BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 37, No. 6, pp. 344–348, 2009

Article

Algorithm for Writing a Scientific Manuscript

Received for publication, July 2, 2009, and in revised form, August 13, 2009

Timothy R. O'Connor‡ and Gerald P. Holmquist‡

From the Biology Division, Beckman Research Institute, City of Hope National Medical Center, Duarte, California 91010

We present an algorithm for the construction of a strong initial draft. It is designed to overcome writer's block and to assist scientists who are not native English speakers. The writing starts with making figures and tables. These suggest several terse summary statements, the few major conclusions or observations the author will present to the scientific community. After identification of the audience, the specific community addressed, materials and methods are written to explain how the tables and figures were generated. Results are initially restricted to describing the logical data relations in each table and figure. The discussion then converts each data relationship into mechanistic cause-and-effect interpretations suitable for the abstract. A brief epilogue deals with the submission and the fate of the final manuscript once submitted. Although other models for the initial draft exist, this model has worked for us and new researchers in our laboratories and addresses problems we encountered while editing manuscripts.

Keywords: Manuscript preparation, Writing, Non-native English writers.

Writing a scientific manuscript is a challenging endeavor, especially for young scientists whose experience may be limited to writing brief laboratory reports for courses. Much time is devoted to collecting scientific data during baccalaureate and doctoral training, but now preparation of scientific manuscripts receives more emphasis, and eventually, online writing instruction can potentially yield significant benefits [1]. However, online writing training is not yet available at most institutions, and often, students are still required to learn scientific writing by themselves. Many examples of methods to attack writing a scientific manuscript exist in the literature, most of which address scientific writing for medical journals (e.g., Refs. 2–4). Two older publications by Kenneth L. Knight and Christopher D. Ingersoll [5, 6] are certainly useful, but they are directed principally toward athletic training and do not outline a simple step-by-step method to pursue scientific writing for novices.

In our experience, there are few simple descriptions of how to write a scientific manuscript that can be easily followed to lead beginning scientists to formulate a complete first draft of a manuscript that allows for subsequent polishing and perhaps further experiments. Based on our personal experiences while writing for biochemical and molecular biology focused journals, while assisting the writing of researchers in our laboratories, and while helping nonnative English language scientists, we have used the algorithm described herein to prepare manuscripts. Each step should permit the writer to proceed in a modular manner through each section to assemble a logically coherent first draft. Therefore, this algorithm is an instruction manual to assemble that draft, so, it proceeds from one section to the next (Fig. 1). This simplified outline is certainly not the only way to proceed with writing a scientific manuscript but is a method that we have used that we seek to share with others.

Construct Figures and Tables to Prepare for Formulation of Summary Statements

Construction of first-draft figures and tables based on data is the critical first step toward preparation of an outline for the manuscript. The programs most used by researchers for this purpose are Microsoft PowerPoint, Microsoft Excel, and Adobe Photoshop. Other useful programs available online without charge for students and not-for-profit institutions are Symyx Draw (http://www.symyx.com) to draw chemical structures and ApE (http://www.biology.utah.edu/jorgensen/wayned/ape/) or pDRAW32 (http://www.acaclone.com/) to draw molecular biological data. Use each figure to convey a single point in the manuscript. These figures do not need to be in final form, but in a format that will enable the evaluation of figure quality.

During the course of writing, modification of figures often occurs based on critical evaluation and comments from other scientists. Paste the figures and tables on a wall in their approximate order of appearance. This will provide the skeleton from which to begin the construction of a more detailed outline and draft.

Write a first draft of figure legends. Make each legend's opening sentence into a figure title to describe

‡To whom correspondence should be addressed. E-mail: toconnor@coh.org or gpholmqui@coh.org.

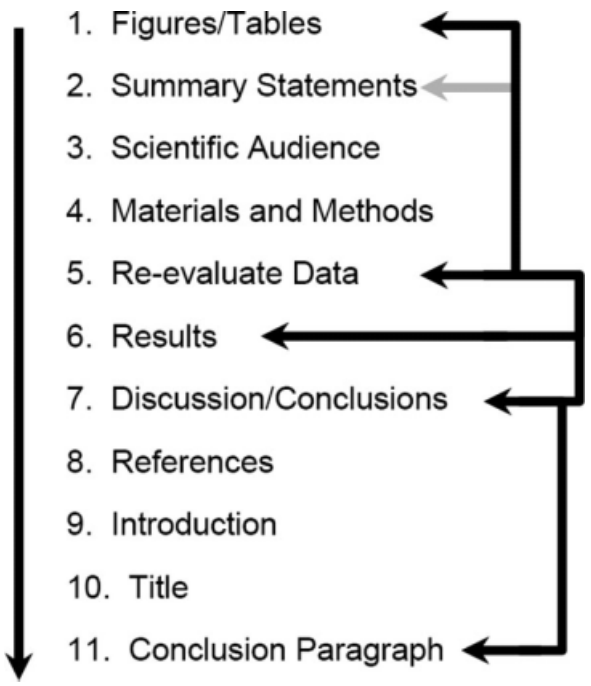

FIG. 1. **Overview of the formulation of a scientific manuscript.** The left arrow shows the linear progression of the process. The right arrows show natural points at which data should be re-evaluated to decide if the data, results, and discussion all point toward the conclusions. The gray arrow refers to the fact that the manuscript should be compared with the accuracy of the Summary Statements and how the other sections support those statements.

the variables compared. For example, for a hypothetical figure, "Figure 10. Exogenous [GTP] and aggregation in wild type *Paramecium*. Cells were grown in various GTP concentrations and aggregation was quantified. . . ."

Write One, Two, or at Most Three Summary Statements

These important statements are conclusions summarizing the major contributions of the manuscript to the scientific community. Do not use flowery clause-ridden sentences studded with conservative caveats and qualifications, but short rigid statements usually containing cause/effect words (Table I) such as;

1. "High extracellular GTP levels are *necessary* for the aggregated phenotype in *Paramecium.*"
2. "Intracellular GTP levels do not *influence* the aggregated phenotype."
3. "DNA damage hotspots *predict* mutation hotspots."
4. "Histone methylation *causes* transcriptional inactivation."
5. "Receptor phosphorylation is *necessary* for to gene expression."

The formulation of these summary statements while evaluating the tables and figures is the crucial first step of manuscript writing, and so much thought is required at this step. In summary statements, strive to distill the essence of what the paper will describe—these are the "take home" messages, the major points that readers should grasp when reading the paper. These statements are a proto-abstract that can evolve depending on the final versions of manuscript. With more experience in writing, formulation of these statements will occur even before making the figures or before designing the experiments.

Identify the Scientific Audience and Journals to Which the Manuscript Can Be Submitted

Once summary statements exist, identify an audience for the manuscript. With the audience identified, look for a journal that has published similar articles to identify a possible place to submit the manuscript. Read the instructions to authors to ensure that the manuscript would fit into the journal's scope. Identify three to five appropriate journals. List the journals along with the internet addresses for the author instructions to aid in making a final decision on the manuscript's destination.

Write the Materials and Methods Section to Supplement and Explain the Figure Legends

This section is generally the initial text in the manuscript. Include as many details as necessary to allow another researcher to duplicate the experiments that were used to generate the tables and figures. If another researcher has described the methods in detail, then reference the publication, but at least outline how the experiment was performed.

Eliminate Data, Tables, and Figures That Do Not Address the Summary Statements

For each figure and table, make a note of which summary statement it addresses. Extra experiments and their results that do not address the summary statement can hinder the initial writing and could confuse others about

TABLE I
Results words for data relations vs. discussion words for logic and mechanism

Results Words (data relation words)	Discussion or Conclusion Words (cause/effect logic and mechanistic words)
Were correlated, were positively correlated	Causes, brings about
A was a function of B; A increased with increasing B	Necessary (strong); mandatory, obligatory, essential
Associated	Necessary and sufficient (very strong)
Accompanied	Influences (weak) (affects)
Interdependent, related, correlated	A brings about a change in B (effects); A influences B
Proportionate, reciprocal, concordant	Consequence, effect, outcome, result
	Elicit, produce, induce, stimulate, consistent with

If English is your second language, grammar can be corrected by English experts, but only correcting the English will usually not correct improper scientific usage of these words.

the message of the manuscript. Temporarily, put these "extra" figures and tables in a drawer or save them in a separate file. Evaluate data for the manuscript as objectively as possible and decide if other experiments are still necessary. This should be done at each step of Results and Discussion.

Write the Results Section

The initial Results section should be brief and direct. Make each figure or table correspond to a paragraph or sub-section in the Results section. Show in the paragraph that how the figure or table supports a summary statement. Results were obtained in the past, hence use the past tense. Results usually use words that describe how mathematical variables were related (Table I), e.g., A and B were associated, A and B were positively correlated. Discussions and your summary statements use mechanistic cause and effect words such as "A causes B." Table I summarizes how to distinguish words for Results from words for Discussions or Conclusions.

The first sentences of each paragraph in the Results section form an outline and should correspond to titles for each figure or table legend. Once each figure and table has a legend describing the mathematical variables compared, then the first sentence of each results paragraph is simply the results of this comparison. Examples are as follows:

1. Increasing exogenous GTP levels *were accompanied* by increased aggregation in wild type *Paramecium* as shown in Figure 8. . ..
2. Exogenous GTP levels and intracellular GTP levels *were* positively *correlated* for wild type *Paramecium* (Figure 2). . ..
3. Figures 3 and 4 show that increasing exogenous GTP levels *were* neither *accompanied* by increased aggregation in mutant #1 Paramecium (Figure 3) nor *accompanied* by increased intracellular GTP levels (Figure 4) in mutant #1 *Paramecium.*
4. Increasing exogenous GTP levels *were* not *accompanied* by increased aggregation frequency in Mutant #2 as shown in Figure 5.

References in the Results section should be few and limited to methods developed in the manuscript or other similar methods described in the literature. If previously reported data are inconsistent with data reported in the manuscript, then defer commenting on these until the discussion section wherein it is possible to comment on discrepancies or why these other data are not applicable to the manuscript's summary statement.

Results Words—The Results section presents the actual data and is not meant to include interpretation. Results words describe data relations, for example, how two mathematical variables are related. Why and how the variables are mechanistically related is for the Discussion. "Associated" is a nice loose results word to describe a nonrandom relation between two different variables especially when one or both are binary, e.g., 0 or 1, + or −, wild type or mutant. "Correlated" is not a word to describe a relation between binary variables. It means that the two variables were quantified, each ascribed to one of many possible different numbers, and $x=f(y)$ or $y=f(x)$ can be plotted, one variable along the X and the other along the Y axes to yield data pair points with a significant correlation coefficient or an apparent slope. If an increase in variable x is accompanied by an increase in y, then x and y are positively correlated (the slope is positive). If a plot of x versus y yields data points that fall along an almost a perfectly straight line, then x and y are strongly correlated. Avoid using correlate transitively, i.e., "x is correlated with y," because it implies a cause and effect relationship with x being the dependent variable. When using results words maintain highly rigorous language and restrain impulses to insert interpretation.

Discussion Paragraphs

Convert by logical arguments the relations of mathematical variables stated in the results section into mechanistic interpretations of cause and effect. Use the present tense to describe these relations because these relations presently exist. Simply restate the data relation from each results paragraph and convert each to mechanistic conclusions such as "High extracellular GTP, but not high intracellular GTP levels, causes. . .aggregation in wt *Paramecium* but is insufficient to cause aggregation in certain mutants." Proper words to distinguish "how parameters are related" in results from "conclusions of causality" in the discussion are important. Causality is argued at three increasing levels, i.e., (i) association (results), (ii) necessity (discussion), and (iii) sufficiency (discussion).

Discussion Words—Discussion words are those that logically infer mechanistic causes and effects from data relations in the Results section. This subsection indicates ways to identify words that describe data interpretation. Below are some wording choice that can help produce a clearer Discussion section.

Effect and affect as verbs should be avoided because the usage of those words is often unclear to readers. A effects B should be written "A brings about a change in B" (if A, then B changes) or "A is necessary for B" (if not A, then not B). It means A is necessary for B to occur; however, a possibility remains that C and D or other things, A→C→D→B, can/may/will also be necessary for B to occur. If these intermediates do not exist, i.e., A→B in all cases, then A is necessary and sufficient for B. "Necessary and sufficient" is a very strong cause/effect statement similar to the logical "if and only if A, then B." A affects B should be written "A influences B." This is a very weak conclusion statement that "A causes some change in B" with the direction, strength, and mechanism of change being undefined. "Associated" is a results word. "Influences" can sometimes be the corresponding conclusion word, but, if the conclusion "A influences B" appears too strong, then just leave the weaker statement "A and B are associated" as the discussion word.

"A sometimes influences B" or "often (greater than 50%) influences B" means that the probability of an influ-

ence is greater than zero and less than one. If the probability that a mutant allele will affect the phenotype is less than 100%, then we conclude that the penetrance is less than 100%, i.e., "A sometimes influences B." In results, A and B are sometimes associated refers to a comparison of A data with B data; the "sometimes associated" may be because of a dirty microscope lens. If it can be concluded that the result, "sometimes associated" has a real mechanical or biological explanation, such as penetrance that was not influenced by dirt on a lens, then elevate the result "sometimes associated" to the conclusion "sometimes influences."

Distinction Between Results and Discussion Words.

When the Results and Discussion sections are completed, re-examine the Results and Discussion sections to identify results and discussion words in each paragraph. Circle or highlight results words in green and discussion words in red. There should be no results words in the Discussion section and no discussion words in the Results section. At this stage, an appreciation of the process of how results are interpreted and step-by-step, logically converted into conclusions will become clear.

In the discussion, a new model can be put forth that is consistent with your conclusions. For example, there may be reasonable data to formulate a model with a GTP receptor on the cell surface and postulate a mutant as deficient in either this receptor or some intracellular signaling pathway from this receptor to the interior. The model should be consistent with the logical conclusions of the Discussion and must also be consistent with the data relations of the Results section.

References

Use a reference library program, e.g., EndNote or Reference Manager, to attach references to statements in the manuscript. Search databases and the literature for references concerning your subject. Using an electronic database will simplify construction of bibliographic data and permit proper journal-specific formatting of bibliographies. When possible, cite the relevant original references and not reviews. The number of references is limited by some journals, so choose them carefully. Also, think about potential referees for the manuscript and try to include relevant references from scientists who could eventually be selected by an editor to review the manuscript.

Go back and Write the Introduction

This is easier because now that Results and Discussion sections exist. First, summarize the subject and review the literature to allow the reader to (i) understand the statements in the Results and Discussion, (ii) understand how the statements fit into the extant scientific body of knowledge, and (iii) Understand that the conclusions are indeed novel, the next step in the knowledge of the subject. Go back and read the introduction and circle or highlight all sentences that do not contribute to understanding the summary statements via points (i), (ii), and (iii). These sentences that can be removed are the superfluous fill that makes editors say "reduce by 50%" or reviewers say "the length does not justify the value of the science."

Generate a Title From the Summary Statements

The title should be a positive statement from the summary statements. Unacceptable titles include "Studies of. . .," "Investigation of. . .." Use titles that entice the reader to actually delve into the abstract.

Add a Conclusion Paragraph

In this paragraph, restate the logical conclusions and explain why these conclusions are important, how they will influence future thinking in this and other fields. In the introduction, these conclusions were the next step. Now discuss the future based on the conclusions in this manuscript or an alternative path to further substantiate the validity of your conclusions. Also, state the relevance of results in the present manuscript to other fields.

Writing the initial draft of the manuscript by using a boring outline with rigid interpretations and conclusions should be finished at this time, so now some of the other aspects of the manuscript need to be evaluated. Examine the extra experiment(s) and figure(s) that were previously set aside to determine if those actually add to the value of the manuscript. As objectively as possible, consider to add those results to the manuscript, to submit them as supplementary data if possible, or to use them for another publication. The conclusions should also be re-visited. If the initial conclusions in the discussion were too strong, then insert qualifying caveat words (e.g., "always→ almost always," "causes→likely causes," "The model in Figure __ summarizes the conclusions and predicts. . ."→ "The model in Figure __ is consistent with the conclusions," "The model in Figure __ explains the data in the previous figures→ "The model in Figure __ is consistent with the data in Figures__ to __"). If discussion words are not justified in the Discussion, then conservatively write the conclusions using results words. If the journal format fuses Results and Discussion into one section, then you can intersperse Results paragraphs with Discussion paragraphs or glue Discussion paragraphs (distinguishing fonts or colors) to the end of Results paragraphs. Rewrite to give this cold, logical skeleton a more warm human feeling. Wait at least 7 days and reread the manuscript. If satisfied, show the complete draft to other people To those individuals who should provide feedback on the science, correct the language, and/or the logic in the manuscript.

Epilogue: Manuscript Submission and Responses

These are not part of the algorithm, but these next steps are necessary for publication.

Final Figures

PowerPoint is generally useful, but other programs may be needed to convert to the proper electronic format, such as a full version of Adobe Acrobat. Fit all figures to one full page (either Letter or A4 size) each. Manipulation of data using image treatment programs is unethical. Many journals now require verification of the integrity of images. Therefore, avoid excessive use of such programs as Adobe Photoshop. For example, do not enhance the band intensity of a particular band in a gel that is actually less intense than other bands. Another example is the removal of background from gels or micrographs. Misuse of such software could result in the refusal of a manuscript and possibly an investigation by the journal or funding agencies. The Journal of Cell Biology has set standards that are generally recognized as clear and present a reference for all scientists [7]. Some journals may require conversion of PowerPoint files to TIFF format or other electronic formats, but that can be performed just before submission. Make sure lines used are thick enough to read, but not too thick (usually 1.5–2.0 point). Normal font size for a figure should be between 20 and 24 and Arial or Helvetica tend to produce the most legible text. Figures with multiple parts (a, b, c, etc.) must be organized to fit all the parts onto a single page. Shrink the figures to see if fonts and line thicknesses are sufficiently large to be readable at the size at which they will be printed. Tables can be constructed in Microsoft Word.

Submitting the Manuscript

A manuscript that is ready for submission must have a cover letter for submission to the journal. The letter will help the editor decide who is competent to review the manuscript. Use the abstract in general terms to describe the most important points in the work. This cover letter of no more than five to six sentences should indicate why this journal is appropriate and inform the editor of people who would be competent to review the manuscript. Do not suggest individuals with whom you have direct collaborations. Do not hesitate to request that scientists with potentially unfavorable opinions be excluded as referees. Most journals have online submission, so the process of review is relatively fast. Journals often indicate the length of time for an editor respond to the manuscript, but there are journals that do not have any stipulated time limit for responding. In that case, if waiting for a response from the editor for a significant period of time (now ~1–1.5 months), it is possible to write the editor and politely ask concerning the manuscript review status.

The Editor's Letter and Referees Comments

After submission, the editor will send a decision letter about the submitted manuscript that includes referees' comments. If the manuscript is accepted, necessary paperwork is required to complete the manuscript publication process, which includes copyright transfers (if not done beforehand), payment for page charges, and possibly other forms. The page charges are generally paid for from grant money, but many journals also waive page fees if the authors cannot pay them.

The editor could also decide that more work on the manuscript or experimental work is required prior to acceptance for publication. Text corrections or changes to the written manuscript can usually be rapidly performed. However, if required to perform the experiments and they would add to the manuscript, agree to carry them out. Contact the editor if the time frame for completion stated by the editor does not allow for completion of the experiments.

Responding to the Editor

When responding to the editor, always thank the editor and the referees for their comments, even if the sentiments is that the manuscript was subjected to intellectual assassination by the editorial team. Number responses in the order the comments appear in the letter even if the referees did not. Clearly indicate what changes were made in the revised manuscript. When disagreements with the referee exist, but there is a cogent argument, make it to the editor, but refrain from comments that are derogatory to the referee. If experiments were performed, describe them briefly. If it is impossible to perform the experiments give reasons for the inability to perform them.

The last possible response is that the manuscript is rejected by the editor. There are two possible options. The first, and easier option, is to pursue another journal for the manuscript. If there were only two referee reports with different opinions, a second possibility is to request that the editor contact another referee. In any case, a logical, well-reasoned reply to the editor is necessary if you choose the second option.

CONCLUSION

These steps are basic steps to construct a draft manuscript, and a brief outline of the process needed for publishing a manuscript. There are other methods, but using this algorithm should provide the insight needed to create a good first draft leading to publication and can also be used to overcome writer's block. We have used such a process for constructing manuscripts, and some members of our groups have also favorably used this method.

Acknowledgment— We would like to thank the Beckman Research Institute of the City of Hope for support.

REFERENCES

[1] A. Phadtare, A. B. A. Shah, R. Pietrobon. (2009) Scientific writing: a randomized controlled trial comparing standard and on-line instruction, *BMC. Med. Educ.* **9**, 27

[2] U. S. Neill (2007) How to write a scientific masterpiece, *J. Clin. Invest.* **117**, 3599–3602.

[3] A. E. Lin (2006) Writing for scientific publication: tips for getting started, *Clin. Pediatr (Phila)*. **45**, 295–300.

[4] M. A. Kliewer (2005) Writing it up: a step-by-step guide to publication for beginning investigators, *AJR. Am. J. Roentgenol.* **185**, 591–596.

[5] K. L. Knight, C. D. Ingersoll (1996) Structure of a scholarly manuscript: 66 tips for what goes where, *J. Athl. Train.* **31**, 201–206.

[6] K. L. Knight, C. D. Ingersoll (1996) Optimizing scholarly communication: 30 tips for writing clearly, *J. Athl. Train.* **31**, 209–213.

[7] M. Rossner, K. M. Yamada (2004) What's in a picture? The temptation of image manipulation, *J. Cell. Biol.* **166**, 11–15.