Table S2. Gen AI Prompts for Teams A and B which used Claude 3.7 Sonnet.

Team	Prompt
A	Looking at these instructions, how would you analyze this in google colab?: ASK: Analyze Eukaryotic Enzyme Length Distribution with regards to the potential preferred sizes. 2. READING: Read TWO key articles: https://www.pnas.org/doi/10.1073/pnas.91.9.4044 and https://www.liebertpub.com/doi/10.1089/153623102760092805. 3. DATA-1: Choose database (data preparation: e.g., EBI or NCBI) and compile target data set (data cleaning). 4. DATA-2: Apply Fourier-transform (data processing) and Exploratory Data Analysis (EDA). 5. HYPOTHESIS: Test hypothesis* and formulate conclusion/s. 6. DESCRIPTION: Provide a detailed description of the entire process.
A	Looking at these instructions, how would you analyze this in google colab? Please use the diverse enzyme .csv for analysis to start in google drive. ASK: Analyze Eukaryotic Enzyme Length Distribution with regards to the potential preferred sizes (to generate our own new Fourier transformation equation and use the cosine equation from the paper that got us ~125aa). Apply Fourier-transform (data processing) and Exploratory Data Analysis (EDA). Please provide secondary analysis using updated approaches including Full Fast Fourier Transform (FFT) that captures both amplitude and phase information * Wavelet analysis for multi-scale detection of periodic patterns * Short-Time Fourier Transform (STFT). Please also provide statistics comparing each analysis with each other and highlight p-values of any positive spectrogram peaks identified.
B	Help me to understand the Spectral Analysis of Protein Distribution using the attached paper as a guide. Create a step-wise approach to understanding this methodology used in the paper and break down each step. Presume that we could use R tidyverse and the R 'spectral' package if necessary.
B	As for the critique of the R script, yes. Adjust mixture model complexity, fixed peaks, parameter constraints, p-value interpretation. Window size limitation estimates, refer to the attached paper for guidance and cross-validation. Requesting to keep the Cosine model per the paper. Script Implementation issues aside, for now we need to fix the SAD and statistical modeling first. Regenerate another R script that addresses these issues, while also taking into account the attached. ALSO an after thought for further discussion - is the protein length normally distributed? What's up with the spikes in protein length in the dataset? Should we somehow take this into account?
B	We've created a complete, fully functional R script that performs spectral analysis on protein length distributions. The script follows the methodology from Kolker et al. and includes significant improvements to ensure reliability.
B	Given the attached R script and attached original study, how does this script compare to scientifically validated research published by Kolker et al? I take this to be the foundation to build upon. The goal is to reproduce as closely as possible the methodologies used by Kolker in the attached.