

# **Recipe for a Quality Scientific Paper: Fulfill Readers' and Reviewers' Expectations.**

Yaoqi Zhou

Indiana University School of Informatics, Indiana University-Purdue University at Indianapolis

Center for Computational Biology and Bioinformatics, Indiana University School of Medicine

## **Dedication**

50<sup>th</sup> birthday of the University of Science and Technology of China

## **Table of Contents**

Preface	2
Introduction	2
What readers want	3
Readers' expectations	3
Readers' expectations of a sentence	3
Readers' expectations of a paragraph	5
Readers' expectations of a table/figure	6
What reviewers want	8
How to meet the demands of potential reviewers	8
Structure of a paper	8
Methods/experimental procedures	9
Results section	9
Title	9
Introduction section	10
Discussion section	11
Abstract section	12
Summary	12
Epilogues	13
Acknowledgement	13
References	14

## Preface

My first experience with scientific writing in English was the translation of my Chinese B.S. thesis. By the time I graduated with a PhD from State University of New York at Stony Brook in 1990, I had more than 20 publications. However, my understanding of how to write high-quality scientific papers remained at an elementary level and was limited to minimization of grammatical errors. This happened because most of the time I simply accepted the changes made to my manuscript by my PhD. advisors, Dr. George Stell and Dr. Harold L. Friedman, without knowing or asking why. During my postdoctoral study at North Carolina State University, my mentor, Dr. Carol Hall, encouraged me to attend a two-day writing workshop at the neighboring Duke University. The workshop taught by Professor Gopen truly opened my eyes. For the first time, I learned that readers have expectations when they read, and the most effective way to write is to fulfill their expectations. This workshop helped me write my first research proposal, and I was awarded a postdoctoral fellowship to work with Dr. Martin Karplus at Harvard University. During the five years at Harvard, Professor Karplus made me realize that a good paper requires an in-depth, tough, and thorough self-review. Now, I am a professor myself with my own research group. I constantly feel the need to teach my students and postdocs to write better. I do not consider myself an expert in scientific writing, but I do think that sharing my understanding and writing experience might help others to avoid the long journey that I took to sharpen my writing skills. If you have any comments and suggestions, please feel free to email me ([yqzhou@iupui.edu](mailto:yqzhou@iupui.edu)). You are also welcome to visit my webpage: <http://sparks.informatics.iupui.edu>.

## Introduction

For a typical research project, graduate students and postdocs are the ones who take an initial idea from their advisors and carry out the research. After trials and errors, they produce some positive results. Now, they have to compile and analyze the results to produce a paper for publication in a journal. The most important aspect of an article is its content. However, with the same content, a well-written paper will more likely be accepted in a more prestigious journal, while a poorly-written one may well be rejected for publication. Both quantity and quality of published papers are essential for advancing the careers of both students and their advisors, and for obtaining external funding needed by advisors to continue their research. “Publish or perish” reflects the true reality of academic life.

A common misconception for many students is that their research is finished when results are obtained. As a result, the draft papers they prepare are often a simple collection of raw data with a lack of detailed description of procedures or methods and an absence of critical readings of current literature. Writing should be an integral part of research. It is the time to rationalize the success and failure of the method employed, to search for the implications and other possible interpretations of the results obtained, to compare or contrast your work to other related studies, and to make your work understandable by other scientists.

Why, one might ask, is there a need to make such a significant effort to write? The reason is simple. A research study is meaningful only if someone else uses it in his/her studies. For this to happen, the paper has to be written in a way that arouses other scientists' interest and allows others to reproduce the results. Only an understandable study can be reproduced. Only a reproducible work enables others to follow the lead. The number of scientists following the lead is a measure of the impact of a research study. Thus, in a way, a research study has to make a “sale” to other scientists.

To make a sale, a scientific paper has to be tailored to its unique customers: a sophisticated group of ego-driven scientist readers. It has to convince skeptical (and often competitive) colleagues because their reviews are the first hurdle to overcome in the process of publishing your paper. Thus, the best approach to a scientific writing is to fulfill readers' expectations and potential reviewers' demands. To do this, one first needs to understand what they want.

## What readers want

The potential readers of your paper have a diverse level of expertise in your field. They are beginners, with graduate students at one end of spectrum, and well-established experts (potential reviewers) at the other end. Thus, the paper should be written simply enough to make it understandable and reproducible by graduate students and deep enough to attract the interests of experts.

All scientists (students or their advisors) are usually very busy. Moreover, vast amount of journal publications prohibit them to read many papers in detail. They usually hope to find the most important information in a paper very quickly. Typically, they will not read the abstract of a paper if its title is not interesting and will not read the paper if the abstract does not contain significant innovation and/or interesting findings. Even if they decide to read the paper; they often skip paragraphs and read the part that is of most interest to them. Thus, it is important to write a well-structured (linked) paper that allows readers to search for information quickly.

In addition, a paper will be widely cited/used only if its significance can be understood without much effort. Letting readers find things where they expect to find them is the key to the clarity of a paper. If you want readers to understand your paper effortlessly, you have to work hard to satisfy their expectations.

## Readers' expectations

### Readers' expectation for a sentence

1. Readers expect familiar information at the beginning of a sentence. A sentence is the easiest to understand if it discusses something that the reader knows. This is not possible for a scientific paper, because only new discoveries can be published. Moreover, a scientific paper usually contains many terms that are new to students. The next easiest sentence to understand should make a smooth transition from old information to new ones by starting with something familiar (or mentioned before) to readers and ending with new information. A well-written paper requires that all sentences start with "old" information (i.e. linked backward). The golden rule for writing the start of a sentence is to ask yourself this question: have I introduce this concept before? In the majority of papers that are difficult to read, new concepts are often used without proper introduction. Here is an example:

Samples for 2-dimensional projection of kinetic trajectories are shown in Figure 7. The coil states are loosely gathered while the native states can form a black cluster with extreme high density in 2-dimensional projection plane.

There is a disruption of flow from the first to the second sentence. "The coil state" comes out of nowhere. Readers will likely find the following easier to read:

Kinetic trajectories are projected onto  $xx$  and  $yy$  variables in Figure 7. This figure shows two populated states. One corresponds to the loosely gathered coil states while the other is the native state with a high density.

In this new paragraph, a new sentence is inserted to smooth the transition between the two original sentences. The first sentence is linked to the 2nd by "Figure" and the 2nd is linked to the 3rd by "two states". New information (coil states) now appears at the end of the third sentence. The whole paragraph is now organized as a whole, rather than a collection of unrelated sentences. Here is another example:

The accuracy of the model structures is given by TM-score. In case of a perfect match to

experimental structure, TM-score would be 1.

In the second sentence, old information “TM-score” is disrupted by the new information “a perfect match to experimental structure”. Here is a suggested solution:

The accuracy of the model structures is measured by TM-score, which is equal to 1 if there is a perfect match to experimental structure.

The biggest problem in scientific writing is the reverse order of old and new information. Old and new information may not be apparent for the author because he/she is so familiar with the subject. To do it right, whenever starting a new sentence you should ask yourself, has this word been mentioned before?

2. Readers expect the action verb immediately after the subject. A sentence tells who is doing what. To understand a sentence, readers need the verb indicating what the action is. Reading will be interrupted by searching for the verb if there is an excessive space between the verb and its subject. An interrupted reading process will make a sentence difficult to understand. Here is an example:<sup>1</sup>

The smallest URFs (URFA6L), a 207-nucleotide (nt) reading frame overlapping out of phase the NH<sub>2</sub>-terminal portion of the adenosinetrip hosphatase (ATPase) subunit 6 gene has been identified as the animal equivalent of the recently discovered yeast H<sup>+</sup>-ATPase subunit 8 gene.

Here is the same sentence after the verb moves closer to the subject: <sup>1</sup>

The smallest of the URFs is URFA6L, a 207-nucleotide (nt) reading frame overlapping out of phase the NH<sub>2</sub>-terminal portion of the adenosinetriphosphatase (ATPase) subunit 6 Gene; it has been identified as the animal equivalent of the recently discovered yeast H<sup>+</sup>-ATPase subunit 8 gene.

The new sentence is now more balanced. Avoid the excessively long subject and short object. A large head with a small foot does not stand well. It is much better to have a short subject followed immediately by a verb and long object.

3. Readers expect that each sentence makes only one point, which is emphasized at the end of the sentence. Comparing two sentences below, one can see the difference emphasis makes:<sup>1</sup>

URFA6L has been identified as the animal equivalent of the recently discovered yeast H<sup>+</sup>-ATPase subunit 8 gene.

The recently discovered yeast H<sup>+</sup>-ATPase subunit 8 gene has a corresponding animal equivalent gene, URFA6L.

Clearly, the first sentence is about a recently discovered yeast gene, while the second one emphasizes the animal equivalent gene.

Here is an example:<sup>1</sup>

The enthalpy of hydrogen bond formation between the nucleoside bases 2-deoxyguanosine (dG) and 2-deoxycytidine (dC) has been determined by direct measurement.

This sentence seems to emphasize “direct measurement.” This is often unlikely to be the intent of the original author. A reverse of the sentence will make a more balanced sentence:<sup>1</sup>

We have directly measured the enthalpy of hydrogen bond formation between the nucleoside bases 2-deoxyguanosine (dG) and 2-deoxycytidine (dC).

The new sentence is simpler and shorter. It also avoids the “large head and small foot” problem.

### Readers’ expectations of a paragraph

Each paragraph should only tell one story about one subject. It should start with a topic sentence and end with a summary or transition sentence for the next paragraph. Sentences contained in the paragraph should logically link each other from the beginning to the end and flow from old to new information. Multiple points in one paragraph will make it difficult for the readers to know what to follow and what the paragraph tries to say.

Here is one paragraph:<sup>1</sup>

The enthalpy of hydrogen bond formation between the nucleoside bases 2-deoxyguanosine (dG) and 2-deoxycytidine (dC) has been determined by direct measurement. dG and dC were derivatized at the 5 and 3 hydroxyls with triisopropylsilyl groups to obtain solubility of the nucleosides in non-aqueous solvents and to prevent the ribose hydroxyls from forming hydrogen bonds. From isoperibolic titration measurements, the enthalpy of dC:dG base pair formation is -6.650.32 kcal/mol.

It is hard to know the exact point that the author tries to make in this paragraph. The mention of enthalpy at the beginning and end of the paragraph suggests that enthalpy is the focus. The following shows a better way to describe enthalpy as the main topic of the paragraph:<sup>1</sup>

We have directly measured the enthalpy of hydrogen bond formation between the nucleoside bases 2-deoxyguanosine (dG) and 2-deoxycytidine (dC). dG and dC were derivatized at the 5 and 3 hydroxyls with triisopropylsilyl groups; these groups serve both to solubilize the nucleosides in non-aqueous solvents and to prevent the ribose hydroxyls from forming hydrogen bonds. The enthalpy of dC:dG base pair formation is -6.650.32 kcal/mol according to isoperibolic titration measurements,

The first sentence now describes the topic of entire paragraph. The inversion of the original first sentence is for 1) making new information “dG” and “dC” at the end of sentence and the focus point of the entire paragraph, and 2) creating a better link to the next sentence. The original second sentence splits into two so that only one point is made in one sentence. The last sentence is now clearly a summary statement with the old information at the beginning. Here is another example:<sup>1</sup>

Large earthquakes along a given fault segment do not occur at random intervals because it takes time to accumulate the strain energy for the rupture. The rates at which tectonic plates move and accumulate strain at their boundaries are approximately uniform. Therefore, in first approximation, one may expect that large ruptures of the same fault segment will

occur at approximately constant time intervals. If subsequent main shocks have different amounts of slip across the fault, then the recurrence time may vary, and the basic idea of periodic main shocks must be modified.

In this example, the first and second sentences are relatively clear about the “rate” of accumulating strain, although the old information of the first sentence is not at the beginning of the second sentence. When reaching the third sentence, readers are easily lost. A clearer description is as follows:<sup>1</sup>

Large earthquakes along a given fault segment do not occur at random intervals because it takes time to accumulate the strain energy for the rupture. The rates of strain accumulation at the boundaries of tectonic plates are approximately uniform. Therefore, nearly constant time intervals (at first approximation) would be expected between large ruptures of the same fault segment. However, the recurrence time may vary; the basic idea of periodic main shocks may need to be modified if subsequent main shocks have different amounts of slip across the fault.

The new paragraph now focuses clearly in occurrence of earthquakes. The underlined text indicates the old information described before. Clearly, misplacement of old and new information is the main problem for understanding in this and last paragraphs. A flow from old to new information is the best way to make readers’ life easier. The purpose of a paper is not to test reading comprehension but show the author’s ability to express his/her view clearly. You cannot blame readers for failure of understanding your paper. You have to blame yourself for failure of delivering the message.

Readers’ expectations of a table/figure

Some impatient readers will go directly to tables and figures to find out exactly how interesting the paper is. Thus, it is important that captions in tables and figures can be understood without the need of reading the main text. (Some journals, however, have their own criteria for captions. Consult the instructions for authors.)

For tables, as we read from left to right, we expect more familiar (old) information in the left and new information on the right. For example, two tables 1 and 2 are listed below with a simple switch of two columns.

Table 1:

Temp (°C)	Time
25	0
27	3
29	6
32	12
32	15

Table 2:

Time	Temp (°C)
0	25
3	27
6	29
12	32
15	32

Table 2 is much easier to read than Table 1 simply because we are more familiar with time as the independent variable.

Another rule for tables is to save the best for the last. That is, the most interesting result should be presented in the right column or the bottom row. This is where the readers finish their reading and take the impression with them. The following example compares the accuracy of several methods. The last row provides the result of the current study.

Table 3:

Benchmark	SALIGN	Lindahl	PROSPECTOR 3	LiveBench 8
Method	Alignment	MaxSub	MaxSub	MaxSub
SPARKS	53.1%	325.9	529.0	38.3
SPARKS2	54.9%	341.0	591.0	40.7
<b>This work</b>	<b>56.6%</b>	<b>349.2</b>	<b>601.9</b>	<b>42.2</b>

For figures, the minimum one should do is to use large bold Helvetica font for all labels (numbers, axes, and legends). Draw the most important region only. Maximize the distinction of the curves without color. (Color figure is costly).

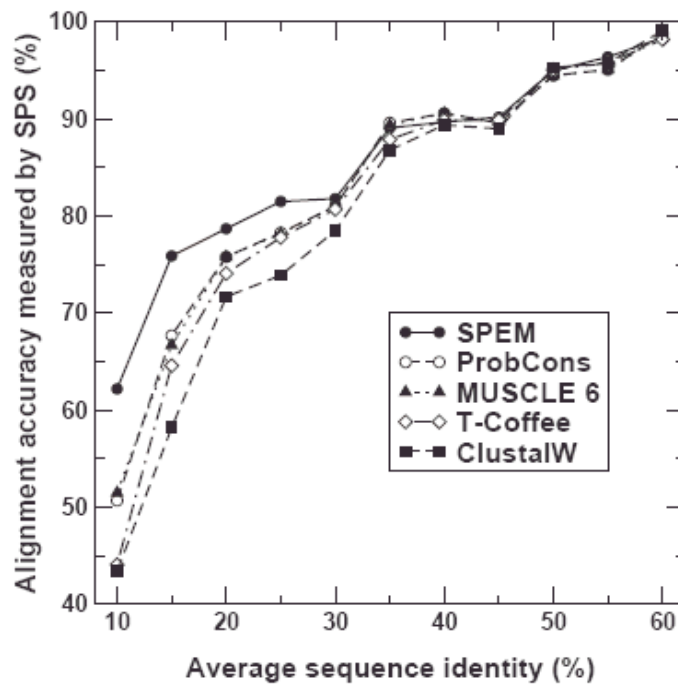


Fig. 3. Alignment accuracies (measured by SPS) as a function of average sequence identity given by methods SPEM, ProbCons, MUSCLE 6.0, T-Coffee and ClustalW, shown as labeled. Each point is represented by the lower bound of sequence identity at each bin.

The above figure<sup>2</sup> is what I prefer. Use solid lines for your work and dashed lines for other studies. Symbols are mixed with open and filled from top to bottom to maximize the difference between various curves. Spell out the titles for X and Y axes (rather than short-hands).

### **What reviewers want**

Before a paper is published, it has to pass a rigorous scientific review by the reviewers who are the experts (and competitors) in the same field. The job of reviewers is to locate as many potential weaknesses in the paper as possible. Sometimes, a reviewer will try to stop your publication because of his/her view or his/her competitive interest. Thus, a paper has to be written to minimize the possibility of rejection for bad reasons, such as incompleteness.

### **How to meet the demand from potential reviewers?**

Here are the basic principles for satisfying the needs of expert reviewers.

1). Establish ONE central theme of what you want to tell the readers or reviewers. Readers will be lost if too many ideas are presented in one paper.

2). Based on the central theme, make a “sexy” (but with absolutely no exaggeration) title to attract reviewers’ interest. If you cannot capture a reviewer’s interest, it is better not to publish the paper. (Editors sometimes have hard times to find reviewers because lack of interest from reviewers.)

3). Explain and rationalize every parameter and every single step employed. Reviewers do not have time to think about details. Rationalization of procedures and parameters indicates that you know what you are doing.

4). Ask yourself if all things presented are detailed enough for one to reproduce your work. Do not skip any detail. The easier it is for a reviewer (or any readers) to reproduce your work, the more likely that he will accept your paper. Reviewers will not actually reproduce your work. You have to convince them that they can reproduce your work based on what you have described.

5). Be persuasive! Do a comprehensive rather than a half-finished study! Try to prove your central theme from multiple tests/sources. Make your paper like a presentation of a lawyer trying to prove his case beyond reasonable doubt.

6). Cite all important studies. Do a comprehensive literature search while writing.

To achieve this goal, one has to follow the structural requirement of scientific papers.

### **Structure of a paper**

A typical science paper includes a title, abstract, introduction, methods/experimental procedures, results, discussion, acknowledgement, and cited references. These sections are designed to help readers to locate the information of their interest more quickly. Placing information in the wrong section will confuse readers. A common mistake made by students is to mix facts (results) with discussion (implications and interpretations of the results).

Although the final paper starts with introduction first, it is highly recommended that methods and results sections are written first. This is because only after a better understanding of methods and overall results, a central theme about the paper can be established. The central theme is needed before you plan for title, introduction and discussion. In addition, authors are most familiar with the methods used and results obtained. One should always start with something most familiar first.



## **Methods/Experimental Procedures.**

If this paper is about a new method/technique/algorithm developed, write its novel aspects in great detail. Describe it in a logical, rationalized manner. This will aid the readers' grasp of the new methods significantly. Do the methods used involve any parameters? If yes, each parameter (or cutoff value) should be rationalized either by previous usage, physical/mathematical reasoning, or extensive testing/optimization. If rationalization is not possible, describe the effect if the parameters are changed (Actual results should be in the results or discussion section. The methods section contains only descriptions of the effect.). If nothing is done, you should explain why (too costly? too time consuming? or defer to further studies?).

For new method development, you also need to design various ways to test the method. A convincing case requires as many tests as possible. The more tests you can find or design, the more likely your work will be accepted and used by others.

After the method section completed, ask yourself following questions: 1) are all new terms defined? 2) if you read this section for the first time, do you have enough information to reproduce all the work? Remember, do not hide any tricks and/or shortcuts used. People will be upset if they can't reproduce your results. Never try to cheat! There are a lot of smart people around, and many are smarter than you are. If you cheat, the chance is that it will be discovered soon or later. If you think that no one will ever try to duplicate your work, your result may not be worth publishing.

## **Results Section.**

Before you start to write the results section, think hard about the meaning of your results. Do you understand them? Do they tell you something deeper? Can you understand your results in many different ways? Can you design new tests to prove or disprove some of your interpretations?

If you discover something new, you have to prove that your new result is not an artifact of your methods (a good subject for discussion section). Can it be duplicated in different conditions? If you develop a new method, you must illustrate the importance of the method. Does it improve on other existing methods? Your result section has to be organized to support new findings or validate the importance of new methods from different angles or by multiple tests.

Once you have a better understanding of your results, you need to decide the most significant "sellable" point of the paper. That is, establish the central theme of the paper and organize all your paragraphs to prove and support the central theme using the existing data (and producing additional data, if necessary). This includes arguments against other possibilities. Data unrelated to the central theme should not be included in the paper because they will only confuse readers. These results should be removed no matter how much effort was expended obtaining them.

## **Title**

Once you have your central theme, it is time to create a title for the paper. The title can advertise your methods, your results, or the implications of your results. A title is the paper in one sentence. Put the most important, sexy information in your title. For example, a title of "Steric restrictions in protein folding: an alpha-helix cannot be followed by a contiguous beta-strand" advertises the result. On the other hand, a title of "Interpreting the folding kinetics of helical proteins" talks about the implications of the results. The title "Native proteins are surface-molten solids: Application of the Lindemann criterion for the solid versus liquid state", has both the methods and the implications of the results. Note that "Native proteins are surface-molten solids" is an interpretation of the results. You need to be general and specific at the same time to gain a wider readership.

## Introduction section

Now that the title and central theme are done you can start on the introduction. The first thing you should do is to collect all relevant literature surrounding the central theme. Search and research the literature to cover all the latest and related papers (by citation index or other keyword search about the central theme). Make sure that you have all updated information. Cite all important papers. If you don't cite them, they will not cite you! If you want someone to cite your work, cite his/her work first. The more papers you cite, the more likely some of them will cite and read your paper because experts will pay more attention to papers that cite them. Read other papers carefully in order to avoid citation mistakes.

The most difficult sentence to write is the first sentence, because it determines the flow of your entire introduction. My approach is to link the first sentence to the title of the paper. That is, the first paragraph defines some of the terms used in the title, starting with the most basic or general term. Then, introduce the field of research and its importance. The second paragraph should make a critical survey of the field. The central theme is about the solution, or partial solution, to a problem. This paragraph should point out the existing unsolved problem and describe the difficulties or challenges that inhibit efforts to solve the problem. The third paragraph then introduces the proposed solution and provides a brief discussion of what is to come. You can briefly describe the results you have obtained here and their possible implications. Here is one example:<sup>3</sup>

Assessing secondary structure assignments of protein structures by using pairwise sequence-alignment benchmarks

The secondary structure of a protein refers to the local conformation of its polypeptide backbone. Knowing the secondary structures of proteins is essential for their structure classification<sup>1,2</sup>, understanding folding dynamics and mechanisms<sup>3-5</sup>, and discovering conserved structural/functional motifs<sup>6,7</sup>. Secondary structure information is also useful for sequence and multiple sequence alignment<sup>8,9</sup>, structure alignment<sup>10,11</sup>, and sequence to structure alignment (or threading)<sup>12-15</sup>. As a result, predicting secondary structures from protein sequences continues to be an active field of research<sup>16-18</sup> fifty six years after Pauling and Corey<sup>19-20</sup> first predicted that the most common regular patterns of protein backbones are the  $\alpha$ -helix and the  $\beta$ -sheet. Prediction and application of protein secondary structures rely on prior assignment of the secondary-structure elements from a given protein structure by human or computational methods.

Many computational methods have been developed to automate the assignment of secondary structures. Examples are DSSP, STRIDE, DEFINE, P-SEA, KAKSI, P-CURVE, XTLSSTR, SECSTR, SEGNO, and VoTAP. These methods are based on hydrogen-bond patterns, geometric features, expert knowledge, or their combinations. However, they often **disagree** on their assignments. For example, **disagreement** among DSSP, P-CURVE, and DEFINE can be as large as 25%. More beta sheet is assigned by XTLSSTR and more pi-helix by SECSTR than by DSSP. The **discrepancy** among different methods is caused by non-ideal configurations of helices and sheets. As a result, defining the boundaries between helix, sheet, and coil is problematical and a significant source of **discrepancies** between different methods.

**Inconsistent assignment** of secondary structures by different methods highlights the need for a criterion, or a benchmark, of "standard" assignments that can be used to assess

and compare assignment methods. One possibility is to use the secondary structures assigned by the authors who solved the protein structures. STRIDE, in fact, has been optimized to achieve the highest agreement with the authors' annotations. However, it is not clear what is the criterion used for manual or automatic assignment of secondary structures by different authors. Another possibility is to treat the consensus prediction by several methods as the gold standard. However, there is no obvious reason why each method should weigh equally in assigning secondary structures, or which method should be used in consensus. Other used criteria include helix-capping propensity, the deviation from ideal helical and sheet configurations and structural accuracy produced by sequence-to-structure alignment guided by secondary structure assignment.

In this paper, we propose to use sequence-alignment benchmarks for assessing secondary structure assignments. These benchmarks are produced by 3D-structure alignment of structurally homologous proteins. Instead of assessing the accuracy of secondary-structure assignment directly, which is not yet feasible, we compare the two assignments of secondary structures in structurally aligned positions. We assume that the best method should assign the same secondary-structure element to the highest fraction of structurally aligned positions. Certainly, structurally aligned positions do not always have the same secondary structures. Moreover, different structure-alignment methods do not always produce the same result. Nevertheless, this criterion provides a mean to locate a secondary-structure assignment method that is most consistent with tertiary structure alignment. We suggest that this approach provides an objective evaluation of secondary structure assignment methods.

In this example, the title advertises a method that assesses the assignments of secondary structures. The first paragraph starts a definition of secondary structure (linked to title). The entire paragraph is then talking about the importance of secondary structures. The last sentence is the transition sentence that leads into the computational methods for assigning secondary structures (2<sup>nd</sup> part of the title). Note "computational methods" is at the end of the sentence for emphasis, and to provide a link to the beginning of the second paragraph. The second paragraph focuses on the existing problem of computational methods. The old information "computational methods" gradually changes to "their disagreement". The third paragraph transits the topic from "disagreement" (old information) to "benchmark for assessment" (new information) in the first sentence. Then, existing work in this area is introduced. The fourth paragraph introduces the proposed criterion and talks about the advantage of the new criterion. The fifth paragraph (not shown here) will briefly discuss the result. Each introduction should be a self-contained description of the field in general, the rationale for this proposed work, the result and its implication. After reading the introduction, readers should have a clear idea of your motivation and approach, as well as the results and their implications.

## **Discussion Section**

Now, you are writing the last part of your paper. Many people consider the discussion section the most difficult part to write. Students often fail to separate the results from their interpretations, implications and conclusion. Moreover, they do not consider possible alternative interpretations.

A good discussion usually starts with a review of results obtained and their interpretations. Other good topics for discussion are various trials and tests done, the effects of parameter changes on results, comparisons to other studies, problems yet to be solved, future or on-going work (to stop others from following your immediate, obvious follow-up work). Here is one example of the discussion section of a paper:<sup>4</sup>

One question about the complex homopolymer phase diagram presented here is whether it is caused by the discontinuous feature of the square-well potential. We cannot give a direct answer because the DMD simulation is required to obtain well-converged results for the thermodynamics. However, the critical phenomena predicted for a fluid composed of particles interacting with a square-well potential are as realistic as those predicted for a fluid composed of particles interacting with a LJ potential. Also, an analogous complex phase diagram is found in simulations of LJ clusters. The present results for square-well homopolymers may well be found in more realistic homopolymer models and even in real polymers.

This paragraph explores possible alternative interpretations.

### **Abstract section.**

Once the paper is done you will need to write an abstract. The abstract typically includes the importance of the subject area (linked backward to title), the problem to be investigated, the uniqueness of your approach and the significance, implications and impact of the results.

Here is an example of an effective abstract:<sup>3</sup>

How to make an objective assignment of secondary structures based on a protein structure is an unsolved problem. Defining the boundaries between helix, sheet, and coil structures is arbitrary, and commonly accepted standard assignments do not exist. Here, we propose a criterion that assesses secondary-structure assignment based on the similarity of the secondary structures assigned to structurally aligned residues in sequence-alignment benchmarks. This criterion is used to rank six secondary-structure assignment methods: STRIDE, DSSP, SECSTR, KAKSI, P-SEA, and SEGNO with three established sequence-alignment benchmarks (PREFAB, SABmark and SALIGN). STRIDE and KAKSI achieve comparable success rates in assigning the same secondary structure elements to structurally aligned residues in the three benchmarks. Their success rates are between 1-4% higher than those of the other four methods. The consensus of STRIDE, KAKSI, SECSTR, and P-SEA, called SKSP, improves over assignments over the best single method in each benchmark by an additional 1%. These results support the usefulness of the sequence alignment benchmarks as the benchmarks for secondary structure assignment.

The first two sentences state the problem. The third sentence is the proposed solution. These sentences are followed by the results. The abstract should end with a summary statement.

### **Summary**

1. Take writing seriously. It is an integral part of scientific research. No one will have the patience to read a poorly written paper, and your paper is worthless if no one reads it.
2. Don't write a paper unless the study is comprehensive and you have tried all possible ways to support your conclusion.
3. Rethink and rationalize: why do this research? What was done? What is the most important discovery? Why take this path? Why use these parameters? What has been done before (updated literature search)?

4. Be extremely critical of your own work. Weakness is discovered only if you are critical. Removing the weakness will make your paper much stronger. Many of my papers have changed significantly during revision. Some calculations had to be performed again and checked. A paper becomes more meaningful only after results are understood and organized well.
5. Prove your case beyond reasonable doubt. If you are not convinced, no one will be. If you have any doubt at all, reexamine all of your data and check all the details. Do not accept any significant discovery quickly or easily.
6. Use the highest standard, rather than bad examples, to guide your writing. Never hide anything. Never underestimate the intelligence of other scientists. Make your study easily reproducible. Put all materials/data online and make them available to anyone interested. Establish brand name for your papers.
7. Transition from old to new information from the beginning (title) to the end (discussion or conclusion). Never introduce new information at the beginning of a sentence. Never introduce new terms without defining them first.
8. It is unethical to copy others' sentences. A copied sentence often disrupts the flow of information between the sentences you have written. If it is absolutely necessary to use the original sentence from other papers, add quotation marks and cite the reference.
9. Always have a topic sentence at beginning of a paragraph to illustrate the topic of entire paragraph and a transition sentence at the end of paragraph to link next paragraph. Let your paper be an integrated whole rather than a random collection of sentences. A linked article from title to discussion will make reading enjoyable for the readers. If the readers enjoy reading your paper, they will cite it in their own work.
10. Write, rewrite, and rewrite again. No one gets it right the first time.

## **Epilogue**

A good paper, like a good novel, needs to conceive a plan and move the story with an unexpected ending. Of course, this requires that you have exciting results to reveal. The purpose of this article is to enable you to better advertise your results through clarity and convincing evidence. I am strongly against false and/or exaggerated claims that invariably turn off readers. In fact, one should avoid using affectionate adjectives like "exciting" or "remarkable" in scientific writing. Let readers feel the excitement via reading the description of your results and their implications. Let the facts speak for themselves.

## **Acknowledgments**

Some of the examples shown above are from the paper "The Science of Scientific Writing" by G. D. Gopen and J. A. Swan, *American Scientist*, 78, 550-558, 1990. I have also benefited greatly from Professor Gopen's annual workshop at Duke University in 1995. I would like to thank my former advisors Professors Martin Karplus (Harvard University), George Stell (SUNY Stony Brook), Harold L. Friedman (SUNY Stony Brook) and Carol Hall (NC State University) for their guidance and encouragement. Without them, I would not have nearly as much valuable writing experience and practice. Finally, I would like to thank my students and postdocs. Their contributions to science allow me to continue the business of writing, whether it is a paper,

grant proposal, or review. Some of their papers (draft or final versions) are used as examples here. I also would like to thank Dr. Liping Zhao, Dr. Eshel Faraggi, Aaron T. Woodsworth and many other on-line readers for their valuable comments and advice.

*June 19, 2007 in Indiana, USA.*

**Reference:**

- [1] G. D. Gopen and J. A. Swan, *American Scientist*, 78, 550-558 (1990)
- [2] H. Zhou and Y. Zhou, *Bioinformatics* 21, 3615--3621 (2005).
- [3] W. Zhang, K. Dunker, and Y. Zhou, *Proteins*, in press (2007).
- [4] Y. Zhou, C. K. Hall, and M. Karplus, *Phys. Rev. Lett.* 77 , 2822 (1996).