Method Sheet 114

Advice for writing up - Cancer biology data analysis projects

Overview

Congratulations on the completion of your data collection and analysis! Now comes the writing up phase. Giving as much care and attention to this process as your experimental phase is essential to help you score the highest possible mark for your final report or dissertation. This document offers some generic advice on how to begin writing up your project report or dissertation. Please check with your university supervisor if this advice is suitable for your own, specific project.

Structuring your report or dissertation

Different universities have slightly different requirements for the structure of a BSc or MSc project dissertation. You should follow the guidelines of your own institution. However, a final report at this level typically comprises the following major elements, in the following order:

- **Title page**
- **Abstract**
- **Acknowledgements**
- **Table of Contents**
- **Abbreviations**
- **Introduction**
- **Methods**
- **Results**
- **Discussion**
- **References**
- **(Appendices - optional extra section)**

Advice on word count

The overall word limit for your final project writeup varies greatly between institutions. However, regardless of the overall length, you will still have to decide how much space to give to each section of your report. Suggestions for word counts for each section that might work for this type of project are shown in the table below, broken down by common total word count limits for typical BSc and MSc research project dissertations. These suggestions assume that your word count limit does not include the references list at the end of the document, or the data contained within tables of the results section.

Please note that these figures are just a guide - they are very flexible. Feel free to alter the lengths of each section to meet your own institution's requirements and project-specific needs. Please check with your supervisor if they will be appropriate for your project.

Section	4,000 word limit	8,000 word limit	12,000 word limit
Abstract	200	250	250
1. Introduction	1,600	3,750	6,000
2. Methods	700	1,000	1,500
3. Results	700	1,500	2,000
4. Discussion	800	1,500	2,250

Advice for the Title page

This will be the front page of your dissertation. Here you should give the title to your project, your name (in most institutions) and your student identification number. Keep your title concise but sufficient to explain what you have done. Insert a ‘page break’ just after the text of this page so that the text from the rest of your document does not overlap onto this front page.

Advice for the Acknowledgements section

This can be a short section of several sentences to offer your thanks to those who have helped you in your project. Unlike the rest of the document, it can be quite personal. For example, you can thank your supervisor, friends and family here. Make sure to use ‘Page breaks’ to present this section on a separate page from the rest of the report.

Advice for section numbering and the Table of contents

This adds a layer of professionalism to your report that may help win additional marks. Method sheet 37 shows how to easily insert a table of contents for your report using a Microsoft Word automated function.

We recommend that you number your major sections as follows:

1. Introduction

2. Methods

3. Results

4, Discussion

5. References

You should split each of these major sections into smaller subsections, each covering a specific topic and with their own subheading with appropriate second level numbering. For example: 1.1 Epidemiology of cancer, 1.2 Overview of breast cancer, etc.

Advice for the abstract

The abstract is a standalone summary of your project, typically between 200 and 300 words (check your institution’s guidelines for your specific word limit). It must explain *why* you did the work, *how* you did it, *what* you found, and *why* it matters. The following is a suggested structure for your abstract, we advise you to give one or two sentences for each of the following:

- **Context**: Introduce the "Big Picture", mentioning the scale of the problem with respect to cancer incidence globally.

- **The gap in knowledge**: Explain why your specific project is necessary. For example, you could mention that there is a requirement to discover new chemotherapy drugs, and that natural products could represent a useful resource for this.
- **The aim**: State the objective clearly and concisely, in no more than one sentence.
- **Key results**: Give space only to the most relevant findings, and no space to minutiae. Be concise and quantitative. For example, instead of saying "some extracts worked", say instead, "Three extracts inhibited tumour cell growth by more than 50% at 256 µg/ml, and the most potent extract (#102, *Cinnamomum verum*) yielded an IC_{50} of 25 µg/ml."
- **Conclusion/Impact**: Try to summarise the benefits your findings bring to the field in a single sentence, and perhaps (very briefly) where to go next with the work.

How to improve your abstract:

- **Do** remember to write in the past tense and the passive voice. For example, instead of saying “I will screen 800 samples for anti-cancer properties”, say instead, “400 extracts were screened for their capacity to inhibit growth of HepG2 cells”.
- **Don’t** write the abstract first. You should only write the abstract after you have completed the rest of your report, so you can summarise it accurately.
- **Don’t** make conclusions beyond what is clearly supported by your data. For example, do not “over-reach” by claiming that you have discovered a new chemotherapy drug, or that it will work well in human studies.

Advice for the introduction

A high-scoring introduction for a BSc or MSc dissertation is not just a summary of facts; it is a persuasive argument that explains to the reader a key knowledge gap in the field, and why your project is necessary to address it. Examiners of project reports give the highest marks for the "Inverted Pyramid" structure in an introduction. This is where you begin with a broad overview of the field, and progressively narrow down the focus to eventually reach your specific research area.

We recommend giving space to at least the following topics in your introduction, in the following order:

1.1 Epidemiology of cancer

Give a basic background to how common cancer is today in your own country and globally. Explain very briefly what cancer is and that there are many types of cancer. Mention the extent of resources given to cancer treatment in our hospital today. Support these statements by referring to studies and statistics from reputable sources (e.g. UN or NHS studies).

1.2 Overview of the specific type of cancer you have chosen to focus on

Choose one of the four cell-lines used to create the *Phytotitre* drug discovery dataset - perhaps the one you find most interesting or where you found the most interesting hits - and write a short section on the cancer type associated with that cell-line (reminder: A2780 - ovarian cancer; HepG2 - liver cancer; MCF-7 - breast cancer; PC3 - prostate cancer). You should give some statistics on its prevalence and prognosis post diagnosis. Give some examples of risk factors for the disease and human behaviours that have increased or reduced the incidence of this type of cancer in recent years. Mention the standard approaches to the treatment of this type of cancer, and any particular challenges it may pose in relation to treatment of other types of cancer.

1.3 Mechanisms of cellular transformation

Explain how cells transition from a state of health, in which rate of division is tightly controlled, into one in which the control of division is lost. Mention some of the risk factors that increase the risk of this type of cellular transformation. Briefly discuss the different stages of cancer during development of the disease.

1.4 Current approaches to cancer therapy

Explain how treatment today involves primarily chemotherapy, radiotherapy, surgery and immunotherapy. Discuss the advantages and limitations of each of these approaches.

1.5 Challenges in the development of new chemotherapy drugs

Mention why it may be helpful to discover a new chemotherapy drug to treat the cancer you have chosen to study. Explain how chemotherapy drugs work (mainly by killing cells that divide quickly) and the limitations of that approach. Explain what is meant by the therapeutic window, i.e. the range of concentrations at which the drug will kill tumour cells but not harm slow dividing healthy cells. Mention the properties a good chemotherapy drug would ideally possess.

1.5 Natural products as structural leads for the development of new drugs

Give a brief overview of the historical successes of natural products in drugs used today, including examples of some drugs in use today that derive from a natural origin. Specifically mention those used successfully to treat cancer, for example taxol, vincristine and doxorubicin. Mention also the successes of the NCI Program for Natural Products Discovery in the field of anti-cancer drug discovery and the approach they have taken to drug discovery.

1.6 Advantages and disadvantages of natural product screening in drug discovery

Explain that natural product screening typically offers a higher hit rate than synthetic compound screening, and often yields molecules with better toxicity / ADMET profiles. Discuss also the disadvantages of natural product screening, such as confounding from pigments or viscosity of extracts, and the requirement for activity guided separation to isolate the active compound. Mention that more work is necessary to figure out how to synthesise natural products and that they are therefore more difficult and time-consuming to patent. Explain why big pharma moved away from natural products to synthetic compound discovery in the early 1990s, and the impact that had on drug discovery. Explain here also the key principles of the Nagoya Protocol.

1.7 Hypothesis and aims

Keep this section very short and simple. The hypothesis (that screening a natural product library may reveal inhibitors of tumour cell-line growth) and each of the aims (maximum 3-4) should be one sentence only. The aims should match the overarching objective of each of the experimental approaches you took.

How to improve your introduction:

- **Do** start simple, and then introduce more complex concepts one at a time
- **Do** cite key articles in the field which support the major points of your introduction, inserting these at the end of the sentence where you make the point
- **Do** give preference to citation of more recent articles, and those which report primary research rather than review articles
- **Do** give space to discussing the pros and cons of various schools of thought if there is controversy in the area you aim to study
- **Do** steer the narrative towards key questions that remain unanswered in the field, specifically those that could be addressed by your project
- **Do** include diagrams which explain the key cellular or molecular mechanisms of the areas you intend to study; we suggest ~1 figure per 500 words of introductory text
- **Do** draw your own diagrams for the introduction - copying an image from the internet will score less marks than if you draw something yourself in PowerPoint or Biorender. (If you do copy an image from elsewhere, remember you must cite the source of the image in the figure legend).

Advice for the methods section

The goal of the Methods section is to enable another scientist reading your report to critically assess how you did the work, and if necessary repeat the experiments or analyses you have performed to replicate your processes exactly. High scoring methods sections will be precise, technically accurate and free of minutiae. We recommend structuring your methods section using subsection titles similar to the following:

2.1. Source of primary data

Explain where you obtained the primary data sets for your study, including the URL, specific file names and the date the files were accessed. Briefly explain the types of experiment that were performed to create the primary data sets, and the format of the data as presented in the original download files.

2.2 Background correction and normalisation

Explain how you performed the background correction and normalisation processes, why this was necessary, and which software you used for this process.

2.3 Calculation of Z' factor values

Give the equation used to calculate Z' factor values for each of your plates, and which software you used for this process.

2.4 Hit criteria and selection

Explain the criteria you used to define a "hit" in the primary screen.

2.5 Correlation analyses

Explain which statistical approach you used to perform correlation analyses on the data.

2.6 Four parameter logistic curve fitting and calculation of IC50

Explain how you performed four parameter logistic curve fitting and calculation of IC_{50} values from the dose response data, and which software you used for this process.

2.6 Statistical analyses

It is essential to include a section on how you performed your statistical analyses - missing this out or explaining it poorly greatly hurts the mark. Mention the main data analysis techniques, and which software you used for each different step of the analysis (e.g. Excel, GraphPad Prism or R). State what type of statistical test you used to analyse each different type of data (i.e. for the primary screening data, correlation analyses, dose curve data). State the p-value cutoff chosen for statistical significance.

How to improve your Methods section

- **Do** remember to write in the past tense and the passive voice (e.g. do not say "First I normalise the data to control", say instead, "Growth values were normalised to the vehicle only control").
- **Do** be precise with quantities and concentrations (e.g. do not say "A small amount of extract was added to each well", say instead, "1 µl of extract was added to 99 µl of culture in each well to yield a final extract concentration of 100 µg/ml."
- **Do** be very clear on how you did your statistical analyses.
- **Don't** give space to the minutiae, which means content that is either not directly relevant to your project, or would be obvious to someone with basic competence in the field. For example, there is no requirement to explain how a pipette works or how to prepare a dilution series - just give the final concentration achieved.
- **Don't** repeat the same content more than once.

Advice for the results section

The results section is where you present summaries of the data you collected in your experiments, or your data analyses, presented in a way that is best able to help the reader understand them. This section will include all of your charts and tables from the analyses. In the main text of the results section, you should only describe the main outcomes of your analyses, highlighting only the most relevant findings in the text. Do not give any space to interpretation of what the findings mean in terms of mechanism, relevance to the field, limitations or comparisons to earlier studies in this section. Those comments are necessary, but they belong in the following discussion section and must be put there.

A high scoring results section will present high quality charts and tables with appropriate description of the main findings and summary statistics in the main body of the text between each figure and table.

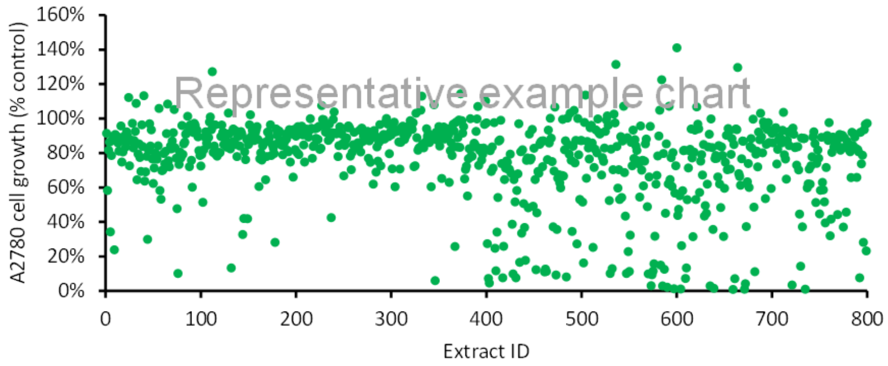
Results to report in this type of project

For a basic *Phytotitre* anti-cancer drug discovery data analysis project, we recommend that you should present the following results in the order shown:

Figure #	Data to report
Figure 3.1	Scatter plot of results from the primary screen of A2780 cells
Table 3.1	List of the top 7 hits for inhibition of growth of A2780 cells
Figure 3.2	Scatter plot of results from the primary screen of HepG2 cells
Table 3.2	List of the top 7 hits for inhibition of growth of HepG2 cells
Figure 3.3	Scatter plot of results from the primary screen of MCF-7 cells
Table 3.3	List of the top 7 hits for inhibition of growth of MCF-7 cells
Figure 3.4	Scatter plot of results from the primary screen of PC3 cells
Table 3.4	List of the top 7 hits for inhibition of growth of PC3 cells
Table 3.5	Z' factor values obtained from the primary screening assays
Figure 3.5	Scatter plot of polar vs non-polar extract activities
Figure 3.6	Scatter plot of correlations between inhibition across different cell-lines
Table 3.6	R-values for pair-wise correlation analyses of inhibition between cell-lines
Table 3.7	P-values for pair-wise correlation analyses of inhibition between cell-lines
Figure 3.7	Bar chart of top hits for inhibition of MCF-7 cell growth
Figure 3.8	Dose response curve of hit extract and DMSO in MCF-7 cells
Figure 3.9	Example of 4PL curve fitting for IC50 calculation
Figure 3.x-y	(*Optional: Data from any additional analyses you performed*)

Results of the primary screen for A2780 cells

Your first chart should present the results of the screen for A2780 cell growth inhibition. This should be a simple scatter plot in which the data are ordered by the Extract ID number, and the chart plots the Extract ID number on the x-axis and the normalised cell viability on the y-axis, with no error bars. This figure should be numbered 3.1, as it is the first figure of your third major section (i.e. the Results section). It should look something like this:



This type of chart is typically favoured in the reporting of drug discovery screening results as it gives an easy visualisation of how variable the responses were across the tested extracts and also the extent of inhibition achieved by the top 'hits' from the screen.

Table of top hits from A2780 cell screen

Present a table of the top 7 hit extracts from the screen for A2780 cells (i.e. those eliciting the lowest normalised growth values). This table should be numbered 3.1 and have the data ordered by greatest extent of inhibition first. Include the extract ID, the type of extract (polar or non-polar), the common and scientific names of the plant, and the percentage of growth measured. For example, your table may look something like this:

Table 3.1: Top hits for inhibition of growth of A2780 cells

ID	Extract type	Common name	Scientific name	% Growth at 24 h
335	Polar	Blood Root	*Sanguinaria canadensis*	0 %
736	Non-polar	Schizandra, Wu Wei Zi	*Schisandra chinensis*	0 %
659	Non-polar	Avocado	*Persea americana*	0.9 %
671	Non-polar	Pepper Powder Black	*Piper nigrum*	1.0 %
735	Non-polar	Blood Root	*Sanguinaria canadensis*	1.0 %
604	Non-polar	Lavender Flowers	*Lavandula angustifolia*	1.1 %
597	Non-polar	Juniper Berries	*Juniperus communis*	1.3 %

Prepare scatter plots and hit tables for the other cell lines

Now repeat the same steps for the other three cell-lines, giving one scatter plot and one table of 7 top hits for the HepG2, MCF-7 and PC3 cell-lines, according to the figure and table numbering shown at the start of this section.

Table of Z' factors for each plate you measured

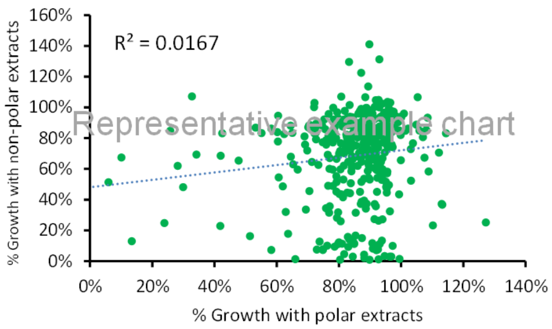
Calculate the Z' factor for all of the plates assayed during the primary screens (see Method sheet 104). Present these values as a table with 5 columns and 11 rows. The top row should begin with 'Stock plate' then 'A2780 cells', 'HepG2 cells', 'MCF-7 cells' and 'PC3 cells'. The left column should list the stock plates that were screened, i.e. Plate 01 through to Plate 10. Type the Z' factor you calculated for each of the 30 plates to three decimal places in the respective remaining boxes. Remember to give a table number, title and footnote. For example, your table might look like this (note that the example numbers shown are not correct, you will have to insert the results of your own calculations):

Table 3.1: Z' factor values obtained from the primary screens for each cell-type

Stock plate	A2780 cells	HepG2 cells	MCF-7 cells	PC3 cells
Plate 01	0.201	0.701	0.501	0.401
Plate 02	0.201	0.701	0.501	0.401
Plate 03	0.201	0.701	0.501	0.401
Plate 04	0.201	0.701	0.501	0.401
Plate 05	0.201	0.701	0.501	0.401
Plate 06	0.201	0.701	0.501	0.401
Plate 07	0.201	0.701	0.501	0.401
Plate 08	0.201	0.701	0.501	0.401
Plate 09	0.201	0.701	0.501	0.401
Plate 10	0.201	0.701	0.501	0.401

Scatter plot of correlation between polar and non-polar extract activity

Working on the data for each of the four cell-lines separately, align the growth values for extracts 1-400 (the polar extracts) with those of extracts 401-800 (the non-polar extracts) in an adjacent column and prepare a scatter plot to reveal whether any correlation exists between the polar and non-polar extracts of the same plants for that cell-line. Remember to insert a linear trendline and show the R^2 value on the chart. Your first chart should look something like this:

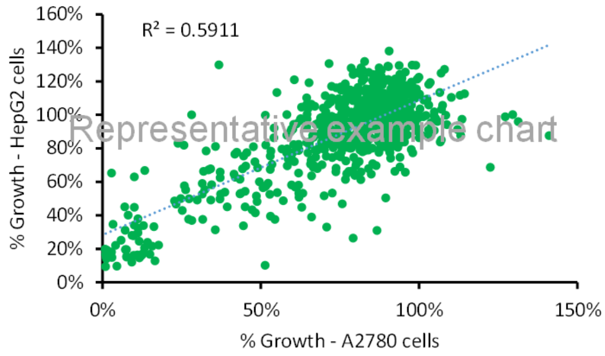


Now prepare the same type of chart for each of the other three cell-lines. Paste them all into the same figure (Figure 3.5) with lettering of 'A' for A2780 cells, 'B' for HepG2 cells, 'C' for MCF-7 cells and 'D' for PC3 cells. Explain the lettering in the legend of the figure.

Calculate the p-value for each of these four Pearson linear regression analyses (see Method Sheet 109) and report both the r-value and the p-value for each cell type in the main text of your results section.

Scatter plot of correlations between inhibition across different cell-lines

Now in a separate Excel file, align the growth values for all 800 extracts in a single column for each of the four different cell-lines side by side. Prepare a scatter plot of the results of all extracts for one cell-line against those of another cell-line. Do this for every possible pairing of two different cell-lines. In other words, you should prepare charts for six different pair-wise correlation analyses: A2780 vs HepG2; A2780 vs MCF-7; A2780 vs PC3; HepG2 vs MCF-7; HepG2 vs PC3 and MCF-7 vs PC3. The first chart should look something like this:



Prepare a separate scatter plot for each of the six pairwise comparisons, and paste them into the same figure, ensuring the entire figure and its legend fits within one page of the report. Label each of the panels A - F, and explain the lettering in the figure legend.

Now use the R^2 value shown on each chart to calculate the 'r' and 'p' values for all six pairwise comparisons (see Method Sheet 109). For example, the correlation analysis of A2780 vs HepG2 cell growth gives an R^2 value of 0.5911, which in turn yields an r-value of 0.769 and a p-value of 4×10^{-157}. Present the results of these analyses in two separate tables, similar to the ones shown below (inserting your own calculated 'r' and 'p'-values where shown):

Table 3.6: r-values for pair-wise correlation analyses of growth inhibition between cell-lines

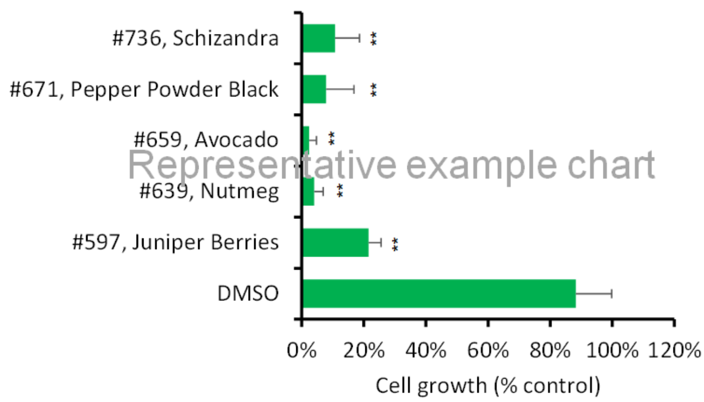
	HepG2	MCF-7	PC3
A2780	0.769	0.xxx	0.xxx
HepG2	-	0.xxx	0.xxx
MCF-7	-	-	0.xxx
PC3	-	-	-

Table 3.7: p-values for pair-wise correlation analyses of growth inhibition between cell-lines

	HepG2	MCF-7	PC3
A2780	4×10^{-157}	$X \times 10^{yyy}$	$X \times 10^{yyy}$
HepG2	-	$X \times 10^{yyy}$	$X \times 10^{yyy}$
MCF-7	-	-	$X \times 10^{yyy}$
PC3	-	-	-

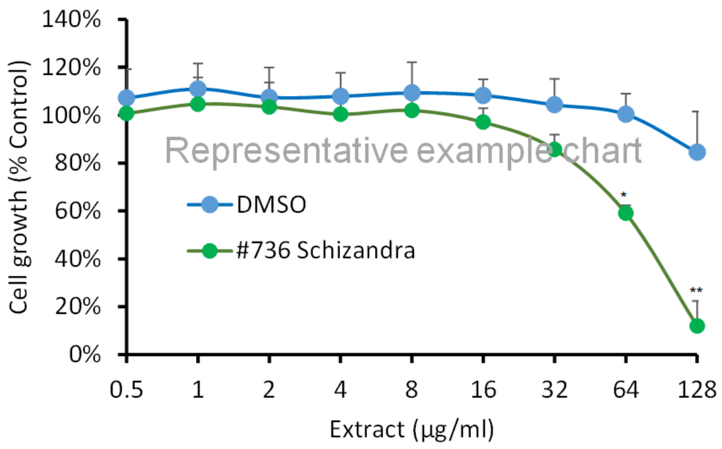
Bar chart of top hits for inhibition of MCF-7 cell growth

Prepare a bar chart of the top 5 hits for MCF-7 cell growth as given in the replication data set for MCF-7 cells (`mcf7_replication.xlsx`). Perform a one-way ANOVA with Tukey's or Dunnett's post-test to compare the response to each extract with the DMSO control and place stars on the chart to indicate significance. It should look like this:



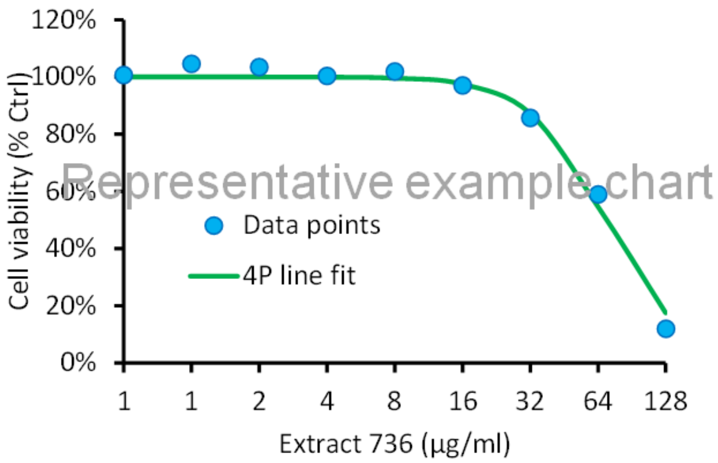
Replication dose curve of a hit for MCF-7 cell growth inhibition

Use the data from the file `mcf7_dose_response.xlsx` to prepare a scatter plot of the dose response curves of DMSO and the hit extract #736, Schizandra with respect to growth of MCF-7 cells. Plot the concentration on the x-axis and the mean normalised growth from the four independent experiments on the y-axis, plus SD error bars in the y-axis direction. Use a log scale with base 2 on the x-axis, but normal (non-log) scale on the y-axis. Remember to plot the results of both the extract and the DMSO control dilution series on the same chart to enable easy visualisation of any dfferences between them. Perform a two-way ANOVA to test whether any significant differences exist between the two curves, and place asterisks (*) above any data point that is significantly higher or lower than the control growth condition (see Method sheet 25). Your chart should look something like this:



Calculation of IC50 values for top hits

Prepare a similar scatter plot to the example given above but with the data for the extract (#736) only. Follow the advice given in Method sheet 112 to add a 4-PL curve fit to the chart and calculate the IC_{50} value for the extract. Show the line fit chart in your result section - it should look something like this:



Don't worry if your line does not reach the x-axis, or the points do not follow the line so clearly as shown this example, but attempt to get the line fit as close as possible to the points in your own chart. Report the IC_{50} values you measured for all of your hit extracts in the main text of your Results section, but only show one example 4PL curve fit chart.

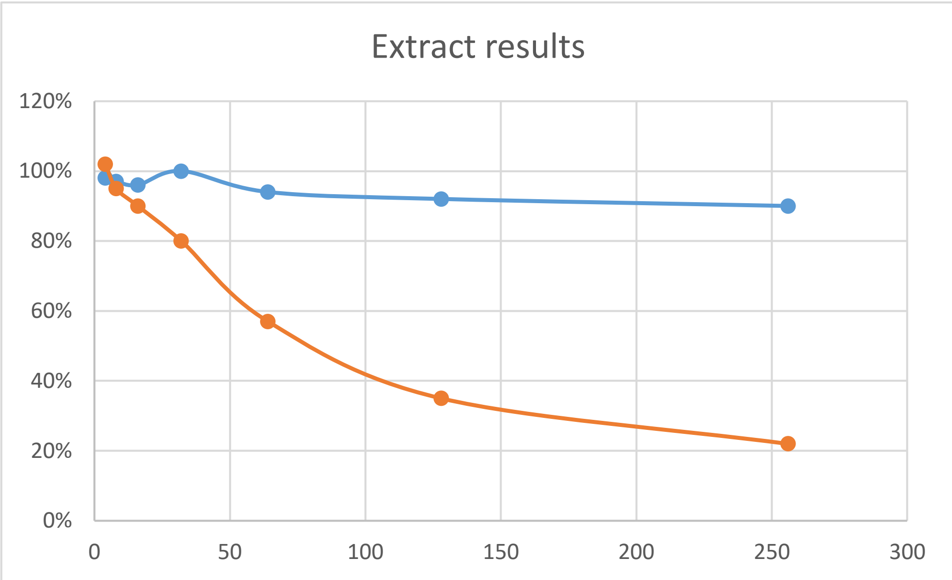
Advice for presentation of results of extended (e.g. MSc) projects

If you have performed additional experiments beyond those suggested in the basic version of the *Phytotitre* project, remember to present these as additional figures following the guidelines shown above.

What makes a good chart?

Good quality charts are essential both for publication of manuscripts in Life Science journals, and for obtaining a high mark in student dissertations. The hallmark of a well-presented chart is one that is clear, uncluttered, and shows the information necessary for easy visualisation of the data. Journals have specific guidelines for chart formatting to meet their requirements for quality. Following these rules in the preparation of your own charts should help lift the mark for your own dissertation significantly.

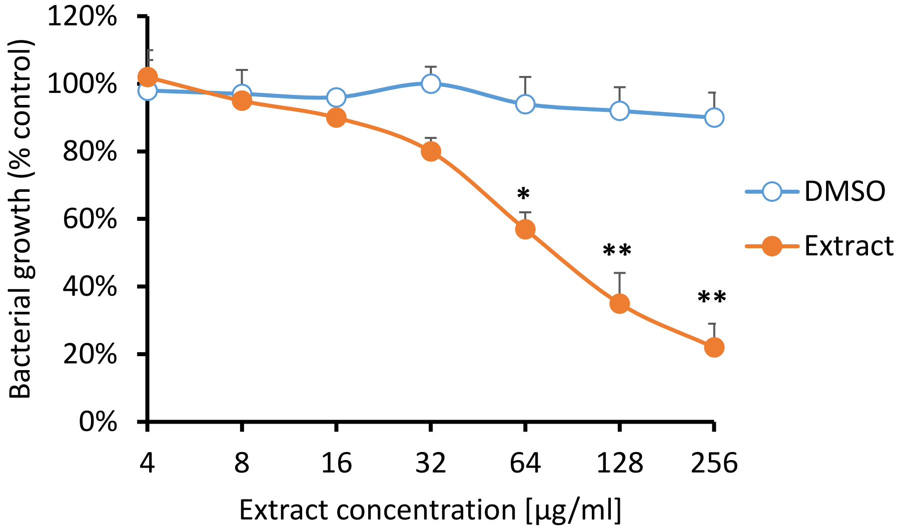
Let's examine what a poorly formatted chart looks like, using a chart prepared with the default settings Microsoft Excel applies to a scatter-plot chart before correct formatting (see below).



Reasons why the above example is a **poorly formatted chart**:

- There are no x-axis or y-axis labels
- There is a title at the top of the chart (Excel inserts these by default)
- There are no error bars to indicate the variability between experiments
- The x-axis is on a linear scale, when it should be on a log scale for dilution curves
- There is no legend on the chart to explain what the different colours of line represent
- There are horizontal and vertical gridlines
- There is a border around the outer perimeter of the chart area
- The font is light grey, a little too small, and difficult to read clearly
- The axis lines are light grey and too thin (insufficient line weight)
- There are no tick marks on the axes
- No asterisks are present to indicate values significantly different from control

Now let's have a look at a well formatted chart plotted from the same data (see example below).



Reasons why the above example is a **well-formatted chart**:

- X-axis and y-axis labels are both present and indicate the correct units
- There is no title at the top of the chart (this will go in the text below your chart)
- Error bars indicating the standard deviation are present
- The x-axis is on a log scale in base 2, which is correct for doubling dilutions
- A legend is present on the chart to explain what the different colours of line represent
- The horizontal and vertical gridlines have been removed
- The border around the outer perimeter has been removed
- The font is solid black, a little larger, and now easier to read
- The axis lines are solid black and 1.5 line weight
- Tick marks are present on both axes
- Asterisks are present to indicate p-values on values significantly different from control

Use the above as a checklist to ensure your charts meet the same level of quality. Your examiners will be looking for all of these things in charts of a high scoring Results section.

What to include in your figure titles and legends

Every figure in your report must have a number, a title and a legend. These must also be formatted correctly to score well. The figure title and legend will appear directly below the chart in your report, not above. They should be in a font or style that is slightly different from the main text to help distinguish the title and legend text from the rest of the document. An example of a poor figure title and legend is shown below:

> Figure: Antibiotic effects of extract
>
> The results are shown in the above chart.

Reasons why this is a poor title and legend:

- There is no figure number
- The title is vague and non-descriptive
- There is no clear explanation of how the experiment was performed
- The legend lacks essential information on experimental replicates (‘n’), statistical tests, error bars and p-value thresholds

Now let’s look at an example of a good figure title and legend:

> <u>Figure 3.2: Dose-dependent inhibition of growth of *Micrococcus luteus* by extract of *Cinnamomum verum*</u>
>
> *M. luteus* cultures were grown for 18 h in the presence of indicated concentrations of *C. verum* extract or equivalent content of vehicle control (DMSO). Growth was measured by absorbance at 600 nm. Means of 3 independent experiments ± SD are shown. Statistical significance was determined using a two-way ANOVA followed by Sidak’s post-hoc test, comparing each extract dose to the corresponding vehicle control (* $p < 0.05$, ** $p < 0.01$).

Reasons why this is a good title and legend:

- There is a figure number (3.1) and a descriptive but concise title that explains clearly what is shown
- The legend gives sufficient explanation of what was done in the experiment so a reader can understand the basics of what is shown without having to refer to the main text
- The identity of the control is made clear
- The number of independent experiments performed to collect the data is shown clearly (it is essential to always report this ‘n’ value in your legends)
- The fact that error bars represent standard deviation is explained
- The type of statistical test used, and the comparator condition, is clearly explained (this too, is essential)
- The meaning of the one star and two star marking on the chart is explained in terms of the p-value thresholds they represent

How to improve your Results section

- **Do** make sure every figure is numbered and cited in the main text by its number
- **Do** check whether your data are parametric or non-parametric before choosing a statistical test to apply
- **Do** explain clearly which statistical test you chose for analysis of each type of data and why
- **Do** present the results in an order that follows a logical progression and tells a story beginning with your hypothesis and following your thought process through each experiment
- **Do** prepare your figures to a high standard as explained in the section above
- **Do** mention any problems you may have seen in your experiments if they may affect the results (e.g. unusual cell morphology, low reproducibility, unexpected results, edge effect etc.), but leave discussion of those issue to the next section of the report
- **Do** show evidence of good reproducibility of experimental work (e.g. small error bars)
- **Do** include appropriate controls in all your experiments (e.g. vehicle alone etc)
- **Do** remember to insert a space between number and unit every time (e.g. 1 µg/ml, not 1µg/ml)
- **Don't** show the raw data - only show the summary data after processing and analysis of data from all experiments of the same type combined
- **Don't** give a chart for each individual experiment - give only a single chart combining all the data from all experiments of the same type combined
- **Don't** show the results from individual experiments - only show the means of multiple repeat experiments post analysis (the exception to this is where you show an image of an agar plate or a microscope field or similar, in which case you should clearly indicate in the figure legend that the image shown is "a representative result from 'n' independent experiments")
- **Don't** give space to description of the minutiae - stick to the main findings and key observations

Advice for the discussion section

Here, the examiner will be assessing your understanding of the project, how it fits in with past work in the field and your understanding of what the results show and mean more broadly. Your results section is where you **describe** the findings. The discussion section is where you **interpret** the findings and **compare** them with previous studies. You should structure your discussion section in the following order:

1) **Brief recap**: Why was your project necessary for the field?
2) **Interpretation**: What do your results mean and can you explain the observations?
3) **Context**: How do your findings compare to previous studies?
4) **Limitations**: What were the weaknesses of your study?
5) **Future Work**: What should be done to take this work forward?
6) **Conclusion**: What is the final "take-home" message?

The brief recap should be at most 1-3 sentences, explaining why your project was necessary. Make sure you link back to your original hypothesis here.

Then try to explain what factors may be responsible for the main findings of your project. Give space particularly to why you think any extracts did, or did not, inhibit the readout chosen for your screen. If you saw any unusual results that are difficult to explain, this is where you will raise the point and offer some possible explanations.

Next, you should discuss what others have reported in terms of the effects of the same herb in previous studies, particularly if those studies look at the same phenotype you did. Mention whether it has been used in the past in traditional medicine, and for what purpose. Give space to discussion of the likely toxicity of the plant - has it been trialled in human or animal studies previously? Mention any active compounds that others have previously reported to be present in the extract, and if these compounds have activity similar to what you have seen.

Then you should discuss the advantages and limitations of your own study. Give space to the pros and cons of natural extract screening. Then mention any specific issues you found in your own project (e.g. the edge effect, contamination issues, etc.).

Insert a paragraph or two on future work. This is where you suggest what direction you would like to take the work forward if you had sufficient funding. Focus specifically on which types of follow-on experiments or development programmes you think would be most valuable for the field.

Finish with a clearly defined ‘Conclusion’ section. This should sum up the broad picture of what your work found and how it contributes to the field. Also mention the potential value to society if you have found some potentially useful bioactivity. But do not over-reach - make sure your claims are entirely supported by your own data. If you speculate that your extract could have therapeutic potential in future, make clear that this is a speculative point and will require further work to validate. Do not introduce any new information or discussion into this closing section - it should be quite brief and mainly a summary of what you have said previously. Some institutions prefer this to be a standalone section after the discussion (e.g. 5. Conclusions) - follow their guidelines if that is the case.

How to improve your Discussion section:

- **Do** cite a wide range of relevant references
- **Do** explain how your work addresses an unanswered problem in the field, and how your approach relates to studies of previous workers

- **Do** give space to discussion of both the advantages and limitations of the experimental approaches you have taken
- **Do** mention the key limitation that your work is *in vitro* only, and that the results may not be applicable to an infection *in vivo*
- **Do** give space to discussion of evidence from earlier studies which support or refute your own observations
- **Do** mention the possibility of activity guided separation as the next step in identifying any active compounds of interest that may be present in your hit extracts
- **Do** explain that progressing to isolation of a natural compound and developing it as a candidate drug in future work will require careful consideration of the Nagoya Protocol
- **Don't** simply repeat the minutiae of the results again, a brief summary of a key finding is fine, but leave most of the numerical results in the results section
- **Don't** be afraid to express your view if your findings challenge existing thought in the field - just be sure to explain clearly why you think you are right and they are wrong
- **Don't** simply ignore any unusual results, or those that are difficult to explain - offering plausible explanations for these can actually win extra marks
- **Don't** be afraid of "negative results" - remember that two projects will score exactly the same mark if they are performed and written up to the same standard, regardless of whether one makes an amazing discovery and the other does not - this is not what you are being assessed on

Advice for the references section

Here, your examiner will assess your grasp of the literature in the field relevant to your project, and your ability to cite previous works correctly. Different institutions prefer different referencing styles. However, most UK universities require you to use either the **Harvard** or **Vancouver** styles. These styles have some key differences you should be aware of. Harvard style involves giving the first author surname and year of publication in the in-text citations (e.g. Smith et al., 2019). Vancouver style replaces this with a simple number, often in square brackets (e.g. [1]). Make sure you format your own references in the main text and in the reference list according to the style favoured by your own institution. These can be inserted and sorted by hand, but using a software application to manage your references will save a lot of time and effort in this process.

Examples of the Harvard and Vancouver referencing styles:

Style	In-text citation	Reference list
Harvard	(Wilson et al., 2020)	Wilson, B.A.P., Thornburg, C.C., Henrich, C.J., Grkovic, T. and O'Keefe, B.R. (2020) 'Creating and screening natural product libraries', Natural Product Reports, 37(7), pp. 893–918.
Vancouver	[1]	Wilson BAP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR. Creating and screening natural product libraries. Nat Prod Rep. 2020;37(7):893-918.

What references should you include?

Cite articles that support the major claims you are making, primarily in the introduction and discussion sections. It is better, where possible, to cite peer-reviewed journals over websites, and primary research articles over review articles. Remember to insert your citations at the end of the same sentence in which you have made a major claim that refers to or requires support from earlier work.

Here are some examples of journal articles that you could read to help you get started with this specific project:

1) Wilson BAP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR. Creating and screening natural product libraries. Nat Prod Rep 37:893-918 (2020)
2) Atanasov AG, Zotchev SB, Dirsch VM; International Natural Product Sciences Taskforce; Supuran CT. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov 20:200-216 (2021)

How to improve your citations and referencing

- **Do** cite previous works that are most relevant to the area of interest, the methods you have chosen, and the overall aim of your project
- **Do** cite studies in reputable, peer-reviewed journals
- **Do** cite mostly primary research articles, with only a handful of review articles
- **Do** place your in text citations at the end of the same sentence where you make a claim or statement, don't save them up to the end of the paragraph or section
- **Do** favour citation of journal articles over textbooks
- **Don't** cite studies that report results that are far beyond the scope of your own project
- **Don't** cite studies from disreputable sources (e.g. unofficial websites, wiki pages etc.)
- **Don't** forget to order your reference list alphabetically if you choose to use Harvard referencing

General advice for your dissertation write-up

- Your writing should be clear, well organised and with excellent grammar and spelling
- Present your work to a high standard in terms of the formatting and presentation of the report (e.g. consistent font style and size for each level of heading, consistent formatting of charts and legends, etc.)
- Use hierarchical numbering for sections and subsection headings (e.g. 1. for Introduction, then 1.1 for the first subheading in your introduction, etc.)
- Use the Microsoft Word headings function to prepare a table of contents page just after your acknowledgements section
- Where possible, draw your own diagrams for explanatory figures in the introduction (e.g. using PowerPoint or BioRender)
- If you do copy an image from elsewhere to use it in your thesis, make sure to cite the source of the image clearly in the figure legend
- Introduce your facts in a logical order that makes sense for the reader
- Sections which jump about randomly between topics without a clear thread score low marks
- Use italics for all genus and species names (e.g. *Escherichia coli*)
- Remember to also use italics for all other Latin terms, such as *et al*., *in vitro* and *in vivo*
- For abbreviations, give the full term at first use explaining the acronym in brackets straight after, then use the acronym on every occasion thereafter (e.g. polymerase chain reaction (PCR) at first use, then simply PCR at every point in the document after that)
- Remember to include the abbreviations you have used in the list near the start of your report
- Unlike the Introduction, Methods, Results and Discussion sections, which are all numbered (1-4), the Abstract, Acknowledgements and References sections do not receive a number
- Use the ‘Page break’ function to ensure every figure is on the same page as its title and legend (which should be placed just beneath it)
- Also insert a ‘Page break’ just before each of the major sections to prevent cluttering