UNIVERSITY OF CAMBRIDGE | Yusuf Hamied Department of Chemistry

Successful Completion of a Research Degree: Report & Thesis Writing

2023-24

Deborah Longbottom, Aruna Reddy, David Morgan & Aleks Reinhardt

Table of Contents

Key contacts

Dr Nick Bampos - nb10013@cam.ac.uk
Deputy Head of Department

Dr Howard Jones - hrnj1@cam.ac.uk
Academic Secretary

Prof. Rosana Collepardo-Guevara - rc597@cam.ac.uk
Director of Postgraduate Education

Prof. Deborah Longbottom - dal28@cam.ac.uk
Head of Postgraduate Education

Dr David Morgan - PGEducationManager@ch.cam.ac.uk
Postgraduate Admissions & Education Manager

Tessa Blackman - PG.Student.Admin@ch.cam.ac.uk
Postgraduate Student Co-ordinator

Olivia Hamlet - postgraduate.education@ch.cam.ac.uk
Postgraduate Education Administrator

Preface

The following document is very subjective and you and your supervisor may disagree with what is written here. This is totally fine, but we ask that you please make the sincere effort to consider the points mentioned in this document.

1.0 Rules and Other Useful Information

All the information you need is provided on the postgraduate student web pages, whether you are studying for a PhD or an MPhil. The links to the guidelines are as follows:

First-year probationary review pathway

PhD pathway

MPhil pathway

Should you wish to find information about the University policies regarding plagiarism, see https://www.plagiarism.admin.cam.ac.uk/

2.0 Preparing to Produce a Report or Thesis

It is very hard to write an entire report/thesis beginning with a blank page. Below are some ideas which you may find helpful in preparing your report/thesis as you go along.

- When you read a journal publication, do not only assess the science: take a moment to look at the format and style, and perhaps ask yourself if this is how you would have presented the results. Become a 'literary critic' and take notes. Write a summary of what you have read.

- Look at other reports/theses in your group and try to understand what makes a good piece of scientific writing.

- Rather than leaving the writing until the final deadline, write up your results and experiments into a 'fantasy paper' every so often to help you question what you are doing and identify any gaps. If you are lucky, you might just create something to send to your supervisor for actual publication.

- Little things like title pages, acknowledgements, list of abbreviations, etc. take time to generate, so when you have a spare 15 minutes, use the time to get them done and out of the way.

- Give anything you write to someone you trust to give an honest opinion. You may not agree with everything your colleagues say, but everybody can benefit enormously from some constructive criticism. Be open to others' ideas.

- When you feel you have something to write (which need not be a complete project), start a draft chapter of your thesis or report: accumulate text to finesse and edit at a later date. Further down the line, it will be easier to remove text from something you have, rather than staring at a blank screen/sheet with the aspiration of churning out chapters just before the deadline. Again, give this to someone to read.

- Once every three months or so, collect what you have written and spend an afternoon thinking carefully about your results, any pictures you need to generate, and the key references.

3.0 Beginning to Write in Earnest: Stylistic Issues and Settings

There are some stylistic choices that you should decide early on to avoid having to correct or to make uniform later.

- Your group may have a preference for the drawing and text settings you should use, so always check before you begin. If not, RSC or ACS settings and style are good for drawings and text. Whatever you go with, make sure it is consistent throughout your report, i.e. decide what to use before you begin and use it throughout.

- Decide on heading formats and the style of referencing and any abbreviations. It helps in the long run if you learn to use the appropriate tools in the software you use to produce your document to apply a uniform 'named' style that, should you need to alter it later, you need only change in one place and it will apply throughout your document.

- In organic and inorganic chemistry, it is usual practice to write in the past tense and the passive voice. If you read out aloud what you have written, make sure it sounds like something you would like to be said. For example:

✗ I mixed the reaction for six hours at 60 °C. It changes colour from yellow to red. I then let it stir for twelve more hours at room temperature.

✓ The reaction was stirred for six hours at 60 °C, during which the colour changed from yellow to red, and then left to stir for a further twelve hours at room temperature.

However, in physical and theoretical chemistry, it is more usual to use the active voice, generally in the first-person plural, and often the present tense for statements which are universally true.

4.0 Using Figures, Schemes and Tables Appropriately

Figures, schemes and tables all have clear designations and must be referred to in the text to connect the discussion.

- Figures are 'pictures' which typically have captions and the reader should be able to look at these and know what they are about without having had an opportunity to read the details in the text. Figures do not normally contain reaction arrows. Recent feedback from examiners suggests that figure captions are something students should pay more attention to so that figures can be read and interpreted without reference to the text.

- If a figure is in fact a graph or other diagram, it may also have a legend (or key) that indicates, for example, which line or symbol corresponds to which property. For example, in a histogram, you might have something like ' compound 1; compound 2'; in a graph with different data points, you might have something like ' product yield; measured viscosity', depending of course on what you are plotting. A legend can be included as part of the graph or inline as part of a caption, but be consistent in your approach to presenting figures. An example of both approaches is shown below.

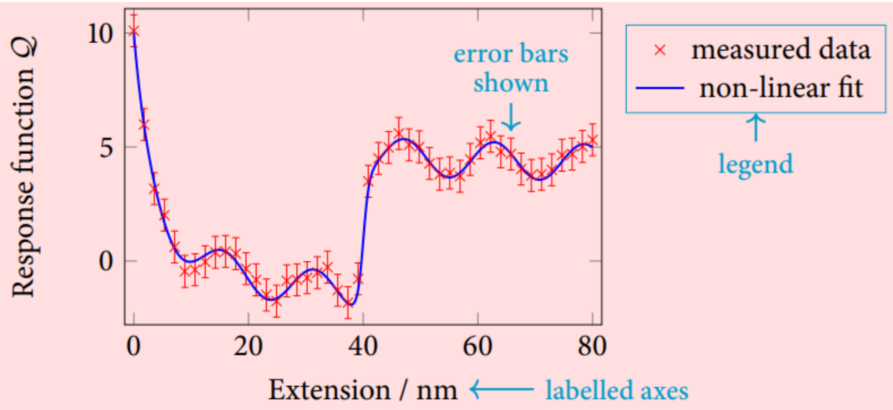

Figure 1: In some journals, figures have a title like this. A description of your figure should follow. This should be concise, but it should explain what you are showing in your figure, including any points of interest and, if appropriate, what the conditions of the experiment were. If your figure includes multiple panels, the caption should contain information about each one. If you prefer, you can include the legend inside the caption, e.g. by saying that measured data points are represented by markers (×) and a non-linear fit by the solid line (——).

figure caption

- It may also be worth noting that it may be clearer to your readers, particularly if they have printed your document in greyscale or if they cannot distinguish between the colours you have chosen, if you label lines or regions of your diagrams directly next to the data, rather than through a separate legend. Having to jump back and forth between the data and the legend and having to decipher a key add (often needlessly) to a reader's workload.

- Schemes are a representation of a synthetic/experimental sequence. They usually contain a reaction arrow and depict going from 'A' to 'B'. They need not have a caption. You may choose to place the 'Reagents and conditions' and the yield below the scheme (as shown in the example below) or on the arrow. Either is fine, but decide before you begin which you are going to do and do it uniformly.

120 **122**

Scheme 31

Reagents and conditions: CuI, HI (57 % aq), 130 °C, 30 min (79 %).

- Tables typically come with a caption.

- In scientific writing, figures and tables generally appear at the top or the bottom of a page or column of text. It is not usual to wrap text around figures or tables.

- **Do not** present figures, schemes etc. that are not referred to in the text, or refer to figures, schemes etc. that do not actually exist in the document.

- **Do** keep a sensible order in the numbering of compounds, figures etc. as the results are presented in the text. It is usually a good idea to number everything (i.e. not just the compounds you refer to in the text): even generic structures containing 'R' groups can be given a number.

- Numbering of compounds is typically one of the final things you do when writing a report/thesis, as it can be time consuming and you only want to do it once!

- You can automatically number molecules, sequentially, in reaction schemes using ChemDraw. Search 'Autonumbering Reaction Schemes' in the ChemDraw user guide for more info. If you use LaTeX and draw compounds with ChemDraw, you can number them automatically using the chemnum package.

5.0 Typesetting Units, Variables and Equations

You should aim to follow standard typographic practice in your documents. When in doubt, chemists can always consult the IUPAC Green Book (Quantities, Units and Symbols in Physical Chemistry), which is based on the ISO 80000 standard, which outlines the conventions normally used in scientific typesetting. The IUPAP standard for typesetting in physics is exactly the same, but somewhat less easy to read.

Here are a few points that are worth emphasising. You can then delve deeper into the guidelines if you feel you need to.

- **Numbers** should be typeset upright (including π, Euler's number e and the imaginary unit i).
- **Variables** and **generic functions** should be *italic* (including capital Greek letters).
- **Operators** and **named functions** should be upright (e.g. cos, sin, log, but also the differential operator d, in e.g. '$\mathrm{d}x$' at the end of an integral, and 'named' functions such as the Dirac delta function δ or the Heaviside step function Θ).
- **Vectors** and **matrices** should be ***bold italic*** (and not just bold), although this is often not how they are typeset in journals in practice.
- **Units** should be upright and preceded by a non-breaking (half) space (see subsection 6.2).
- **Chemical formulae** should be upright (e.g. $\mathrm{N_2(g)}$ or $\mathrm{Cl^-}$ (aq)).
- **Any other text** should be upright, e.g. for a rate constant k_{off}, the k should be italic, but the subscript 'off' should be upright. It is easy in LaTeX to make subscripts italic in maths mode and this is what people often do in practice. However, this makes *o* and *f* look like variables and can fundamentally change the meaning, particularly if it is a single-symbol subscript. For example, G_{i} could be the 'initial' Gibbs energy, whilst G_i could be the Gibbs energy of species *i*, depending on whether the *i* is upright or italic.

Of course, these are not strict rules. Many people flout some of these conventions, sometimes on purpose (which is fair enough), but often through ignorance, simply because they do not know the conventions or because they do not wish to learn how to make their typesetting software do the right thing. Remember that your typesetting software cannot know what you mean when you write something.

There are reasons behind these conventions, so even though they are ultimately largely arbitrary, they are self-consistent and have been around for such a long time that people are used to them. If you follow conventions, this therefore often helps the reader to understand intuitively what sort of quantities and operations they are looking at without necessarily reading all of your document. It is thus important to follow conventions to ease comprehension and minimise any possible confusion that might arise, unless you have legitimate reasons not to.

5.1 Writing Equations

Displayed equations (equations that are shown in their own line) should be numbered if they are important or if you refer to them later in the text. They should be treated as any other part of the sentence, and so should be punctuated accordingly (e.g. with commas or full stops – not colons!). If the sentence then continues, particularly with a relative clause that is non-restrictive [e.g. 'where k_{off} is the rate of dissociation' or some other similar kind of explanation of what the symbols in the equation represent], add a comma at the end of the equation if appropriate. Avoid starting such relative clauses with a capital letter when they are part of the preceding sentence.

For example, we might write the relation

$$\cos^2 x + \sin^2 x = 1 \qquad (1)$$

as a displayed equation. If it came at the end of a sentence, however, we might wish to write it as:

$$\cos^2 x + \sin^2 x = 1. \qquad (2)$$

Notice the presence of a full stop at the end of equation (2).

To give an example of typesetting conventions, we might write

$$I = \int_{-5}^{5} \frac{\log(ax) + 2\cos(x + \mathrm{\pi})}{\mathrm{e}^{3x}} \,\mathrm{d}x, \qquad (3)$$

where numbers (2, 3, 5, π and e) and operators (log, cos, d) have been typeset in upright style, and variables and functions (I, x and a) in italic style. Similarly, when we write about vectors, we could write

$$\boldsymbol{a} = 5\boldsymbol{i} + 6\boldsymbol{j}. \qquad (4)$$

Vector operators are often typeset in (upright) boldface, such as

$$f = -grad\ U = -\nabla U \qquad (5)$$

The laplacian, by contrast, is written as ∇^2, since it is a scalar.

If you use delimiters, make sure they are appropriately scaled and spaced. For example, we might write

$$\mathrm{artanh}\left[\frac{a + b}{c}\right], \qquad (6)$$

rather than

$$\operatorname{artanh}[\frac{a+b}{c}], \quad \text{or} \quad \operatorname{artanh}\left[\frac{a+b}{c}\right], \tag{7}$$

where the delimiters are too small in the first, and the spacing before the brackets is too large in the second example.

If you use LaTeX to typeset your documents, you may some of the following packages helpful:

- siunitx can be used to typeset physical quantities with units.
- mhchem can be used to typeset chemical formulae.
- mleftright can be used to ensure the correct spacing before large delimiters following operators.
- fixmath can be used to give you access to bold italic maths fonts (using the 'mathbold' command), depending on the fonts you use.
- mathastext can be used to give you more consistency between maths and text fonts.
- upgreek with the slantedGreek option can be used to give you access to upright lowercase Greek letters and italic uppercase Greek letters, although this is not necessary with all fonts.
- microtype can help you to make text appear visually more balanced.

5.2 Quantity Calculus

A physical quantity is a product of a numerical part and a unit. For example, we might say that the heat capacity of water at room temperature and pressure is C_P = 4.18 kJ kg^{-1} K^{-1}. The numerical value here is 4.18, whilst the unit is kJ kg^{-1} K^{-1}. However, we could also express C_P as

$$C_P = 4.18 \text{ kJ kg}^{-1}\text{ K}^{-1} = 4.18 \text{ J g}^{-1}\text{ K}^{-1} = 1.00 \text{ kcal kg}^{-1}\text{ K}^{-1}. \tag{8}$$

It is unhelpful to think of units as something you simply add on to your final result at the end of a calculation; this is very prone to errors. The numerical value of a physical quantity will vary depending on what units we use, so just writing an energy of E = 4.18 without stating the units is completely meaningless. An energy of 4.18 kJ can equally be expressed as 1.00 kcal, 2.61 x 10^{22} eV, 1.16 kWh, 4.18 x 10^{10} erg or 9.59 x 10^{20} E_h. Units are clearly crucial to include. Including units in your calculations also allows you to use the power of dimensional analysis to check whether your answer is dimensionally consistent, and so can allow you to spot mistakes very readily.

Once we begin to think of a physical quantity as being a product of a numerical value and a unit, dealing with units becomes straightforward. For example, rather than write E = 4.18 kJ, we might write E/kJ = 4.18, i.e. we have divided both sides of the equation by the unit, kJ. This is convenient because we can only apply mathematical functions to numbers, not units. Expressions like ln E are rather meaningless, because what *is* the natural logarithm of (say) a kilojoule? By contrast, we can easily calculate ln(E/kJ), since E/kJ is, as we have seen, a dimensionless number.

Within the framework of quantity calculus, it is not correct to say that 'the energy E is measured in kilojoules' and then treat E as a dimensionless quantity. What this phrase really means is that we have defined a reduced quantity $E^\star \equiv E$/kJ, which is dimensionless. However, the energy itself is a physical quantity and it can be measured in any energy unit we wish; for example, if we plug in E = 500.0 cal into our expression for $E^\star$, we get $E^\star$ = 500.0 cal/kJ. Since 1 kJ 239.0 cal, we have

$$E^\star = \frac{500.0\ cal}{239.0\ cal} = 2.092, \tag{9}$$

which is still dimensionless. The formalism of this argument is helpful because it makes unit conversions much clearer.

We often wish to plot or tabulate experimental data; using quantity calculus, this again makes intuitive sense. We can only plot pure numbers, which we can obtain by dividing the physical quantity by some dimensionally correct choice of unit. This is why you will typically see things like T/K on axis labels: this means that we have divided the temperature T – a physical quantity – by the unit kelvin to obtain a dimensionless number that we can plot. Alternative conventions are commonly used (e.g. 'T (K) ' to signify that the temperature is expressed in kelvin), but they are arguably incorrect in the framework of quantity calculus.

Of course, one should be pragmatic about these things, and quantity calculus may not always be appropriate; however, it is very helpful to understand the underlying ideas even if you end up using something like atomic units in your work.

Some further examples of quantity calculus can be found in Chapter 7 of the Green Book.

6.0 Writing up the experimental section

Experimental write-ups are tedious and time-consuming, so once you have characterised a compound, document the details while the information is fresh in your mind. This will also help you to identify what characterisation is missing. Use all the information in 'Successful Completion of a Research Degree: Guidelines for Experimentalists' to make sure that you are doing this as fully and correctly as required.

7.0 References

Whatever style of reference listings you adopt, be consistent throughout. For example, you could use the format given below, but don't mix your styles and do abbreviate the journal title.

30. J. L. Sessler, **Brucker E.A.**, V. Lynch, M. Choe, S. Sorey and E. Vogel, *Chemistry – A European Journal* 1996, **2**, 1527–1532.

30. J. L. Sessler, E. A. Brucker, V. Lynch, M. Choe, S. Sorey and E. Vogel, *Chem. Eur. J.* 1996, **2**, 1527–1532.

It is becoming increasingly more common to include full article titles in bibliographies, which can help your reader to decide whether a given reference is worth following up on.

You can find standardised journal abbreviations in the chemical sciences (very broadly interpreted) using the CASSI search tool, available at cassi.cas.org.

You will almost certainly want to use some kind of reference management software/suite to help you to deal with references in a consistent style.

- Mendeley
- Zotero
- EndNote
- BibTeX (Brief intro included in our local training course each year)