

Project Outline 05 - Data analysis for antibiotic drug discovery

Background to the project





Antimicrobial resistance (AMR) is one of the most significant threats to healthcare systems globally. As pathogenic bacteria increasingly evolve resistance to existing antibiotics, there is an urgent need for the discovery of either new antibiotics, or drugs with potential to restore the sensitivity of pathogens to existing antibiotics by blocking resistance mechanisms.

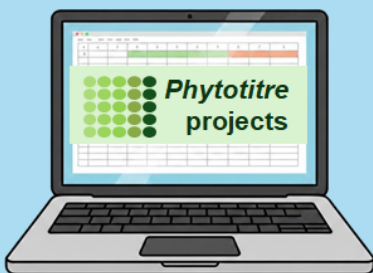
Historically, the discovery of new antibiotics has been achieved mainly by the large-scale screening of natural products. This approach has also given rise to many of the most widely used and successful drugs in other categories, including anti-cancer, anti-inflammatory, anti-hypertensive and cholesterol lowering drugs.

These discoveries were enabled by the vast chemical diversity present in the Natural world, a uniquely powerful resource which retains great potential for further drug discovery. By systematically testing hundreds or thousands of different natural extracts or compounds, researchers aim to discover and isolate a small handful of rare 'hits' which possess the desired biological activity against a target of interest. This method, called **high throughput screening** (HTS), is a critical step in the modern drug discovery process. Your project will involve the analysis of datasets arising from screens of two large natural product libraries for novel molecules with potential to inhibit the growth of model Gram-positive and Gram-negative bacteria.



Project at a glance

-  Data analysis project
-  8 weeks of analysis
-  BSc / MSc level
-  Antibiotic discovery



Benefits of choosing this project

Through this project, you will gain experience of several key data analysis and statistical techniques which are in high-demand from life-science employers. These include management of large datasets, background correction, normalisation, plate mapping, correlation analyses, ANOVA, and four parameter logistic (4PL) curve fitting. Beyond gaining key industry-relevant skills, your analyses will contribute to the search for new antibiotic leads.

How long will it take?

The basic project outlined below can be achieved typically within an **8 week** period of analytical time (not including the time necessary for writing of the dissertation). However, this can be extended to over **16 weeks** if necessary by following the suggestions for further, more in-depth analyses given in the later sections. Please discuss with your supervisor how to modify the project to achieve your aims while fitting within the available timeframe.

What will I do on this project?

The core activities you need to complete the basic version of this project are listed below:

- 1) Perform background correction and normalisation of a large HTS dataset
- 2) Calculate Z'-factor scores for each assay plate from a large screen for new antibiotics
- 3) Map assay values to extract and compound IDs using plate maps
- 4) Identify the top "hits" from the screen for further analysis
- 5) Perform correlation analyses to seek insight into the specificity of hit extracts
- 6) Explore correlations between compound structural features and biological activity
- 7) Use 4 parameter curve fitting to calculate IC₅₀ values for top hits
- 8) Use these analyses to prepare your final report or dissertation

What skills will I gain through this project?

Completing this project should give you experience of the following core employability skills:

Data handling skills		Statistical analyses	
Standard Microsoft Excel techniques	✓	Calculation of assay Z' factor	✓
Data Management and Integrity	✓	One way ANOVA	✓
HTS data handling	✓	Two way ANOVA	✓
Background correction & normalisation	✓	4-parameter dose response curve fitting	✓
Mapping assay values to plate maps	✓	Linear regression for correlation analysis	✓
Creation of appropriate charts	✓	Applying corrections for multiple testing	✓

What support is available?

Comprehensive method sheets, which describe how to perform each step of the data analysis protocols are available from the [downloads page on our website](#). These are free to download without registration. Troubleshooting advice to help solve common problems that may arise during the project is also given towards the end of this project guide. Your university supervisor will be your primary point of contact and will be responsible for your day-to-day supervision.

Instruction list

Complete these tasks in the following order for a basic computer-based data-analysis project in Antibiotic Discovery. These should be completed in the order shown below.

The preparation phase

Step 1: Read the introduction to the *Phytotitre* and *Puretitre* libraries

The *Phytotitre* library is a collection of natural extracts that has been developed to support drug discovery projects in academia and industry. The full-size kit comprises both polar and non-polar extracts of 400 plants, for a total of 800 extracts provided in ten 96-well plates. This collection comprises almost entirely of traditional herbs or medicines with a history of oral use in humans. By focussing on such plants, the aim is to increase the likelihood of identifying leads with both a high hit rate for biomedical targets, and a favourable safety profile. A smaller version of the *Phytotitre* collection, comprising 400 extracts, is also available to support student research projects. The *Puretitre* library is a collection of 200 pure natural compounds, most of which are also components in traditional herbs or medicines.

To learn more about the background to the *Phytotitre* and *Puretitre* libraries, please read **Method sheet 100**.

Step 2: Familiarise yourself with relevant Health and Safety information

Although you will not be performing any laboratory work during a data analysis projects, you should still be aware of the risks associated with the collection and analysis of data. Please read our '**Safety First**' document to understand the primary risks associated with a computer-based data analysis project in this area.

Every institution has a slightly different way of training and recording student health and safety inductions. However, the following are some common steps that you may have to complete before starting your data analysis project:

1) Read the project risk assessment

Every new project should always begin by writing a **Risk Assessment** document to cover the anticipated risks associated with the project. You should discuss with your supervisor who will prepare this, then read and sign the risk assessment associated with your project. Take note in particular of what the key risks are, and how to mitigate them (e.g. adjusting your seating height and keyboard placement, etc.).

2) Be aware of and follow your local health and safety rules

In addition to reading the relevant risk assessment documents, there may be other procedures necessary to complete before you are able to work in the laboratory. Please ask your supervisor for advice on what else may be necessary to complete before starting.

Step 3: If necessary, complete a Research Ethics application

Some universities require students to submit an Ethics Application to address any potential ethical issues the project may raise before you start your project. The format of such applications varies greatly between institutions. However, if you are required to submit an ethics application for your project, we advise that you cover at least the following points in your submission:

1: No use of human samples or personal data

It will be helpful to state near the start of your application that the project will be based on the analysis of publicly available data only. Mention that there will be no laboratory work performed,

which means there will be no biological, genetic or chemical risks associated with the project. You should also state that you will not be collecting any data from volunteers, or analysing any personal or identifiable data. It may help to mention that as your project be *in silico* only, with no use of animals or animal products, your project aligns well with the modern movement towards reducing animal use in research (3Rs - Refine, Reduce, Replace).

3: Compliance with the Nagoya Protocol

The Nagoya Protocol is an international agreement under the Convention on Biological Diversity (CBD) that ensures the fair and equitable sharing of benefits arising from the use of genetic resources (such as plants) or related traditional knowledge in certain ways. Your ethics application should acknowledge that (i) the Nagoya Protocol is relevant as the experiments from which the data were obtained involved the screening of plant extracts from diverse countries and (ii) the experiments as designed are compliant with the Nagoya Protocol, as discussed in more depth on [our website](#).

Reminder: Always show your draft ethics application to your supervisor before submission, as they are the 'Principal Investigator' (PI) responsible for your supervision.

Step 4: Make sure the necessary IT resources are in place

Our data analysis projects are designed in such a way that students should be able to complete them without requirement for expensive specialist software or subscriptions. We recommend that you use Microsoft Excel for the majority of the work, but any equivalent spreadsheet software should work as well, including freeware alternatives. The main requirements for equipment are as follows:

Equipment

- Any PC or laptop capable of running standard spreadsheet software

Software

- *Recommended:* Microsoft Excel (any version)
- *Alternative:* Any equivalent spreadsheet software

Step 5: Plan strategies for Good Laboratory Practice, Data Management and Integrity

A key employability skill for those seeking work in the life-sciences is knowledge and experience of working to the standards of **Good Laboratory Practice (GLP)** and **Data Management and Integrity**. You should therefore aim to practise these principles throughout your project. Begin by reading **Method Sheet 17**. This document will give you an overview of how to maintain your laboratory notebook and data files to comply with GLP and Data Management standards, which are essential for workers in the life science industries.

The data analysis phase

Please make sure you have completed all necessary activities in sections 1-5 before beginning the analysis phase of your project.



Step 6: Download the necessary Excel files and Method sheets

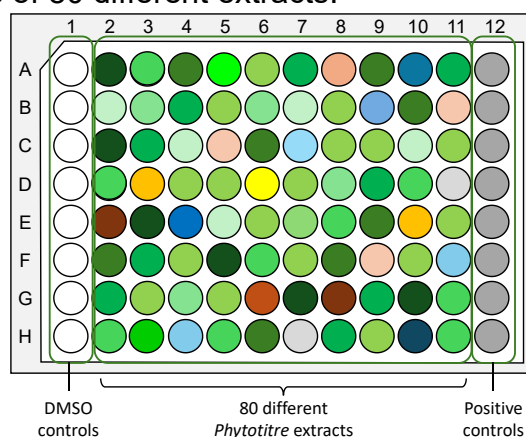
You should begin by downloading the necessary Excel files from the [downloads page on our website](#). These files contain the raw data which you will analyse. As soon as you have downloaded the files, place them in a folder marked 'Primary data files', and remember to never alter these files. Instead, you should copy the Excel files into a new folder, marked 'Data analysis', and give each file a new name (e.g. "Phytotitre analysis 01"). In this way, you keep the original data safe, minimising the risk of accidentally altering it during the analysis. This is an essential principle of Data Management and Integrity. It may also be helpful to download all of the relevant Method sheets, and save them in a separate folder.

At this stage you should also read **Method sheet 101**, which gives advice on the basic data manipulation methods you will need to be familiar with to be able to complete the project. Practise some of the examples yourself until you feel confident performing all of the main data processing functions described in this method sheet.

Step 7: Perform background correction of plate data

Open the working version of the *Phytotitre* dataset relating to the screen for inhibitors of growth of *Escherichia coli* (*E. coli*). The first tab explains the methods used during the screen, the second gives the plate maps, and the third gives information on each plant contained within the library. The following tabs contain the raw data from the screening experiments. As each stock plate contains 80 different extracts, and there are 10 different stock plates in the *Phytotitre* library, there are 10 tabs containing data, each obtained from assays of 80 different extracts.

The raw absorbance values of the bacterial culture plates have been measured at 0 hours (the baseline measurement), 8 hours and 21 hours. The data are arranged in the same format as the 96-well plates as they were screened. Each stock plate was assayed on three separate occasions (i.e. 3 independent experiments were performed). These are shown as three tables of values in 12 x 8 cell format, arranged side by side horizontally. Every plate was challenged with treatments according to the map shown at right.



Your first task will be to perform **background correction** of the absorbance values for every plate shown in each tab. This is necessary to account for any effects of coloured plant compounds on absorbance measurements that could otherwise confound interpretation of the results. To do this, follow the advice given in **Method sheet 102**.

Step 8: Perform normalisation of the data

Once all the data are background-corrected, you should proceed to **normalising** the data from every plate. This means transforming the raw absorbance values into a percentage of the negative control values (i.e. the DMSO controls in column 1 of the plate). This is necessary to enable correction for day to day variation in the maximum absorbance reached by cultures between different experiments. Do this for all three experiments at both the 8 hour and 21 hour timepoints. You can do this by following the advice given in **Method sheet 103**.

Step 9: Calculate Z' factor values for each of your plates

Now that you have normalised the data from each plate, the next step is to test how reliable and reproducible the results of the screening assay were. This is a necessary step in the drug discovery process, as it helps to establish whether the observed results reflect real differences, and are not just artefacts or “chance events”. The industry standard approach is to calculate a metric called the Z' factor (pronounced “Z-prime factor”) for every plate tested during the primary screen. Z' factor values greater than 0.5 indicate that the positive and negative controls are reproducible and well-separated, raising confidence that any hits discovered on that plate are likely to be real. Values between 0 and 0.5 indicate a less reproducible assay and less confidence in the hits. Values below zero indicate that the assay has far too much variability and cannot be relied upon to identify hits.

You should therefore calculate the Z' factor value for every plate measured in the screen, using the normalised absorbance data. Do this by following **Method sheet 104**.

Step 10: Mapping experimental data to plate maps

Now that you have gauged how reproducible the results of the screening experiments were, you should map the normalised data to the respective extract ID and plant name. To do this, insert a new tab in your Excel file (e.g. '8 hour results') and collate all of the results from the preceding 10 tabs for the 8 hour timepoint to yield a table in which you align every normalised absorbance value (i.e. the percentages) to the corresponding extract ID number and plant name, according to the plate maps shown in the second tab of the Excel file. Please take great care with the placement and copying of your data at this stage, as any error in the alignment will prevent the correct identification of “hit” extracts. Once you have aligned all 800 ID numbers with the normalised absorbance at the 8 hour timepoint, you should prepare two scatter plots, one for all the polar extracts (IDs 1-400) and one for all the non-polar extracts (IDs 401-800). Place the extract IDs on the x-axis and the normalised absorbance values on the y-axis. Once these charts are complete, repeat the process for the 21 hour results in a new tab (naming it, e.g., '21 hour results'). For advice on how to complete these tasks, please follow **Method sheet 105**.

Step 11: Identify hits from the primary screen

Now that you have aligned the data from both timepoints against the extract IDs, you are ready to identify the “hits” from the screening experiment. For your project, these will be the extracts that show the greatest potential to inhibit bacterial growth. To do this, insert a new tab into your working Excel file (e.g. ‘8 hour hits’). Now copy the whole table of normalised values, including the extract ID and plant names, from the previous ‘8 hour results’ tab. Paste the values only of this table, with no equations or cell-references, using the Excel function: Paste-special → Values. Now select the whole table of values, including all columns containing data or labels, and use the Excel ‘Sort’ function to order the table by ascending absorbance values. Those extracts that appear near the top of the table will be those in which there was the lowest growth of bacteria. These are the “hits” from this screen.

Take a note of the identities of your top 7 hit extracts, including the extract ID number, the genus and species names, and the common name of the plant. Now repeat the whole process for the 21 hour timepoint data, pasting this data into a ‘21 hour hits’ tab. For advice on how to complete these tasks, please follow **Method sheet 106**.

Step 12: Plot a bar chart of your top hit extract(s)

You should have 3 normalised data points for every plant extract at both timepoints. Prepare a data table in a new tab that collates the data from your top 7 hits only (none of the other extracts) and the DMSO control at the 8 hour timepoint. Plot a bar chart with the plant names as the bar labels, and error bars indicating the Standard Deviation of each set of three experiments, as shown in **Method sheet 107**.

Next, perform a one-way ANOVA with Tukey’s post-hoc test (or preferably a Dunnett’s post-hoc test if available) to determine if bacterial growth was significantly lower than the DMSO control in the presence of any of the top 7 hit extracts. Repeat this process separately for both the 8 hour and 21 hour hits. For advice on how to do this, please follow **Method sheet 108**.

Step 13: Explore correlations within the HTS dataset

Insert a new tab in your Excel file, and name it ‘Correlation analyses’. Paste as values the 8 hour and 21 hour normalised data, alongside the extract IDs, into a new data table. The first analysis you should perform is to test whether there is any significant correlation between the effects of the extracts at the 8 hour and 21 hour timepoints. Align the data so that they are all numbered by extract ID, from 1 to 800, top to bottom. Then pair the 8 hour and 21 hour data in two columns side by side. Prepare a scatter plot from these two columns and insert a linear trendline, checking the option to show the R^2 value on the chart. Take the square root of this value to calculate the coefficient of correlation, r . Then follow the advice given in **Method sheet 109** to test whether the coefficient reaches statistical significance for this particular test.

Next, align the first 400 values from the 21 hour timepoint data, i.e. relating to extracts 1-400 side-by-side with the following 400 values from the 21 hour timepoint data, i.e. relating to extracts 401-800. You should now have two columns of data, in which Extract 1 is paired with Extract 401, and Extract 2 with Extract 402, etc. In this way, the polar extract of each plant is paired with its non-polar extract counterpart. Use these two columns to test if there is any correlation between the growth observed in the presence of the polar and non-polar extracts of the same plants.

Step 14: Repeat the above steps for the other bacteria

Now that you have completed an analysis of the *Phytotitre* screen for inhibitors of *E. coli* growth, you should repeat steps 7 to 14 using the *Phytotitre* dataset for *S. aureus* growth inhibition. Prepare a fresh Excel file for analysis of these new data. A key question to answer for your report will be, is there any correlation between the extent of growth inhibition achieved by the extracts for these different organisms?

Step 15: Analyse the results of a pure compound screen

Natural product libraries typically comprise either of natural extracts, each of which contains many hundreds of different compounds in a complex mixture, or pure compounds, which contain only one well-characterised compound per well. Both types of collection have their advantages and disadvantages, as explained in **Method sheet 100**. So far you have explored the impact of natural extracts on bacterial growth. Now you should progress to exploring the hit-rate for isolated natural compounds.

A library comprising 200 pure natural compounds (the *Puretitre* collection) was also screened to seek inhibitors of growth of *E. coli* and *Micrococcus luteus*. Download the Excel files for these screens and create new working files as described above. Then, repeat steps 7 to 15, making adjustments to the protocol where necessary to account for the slightly different experimental design. We recommend that while you background correct against the 0 hour timepoint as normal, the two timepoints on which to perform your initial analysis should be the 12 hour and 24 hour timepoints, for these particular experiments. Ignore the other timepoints shown, you can return to examine these in later work if you have time at the end of your project. Complete these tasks for both the *E. coli* and *M. luteus* *Puretitre* screens. Is there any correlation between the results obtained for both organisms across the entire collection? After preparing a list of the top 7 hits for each organism, prepare a list of those hits which are not already used as antibiotics clinically.

Step 16: Explore potential correlations with compound structural features

Now you should explore if any significant correlations exist between major structural features of the *Puretitre* compounds and their impact on growth of the *E. coli* or *M. luteus*. To do this, please follow the advice given in **Method sheet 110**.

As you have performed multiple correlation tests for the various compounds, proceed to check whether the p-values you have calculated survive correction for multiple testing, by following the advice given in **Method sheet 111**.

Step 17: Examine dose response assays of individual hits

The next step in drug discovery after completion of a screening experiment is to attempt **replication** of the top hits. This is the process where the highest scoring extracts only, not the full collection, are tested in further experiments to confirm that they have biological activity against the target of interest. These tests aim to validate that the hits are genuine, and not false positives, which may arise from artefacts in the screening process, statistical error arising from variability in the assay, or mis-calculation during analysis of the original dataset.

One of the top non-antibiotic hits from the *Puretitre* screen has been tested for its ability to inhibit the growth of *E. coli* at different doses (this is called a dose-response experiment). Your next task is to analyse these data to calculate the concentration of compound necessary to achieve a 50% inhibition of bacterial growth. This value, referred to as the IC₅₀, is a key metric in the drug discovery process as it gives an indication of how potent your hit extract or compound is. IC₅₀ can be calculated by fitting a 4 parameter logistic (4PL) curve to the dose response data. This can be easily achieved using specialised data analysis tools such as GraphPad Prism or R. However, it can also be calculated using Microsoft Excel, as explained in **Method sheet 112**.

Download the `e.coli_replication.xlsx` file from the Downloads page to also perform a two way ANOVA to test if any differences in cell growth seen between the compound dose curve and the DMSO dose curve are statistically significant, as explained in **Method sheet 25**.

Got this far and still have time left in the analysis phase of your project? Please see the following section for suggestions of optional, additional analyses you can try to extend your project if you'd like to do so.

Writing up: Advice for your project dissertation

Congratulations on the completion of your data analysis! Now comes the writing up phase. Giving as much care and attention to this process as your analytical phase is necessary to help you score the highest possible marks in your dissertation. If you would like some advice on how to begin writing up your project report or dissertation, please refer to our notes in **Method Sheet 113**.



Suggestions for project 'Extension' or upgrading to 'Level 7' (MSc) work

It should be possible to complete the analyses suggested above for the basic antibiotic discovery data project within about **8-10 weeks**. However, if you have managed to complete all of these analyses, and still have more time available, please discuss with your supervisor whether it would be appropriate for you to attempt one or more of the additional analyses shown below to further explore your hits and extend your project.

If you are completing a project at Masters level (i.e. Level 7), or have much more time available, you should aim to complete at least two of the additional analyses shown below. These additional analyses do not have to be performed in any particular order, and you are of course free to modify them with the approval of your supervisor to better fit your own project aims.

Project variation 1 - Explore effects of compounds on rate of growth

The instructions for the basic project described above involve analysis of data from the 12 hour and 24 hour timepoints for the *Puretitre* screens. However, data are also available for the same plates at the 2, 4, 6 and 8 hour timepoints. You could expand your analyses of the *Puretitre* screens by repeating the steps outlined above to explore whether any extracts alter the rate of growth between different timepoints. For example, you could compare the rate of growth between the 0 hour and 6 hour timepoint, with that between the 8 hour and 12 hour timepoints, or any other combination you may find of interest.

Project variation 2 - Test if the hit extracts are toxic to mammalian cells

A key element for the successful clinical use of a new antibiotic drug is that it should be toxic to bacterial cells, but non-toxic to mammalian cells. You could explore this by beginning the first part of Project 06, the data-analysis project for cancer cell killing, to explore whether the extracts you have identified as hits from the *Phytotitre* screen for either micro-organism, are also toxic to four different types of mammalian cell-line.

Project variation 3 - Use AI or machine learning to predict compound activity

Use the smiles structural data, and associated structural feature data, given in the *Puretitre* sdf or Excel files, paired with your calculated bacterial growth values, to train an AI or machine learning tool to predict the likely anti-bacterial properties of any other molecule. Use the data from 140 of the compounds to train the regression model, then test its ability to predict the anti-bacterial activity of the next 60 compounds. How good was the model's prediction? How do you think it could be improved?

Troubleshooting FAQ

Science does not progress without challenges! It is very unusual for a student project to reach completion without facing some difficulties along the way. A key element of your training is learning how to identify problems and apply solutions as you encounter them. The FAQ section below lists some issues that are commonly seen in such projects, and approaches you can try to remedy them.

1) Some of the growth values are greater than 100 % after normalisation of the data - why is this?

Probable Cause: This is commonly seen in natural extract library screens. In this example, while some extracts may inhibit bacterial growth, others might encourage it by acting as a rich source of nutrients to support their metabolism.

Suggested Action: Do not discard or attempt to modify these data, keep them within your dataset and include them as they are in the following analytical steps.

2) Some of the growth values are below zero after normalization - is that possible?

Probable Cause: These values do not mean there was negative growth, it just means the test well had a lower absorbance than the mean of your positive (100% inhibition) controls.

Suggested Action: Values only slightly below zero are common and statistically normal due to minor pipetting variations. Values that are substantially below zero typically arise because of loss of extract colour over time as certain bacteria can metabolise some plant pigments. You should retain the negative values within the dataset for your mean/ANOVA calculations to maintain statistical integrity, but for your final IC50 graphs, you can treat them as “0% growth”.

3) Some of the Z' factor values are very low at the earlier timepoints - why is this?

Probable Cause: There is always much more variability in bacterial growth between wells at the earliest timepoints because the absorbance values are still quite small, so more prone to measurement errors. Also, the cultures often differ substantially between wells in terms of which exit the lag phase to begin the exponential growth phase first, since even minor variations in well temperature can have a large impact on the growth achieved between wells at this early stage.

Suggested Action: As this source of variation diminishes greatly at later timepoints, you can have greater confidence in the results from the timepoints beyond about 8 hours. Report the Z' factor values in your report and point out that any hits identified on plates with very low Z' factor values would be discarded on the basis of excessive variability for that plate. Mention also that as *M. luteus* is a slow growing organism, it may not have reached the plateau phase by the 24 hour timepoint used for the final measurements in these experiments.

4) I can't get the sort tool to work in Excel, the values are still in the same order as before or not sorting properly - why is this?

Probable Cause: Excel can only sort numerical values, not cell references.

Suggested Action: Ensure you paste the data table as 'Values' (Paste special - Values), rather than a simple “Paste”, or Ctrl-V alone before you attempt a sort function. If you paste without this 'Values' distinction, the cell references will be copied and pasted into the destination cells. Click on one of the cells in your table. If it says something like “=AVERAGE(C8:C16)”, or any other cell-reference, you have to delete the table and start again, pasting the table as values only. Once every well has a value (e.g. “0.1158”), then you will be able to sort the values appropriately.

Remember also when sorting to select the whole table, not just individual columns, before sorting. Failing to do this can also result in an incorrect sorting result.

5) Some of my “hits” at the 8 hour timepoint are different from those at the 21-hour timepoint - why is this?

Probable Cause: Metabolism of the inhibitory natural compounds by live bacteria.

Suggested Action: This is a common observation in natural product screening. Some compounds or extracts may slow bacterial growth initially, but some bacterial species are able to degrade these compounds, which means they lose their activity after a certain time. This does not mean that hits found at the early timepoints have no value. Medicinal chemists may be able to alter their structure so that they retain the growth inhibition properties, but become resistant to degradation by the bacterial enzymes. If you make such observations, they will be useful to bring out in your discussion section.

6) My ANOVA gives “significant” results, but the bar chart error bars overlap heavily - is this possible?

Probable Cause: ANOVA tests analyse the variance across the whole experiment, not just between two means.

Suggested Action: Overlapping standard deviation error bars do not always mean “not significant”. If you have used ANOVA with Dunnett’s test (comparing each value to a control) or Tukey’s test (comparing each treatment to every other treatment), it is quite common to find statistically significant differences between treatments even if there is some overlap between the error bars.

7) My IC50 curve is a flat line or looks like a “V” shape - what happened?

Probable Cause: Incorrect data pairing or non-logarithmic scaling.

Suggested Action: First, ensure your x-axis is set to a Logarithmic scale (the data will not plot well on a linear scale). Second, check that you haven't accidentally included the “0 concentration” (Control) point in the log-plot, since zero values cannot be plotted on a log-scale axis. Finally, ensure your concentrations are paired with the correct absorbance values; a “V” shape often means the data was sorted incorrectly.

8) I can't get the 4-parameter logistic (4PL) formula to work in Excel - why is this?

Probable Cause: Poor initial “guesses” for the slope and IC50 values.

Suggested Action: When using Excel to calculate IC50 using a 4PL model, it is necessary to begin by estimating the IC50 and slope values. Try to make sure you make a good first guess with the IC50 value by looking at where the 50% inhibition value lies on your x-axis. Then, try various values for the slope between -1.0 and -3.0. Keep adjusting these until the curve fits reasonably well between the points. After that, adjust the values more finely by adjusting the slope and IC50 estimates in a stepwise fashion until the sum of residuals is as low as you can make it. More advice on these functions are available in **Method sheet 112**.

Project Progress Checklist

The following checklist table will enable you to keep track off the tasks you should aim to complete for the basic project before entering the writing-up phase.

Step	Task	<input checked="" type="checkbox"/>
1	Read the introduction to the <i>Phytotitre</i> library (Method sheet 100)	<input type="checkbox"/>
2	Familiarise yourself with relevant Health and Safety information	<input type="checkbox"/>
3	If necessary, complete a Research Ethics application	<input type="checkbox"/>
4	Make sure the necessary IT resources are in place	<input type="checkbox"/>
5	Plan strategies for Good Laboratory Practice, Data Management and Integrity	<input type="checkbox"/>
6	Download the necessary Excel files and Method sheets (Method sheet 101)	<input type="checkbox"/>
7	Perform background correction of plate data (Method sheet 102)	<input type="checkbox"/>
8	Perform normalisation of the data (Method sheet 103)	<input type="checkbox"/>
9	Calculate Z' factor values for each of your plates (Method sheet 104)	<input type="checkbox"/>
10	Map experimental data to plate maps (Method sheet 105)	<input type="checkbox"/>
11	Identify hits from the primary screen (Method sheet 106)	<input type="checkbox"/>
12	Plot a bar chart of your top hit extracts (Method sheets 107 & 108)	<input type="checkbox"/>
13	Explore correlations within the HTS dataset (Method sheet 109)	<input type="checkbox"/>
14	Repeat the previous steps for the other bacterial species	<input type="checkbox"/>
15	Analyse the results of a pure compound screen	<input type="checkbox"/>
16	Explore correlations with compound structural features (Method sheets 110 & 111)	<input type="checkbox"/>
17	Examine dose response assays of individual hits (Method sheets 108 & 112)	<input type="checkbox"/>
18	Write the dissertation - hints, tips and advice available in Method sheet 113	<input type="checkbox"/>

Found a Bug?

We care greatly about the utility of our resources for customers, and are very responsive to user feedback. If you spot an error, incorrectly numbered method sheet, or a broken link, please feel free to report the issue by sending an email to: contact@caithnessbiotechnologies.com.

Disclaimer

These resources are provided for educational purposes only. The user's University Supervisor remains the Principal Investigator and the sole party responsible for the safe conduct, risk assessment, and ethical oversight of all laboratory work. Caithness Biotechnologies Ltd. accepts no liability for any injury, loss, or damage resulting from the application of the advice or protocols provided herein. Copyright © 2026, Caithness Biotechnologies Ltd. All Rights Reserved.