3 Methodology

The Garrido-Oter lab aims to develop a high-throughput phycosphere modeling system to study synthetic communities (henceforth, SynCom(s)) or rationally designed compositions of microorganisms[39]. The proposed system makes use of the ubiquitous soil algae *C. reinhardtii* (Chlamydomonas) and bacterial strains (henceforth, Strain(s)) from the Indexed Chlamydomonas-associated bacterial Library (ICL). Through the successful development of this system, the group hopes to uncover fundamental rules and processes, which determine community establishment, dynamics and stability. Absorbance and fluorescence spectroscopy are used to probe the system as an efficient way of quickly iterating over hypotheses and experiments. The two main instruments employed in the generation of the data used in this research are the Tecan M200 Infinite Pro and the Photon Systems Instrument Multi-Cultivator 1000-OD (henceforth Tecan and PBR, respectively). A third instrument, the Beckman Coulter Counter Multisizer 3 (henceforth, Coulter) is employed to obtain cell count estimates of PBR experiment samples. Two main data sets were created for analysis: Screening and PBR absorbance data. Some parts of each kind were generated prior to the commencement of this work, other parts were created specifically for it. These data are subsequently analyzed through predictive models based on the Beer-Lambert law and are derived from Tecan-measured calibrations.

3.1 Growing Chlamydomonas & Strains

Before any experiment can be conducted, Strains and Chlamydomonas are grown to quantities allowing experimentation; this typically means stock cultures of optical densities between $0.1 > OD > 1.0$. Chlamydomonas and Strains are individually obtained from stored samples in the ICL or other bacterial collections housed at MPIPZ. After thawing, the microorganisms are placed in respective pre-growing vessels until a desired cell density has been achieved and subsequently co-inoculated with respect to individual experimental designs. Pre-growing vessels are either tryptic soy agar (50%) plates for Strains or TAP medium for Chlamydomonas. Once SynComs have been designed and their strains pre-grown, individual members are left to grow continuously whenever possible.

3.1.1 Media

Several growing media are used during Tecan experimentation. Tris-acetate-phosphate (TAP) medium is the standard (full) growing solution for Chlamydomonas: Acetate and ammonium serve as Carbon and Nitrogen sources, respectively, and Tris helps buffer the pH [40]. Chlamydomonas is a heteroautotroph, hence it can grow via photosynthesis or use external organic carbon sources. In order to force Chlamydomonas to photosynthesize, Tris-phosphate medium (TP) is used; however, this requires CO_2 to be introduced into the system. TAP-N, is TAP medium without Nitrogen and TP-N has no acetate or Nitrogen. Chlamydomonas is unable to grow in any of the -N media and these are used as controls as needed**.** Previous work has determined which Strains can and cannot grow in full media; and given that the ICL was generated from bacterial strains grown alongside Chlamydomonas, it thus stands to reason that those which cannot grow in TAP by themselves must receive some nutrition either from Chlamydomonas or through other favorable metabolic exchanges with other community members. This notion helps guide SynCom design and experimentation.

3.2 Experiments & Data

Two general data sets were obtained for this project: Screening experiments absorbance and fluorescence data, and photobioreactor experiments absorbance data with its sample's Tecan measurements, cell count estimations and 16S profiling relative abundances. Tecan data are generated as excel files while PBR data is continuously measured and stored in CSV files. Processing of this data is discussed in section 3.5. The screening experiments were conducted to obtain Strain-specific or SynCom-specific interaction information as well as to test parameters for phycosphere modeling. PBR experiments were conducted to study the structure and dynamics of the specific SynComs themselves across longer time spans. The data is provided upon request as is referred to in this through this work as "Supplementary Material".

3.2.1 Tecan & Photobioreactor

The Tecan and PBR instruments employ the same fundamental physics properties to measure absorbance values and by doing so allow the Beer-Lambert law to be leveraged on their data. While the PBR is sold for the purpose of continuous production of microorganism under fixed parameters, in the lab the PBR serves to model the phycospheres over extended periods of time. Conversely, the Tecan serves to prototype and test inoculation mixtures and their stability, as well as to iterate over and test hypotheses without the time overhead that is required to prepare and set-up PBR experiments. The Tecan's measurements are made over

standard flat-bottomed Greiner 96-well plates (henceforth, microplates) which hold up to 300 μL of sample per well. Tecan experiments last no longer than two weeks and measurements are conducted daily. In comparison, the PBR hosts eight tubular vessels, each capable of housing 80 mL samples; measurements are taken continuously over user-specified time intervals across several weeks to months: the latest experiment (PBR4) is still ongoing. The Tecan serves to generate the screening data as well as the calibration data that allows for the derivation of Beer-Lambert law based models. Samples collected from some of the PBR experiments are also measured with the Tecan which allows testing the model's transferability across these two systems. While it is possible to generate models using the PBR, this would not only require greater quantities of growing media, Strains, and Chlamydomonas, but also time: the Beer-Lambert law and the Tecan minimizes these overheads. Fluorescence is measured during some of the Tecan experiments (the PBR is unable to measure this), but these values are discarded for reasons discussed in section 3.4.3.

3.2.2 Screening Experiments

Screening experiment 1 (S1) is conducted to obtain information on binary interactions between specific Strains and Chlamydomonas in the four growing media previously mentioned. Strain-only wells serve as controls in this and subsequent screenings as do the -N media. Previous analysis of S1 concluded that no Strain could grow in TP-N media thus subsequent screens no longer made use of these media. The specific strains used in each screening experiment are provided in the Supplementary Material. Absorbance measurements were collected at 600 and 750 nm wavelengths (A_{600} and A_{750} , respectively), and fluorescence was measured at 680 nm wavelength after 480 nm excitation (henceforth, F_{680}). S2 is conducted over three conditions in TP media: Strain-only controls, Strain and Chlamydomonas binary interactions, and Strain and Chlamydomonas in the context of a defined SynCom (denoted as B, B+C, SC+C, respectively). Absorbance measurements were taken at A_{750}, and fluorescence at F_{680}. S3 is conducted to obtain information on whether the initial inoculum ratios of Chlamydomonas-to-Strain in TP media make an impact on the overall growth. These ratios were 1:1.6, 1:16, and 1:160 as measured by optical density. Absorbances A_{600} and A_{750} were measured as well as F_{680}. S4 is conducted to determine if SynCom growth necessitated Chlamydomonas or if Chlamydomonas supernatant is sufficient for their growth and is meant to provide further evidence for the modulating effect of Chlamydomonas. Ratios were used to determine growth impact as in S3. Absorbances were measured at A_{680} and A_{750} and fluorescence at F_{680}. The screening experiments 1-3 (S1-S3) were conducted prior to commencement of this project while Screening 4 (S4) was conducted

for this project. The cell fractions of the previously conducted experiments S1-S3 had not yet been resolved and thus are given for analysis along with S4 data for this project.

3.2.3 PBR Experiments

Given that the phycosphere modeling system is still under development, the PBR generated data does not only serve to elucidate SynCom properties, it also informs on the conditions that best keep the system from crashing. This is especially true for experiments one and two (PBR1-2). This data was generated before the commencement of this work and helped guide the growing conditions for PBR3, which was conducted directly for this project. Unfortunately, PBR3 failed two weeks post-inoculation: The resulting data was nevertheless analyzed. This particular experiment was restarted as PBR4 after modulating growing conditions and has been continuously growing and generating data since. The PBR is making continuous measurements over specified time intervals and storing the information on a CSV file. The information can be accessed live via instrument-specific software or downloaded to be analyzed.In order to ensure the data is modeled properly, the PBR data has to be devoid of the medium's absorption effect.As it is necessary to perform the inoculation of the phycosphere vessels in a laminar airflow hood blank measurements are not taken to correct for the s. However, it makes sense that the measurements immediately post-inoculation are dilute enough to use them as a blank correction. This ensures that the predicted measurements are not over-predicted because of the medium's absorption. These correcting post-inoculation blank values are 0.066003 and 0.037617 for $A680$ and $A720$ measurements respectively.

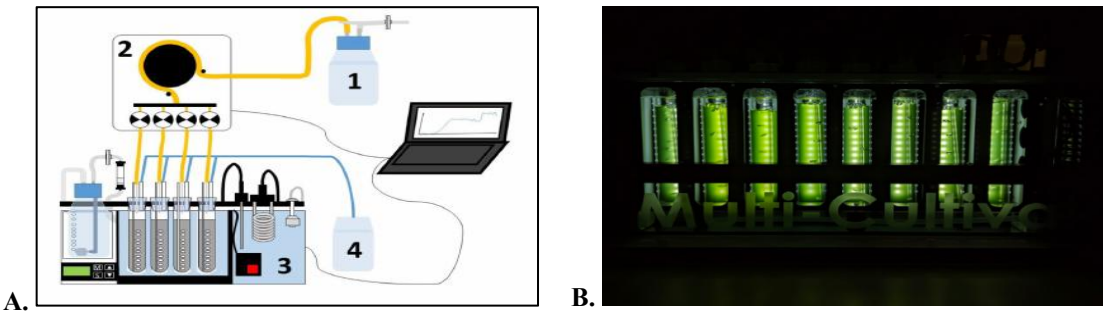


Figure 3.1: Photobioreactor & Accessories **A:**1. Media; 2. Turbidostat; 3. Photobioreactor 4. Waste Tank. **B.** Photobioreactor at night cycle shining green due to Chlamydomonas presence during a measurement.

3.2.4 PBR Correlating Data & Usage

As noted earlier, PBR experiments are not only measured for absorbance values; it is rather that each vessel—i.e synthetic phycosphere—is sampled and subsequently has its cell contents estimated with the Coulter and has its 16S sequences profiled. This data will be compared against concentration predictions. As the system gradually becomes more robust

and consistent, the group aims to perform sequencing less frequently. The general sampling steps are as follows: First, a 1 mL sample is taken from each of the eight vessels. Second, a 250 µL aliquot from each sample is transferred on to a microplate and measured with the Tecan at A_{680} and A_{750} wavelengths and the data is stored. Another 24µL aliquot is taken from each sample and transferred into its respective polymerase chain reaction (PCR) tube. Then 40 µL of a DNA buffer is introduced into each tube and thermocycled for 30 minutes at 94 °C. After thermocylcing is complete, 40 µL of a different DNA buffer is pipetted into each tube and subsequently frozen (as are the remainder of the samples). Once either an arbitrary number of daily samples has been collected or the experiment is considered complete, the frozen samples are thawed and prepared for 16S sequencing and relative abundance measurements as noted reported by Duran, et al[33].

Typical analyses of 16S sequencing results are performed over the type of bar graph shown in Figure 3.2. For this project, the relative abundances (RA) of the bacterial strain's Class and the Chlamydomonas are compared against the predictions of the models as a form of validation for the efficacy of this method. Each vessel's 16S sequencing reveals the composition of the community in terms of taxonomic Class relative abundances for any given day. The resulting sum of the bacterial Class RA is then compared against the Beer-Lambert model predictions for bacterial content as are the Chlamydomonas's RA to their predicted fraction.

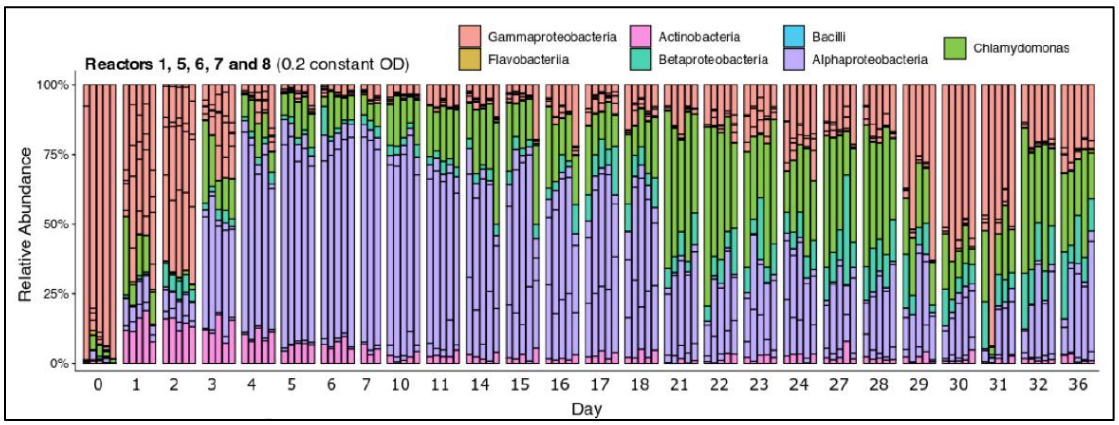


Figure 3.2: Relative abundances bar graph of 16S sequencing from PBR4. Dr. Jose Flores and Dr. Ruben Garrido-Oter provide the preparation for sequencing, sequencing and the visualization of this data.

3.3 Modeling I

The Beer-Lambert law states that the absorbance of a substance is proportional to the concentration of the sample and the path-length through which the light travels. The systems,

which employ this law often, standardize the length of the vessel through which light travels to 1cm, and by doing so simplify the procurement of the extinction coefficient (ϵ_λ), which typically has units of ($M^{-1}cm^{-1}$). This standardization of the path-length transforms the Beer-Lambert law to Equation 8 and facilitates the derivation of the constant k_λ as it would be equal to the slope of the resulting calibration curve having units of molarity (M^{-1}) instead. The Tecan, however, does not measure absorbance across 1cm sample vessels, and thus employs a path-length correction algorithm that utilizes an internal standard[41]. The magnitude of the slope of the lines from these resulting calibration curves can then be written to conform to standard extinction coefficient units ($M^{-1}cm^{-1}$) as needed, and facilitate the portability of the extinction coefficients to other systems employing absorbance.

$$A_\lambda = \epsilon_\lambda \bullet c \bullet l \quad \text{Equation 7}$$

$$A_\lambda = k_\lambda \bullet c \quad \text{Equation 8}$$

3.3.1 Extension of the Beer-Lambert Law

Commonly, the Beer-Lambert law is employed to generate simple regression curves for single component solutions of chemical compounds or macromolecules (e.g. proteins) allowing researchers to quickly estimate the concentration of their sample and move onto experimentation and analysis[36]. Additionally, a more elegant and expanded form of this law exists, which can be employed to resolve the concentrations of individual species in multiple component solutions and serves as the foundation of this study. The extension of the Beer-Lambert law states that in a multiple component system, its measured absorbance is equal to the sum of the individual measured absorbance of each species (Equations 9 & 10).

$$A_{\lambda total} = A_{\lambda 1} + A_{\lambda 2} + \cdots A_{\lambda n} \quad \text{Equation 9}$$

$$A_{\lambda total} = k_{1\lambda} \bullet c_1 + k_{2\lambda} \bullet c_2 + \cdots + k_{n\lambda} \bullet c_n \quad \text{Equation 10}$$

$$A_\lambda = k_{1\lambda} \bullet c_1 + k_{2\lambda} \bullet c_2 \quad \text{Equation 11}$$

Consequently, in order to determine the concentrations of N species, N wavelengths must be measured, and an N equation system needs to be solved for each species' concentration. In this theoretical system, there are N^2 coefficients which would first need to be determined as well. The phycosphere modeling system can theoretically be modeled using N species if every organism of the SynCom can have its extinction coefficient is determined. The goal undertaken for this project, however, is to ensure that the resolution between Chlamydomonas and SynCom bacterial strains is possible. Consequently, this project only considers a two

component system (Strains and Chlamydomonas) with four extinction coefficients to be determined.

3.3.2 Forced Constraints on Modeling

The Tecan is capable of taking absorbance measurements between 200 and 1000 nm, but the design of the Tecan screening experiments that were provided for analysis and the component specifications of the PBR dictate which wavelengths are to be used for modeling. S1 and S3 measurements were taken at A_{600}, A_{750} and F_{680}. S2 was measured at A_{750} and F_{680}, and S4 at A_{680}, A_{750} and F_{680}. Moreover, the particular photobioreactor model currently serving as phycosphere modeling system only measures absorbances at A_{680} and A_{720}. Lastly, the Tecan measurements of PBR 2 & 3 samples were measured at A_{680} and A_{730}. In order to probe the data provided, the models must consequently be generated using those wavelengths as regressands.

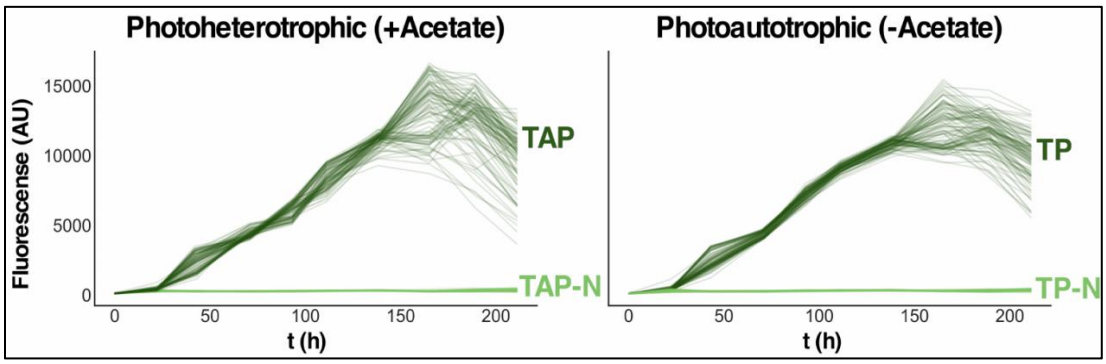


Figure 3.3: Fluorescence measurements of Chlamydomonas. Increasing intensity indicates growth initial growth. Flat curves indicate no observed growth. Figure is provided by Dr. Jose Flores

3.3.3 Exclusion of Fluorescence Measurements & Models

During the initial development phase of the phycosphere modeling system, daily fluorescence measurements were found to be a good proxy for measuring the growth of Chlamydomonas alone. This is due to Chlamydomonas's chlorophyll being a fluorophore and fluoresces near 680 nm [42]. As these cells grew, the signal became stronger than expected, however, and paradoxically, chlorophyll also absorbs at this wavelength and likely has the effect of underestimating the actual Chlamydomonas content of those measurements. An example of these initial measurements is shown in Figure 3.3. This overlapping of absorbance and fluorescence wavelengths causes the Inner Filter Effect (IFE)[43]; in essence, as one chlorophyll molecule becomes excited and fluoresces, the neighboring molecule is likely to absorb the emitted photon. As the system is rich in Chlamydomonas's chlorophyll, the probability of emitted photons being re-absorbed is high. As noted in the Background, the

derivation of fluorescent regression models is best employed when samples for these types of measurements are very dilute. While this technique is highly sensitive and allows for the determination of very dilute concentrations as low as 10^{-12}M[37,38], it does not work well in a context where absorbance measurements are OD > 0.1. Furthermore, when creating models to deconvolve these signals, the fluorescent measurements showed another behavior, which is dependent on bacterial load. In Figure 3.4 the overall signal is observed to decrease as the bacterial load increases. This dependence on the bacterial concentration, as well as the IFE and the fact that research is conducted has yielded measurements of absorbances greater than what is recommended for fluorescence regression models, makes the creation of these models unreasonable. Therefore, fluorescence measurements are disregarded altogether.

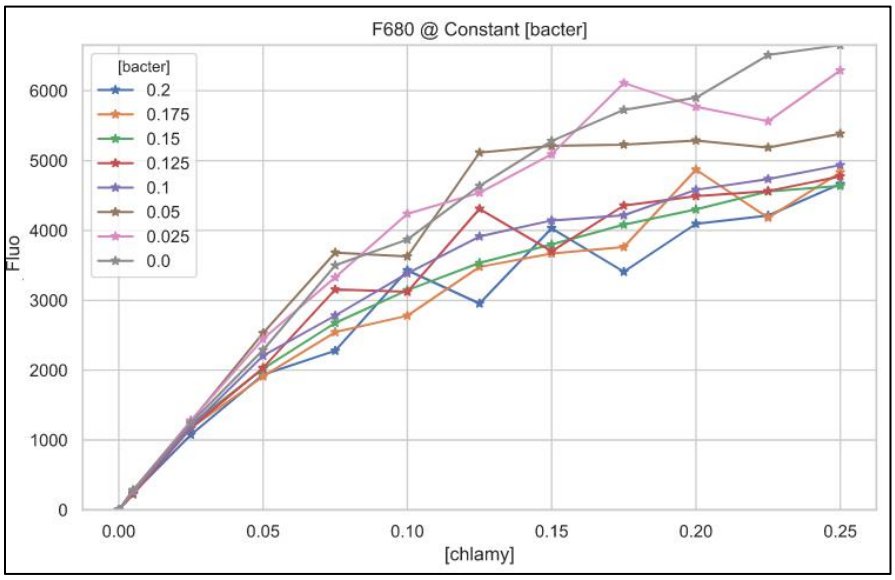


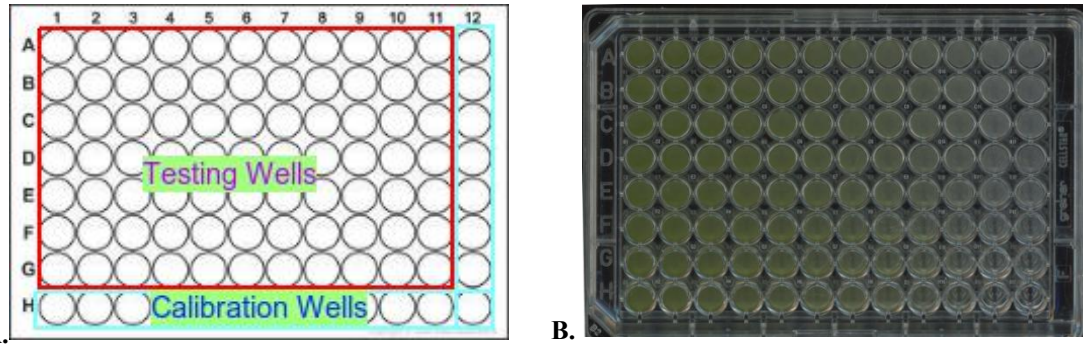
Figure 3.4: Fluorescence measurements of single strain calibration (SS). Fluorescence measurements decrease as bacterial [bacter] concentration increases

3.3.4 Calibration Samples Preparation for Modeling

Two types of calibration data were obtained to generate the models: a single-strain calibration and synthetic community calibration (henceforth SS and SC calibrations, respectively). Both of these calibration data are expected to behave similarly and are also expected to produce similar predictions; but were made at different times during this project and are discussed later. Chlamydomonas and bacteria samples were obtained from continuously growing stocks. For the SC calibration, several bacterial stocks were mixed in equal parts to compose SynCom007 stock (Strain lists are provided in Supplementary Material), whereas for the SS calibration, only one of these strains was used. From these stocks, 100µL & 20µL aliquots were taken and diluted into TP media to compose 1:10 and 1:50 ratios of 1mL samples; these

were subsequently measured using A_{750} to estimate the total concentration in absorbance units of the original stocks (cell count measurements are also measured). From these determined original stock concentrations, serial dilutions of 1mL are prepared to range between 0.0 to 1.0 theoretical (i.e. unmeasured) absorbance units for both Chlamydomonas and Strain(s). Theoretical absorbance concentration units are obtained using the serial dilution equation (Equation 12. Note that these values are irrelevant and are used only to ensure serial dilutions in accordance with the specified range for the calibration.

$$c_1 \bullet v_1 = c_2 \bullet v_2 \qquad \text{Equation 12}$$



A. **B.**

Figure 3.5: Microplate A. Schematic showing the relative positions of the wells used for calibration and those used for testing(Google search: 96 well microplate). **B.** Scanned microplate showing the directionality of the gradient; these increase in opaqueness right-to-left, bottom-to-top (Curtesy of Dr. Jose Flores-Uribe)

Next, the serial dilutions are transferred into the microplates and mixed with their corresponding partner, either to Chlamydomonas or bacterial solution, to a final volume of 300µL. Note the labeling of the microplates in Figure 3.4A, numbers are used (1-12, left-to-right) for columns of wells and letters for the rows of wells (A-H, top-down)**.** From each of the twelve serial dilutions of Chlamydomonas, 150µL are pipetted into row H, beginning with the lowest dilution at H12 and going up in gradient to H1. The same process is repeated for each row resulting in every column containing the same concentration of Chlamydomonas. Then, from each of the eight serial dilutions of SS or SC samples, 150µL are introduced at column 12, beginning with the lowest dilution at H12 and going up in gradient to A12 using a multi-channel pipette. The same process is done for every column resulting in every row containing the same bacterial load. H12 contains 300µL of TP medium only and is used as a means of correcting for the absorbance of the medium in all wells. Figure 3.5B shows a scanned image of the calibration plate. The calibration plates are subsequently measured using the Tecan: Row H and column 12 serve to generate the calibration models and all other rows serve to test the models before they are applied to the screening and PBR data. For the

SS calibration, the Tecan is set up to measure the microplate between 300 and 800nm wavelengths using a 5nm step-size. Fluorescence measurements were also taken at F_{680}. For the SC calibration, the Tecan is set up to measure the microplate between 500 and 750nm using a 10nm step-size without taking fluorescence measurements. The SC calibration data have multiple measurements for A_{600}, A_{680}, A_{720}, and A_{750} and are averaged as used to make the required models. Single measurements are made at all other wavelengths in both SC and SS.

3.4 Data Processing

All Tecan generated data are measured over several days and produce an excel file per condition in any given experiment including the calibration data. Tecan data thus result in a variable number of files for any given experiment (all data are provided in Supplementary Material) and must be parsed into their own manageable (comma separated values) CSV files for efficient use. The resulting CSV files and those of the PBR are then pre-processed. Pre-processing of the screening experiments is done extensively in order to facilitate model application and time-series analysis; details are provided in their respective Jupyter Notebook. The PBR data processing and requires only the removal of empty values and formatting for model application. All work was accomplished using Python (v. 3.7.3) libraries on Jupyter Notebooks (v. 1.0.2). The use of Pandas (v. 0.24.2) is extensive for the majority of the work conducted. Modeling relies heavily on Numpy (v. 1.16.4), Scikit-Learn (v. 0.21.2) and SciPy (v. 1.3.0) libraries. Visualization functions rely mostly on MatplotLib (v 3.1.0), but Seaborn (v. 0.9.0) is used throughout for quick visualizations as needed. Jupyter Notebooks were utilized for the coding to create the models and the visualization of the data and predictions, and are provided as Supplementary Materials upon request.

3.5 Modeling II

To create predictive models, the extinction coefficients need to be extracted from the calibration data. However, at the onset of this project, it was not self-evident how best to achieve this given that measurements of the concentration were ‘theoretical’ and unit-less and not in molarity (M) as expected. Nevertheless, the imposed constraints of the existing data serve as a good starting point and scatter plots at the imposed wavelengths (A_{680} & A_{750} for S4) are made against the theoretical concentrations of the SS calibration’s row H and column 12, as seen in Figure 3.5A. Ultimately, any wavelength measured can serve as both regressor and regressand, and therefore all measured wavelengths can serve as a proxy for concentration to regress back to.

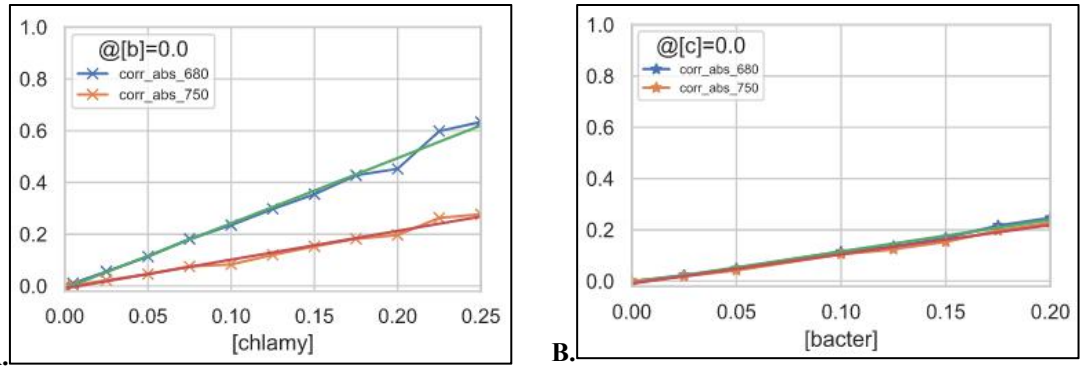


Figure 3.6: Typical simple linear regression on scatter plots. Represented here are the microplate's row H (green curves) and column 12 (red curves). The [bacter] and [chlamy] labels refer to the concentrations of single bacteria (or community) and Chlamydomonas concentrations.

3.5.1 Obtaining the Extinction Coefficients

Linear Regression is used to obtain the coefficients needed for the models. This statistical analysis technique is used to find the best fitting line through a set of points and takes the general form of Equation 13 [44]. To apply the Beer-Lambert law, however, only Simple Linear Regressions (SLR) need to be solved as each regressand (A_λ) can only depend on one regressor at a time. The regressors are the absorbance values at any measured wavelength, representing concentrations of either Chlamydomonas or Strain(s) in the x-axis.

$$f(X) = \beta_0 + \sum_{j=1}^{n} X_j \beta_j \qquad \text{Equation 13}$$

$$f(x) = \beta_0 + x\beta_1 \qquad \text{Equation 14}$$

Equation 14 is similar to the Beer-Lambert law (Equation 8) where $f(x)$ represents A_λ, $x\beta_1$ corresponds to $c \bullet k_\lambda$, and β_0 represents the error in each prediction. To obtain these coefficients, Scikit-Learn's Linear Regression class was used. Scikit-Learn's implementation uses Ordinary Least Squares (OLS) to find the slope—and hence extinction coefficient—of the line that best fits the data. The OLS algorithm solves Equation 15 to estimate the slope (β_1).

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \underline{x})(y_i - \underline{y})}{\sum_{i=1}^{n} (x_i - \underline{x})^2} \qquad \text{Equation 15}$$

$$\beta_0 = \underline{y} - \beta_1 \underline{x} \qquad \text{Equation 16}$$

For the calibration data sets SS and SC, the wavelengths mentioned in section 3.4.2 are modeled against all other measured wavelengths, and the obtained extinction coefficients for each of these regressions are provided in the Supplementary Material.

3.5.2 General Form of Predictive Models

The extended form of the Beer-Lambert law is used to create the predictive models, which will be applied to the collected absorbance data. To create these models, extinction coefficients are first derived as described above and are put into a two-equation system, which is then solved for the respective variables representing bacteria and Chlamydomonas concentrations. The derivation of the models is illustrated in Equation 10 for wavelengths A_{680}, A_{750} which are used to analyze S4 data. Equation 10 is first rewritten for these specific wavelengths and new variable symbols are chosen to represent bacteria ($[B]_\lambda$) and Chlamydomonas ($[C]_\lambda$) concentrations as measured at an arbitrary absorbance wavelength, A_λ. Notation 1 is employed when referring to the system of equation that corresponds to a model.

$$A_{\lambda 1} \sim A_{\lambda 2} + A_{\lambda 3} \qquad \text{Notation 1}$$

$$A_{\lambda total} = k_{1\lambda} \bullet c_1 + k_{2\lambda} \bullet c_2 + \cdots + k_{n\lambda} \bullet c_n \qquad \text{Equation 10}$$

$$A_{680} = k_{C680}[C]_\lambda + k_{B680}[B]_\lambda \qquad \text{System 1}$$

$$A_{750} = k_{C750}[C]_\lambda + k_{B750}[B]_\lambda$$

$$\frac{A_{680} - k_{C680}[C]_\lambda}{k_{B680}} = [B]_\lambda \qquad \frac{A_{680} - k_{B680}[B]_\lambda}{k_{C680}} = [C]_\lambda$$

$$\frac{A_{750} - k_{C750}[C]_\lambda}{k_{B750}} = [B]_\lambda \qquad \frac{A_{750} - k_{B750}[B]_\lambda}{k_{C750}} = [C]_\lambda$$

$$\frac{A_{680} - k_{C680}[C]_\lambda}{k_{B680}} = \frac{A_{750} - k_{C750}[C]_\lambda}{k_{B750}} \qquad \frac{A_{680} - k_{B680}[B]_\lambda}{k_{C680}} = \frac{A_{750} - k_{B750}[B]_\lambda}{k_{C750}}$$

$$A_{680}k_{B750} - k_{B750}k_{C680}[C]_\lambda = A_{750}k_{B680} - k_{B680}k_{C750}[C]_\lambda$$

$$A_{680}k_{C750} - k_{C750}k_{B680}[B]_\lambda = A_{750}k_{C680} - k_{C680}k_{B750}[B]_\lambda$$

$$A_{680}k_{B750} - A_{750}k_{B680} = k_{B750}k_{C680}[C]_\lambda - k_{B680}k_{C750}[C]_\lambda$$

$$A_{680}k_{C750} - A_{750}k_{C680} = k_{C750}k_{B680}[B]_\lambda - k_{C680}k_{B750}[B]_\lambda$$

$$A_{680}k_{B750} - A_{750}k_{B680} = (k_{B750}k_{C680} - k_{B680}k_{C750})[C]_\lambda$$

$$A_{680}k_{C750} - A_{750}k_{C680} = (k_{C750}k_{B680} - k_{C680}k_{B750})[B]_\lambda$$

$$\frac{A_{680}k_{B750} - A_{750}k_{B680}}{(k_{B750}k_{C680} - k_{B680}k_{C750})} = [C]_\lambda$$

$$\frac{A_{680}k_{C750} - A_{750}k_{C680}}{(k_{C750}k_{B680} - k_{C680}k_{B750})} = [B]_\lambda$$

$$\widehat{[C]}_\lambda = \frac{A_{680}k_{B750} - A_{750}k_{B680}}{k_{B750}k_{C680} - k_{B680}k_{C750}} \qquad \widehat{[B]}_\lambda = \frac{A_{680}k_{C750} - A_{750}k_{C680}}{k_{C750}k_{B680} - k_{C680}k_{B750}} \qquad \text{Predictive Model 1}$$

The solutions to System 1 states that given absorbance measurements at (A_{680}) and (A_{750}) of a multiple component system (Chlamydomonas + Strain(s)), predictions, represented with the hat (^) symbol, can be obtained for each component in the system. Furthermore, it follows from the Beer-Lambert law that since each component's prediction is actually an absorbance prediction, they must also obey the law and be equal to the empirical measurements of the microplate wells at the respective absorbance wavelength (A_λ). Therefore, in order to evaluate the predictive power of the models, the sum of the predictions $\widehat{[C]}_\lambda + \widehat{[B]}_\lambda$ can be compared against the empirical measurements of wells A-G:1-11 of the calibration plates. Furthermore, since the concentration of each component introduced into the wells A-G:1-11 is known, each prediction $\widehat{[C]}_\lambda$ or $\widehat{[B]}_\lambda$ at each of these wells must also equal the known concentrations. Model application is not as trivial as in their application in compound or macromolecule predictions and required thorough analysis for its application in these systems; thus, the models are further discussed in Results.

3.5.3 Model Application to PBR Data

In order for the Tecan-generated models to be applied to PBR data, they must be converted to fit its path length. In practice, this is done at the moment of model application after the extinction coefficient has been determined and is only a matter of scaling these coefficients accordingly. Recall that the general form of the Beer-Lambert law is as Equation 7 states. For this equation to hold true, the path-length (l) of the vessel must be known and the extinction coefficient must have units which cancel out with respect to the concentration and path-length units. The extinction coefficients constants k_λ obtained here are unit-less since they are produced by regressions that use unitless absorbance values as concentration. However, the standardization of the values by the Tecan means that the magnitudes of the ϵ_λ and k_λ are equal and we denote them as having cm^{-1} units. Thus, model application to the PBR measurements only require that we scale these extinction coefficients by the path-length which is 2.7cm [45]. This will ensure that the returned predictions are also in unit-less absorbance values.

3.6 K-Space & Model Optimization

Two key results are analyzed in Results & Analysis. These are the computation of K-Space yielding an optimal model and the optimization of these models given the observed parabolic dependencies Results & Analysis 4.3. Discussed here are the technical steps of those findings.

3.6.1 K-Space Calculation

K-Space is comprised of the computed denominator K from the coefficients of all possible models regression that the given SC calibration data can generate. Three wavelengths are required to create each model and since there are twenty-six of them, 15600 permutations can be calculated for their respective K. The computation of K-space (Figure Results & Analysis 4.8) took 30 minutes on an Ubuntu 18.04.2 machine running on an Intel Core i7 CPU. The four coefficients are derived as describe in section 3.6.1 and then calculating the denominator of the models. Each coefficient is tabulated to its model $(A_{\mathrm{Z}} \sim A_{\mathrm{X}} + A_{\mathrm{Y}})$ as an index. The highest yielding value is hypothesized to be the best performing model.

3.6.2 Theory of Model Optimization

It is observed that Strain concentrations are under-predicted by the models proportionally to the known bacterial concentration and Chlamydomonas concentration. After fitting the prediction with a curve it is observed that the Strains predictions are non-linearly dependant on the concentration of the Chalmydomonas, rather the fitted curves are parabolic (Equation 11) in nature with an average breadth or a_2 value of -0.4187. It is also observed that the parabolic curves (one curve per row of the microplate) shift their vertices up and leftwards proportionally to the true concentration of the bacteria (relationship explained in Results & Analysis). This x-coordinate of the fitted parabola's vertices (h, k) [47] can be calculated numerically using Numpy and since these h are observed to have a relationship with the true concentration of the bacteria, their relationship can be fitted to obtain the constant or slope $[B]/h = H$, where h is a theoretical concentration of Chlamydomonas representing the x-coordinate of a determinable parabola, and $[B]$ is the known concentration of bacteria. This results in a linear relationship (Equation 12) that can be used to find a the theoretical h belonging to that predicted p[B]. The yielded theoretical h from Equation 12 is used to determine $\boldsymbol{a_1}'$ using the relationship in Equation 14. Given that when the parabolas are truncated to have an y-intercept value of zero, they maintain the same parabolic form (Equation 13) and represent how much signal Chlamydomonas is responsible for taking, these resulting curves given any ${}_{\mathrm{p}}[B]$ and ${}_{\mathrm{p}}[C]$ can be used can be used to predict values which are nearly equivalent to how much the bacterial concentration is under-predicted at the given $({}_{\mathrm{p}}[C])$. The final yielded result is a good first approximation to correct the signal because the Chlamy predictions themselves are highly accurate (Figure 3.7).

The parameters h must be found numerically from each of the fitted parabolas as the set of returned parameters [a_2, a_1, a_0] from those fittings do not yield an h using the analytical function to obtain the vertex of a parabola (h,k). Through Nympy's poly1d function, which takes in a curves parameters to create a predictive linear model, a list of 100K numbers, representing a range of h values in a range [-1 1], is used to obtain the h value at each apparent vertex (using the argmax property of Nyumpy's lists). The large list of numbers is necessary approximate the true h as best as possible. As described previously these h values are then plotted against known B concentration of each row and results in a unique Equation 12 depending on the model.

$$f(x) = \boldsymbol{a_2}x^2 + \boldsymbol{a_1}x + a_0 \quad \text{Equation 11}$$

$$h((\mathrm{p}[B]) = H \bullet (\mathrm{p}[B]) + b \quad \text{Equation 12}$$

$$f'((\mathrm{p}[C]) = \boldsymbol{a_2}(\mathrm{p}[C])^2 + \boldsymbol{a_1}'(\mathrm{p}[C]) + 0 \quad \text{Equation 13}$$

$$\boldsymbol{a_1}' = -2\boldsymbol{a_2}h \quad \text{Equation 14}$$

3.7 Absorbances to Cell-Counts

The majority of soil bacterium measures under 1.2μm while Chlamydomonas cells conservatively measure 10μm [40,46]. If the cells are modeled as spheres, this would mean that Chlamydomonas has 500 times the volume of a typical bacterial cell and will absorb much more light per unit cell than the Strains. To abstract away this layer, Coulter cell count estimates are regressed against the measured spectrum of the SC calibration data (400 - 750 nm) using SLR as described above. The cell count regression for A_{560}.predictions are found in Appendix 1. For quick reference to predicted absorbances-to-cell counts and cell-cell ratio estimates, a table from these regressions has been devised and provided in Appendix 2. The regression is used to compare cell-count relative abundances to the 16S relative abundances provided.

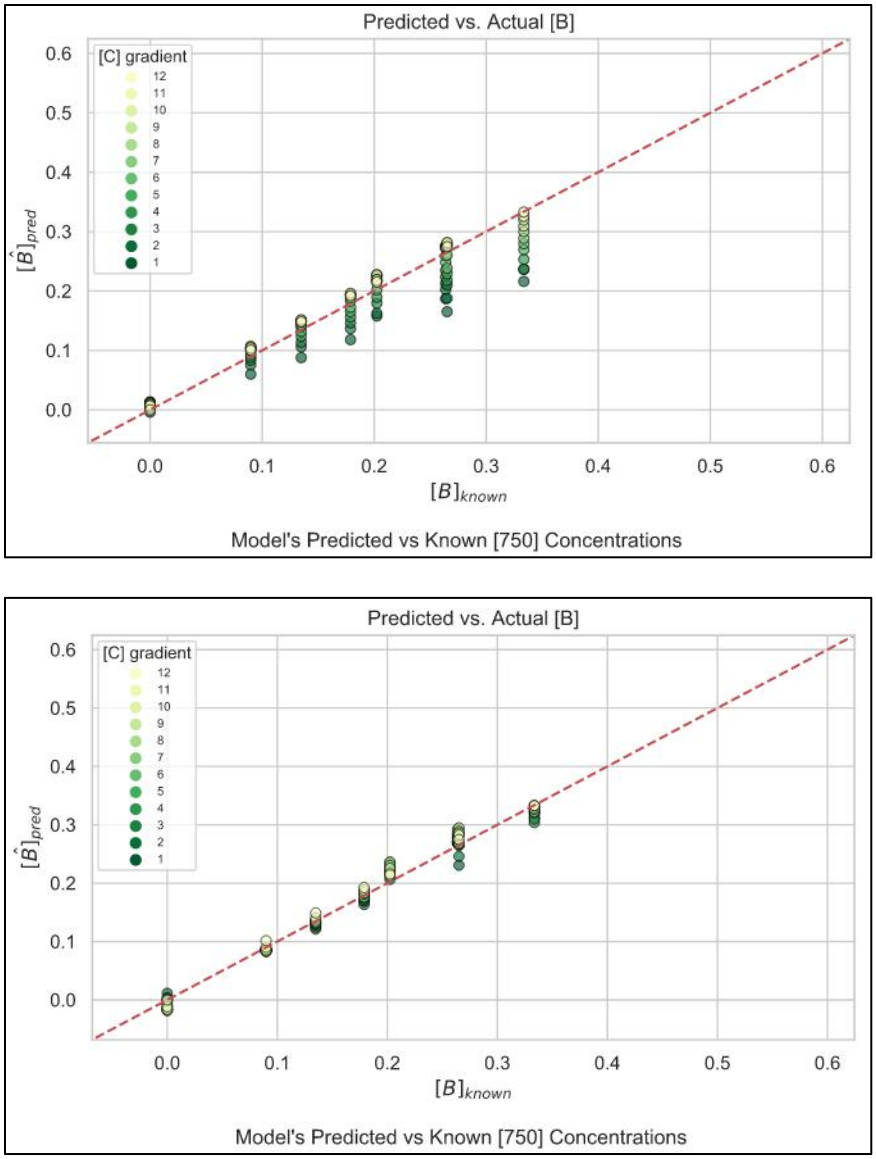


Figure 3.7: Model S results before (top) and after(bottom) model optimization.