Efficient manuscript writing in six steps

Stefan T. Arold

King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Biological and Environmental Science and Engineering (BESE), Thuwal, 23955-6900, Saudi Arabia.

Contact: stefan.arold@kaust.edu.sa

INTRODUCTION:

When it comes to manuscript writing, many students and junior scientists tend to face the same difficulties and make the same mistakes because they lack a structured approach and basic guidelines. This situation leads to a highly inefficient and frustrating process, where students spend much time producing an initial draft that has fundamental structural and stylistic flaws. The mentor then either needs to spend an equally excessive amount of time explaining all the mistakes to help the student correct them; or the mentor resolves to make the corrections him/herself, which is also time-consuming and provides only little learning experience for the student.

Therefore, I propose a structured approach that breaks down the writing process in six simple steps. Mentor feedback is integrated into this process to catch structural and fundamental problems before they can propagate through the manuscript. In case there is no mentor, a friend or a fresh 'next-day' mind might be used instead. Through the combination of structure and feedback, this approach should allow completing the manuscript swiftly and efficiently, with minimal efforts for both junior scientist and mentor.

Of course, there are many other excellent resources for manuscript preparation. In particular, I recommend 'How to write and publish a scientific paper' by R.A. Day and 'The grant application writer's workbook' by S.W. Russell and D.C. Morrison. However, these are (rightfully) book-format resources, and many students or scientists might be (wrongly) reluctant to commit to a long read. Other, shorter resources typically only provide rules for style and content but do not help scientists in structuring their writing. Such rules are essential, of course, and I have added a section compiling those that I found most helpful, or most often breached.

I wrote the current version of this six-step guide during the COVID-19 lockdown, where manuscript writing was one of the few options left for many scientists. Nonetheless, this guide is the result of more than two decades of work in multidisciplinary and international environments, with much time spent on reflecting on how to best write and correct scientific manuscripts.

I initially conceived this six-step guide to help students writing thesis-type documents. Yet, the same approach can be used for a journal publication or grant proposal. I also recommend following the same steps for preparing seminars. In fact, prior oral presentations (and the feedback obtained) are often of great help for preparing a manuscript.

Please read through the whole document carefully before you start with the manuscript.

SECTION I: SIX-STEP-GUIDE

FIRST:

Don't write anything, but gather all information on the length, format, scope and readers of your manuscript. This is particularly important for grants (an out-of-scope grant is already rejected). What do your readers/reviewers know, look for, are interested in? What sections do you need?

SECOND:

Take a pen and a piece of paper (or electronic versions thereof). Decide on the rough volume of each of the required sections. Decide on how many pages for introduction, for methods, for results, for discussion, given the total volume you aim for. What is the total amount of pages you want to have as text? 3-12 (journal publications) or 40 (MS thesis) or ≥100 (PhD)? Then spend time to assign numbers of pages to each chapter, section and subsection (i.e. in Methods, these might be 'crystallisation', 'MST'). It might be good to take five minutes to discuss your conclusions up to now with your mentor.

THIRD:

Continue with your piece of (electronic) paper. Collect all the experiments you have. You may also choose to include experiments collected in our group by colleagues; indeed most scientific work is a team effort, and needs to be presented in this context. Of course it is important that you clearly identify (upfront and throughout) what you have done and what has been contributed by others.

Now, looking at all the experiments available, select the ones you want to show in your manuscript and/or presentation. There is no need to show all experiments you have compiled (a PhD thesis is often only reporting 20-40 % of what you have done). But you need to show all those that support your claims. To make this decision it is best to assemble the experiments into groups that support the same claim/statement (i.e. 'dimerization enables DNA binding of protein X').

B.t.w. If you are also preparing an oral presentation, then having this conclusion as a header is much better than, say, 'biophysical analysis of protein X'.

Take these groups then as pillars on which you develop your story (next point). But before proceeding to the next step, discuss and check your choices with your mentor.

FOURTH:

Now you need to *outline* your way of presenting your data. You can change the timeline of your work if this makes the story more fluid, clear and logical (you can say "given result A, we next wanted to test B" even though you might have performed experiments B before A). Of course, you cannot, never, change the data.

With this in mind start writing down bullet points.

Start with the results section first. For a thesis, this Results section typically has several 'chapters', each with a particular research focus, and sub-chapters. For a manuscript the Results typically only have paragraphs with different headings.

It is best not to write 'biophysical analysis of protein A' as a bullet point, but rather the take-home message 'Protein A is dimeric in solution'.

Lay out the whole story as a sequence of 'take-home message' bullet points. By reading the bullet points only, you should be able to get an abstract of the full story. Particularly in thesis manuscripts, you need to find a balance between giving a feel for the difficulties (through mentioning failed trials) and keeping the reader engaged (through presentation of new data). Don't show your single most successful experiment first and then just add other failed or inconclusive ones. But also don't hit the reviewer with 30 pages of failed experiments before the first interesting result is shown.

Build up the story starting with experiments that lay the fundament of your work. These experiments should be solid and completed. If you have experiments that are exciting but may still need some additional future work, then put them at the end of a thesis.

Always put yourself into the position of the reader/reviewer. Since this might not be easy, you absolutely need to show your bullet point outline to a friend and see his/her reaction. Is all clear? Then show it to your mentor.

Only once the Result bullet points are established, you should to do the same with the Introduction and Discussion/Conclusions.

For the Introduction, you should not simply pool all information that you think is interesting in the field, but provide exactly and only the information that is needed for your reader/reviewer to understand and appreciate (!) your work. This step also includes highlighting gaps of knowledge or controversies (which you would have addressed/clarified in your work). This part should be formulated in a 'funnel' approach, i.e. you start with the most general information and then proceed towards more and more specific details. For large and diverse documents (thesis, grant proposal), you can also choose to have a general introduction, and then individual more specific sub-introductions for each chapter/aim. Although the Introduction in scientific publications can end by giving a high-level overview of the key results, an Introduction should normally not contain your current results where it describes the context of your study.

The Discussion/Conclusion is often the most difficult and neglected part of a thesis, because it requires deep reflections and understanding of both your advances and shortcomings. But that's also where outstanding students/researchers distinguish themselves from others. The role of a good Discussion is to critically (!) evaluate your work. Where are your results certain, uncertain, leave space for different interpretations? How do they fit in with published/established knowledge? Note that you can adopt a structure where you combine Results and Discussion (and then only finish with Conclusion). But a Discussion should not simply be a summary of the results. However, a Conclusion can be a summary of your achievements (aimed at the reader who has read your manuscript—so who knows more than when s/he read the Abstract). A Conclusion can also contain suggestions for where future works should aim at.

For theses, write the 'Objectives/Aims of the work' last. Of course you had a particular aim when you started the thesis, but often parts didn't work out, whereas others gave unexpectedly interesting results. Tailor your Objectives to what you have achieved (but you can also admit defeat in some points). For example, if your objective was to produce the high-resolution cryo-

EM structure of complex A, but you finally only managed to produce a model, combining X-ray, NMR, SAXS, modelling and biophysical experiments, then your Objective would be 'structural and biophysical characterisation of complex X' and not 'high-resolution cryo-EM structure of complex X'. Better even would be to aim at 'understanding the molecular mechanism for how complex X does Y'.

The Objective should be mirrored in the Conclusion. So you may choose to write the Objectives before the Conclusions. But don't forget that for now, you only write bullet points!

However, it is good to formulate the Objectives as bullet points even in the final thesis. In this way you can go back to these bullet points and discuss them in the 'Discussion' section (How well have you addressed each Objective?).

For the Methods section you can also start compiling which methods you show in your writing.

FIFTH:

Now write a few key sentences under each bullet point. These should state the heart of the data/message that would need to appear.

For the Results section, you should now group your images so that they provide a multi-panel image that (ideally) supports one particular point. E.g. 'Protein A dimerises in solution'. In most cases this is better than grouping figures according to methods (e.g. 'SAXS data on all-my-proteins'). Even in a thesis, it is normally better to group figures, rather than displaying a figure, have a bit of text, a figure, a bit of (probably very repetitive) text. Other data (Kd's, SAXS) can be also compiled and presented as Table. This is also a good moment to read about 'significant figures', unless you know exactly what this is.

Figure legends should be as short as possible so that the reviewer/reader can understand and evaluate the data presented. Again, you need to put yourself into the place of the reviewer for this. If values are presented in a Table, you don't need to put them in the text. For example it is not good practice to write: 'protein A binds to protein B with a Kd of 2.34 ± 0.4 µM, and protein A binds to mutant protein Bmut with a Kd of 0.19 ±0.2 µM (Table 1).' Rather write: "the mutation in B increased its affinity for A more than tenfold (Table 1)." Which is also a good key sentence to be put at this stage. *Remember, you still don't write the full text*.

Do the same for the introduction. You can now also start looking into the figures you best need to illustrate the introduction.

After this stage, again, grab your friend and go through it once more. Then show your mentor.

SIXTH:

Now start to flesh-out your key sentences, and write full paragraphs and text. Since everything is logically laid out, you can write what suits you best at a given moment. See how you feel. Sometimes, when you are tired, or when you have a writer's block, it's easier to write figure legends of Methods. When you are energetic and clear-minded, write the Discussion or Introduction, or acknowledgements. Write the Abstract at the end.

I found it extremely helpful to avoid thinking that your aim is to "finish my thesis"; rather, do everything with the idea of 'best explaining to a good friend what you did". Don't just want to finish this manuscript. You need to want to explain your work.

For writing the convention is as follows: Published observations/facts are presented in present tense: e.g. "FAK mutants lacking the auto-phosphorylation site result in embryonic lethality (XX and YY et al 2009)". Your own unpublished data however need to be given in the past tense. "FAK autophosphorylation was significantly enhanced by addition of 10 mM zinc sulfate (Figure 14)". If you combine both, still keep the tenses: "ADH-D14 was stable for 4h, whereas previously reported ADHs are only stable for 2h (XX et al. 2007)".

Before you start, please carefully read the following tips and rules for writing and making figures. And once you finished writing, but before you give the document to your friend or mentor for final proof-reading, go through the tips once more and correct your work where needed. This saves a lot of time and frustration, because proof-readers can concentrate on what matters.

SECTION II: RULES FOR STYLE AND CONTENT

TIPS FOR MAKING FIGURES:

- Try to be consistent with colours: if you have 3D crystal structures and schematic drawings of your FAT protein domain, then always choose the same colour throughout. If you have a similar version for the same domain, then you can choose similar colours, i.e. magenta for human FAT, pink for rat FAT.

- As much as possible, always keep your 3D protein structure in the same orientation. Often there are 'consensus' orientations for proteins in the field (e.g. for kinases, it is the C-terminal lobe on bottom and N-terminal lobe up). If you need, then do 90 degree rotations around vertical or horizontal axes.

- If you make your figures for a manuscript, then it is good to already choose your PPT page layout in this format (portrait, not wide-screen)

- Make the labels big enough to be read even when the figure is scaled down to fit the available space. You can also colour-code the figure labels to fit the colour of the labelled entity (e.g. protein domain). If the domain is in light colours, you can choose a slightly darker version for the label for better visibility. You can have the labels either directly next to the feature (e.g. Asp235) or, if this gets too crowded, linked with an arrow or line. In addition you can label particular features with, say, an asterisk or other labels that can then be explained in the figure legend.

- A figure legend should contain the minimum information needed for the reader to interpret and judge (for him/herself) the result/message. For this you need to put yourself into the mind of the reader and see what information you need to give. You can give in the figure some key details about the methods if they are relevant to understanding/appreciating the result (e.g. DNA binding at 50 and 500 mM NaCl [assuming that this is important to understand the message]), but extensive method details can remain in the methods section. No need to state the obvious (e.g. "important residues for binding are shown in stick representation and are labelled according to their residue number" can be simply: "key residues are highlighted").

- A good figure legend title should be summarising the message, rather than just the method (e.g. 'Crystallographic structure of protein X').

TIPS FOR WRITING STYLE:

- Write simply. Avoid jargon.

- Choose either British or American English spelling for your manuscript, and stick to it.

- Don't write 'It is worth noting that'. Just delete and put the statement, if it is worth noting. If not, delete the statement.

- Add references after a statement (if from literature) or refer to a figure (if the data are from you). However, don't make the adding of references or figures too disruptive to the sentence. E.g.: We defined a 10 base-pair genetic 'barcode' of SARS-CoV-2 (Figure 1A) that characterized with high sensitivity and specificity the five major clades of the virus populations circulating (Supplementary Table 2) on March 31th (Figure 2B). Could be written as: We defined a 10 base-pair genetic 'barcode' of SARS-CoV-2 that characterized with high sensitivity and specificity the five major clades of the virus populations circulating on March 31th (Figure 1A, 2B and Supplementary Table 2).

- Abbreviations: Once you define a particular abbreviation [e.g. focal adhesions (FAs)], then use it throughout the text. And only write out the abbreviation in the first occurrence in the text. I propose to write the full text without writing out the abbreviations, to avoid having several or late instances of spelling out abbreviations. Only in the last reading, add in all first occurrences.

- The convention for introducing an acronym is that you first write out the full name, then put the abbreviation in brackets. Correct: "the focal adhesion kinase (FAK) also localises to the nucleus", incorrect: " FAK (focal adhesion kinase) also localises to the nucleus". In rare cases this can be inverted, if it is critical for clarity.

- Avoid typos and inconsistent spelling, numbering, paragraphing, subdividing, policies, or fonds. Sit down with yourself before writing and decide on abbreviations. Is it p85nicSH2 or p85_nicSH2 or $p85_{nSH2iSH2cSH2}$ – you decide, but do so once for all.

- You need to proofread your manuscript for logic and content. Don't say the same thing twice in the same sentence, or in two following sentences. A particular fact should be mentioned only once, not twice or more. While there can be some repetition between Abstract, Introduction and Conclusion, it should not be verbatim

- Include time to allow for English proofreading. This should be done once the manuscript is considered final in its structure and content (to avoid adding large paragraphs after proofreading). Typos are unnerving for reviewers and make them frustrated. Why should they go through the manuscript carefully if you didn't?

- In the Introduction you don't have to repeat 'Studies show' each time. Just write the result and add references. Example: Nef binds to the SH3 domain (Whoever et al. 2001)

- Be careful with tenses: verbs in the past tense generally have 'ed' at the end; and be more careful with the plural/singular 's'. proteinS bind, but a protein bindS.

- You need to create some sort of awareness for what I/we/your group published in the past. Cite my/our references (if justified and relevant!), and say that 'We/my group and colleagues' have shown that …'

- For the methods, you need to give the minimum information that allows a skilled researcher to reproduce the data.

- Read all comments/corrections that your mentor made to your manuscript very carefully, please. Try to understand for each change why s/he made it. If you don't, then ask (mark it in the text, and discuss collectively). You need to understand your mistakes and learn from the hours your mentor or friend spent correcting your text.

- Use less (or no) passive voice in the Results. Say 'we found that protein A did X' rather than 'protein A was observed to do X'. Or: or use 'we speculated' instead of 'It is also speculated'

- Don't ignore the verb. Often a sentence can be shortened and clarified by using verbs instead of substantives. E.g: 'The Arg633Pro mutation found within the CHCR likely leads to distortion of the alpha-helix' could be written as: 'The Arg633Pro mutation found within the CHCR likely distorts the alpha-helix.

- Repetitions of names/acronyms are OK if they enhance clarity. You don't have to vary between ADH/D1, 'the protein' and 'the enzyme'.

- Either write 'the p85 protein [dimerises]' or just 'p85 [dimerises]'. Don't write 'the p85 [dimerises]'.

- Most people will read your manuscript often many years after it has been published. Hence it is better to refer to events in the past with a date, rather than with a relative time: E.g. "The phosphoinositide 3-kinase (PI3K) is a large family of kinases that were discovered more than 30 years ago" is better written as "The phosphoinositide 3-kinase (PI3K) is a large family of kinases that were discovered in 1985".

- Only use 'surprisingly' and 'interestingly' 1-3 times per manuscript, not 30 times.

- Never use 'This' at the beginning of a sentence without a noun following. Wrong: 'This also explains'. Correct: 'This chemical shift difference also explains'.

- Give numbers instead or in addition to quantitative statements. "ADH was stable up to 70degC" is better than "ADH was stable at high temperature". "Only 3 out of 24 characterized ADH are halophiles" is better than "only few ADHs are halophiles". This helps to maximise the information content of each sentence. E.g. write 5-fold larger rather than 'much larger'.

- Don't just copy blindly the values that an instrument outputs. For example, stating a Kd of 235.3575 ± 13.7523 µM does not make sense because you cannot measure these figures so precisely. 235 ± 13 µM might be appropriate, depending on the method. However 0.35 ± 0.15 µM is of course ok.

- Don't mix up 'significantly' (which means 'statistically valid') with 'substantial' (which means a lot). 'Significantly more' might be very little in absolute terms if the errors are tiny. 'Significant' should only relate to the results of a statistical analysis. Do not use it as a replacement for 'important' or 'substantial'.

-'Data' is a word in the plural (its singular form is 'datum'). Hence your data *were* confirmed, not your data *was* confirmed. Similarly: 'media' is plural, 'medium' is singular. Same for 'spectra' and 'spectrum'.

- Only use 'has been reported' (or similar) if you want to express doubt or credit a particular fact (timing or others).

- Avoid repeat statements about methods in the Results section; these can go into labels (if different in details) or Methods (if always the same).

- Always see if you can delete words in a sentence without losing its impact and meaning. Example: "it has been shown recently that FAK is a very important promoter of cancer cell invasiveness" can be said just as powerfully like this: "FAK promotes cancer cell invasiveness (REFERENCE).". Unless you want to stress, for example, that FAK had been shown *only recently* that it was involved in cancer, and not suspected to do so previously.

- Express things as simply as possible. E.g. rather than "FAK has been also shown to be involved in" write "FAK is involved in". Or "where they regulate different proteins" instead of "where they have been found to regulate different proteins".

- Make short sentences, where only one logic thought is expressed. Don't make circular arguments

- Don't use 'is key', or 'key role' or similar ('essential, crucial, central, primordial, …') too often.

- Try to avoid using 'as well as'. This is just useful in very particular cases where you have already several 'and' – linked enumerations.

- 'While' means 'at the same time'; what you normally mean, is 'whereas'.

- Same for 'since', which means 'relative to a moment in time'. What you normally mean is 'given that'.

- Either say 'the FAT domain' or 'FAT' in a sentence. Not ' The structure of FAT domain'.

- In theses, there is often a philosophical divide over using 'I' or 'we'. Please check with your mentor which one you should use, and then be consistent.

- Don't use "etc…" in scientific writing. Write: "such as, A, B and X" or "Including A,B and X".

 -Only use 'on the other hand' if it is preceded by a sentence/statement starting with 'on the one hand'.

- Be consistent: You write Paxillin, but also talin. Upper of lower case? You need to decide if you write Kinase or kinase. Normally it's lower case.

- Cut to the chase, especially if you talk about >10 year old research.

- Put the figures close to the text they are supposed to illustrate.

- A Discussion is a critical (!) discussion and not a Conclusion, which is a summarising Conclusion (and possibly outlook). Hence, Discussion and Conclusion are only minimally overlapping.