021

Introduction

NaturalCatalyst set us the problem of constructing an algorithm to analyze kinetic energy data for their new enzymes, provide an error analysis that determines the accuracy of our approach, and finally develop a recommendation about what they can promote about their new enzymes. The deliverable for this project was a MATLAB function that can analyze large enzyme datasets and return the v0, Vmax, and Km parameters for each enzyme as well as each enzyme's Michaelis-Menten plot. The criteria for success were that our program should be versatile and that our calculated parameters should be accurate. The constraints were that we had to code out function in MATLAB and that we had to test our function using the enzyme dataset that NaturalCatalyst provided for us.

Our algorithm first accepts either only original enzyme test data or original and duplicate test data as input and parses the data into more usable chunks. It then calculates the initial slope (v0) at each substrate concentration, and uses these v0 values to determine Vmax and Km.

Our first decision was to linearize the v0 and substrate concentration data to find Vmax and Km using the Hanes-Woolf plot. According to the journal article by Marasović et al. (2017), the Hanes-Woolf plot produces Vmax and Km parameters that are more accurate than both the Lineweaver-Burk plot and Eadie-Hofstee plot. To linearize data according to the Hanes-Woolf plot, we plotted (substrate concentration [S]/reaction velocity) against substrate concentration[S]. The Vmax was 1/slope, and the Km was the y-intercept * Vmax. Using the Hanes-Woolf plot allowed us produce Km and Vmax values that were very similar to the reference values for the PGO-X50 enzyme.

Our second decision was to use the "smoothdata(data, 'rloess')" function in MATLAB to filter out noise from the enzyme dataset. Noisy data severely affects the accuracy of any analysis; hence it is important to filter it out. After extensive experimentation, we found that using the above smoothening function helped us derive the most accurate v0 values from the product concentration data.

Our third decision was to calculate v0 by determining the average (mode) slope of the first 39 points of the smoothened data. Taking the average slope of multiple points instead of just one reduces the margin of error of our v0 values. Additionally, we also found that by using the mode average instead of the mean, we were able to output Km and Vmax parameters that were more accurate when compared to the reference values of the PGO-X50 enzyme. This is likely because the mode takes the most frequently occurring slope as the average, removing the possibility of extreme values affecting v0.

Parameter Identification Procedure

1. Accept original enzyme data, duplicate data, and substrate concentration as input.
2. If there is no duplicate data present, the user is expected to enter "0" in the duplicate data input section.

3. Use a nested for loop to parse out product concentration data at a specific substrate concentration and smoothen it.
4. Calculate the mode average slope of the first 39 points of the smoothened data, excluding the origin.
5. The average slope calculated in step 4 will be the v0 value at a specific substrate concentration.
6. The loop will iterate through steps 3-5 until an array of 10 v0 values is produced for each substrate concentration.
7. If duplicate data is detected (not equals 0), a selection structure will repeat steps 3-6 with the duplicate data.
8. The algorithm then averages the v0 values of the original data and duplicate data to get a final array of 10 v0 average values.
9. If no duplicate data is detected, the algorithm sets the final v0 values as the v0 values calculated from the original data.
10. Set substrate concentration [S] as the x variable, and (substrate concentration [S]/reaction velocity) as the y variable of the Hanes-Woolf linear plot.
11. Determine the slope and y-intercept of the Hanes-Woolf linear plot.
12. The Vmax will then be 1/slope and Km will be Vmax * y-intercept.

Results

- Table 1 shows us the Vmax and Km parameters of each of the 5 enzymes.
- Figures 1 to 5 show us the Michaelis-Menten Plot of enzymes A to E respectively.
- Table 2 shows us the SSE and Mean Absolute Percent Error between the calculated v0s and Michaelis-Menten model
- Table 3 shows us the algorithm performance compared to reference values of the PGO-X50 enzyme

Interpretation

Q1: We can calculate the SSE and the mean absolute percent error (Stephanie, 2021) between the Michaelis-Menten model (created using our Vmax and Km) and our v0 to characterize the error. Referring to Table 2 in the appendix, we found that the SSEs between our model and data were relatively small, and more importantly the mean absolute percent errors were less than 0.1% for each enzyme. Additionally, Table 3 shows us that our parameters for the PGO-X50 enzyme were fairly similar to the reference values. Therefore, we can conclude that the experiment themselves were conducted with precision, and that our algorithm accurately calculates the Vmax and Km parameters.

Q2: NaturalCatalsyst can say that their NextGen-E Enzyme is their best enzyme. According to Table 1 below, the NextGen-E enzyme has the highest Vmax and the second lowest Km, making it the best performing enzyme out of the five. On the topic of manufacturing consistency, the calculated v0 in the Michaelis-Menten plots in Figures 1 to 5 below seems to match the model fairly accurately. The SSEs and mean absolute percent error between each enzyme's v0 and model are also relatively small (Table 2). These factors tell us that the enzymes are manufactured with consistency and that the enzymes themselves have highly predictable behavior when degrading stains.

References

Marasović, M., Marasović, T., & Miloš, M. (2017, March 5). *Robust Nonlinear Regression in Enzyme Kinetic Parameters Estimation*. Journal of Chemistry. https://www.hindawi.com/journals/jchem/2017/6560983/.

Stephanie. (2021, April 26). *Mean Absolute Percentage Error (MAPE)*. Statistics How To. https://www.statisticshowto.com/mean-absolute-percentage-error-mape/#:~:text=The%20mean%20absolute%20percentage%20error,values%20divided%20by%20actual%20values.

Appendix: Figures and Tables

Table 1: Vmax and Km Parameters of each of the 5 enzymes

Enzyme	Enzyme Parameters	
	V_{max} (µM/s)	K_m (µM)
NextGen-A	0.9290	165.0574
NextGen-B	0.8238	347.2961
NextGen-C	1.1673	192.6289
NextGen-D	1.4639	280.6448
NextGen-E	1.5567	169.7266

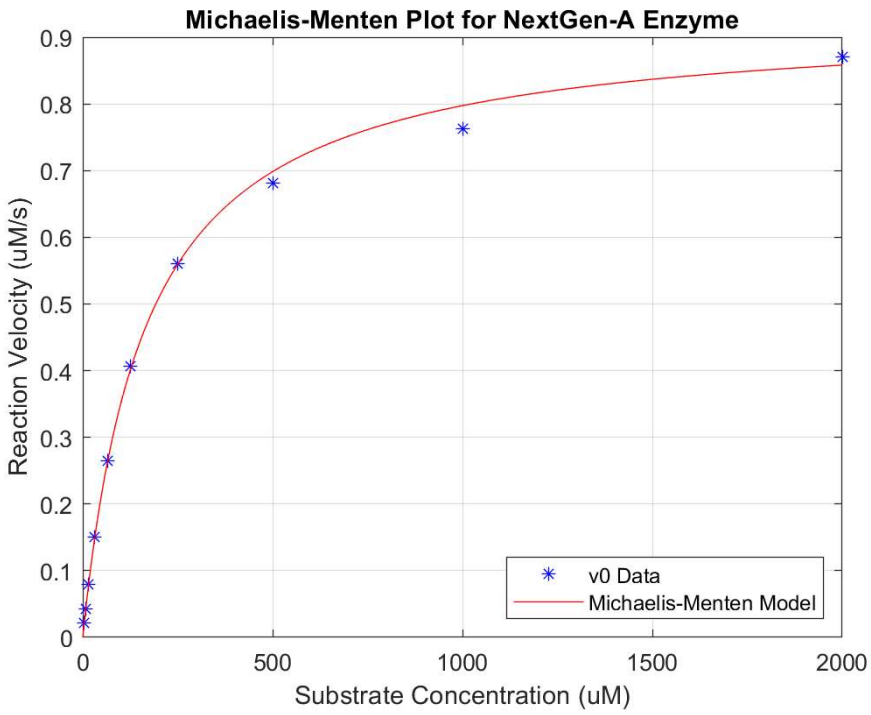

Figure 1: Michaelis-Menten Plot of the NextGen-A Enzyme

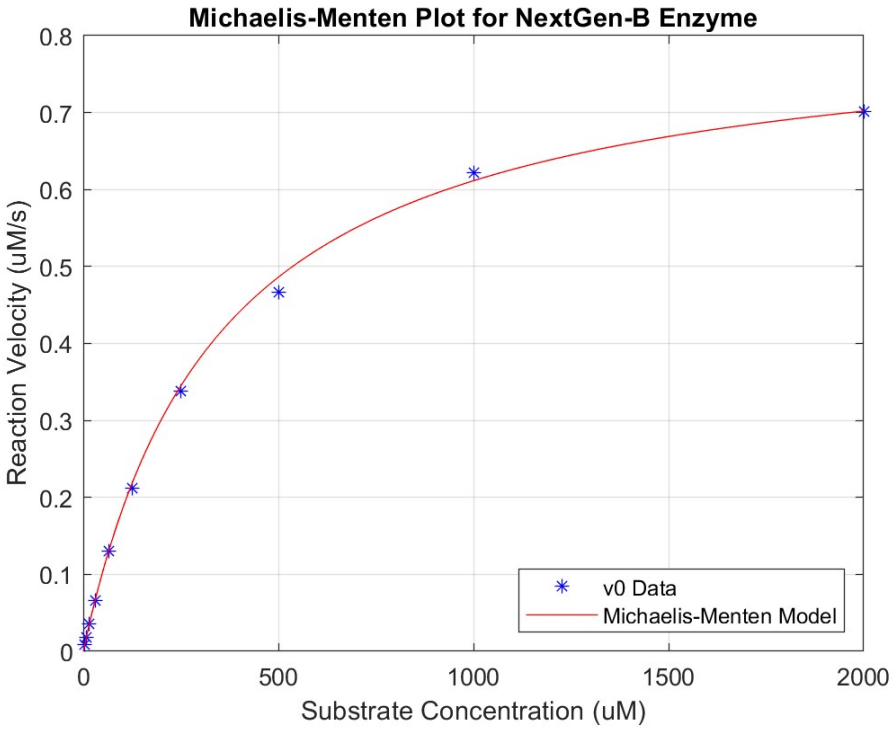

Figure 2: Michaelis-Menten Plot of the NextGen-B Enzyme

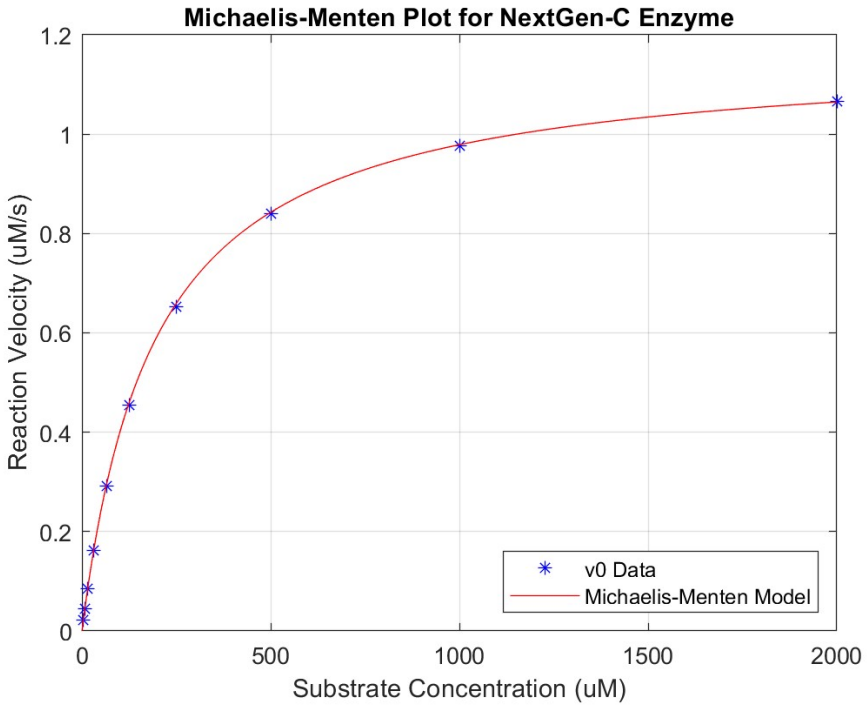

Figure 3: Michaelis-Menten Plot of the NextGen-C Enzyme

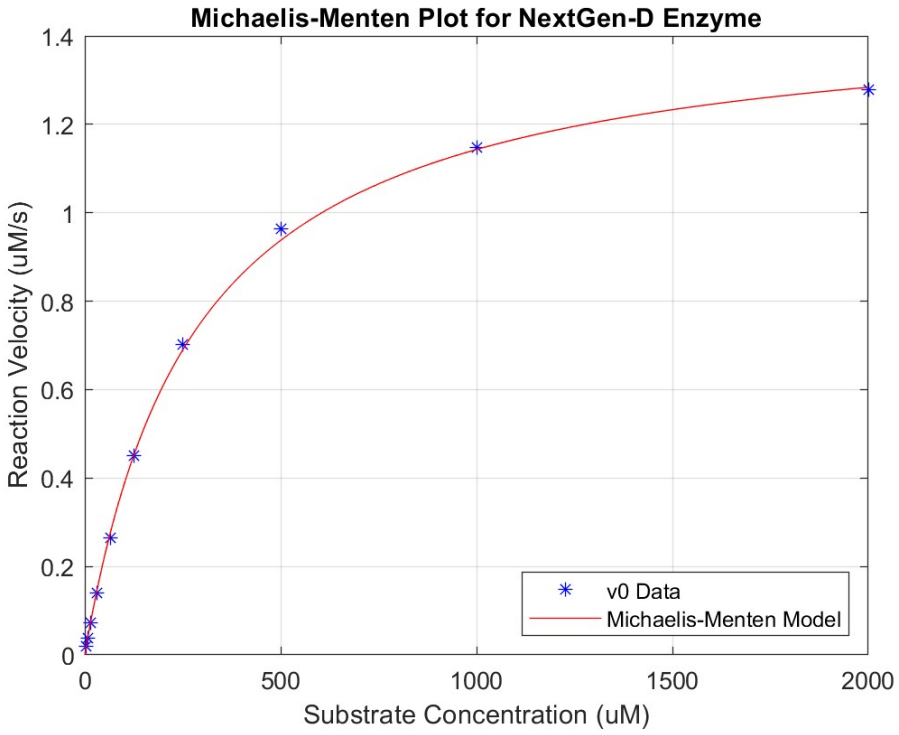

Figure 4: Michaelis-Menten Plot of the NextGen-D Enzyme

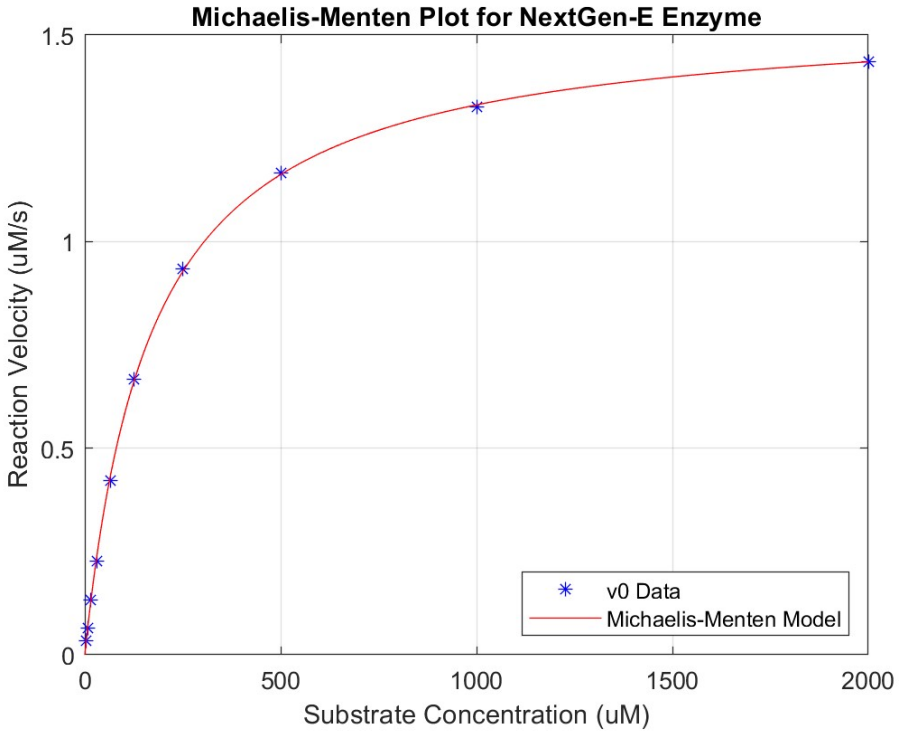

Figure 5: Michaelis-Menten Plot of the NextGen-B Enzyme

Table 2: SSE and Mean Absolute Percent Error between calculated v0 and Michaelis-Menten model

Enzyme	SSE ($\mu M/s)^2$	Mean Absolute Percent Error (%)
NextGen-A	0.0017	0.0924
NextGen-B	0.0006	0.0443
NextGen-C	0.0001	0.0292
NextGen-D	0.0010	0.0637
NextGen-E	0.0003	0.0042

Table 3: Algorithm performance compared to reference values of the PGO-X50 enzyme

Parameter (µM/s)	PGO-X50 Reference Values	Algorithm
v_{0_1} (uM/s)	0.025	0.0254
v_{0_2} (uM/s)	0.049	0.0491
v_{0_3} (uM/s)	0.099	0.0985
v_{0_4} (uM/s)	0.176	0.1729
v_{0_5} (uM/s)	0.329	0.3384
v_{0_6} (uM/s)	0.563	0.5366
v_{0_7} (uM/s)	0.874	0.8395
v_{0_8} (uM/s)	1.192	1.1754
v_{0_9} (uM/s)	1.361	1.3338
$v_{0_{10}}$ (uM/s)	1.603	1.5667
V_{max} (uM/s)	1.806	1.7609
K_m (µM)	269.74	269.6739
SSE $(\mu M/s)^2$	0.0048	0.0046