

The art of writing science

Kevin W. Plaxco*

Department of Chemistry and Biochemistry, Interdepartmental Program in Biomolecular Science and Engineering, University of California Santa Barbara, Santa Barbara, California 93106

Received 25 August 2010; Revised 19 September 2010; Accepted 21 September 2010

DOI: 10.1002/pro.514

Published online 15 October 2010 proteinscience.org

Introduction

The value of writing well should not be underestimated. Imagine, for example, that you hold in your hand two papers, both of which describe precisely the same set of experimental results. One is long, dense, and filled with jargon. The other is concise, engaging, and easy to follow. Which are you more likely to read, understand, and cite? The answer to this question hits directly at the value of good writing: writing well leverages your work. That is, while even the most skillful writing cannot turn bad science into good science, clear and compelling writing makes good science more impactful, and thus more valuable.

The goal of good writing is straightforward: to make your reader's job as easy as possible. Realizing this goal, though, is not so simple. I, for one, was not a natural-born writer; as a graduate student, my writing was weak and rambling, taking forever to get to the point. But I had the good fortune to postdoc under an outstanding scientific communicator, who taught me the above-described lesson that writing well is worth the considerable effort it demands. Thus inspired, I set out to teach myself how to communicate more effectively, an effort that, some fifteen years later, I am still pursuing.

Along the way I have learned a thing or two that I believe make my papers easier to read, a few

of which I am pleased to share with you here. Before I share my hard-won tips, though, I have an admission: there is no single, correct way to write. In fact, there are a myriad of solutions to the problem of writing well (see, e.g., Refs. 1–4). The trick, then, is not to copy someone else's voice, but rather to study what works—and what does not—in your own writing and that of others to formulate your own guide to effective communication. Thus, while I present here some of my most cherished writing conventions (i.e., the rules that I force on my own students), I do not mean to imply that they represent the only acceptable approach. Indeed, you (or your mentor) may disagree strongly with many of the suggestions I make below. This, though, is perfectly fine: my goal is not to convince you that I have found the one true way, but instead simply to get people thinking and talking about writing. I do so in the hope that this will inspire a few more young scientists to develop their own effective styles.

The Elements of Scientific Style

Words

Word choice directly impacts the readability of your writing. First and foremost is the observation that jargon is one of the greatest enemies of clear scientific writing. It is a crutch that narrows one's potential audience and should be avoided like the trap that it is. The rules for this are simple. First, do not invent new words or phrases if there is already a commonly used word or phrase that accurately fits the bill. Every new definition that your reader has to learn to follow your work adds to his or her

*Correspondence to: Kevin W. Plaxco, Department of Chemistry and Biochemistry, Interdepartmental Program in Biomolecular Science and Engineering, University of California, Santa Barbara, CA 93106. E-mail: kwp@chem.ucsb.edu

burden. Second, if at all possible avoid using abbreviations. Consider, for example, writing the phrase “the forward rate constant” rather than, say, using the abbreviation k_f . The latter may make your job as a writer easier, but it makes your audience’s job as readers that much harder. If you must use an abbreviation, make it meaningful. For example, use k_f for the forward reaction rate, rather than, say, k_1 as your readers are more likely to remember what the former means without having to refer to some scheme or figure elsewhere in your paper. Likewise, if you must use a large number of abbreviations or other identifiers, avoid arbitrary names in favor of names so logical that your readers can effortlessly keep track of them. For example, “protein dilution buffer” and “4-amino-1,10-phenanthroline” are far easier to understand and keep track of than “buffer A” or “compound 2.” If this is not possible, the third best approach (a distant third at that) is to provide a table that your readers can conveniently refer to. But I have to warn you that I am lazy enough that I am likely to set a paper down if I am forced to break from reading to look up anything more than twice.

Word choice also impacts the precision of your writing, which affects the ease with which it is read: any ambiguity in your writing forces your reader to work to understand your meaning. For this reason, when faced with the choice between a colloquial, but imprecise, way of saying something and a more formal, but more precise, way of saying the same thing, I opt for the latter. For example, “these data show” is less precise than “these data indicate” or “these data demonstrate”; although “show” can denote both “indicate” or “demonstrate,” the two words are not interchangeable as the former is synonymous with “suggest” and the latter with “prove”. Using the word “show” thus forces your readers to pause, break their stream of consciousness, and waste effort deciding which you meant.

Think hard about your tenses. We all agree that some observations are truly past tense (any specific supernova, for example, explodes only once). But, while not all authors agree with me on this, I believe most observations reported in the scientific literature are better described in the present tense because, assuming your experiments are reproducible, *the observation remains true even after the experiment was performed*. As an added bonus, present tense is more engaging than past tense. For example, I prefer “the phylogenetic data indicates that Chimps and Humans are related” over “the phylogenetic data indicated that chimps and humans are related”; although both phrasings are equally true, the former is more general, and sounds more exciting, than the latter.

Sentences

A well-written sentence usually leads with the action. Passive language, although extremely com-

mon in scientific writing, is a bore that your readers’ eyes will tend to glide over without taking in. Sometimes passivity is easy to spot, sometimes less so. But it is almost always dull. “Sequences were searched” is easy to recognize as passive (and should be cut! “We searched the sequence database using...”). But here is a more pernicious example: “The level of sequence identity between Chimp and Human genes is large, which illustrates the relatedness of these species” is more passive than “The close relatedness of Chimps and Humans is illustrated by the high degree of sequence identity shared by their common genes”, or “The high level of sequence identity between Chimp and Human genes illustrates the relatedness of these species”. The latter two sentences shift the focus of their action more firmly onto their subjects, making for more engaging reading.

Beware of long sentences. While it is good to vary your sentence length across a paragraph (this mimics the cadence of spoken language and is thus easier to follow), longer sentences are generally harder to parse than their shorter brethren. Indeed, average sentence length is a key predictor of reading difficulty in most of the commonly employed measures of “readability” (e.g., Ref. 5).

Symmetry of language is often helpful. Parallel structures within a sentence, or within adjacent sentences or paragraphs, can often be used to good effect to hammer a point home. In this context, it is more than acceptable to repeat words or phrases in a sentence. (“While it is largely accepted that Chimps and Humans reside in the same taxonomic order, there has been much debate as to whether or not they reside in the same taxonomic family”). Conversely, missed opportunities for symmetry can create a void or hiccup in the text, as the reader’s brain screeches to a halt after failing to find an anticipated parallel.

Paragraphs

In my experience, paragraph structure is both the most difficult element of writing to master, and the most important. Here are my favorite paragraph rules.

The first sentence of each paragraph should tell the reader what you expect them to get out of the paragraph that follows, which makes their job of following it far easier. Put another way; use the opening sentence of your paragraph to state your argument, and the rest of the paragraph to make your argument. This issue is so important that I regularly test my students by reading only the first sentence of each paragraph in their drafts, scrutinizing whether this alone is enough to communicate their entire story. If not, it is back to the drawing board. (Go back and re-read the first sentence of

each paragraph above and you will see that I hold myself to the same standard).

A paragraph should discuss only a single idea and thus should have a single, unifying theme running throughout it. This said, sometimes a paragraph can mix ideas that might seem, at first glance, separate concepts (“Our sensors are rapid. Our sensors are specific. Our sensors are selective.”). *But only if they can be tied together.* A test is whether you can tie all of the concepts together in a single opening sentence (“Our sensors exhibit many positive attributes”). If not, *you do not have a single paragraph.*

Conceptually, a paragraph should also stand on its own two feet. For example, the first sentence of a paragraph generally should not refer to an idea in the paragraph above without, at the least, restating that idea. For this reason, I avoid starting paragraphs with phrases such as “However”, or “In contrast”. If I must refer to an idea from the paragraph above, I will typically employ constructs that make it obvious how the paragraph following is related to those that preceded it. (“In support of the above arguments...”) In short, given your reader everything they need to understand a paragraph in that paragraph.

Flow is an equally critical element of good paragraph structure: every sentence in a paragraph should arise logically from the sentence before it, and transition logically into the sentence that follows. Lead your readers by the hand—do not leave them struggling to figure out how the idea embodied in a sentence is linked to the idea that preceded it. And do not scrimp on your transition elements (“Because of this...” “This, in turn, leads to...”) in the name of reducing your word count during editing; doing so shortens the text, but usually only at a cost to flow and readability.

The Elements of a Scientific Paper

A scientific paper is composed of specific elements: a title, an abstract, an introduction, methods, results, figures and captions, and, often, a stand-alone discussion section. Here, I discuss some of the rules I have picked up for tackling each of these elements.

The title

The title should headline the main result described in the paper, not the main effort. That is, a title that describes what you did (“A phylogenetic analysis of humans and chimps”) will likely attract far fewer readers than a title that describes what you learned (“Phylogenetic evidence indicates an exceptionally close relationship between humans and chimps”). Above and beyond this rule, I try to use as broad a title as I can reasonably get away with. For example, if no one has ever studied the relatedness of chimps and humans using any gene, do not mention

the gene you used in your title: not “Phylogenetic analysis of the GyrG gene indicates that humans and chimps reside in the same taxonomic family” but instead the shorter, broader, sexier “Phylogenetic analysis indicates that humans and chimps reside in the same taxonomic family”. If your work can be generalized then do not limit the breadth of your potential audience by being overly specific in the title.

The introduction

The introduction has precisely one purpose: to convince your readers that they should read your paper. To do so, it must first convince your readers that you have identified an important, open scientific question that they should care about. Following this, it should prime your readers to expect an answer to that question, encouraging them to read on.

Keep the introduction short! Here is another thought experiment for you: imagine you have in hand a paper with an introduction stretching some twelve paragraphs long. How many of these paragraphs will you dutifully wade through before you skip ahead to the results section? For me, the answer is three paragraphs or, rarely, four. It is probably wise to expect that your readers are similarly impatient. The introduction should only be long enough to convince the reader that the question you are addressing is real, that they should care about it, and that they should expect to learn the answer to it; the introduction is not the place for an exhaustive review of the literature because, to be blunt, few of your readers will invest the time to read a literature review until after the sheer compelling force of your results has convinced them they should.

Your high-school English teacher was right: the ideal introduction forms an inverted pyramid, starting broad and progressing to specific details. The first paragraph places the work in the broadest possible context, serving to convince the readers that you are working in an interesting and important area. (“It is clear that Chimps and Humans are both primates”). The second paragraph describes the specific topic addressed in your paper; it should convince the reader that you have identified an important, open question in the broader subject area. (“It remains contentious whether Chimps and Humans are in the same taxonomic Family”). The third paragraph (or the end of the second) describes exactly what was done in your paper. (“Here we report novel phylogenetic data that address the question of whether Humans and Chimps belong to a common family”).

The introduction need not, and generally should not, describe the conclusions of the study. Novelists, after all, never spoil their climax by spilling the beans on the first page, so why should you? (And you do share a important goal with the novelist; neither of you will get anywhere if your work is not

read.) Instead the introduction should serve as your hook, telling the reader what question they should expect the paper to address and motivating them to read on to discover the answer for themselves.

Results

The results section should typically start with a paragraph describing your approach in broad terms, as most of your readers will initially skip over the methods section and jump right to the “meat”. You do this when reading a paper for the first time, don’t you? I know I do. Because of this, it is critical that the results section provide your readers with a broad enough outline of your approach that this alone gives them sufficient context to understand your results. It is also important, however, that you not go into too much detail here; give them the conceptual tools to understand your work but not more. How much is enough? If it affects the interpretation of your results, then it needs to be included, but if the presumption is that it does not (e.g., who supplied your reagents) then move it to the methods section. The tiny fraction of your readership that will want to see these gory details can always turn to the methods section to satisfy their strange desires.

It is generally acceptable to include some “discussion” in the results section even if the paper also contains a stand-alone discussion section (an approach I champion below). Reasonable examples are individual sentences that compare a given result with literature precedents. (“Similar to the results of Smith and coworkers, we find. . .”). Likewise, if a second set of results bolsters an earlier set of results I will typically point this out as I am describing them, while everything is still fresh in the reader’s mind.

Discussion

In my view, it is usually better to maintain separate results and discussion sections. A separate discussion section allows you to reiterate your point bluntly and concisely and thus provides an additional forum in which to get your point across. It is also the only place you are allowed to speculate widely (and perhaps wildly) on the implications of your work. Why surrender such a great opportunity?

The discussion section should be constructed like a (normal, upright) pyramid, stating your most specific research conclusions at the top before broadening out to encompass wider and wider ideas. The first paragraph answers the question: what are your key results? (“The phylogenetic data presented here indicate that Chimps and Humans are in the same taxonomic family”). Following this, it is often effective to lead your readers through an overview that describes how your results fit into the bigger picture. (“Our findings support previous work by Jones and coworkers, who have argued. . .”).

As was true for the introduction, keep the discussion section short; more than about five or six paragraphs and you risk boring your readers and diminishing the impact of your work. That said, any reader who has gotten this far is obviously enjoying your paper and can be counted on to stick with you a little longer. Indeed, if they have made it this far they will likely find it acceptable—even entertaining and engaging—if you close with some broad speculations. Speculations, though, should be clearly identified using explicit phrases, such as “we speculate” or “this suggests”. (“The discovery that Chimps and Humans reside in the same family suggests that we can all live together in peace and harmony”).

Methods

As I noted above, many journals put the methods section at the end of the manuscript. Even if the journal you are writing for does not, it is often best to write as if your readers have not read the methods section, since most will have skipped it. Never direct your readers to look in the methods section for something that is critical to understanding the paper; present everything the reader needs to understand your claims directly in the results section.

What, then, goes in the methods section? Every detail necessary were someone to attempt to reproduce your work, and then some. For example, as in the results sections, methods sections often profit from at least limited “discussion”. Drawing analogies/contrasts to existing experimental precedent in the methods section can serve to clarify how your experimental approach relates to, or differs from, those of previous studies.

A final note about the methods section: you should watch your tenses here, too. That is, while I argued above that most observations should be described in the present tense (assuming they remain true after they were made), this does not hold for your experiments, which really were conducted at some specific time in the past. Thus, “we determined the level of sequence identity using ULTRA-BLAST” rather than “we determine. . .”

Figures and figure captions

Good writing involves more than just writing: the vast majority of scientific papers employ figures to illustrate important concepts and present supporting data. If crafted with skill and care, these can boost the impact of your work nearly as much as the entirety of your text. Before I talk about figures and figure captions, though, I want to share a few words about how I like to cite figures in the text of my papers. Specifically, my writing is almost always stronger when I write as if my readers do not have access to the figures. This exercise forces me to describe the main result presented in each figure so that the reader does not have to break from reading

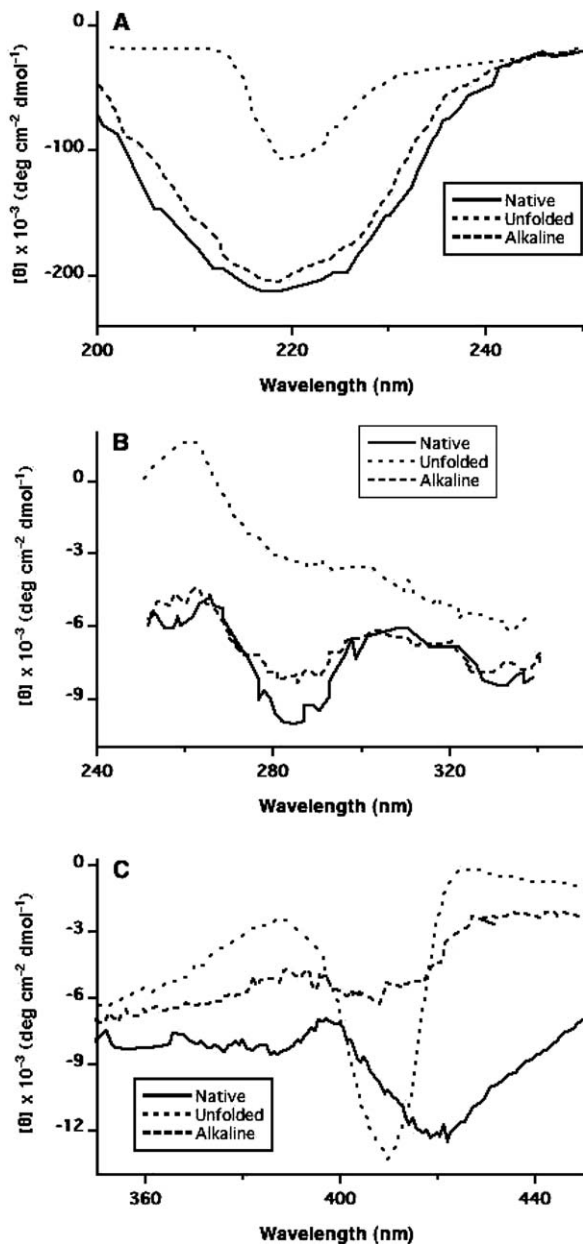


Figure 1. Far-UV (A), near-UV (B), and Soret (C) circular dichroism spectra of native cyt c (solid line), the alkaline form (dashed line), and the unfolded protein (dotted line). (Adopted from Ref. 6, with permission from ACS Publications).

and find the relevant illustration. In short, I try to paint a picture with words. Consistent with this, I almost never refer directly to figures in the text. That is, instead of writing “In figure X we present the phylogenetic relatedness of Chimps and humans” I describe what is in the figure and then parenthetically note the figure number for those readers who might want to check up on what I have written: “The phylogenetic relatedness of Chimps and Humans approaches 98% (Fig. X)”.

The figures, which provide a second mechanism to communicate your story, are at their strongest

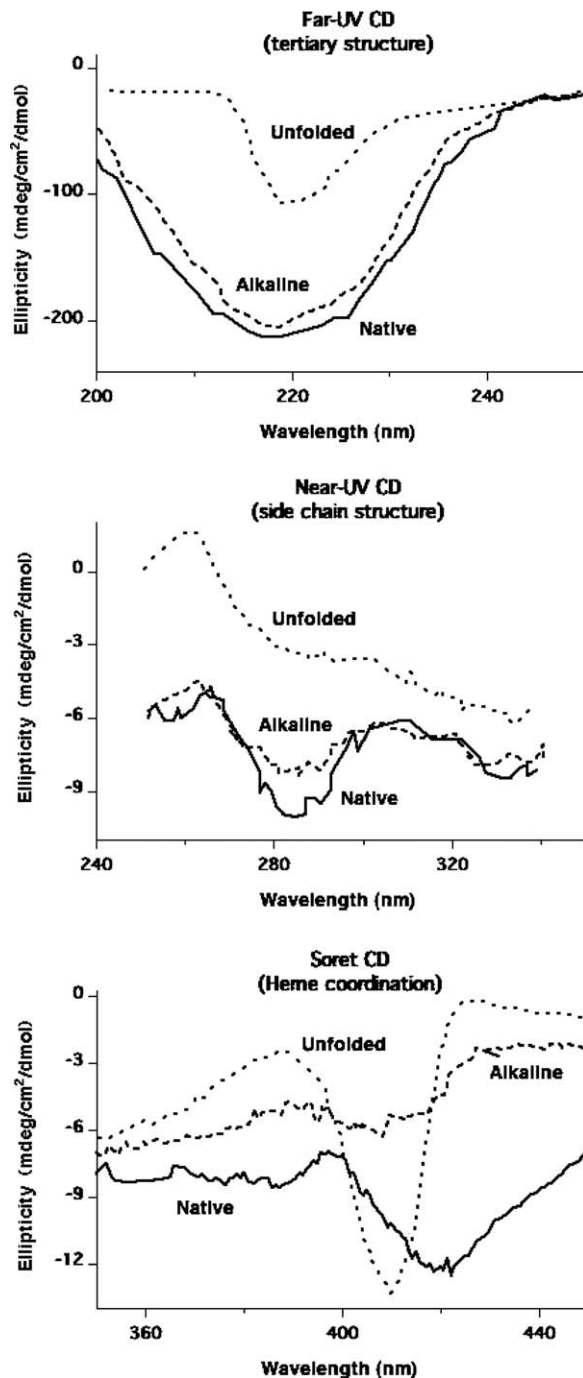


Figure 2. Circular dichroism (CD) spectroscopy demonstrates that, although the structure of cytochrome C remains largely unchanged at alkaline pH, its heme coordination is significantly altered. Specifically, while the (top) far- and (middle) near-UV spectra of alkaline cytochrome c, which are indicative of tertiary structure and side chain packing, respectively, are close to those of the native protein, its (bottom) Soret band CD, which is sensitive to heme coordination, shifts dramatically. (Adopted from Ref. 6, with permission from ACS Publications).

when the reader can take in that story just by looking at them. It is easy to test this: give your figures to someone who vaguely knows your work, but does not know the exact story at hand; can he or she

describe your story simply by looking at the figures? (For an example of what not to do see Figure 1, which, embarrassingly, I adopted from one of my own papers with only slight modification.) In keeping with this, if at all possible, the first figure should explain the overall goal of the paper. A cartoon often serves this purpose well. Likewise in keeping with this, an ideal figure is interpretable *without reading the figure caption or referring to legends* (Fig. 2 approaches this ideal). For example, label lines in graphs directly in the image rather than referring to the “dot-dashed line” in the figure caption or even in a legend. Similarly, consider using meaningful titles (e.g., “Chimpanzee”, “Human”) for individual panels rather than meaningless titles such as “A” and “B”.

What holds for figures also holds for figure captions. Namely, I believe that the reader should be able to ascertain the entire story just by reading the figure captions (compare, for example, the captions of Figs. 1 and 2 in this article). To help with this, the first sentence of the figure caption should be, in effect, a title for the figure (although, unless the journal says otherwise, it should remain a complete sentence) that clearly and concisely *tells your readers what you expect them to learn* from the figure. This will make their job interpreting the figure much easier, and thus maximize the readability of your work.

The abstract

I have described the abstract last rather than first because, of course, the abstract is the entire paper in brief. I also placed discussion of the abstract at the end to emphasize that it is perhaps the single most important element in a paper; it will, after all, be read far more often than the paper it describes. Indeed, a badly written abstract almost guarantees that no one will read the paper itself.

Because it recapitulates the entire paper, the abstract should start with a very blunt sentence describing the broad field. (“It is widely accepted that Chimps and Humans reside in the order primata”). This is usually followed by a sentence or, perhaps, a second clause in the first sentence that describes the specific problem at hand. (“It remains contentious, however, whether or not they reside in the same taxonomic family”). The next sentence states what the reader should expect to learn upon reading this article. That is, this sentence recapitulates the last paragraph of the introduction. (“Here we use phylogenetic analysis of the GyrG gene to address this question”). The body of the abstract should be two to three sentences about the results

that clearly tell the story of the paper. These sentences will closely parallel the first paragraph of the discussion. Indeed, they can often be pulled from that paragraph almost verbatim; this sort of parallel structure often makes a paper easier to read. Finally, abstracts need closure (i.e., concluding sentences) too. Often an effective means of achieving this is if the last sentence in your abstract is a restatement of the last or second to last paragraph in your discussion. And although I cautioned you about showing your hand when writing the introduction, do not worry about spoiling the punch line in the abstract: the abstract should be considered a stand-alone piece of literature, which, again, is how it is often read.

The Most Important Rule

The most important rule is simple: ignore any and all other rules *if doing so makes the paper easier to read*. Writing the clearest, easiest to read papers possible is the one-and-only goal. And there is no single “right way” to do this that fits all stories under all circumstances. Remember: *writing is an experimental science*. Just keep experimenting until you find a way that works for the task at hand. Good luck!

Acknowledgments

The author thanks his mentor, Prof. Chris Dobson, who motivated him to do this work; he thanks his past and present students for the countless hours they spent polishing drafts and hashing out these rules by trial, error, and endless discussion, and he also thanks his friends and colleagues who provided encouragement and editorial input, and a special thanks to Federico Rosei for planting the seed of this idea.

References

1. Gopen GD, Swan JA (1990) The science of scientific writing. *Am Sci* 78:550–558.
2. Whitesides GM (2004) Whitesides’ group: writing a paper. *Adv Mat* 16:1375–1377.
3. Thrower PA (2007) Writing a scientific paper: I. Titles and abstracts. *Carbon* 45:2143–2144.
4. Schulman RE (1996) How to write a scientific paper. *Ann Improb Res* 2:8–12.
5. Spache G (1953) A new readability formula for primary-grade reading materials. *Elem Sch J* 53:410–413.
6. Fan C, Gillespie B, Wang G, Heeger AJ, Plaxco KW (2002) Spectroscopy and electrochemistry of the pyridine-chytochrome c complex and the pyridine-induced ‘alkaline-like’ conformation. *J Phys Chem B* 106:11375–11383.