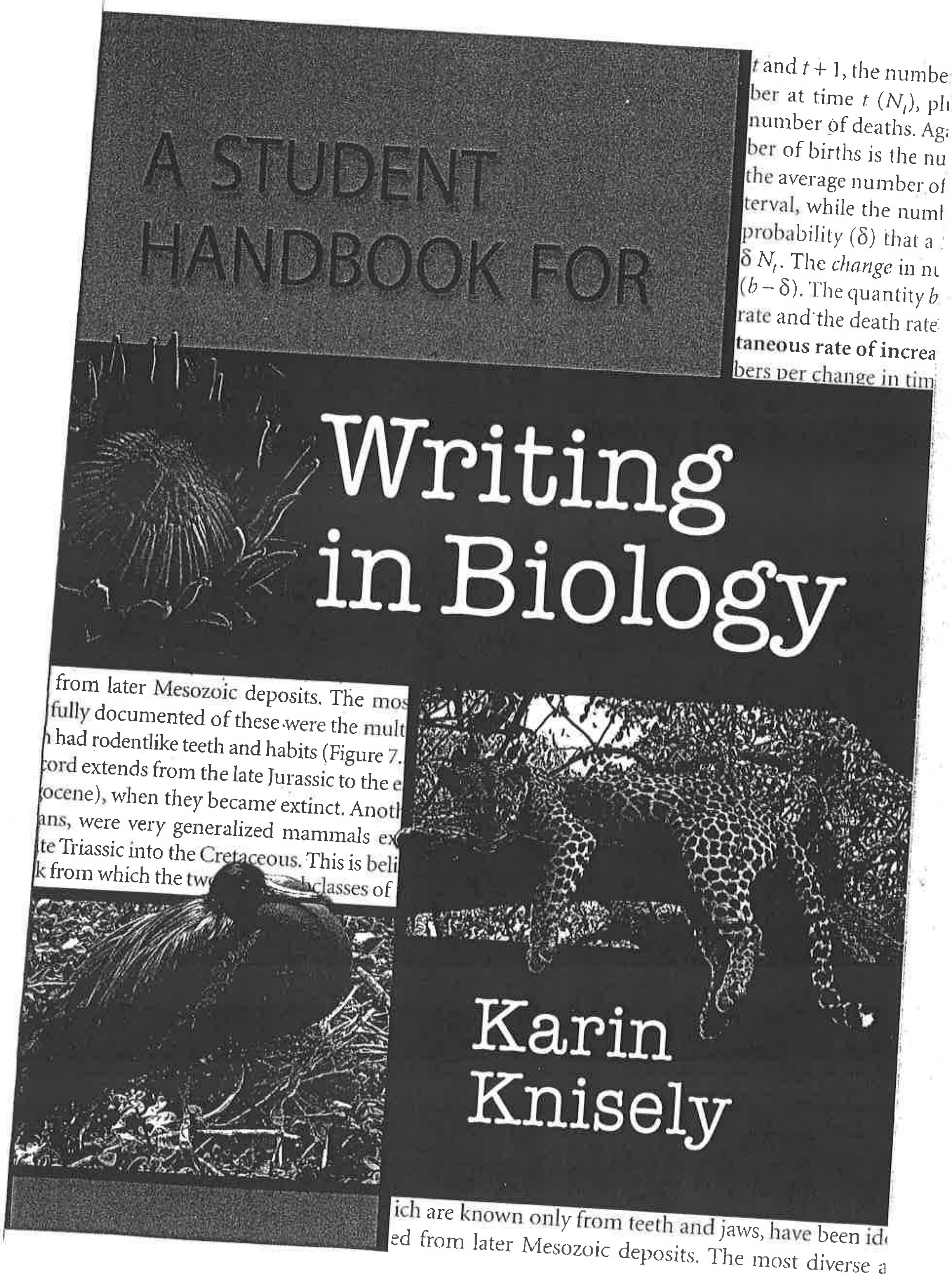
A STUDENT HANDBOOK FOR
Writing in Biology
Karin Knisely

Chapter 1

# The Scientific Method

Trying to understand natural phenomena is human nature. We are curious about why things happen the way they do, and we expect to be able to understand these events through careful observation and measurement. This process is known as the scientific method, and it is the foundation of all knowledge in the biological sciences.

## An Introduction to the Scientific Method

The scientific method involves a number of steps:

- Asking questions
- Looking for sources that might help answer the questions
- Developing possible explanations (hypotheses)
- Designing an experiment to test a hypothesis
- Predicting what the outcome of an experiment will be if the hypothesis is correct
- Collecting data
- Analyzing data
- Developing possible explanations for the experimental results
- Revising original hypotheses to take into account new findings
- Designing new experiments to test the new hypotheses (or other experiments to provide further support for old hypotheses)
- Sharing findings with other scientists

Most scientists do not rigidly adhere to this sequence of steps, but it provides a useful starting point for how to conduct a scientific investigation.

## *Ask a question*

As a biology student, you are probably naturally curious about your environment. You wonder about the hows and whys of things you observe. To apply the scientific method to your questions, however, the phenomena of interest must be sufficiently well defined. The parameters that describe the phenomena must be measurable and controllable. For example, let's say that you learned that:

> Dwarf pea plants contain a lower concentration of the hormone gibberellic acid than wild-type pea plants of normal height.

You might ask the question:

> Does gibberellic acid regulate plant height?

This is a question that can be answered using the scientific method, because the parameters can be controlled and measured. On the other hand, the following question could not be answered easily with the scientific method:

> Will the addition of gibberellic acid increase a plant's sense of well-being?

In this example, "a sense of well-being" is not something that can be measured or controlled.

## *Look for answers to your question*

There is a good chance that other people have already asked the same question. That means that there is a good chance that you may be able to find the answer to your question, if you know where to look. Secondary references such as your textbook, encyclopedias, and information posted on the websites of university research groups, professional societies, museums, and government agencies are usually easier to comprehend than journal articles and may be good places to begin finding answers (see the section "Understand your topic" in Chapter 2). Curiously, attempts to answer the original question often result in new questions, and unexpected findings lead to new directions in research. By reading other people's work, you may think of a more interesting question, define your question more clearly, or modify your question in some other way.

## *Turn your question into a hypothesis*

As a result of your literature search or conversations with experts, you may now have a tentative answer to your original (or modified) question. Now it is time to develop a hypothesis. A **hypothesis** is a possible explanation for



something you have observed. **You must have information before you can propose a hypothesis!** Without information, your hypothesis is nothing more than an uneducated guess. That is why you must look for possible answers before you can turn your question into a hypothesis.

A useful hypothesis is one that can be tested and either supported or negated. A hypothesis can never be *proven* right, but the evidence gained from your observations and/or measurements can *provide support for* the hypothesis. Thus, when scientists write papers, they never say, "The results prove that... ." Instead, they write, "The results suggest that..." or "The results provide support for... ."

You might transform your question "Does gibberellic acid regulate plant height?" into the following testable hypothesis:

> **Good:** The addition of gibberellic acid to dwarf plants will allow them to grow to the height of normal, wild-type plants.

This hypothesis provides specific expectations that can be tested. In contrast, the following hypothesis is not specific enough:

> **Vague:** The addition of gibberellic acid will affect the height of dwarf plants.

## *Design an experiment to test your hypothesis*

In an **observational study**, scientists observe individuals and measure variables of interest without trying to control the variables or influence the response. While observations provide important information about a group, it is difficult to draw conclusions about cause and effect relationships because multiple factors affect the response. That's the main reason why scientists conduct experiments. **Experiments** are studies in which the investigator imposes a specific treatment on a person or thing while controlling the other factors that might influence the response.

The first step in designing an experiment is to determine which variables might be influential. Of those variables, only one may be manipulated in any given experiment; the others have to remain constant. The individuals in the experiment are then divided into treatment and control groups. The treatment group is subjected to the independent variable and the control group is not; all other conditions are the same for the two groups. If the hypothesis is supported, the individuals in the treatment group will respond differently from those in the control group. If there is no difference in response between the treatment and control groups, the so-called null hypothesis is supported. Having enough replicates lends assurance that the results are reliable.

**Define the variables.** Variables are commonly classified as independent or explanatory variables, dependent or response variables, and controlled variables. The *one* variable that a scientist manipulates in a given experiment is called the **independent variable** or the explanatory variable, so-called because it "explains" or influences the response. It is important to manipulate *only one* variable at a time to determine whether or not a cause and effect relationship exists between that variable and an individual's response. The other variables that may affect the response must be carefully controlled so that they do not confound the relationship between the independent variable and the dependent variables.

**Dependent variables** are those affected by the imposed treatment; in other words, they represent an individual's response to the independent variable. Dependent variables are variables such as size, number of seeds produced, and velocity of an enzymatic reaction, which can be measured or observed.

The hypothesis proposed earlier involves testing whether there is a cause and effect relationship between gibberellic acid (GA) treatment and plant height. GA level is the variable that will be manipulated; plant height is the response that we'll measure. Because plant height is affected by many other factors such as ambient temperature, humidity, age of the plants, day length, amount of fertilizer, and watering regime, however, we must keep these controlled variables constant so that any differences in response can be attributed to the GA treatment.

**Set up the treatment and control groups.** The individuals in the experiment are assigned randomly to either a treatment group or a control group. Those in the treatment group will be subjected to the independent variable, while those in the control group will not. Depending on the hypothesis, the control group may be subdivided into positive and negative controls. Negative controls are not treated with the independent variable and are not expected to show a response. Positive controls are individuals not treated with the independent variable, but represent a reference for treatment groups that demonstrate a response consistent with the hypothesis.

**Hypothesis:** Adding GA to dwarf plants will allow them to grow to the height of normal, wild-type plants.

**Treatment group:** Dwarf plants + GA

**Control groups:**

**Negative:** Dwarf plants + no GA (substitute an equal volume of water)

**Positive:** Wild-type plants + no GA



**Determine the level of treatment for the independent variable.** How much GA should be added to the dwarf plants in the treatment group to produce an increase in height? Too little GA may not effect a response, but too much might be toxic. To determine the appropriate level of treatment, consult the literature or carry out a preliminary experiment. The level may even be a range of concentrations that is appropriate for the biological system.

**Provide enough replicates.** A single result is not statistically valid. The same treatment must be applied to many individuals and the experiment must be repeated several times to be confident that the results are reliable.

**Make predictions about the outcome of your experiment.** Predictions provide a sense of direction during both the design stage and the data analysis stage of your experiment. For each treatment and control group, predict the outcome of the experiment if your hypothesis is supported. You may also choose to propose a **null hypothesis**, which states that the treatment has no effect on the response.

**Hypothesis:** Adding GA to dwarf plants will allow them to grow to the height of normal, wild-type plants.

**Treatment group:** Dwarf plants + GA

**Prediction if hypothesis is supported:** Dwarf plants will grow as tall as wild-type plants + no GA.

**Null hypothesis:** Dwarf plants will not grow to the height of wild-type plants.

**Negative control:** Dwarf plants + no GA

**Prediction:** Dwarf plants will be short.

**Positive control:** Wild-type plants + no GA

**Prediction:** Wild-type plants will be tall.

## *Record data*

Scientists record procedures and results in a laboratory notebook. The type of notebook (bound or loose leaf, with or without duplicate pages) may be prescribed by your instructor or the principal investigator of the research lab. More important than the physical notebook, however, is the detail and accuracy of what's recorded inside. For each experiment or study, include the following information:

- Investigator's name
- The date (month, day, and year)
- The purpose
- The procedure (in words or as a flow chart)
- Numerical data, along with units of measurement, recorded in well-organized tables
- Drawings with dimensions and magnification, where appropriate. Structures are drawn in proportion to the whole. Parts are labeled. Observations about the appearance, color, texture, and so on are included.
- Graphs, printouts, and gel images
- Calculations
- A brief summary of the results
- Questions, possible errors, and other notes

When deciding on the level of detail, imagine that, years from now, you or someone else wants to repeat the experiment and confirm the results. The more information you provide, the easier it will be to understand what you did, what problems you encountered, suggestions for improving the procedure, the results you obtained, how you summarized the data, and how you reached your conclusions.

### Summarize numerical data

The raw data in lab notebooks are the basis for the results published in the primary and secondary literature. Published results, however, usually represent a *summary* of the raw data by the author, who is both knowledgeable about the subject and intimately familiar with the experiment. We rely on the author's experience and integrity to reduce the original data to a more manageable form that is an honest representation of the phenomenon and which lends itself to interpretation.

How the author presents data in the Results section depends in part on the scope of the question asked at the beginning. Broad questions about a population involve **statistical inference**, whereby results from a sample or subset of the population are applied to the whole. Because a different sample may produce different results, the author includes a statement about the reliability of his or her conclusions using appropriate statistical language. On the other hand, narrower questions about a specific situation may be answered from the data at hand. For example, questions such as "Which fraction of a purification procedure contains the most enzymatic activity?" or "Which medium produces the highest concentration of bacteria?" can be



answered from the collected data and require no inference about a larger population. When the data are consistent from one experiment to the next, scientists gain confidence that their conclusions are valid.

When *you* are given the task of summarizing the raw data, first distinguish between trustworthy and erroneous data. Erroneous data include results obtained by dubious means, for example, by not following the procedure, using the equipment improperly, or making simple arithmetic errors. Trustworthy data include results obtained legitimately, but which may still have quite a bit of unexplained variability. If time permits, repeat the experiment to determine possible sources of variability and make changes in the procedure if necessary.

Once you've identified which data are reliable, graph them. It is easier to spot patterns and outliers on a graph than in a table. Furthermore, graphs are used to check assumptions for certain statistical methods. Use bar graphs when one of the variables is categorical (i.e., it has no units of measurement). Use scatterplots and line graphs when both variables are quantitative. Look for an overall trend as well as deviations from the trend. Reduce the data by taking the average (mean) and express variability, where appropriate, in terms of standard deviation or standard error. Never eliminate data without a good reason.

### Analyze the data

Once you have a visual summary of the raw data, look for relationships between variables. Do the results match the predictions if the hypothesis is supported? If so, then compare your results to those in the primary references you consulted to develop your hypothesis in the first place. Comparable data from different studies help researchers gain assurance that their conclusions about a particular phenomenon are valid. When analyzing data, however, do not let your predictions affect your objectivity. Do not make your results fit your predictions—instead, modify your hypothesis to fit your results. What is learned from a negated hypothesis can be just as valuable as what is learned from a "successful" experiment.

Keep in mind that there may be no difference between the control and the experimental treatments. If there was no difference, say so, and then try to develop possible explanations for these results.

### Try to explain the results

Once you have summarized and analyzed the data, you are ready to develop possible explanations for the results. You previously found information on your topic when you developed your hypothesis. Return to this material to try to explain your results. Do your results agree with those of other researchers? Do you agree with their conclusions? If your results do not agree, try to determine why not. Were different methods, organisms, or conditions employed? What were some possible sources of error?

You should realize that even some of the most elementary questions in biology have taken hundreds of scientists many years to answer. One approach to the problem may seem promising at first, but as data are collected, problems with the method or other complications may become apparent. Although the scientific method is indeed methodical, it also requires imagination and creativity. Successful scientists are not discouraged when their initial hypotheses are discredited. Instead, they are already revising their hypotheses in light of recent discoveries and planning their next experiment. You will not usually get instant gratification from applying the scientific method to a question, but you are sure to be rewarded with unexpected findings, increased patience, and a greater appreciation for the complexity of biological phenomena.

### *Revise original hypotheses to take new findings into account*

If the data support the hypothesis, then you might design additional experiments to strengthen the hypothesis. If the data do not support the hypothesis, then suggest modifications to the hypothesis or use a different procedure. Ideally, scientists will thoroughly investigate a question until they are satisfied that they can explain the phenomenon of interest.

### *Share findings with other scientists*

The final phase of the scientific method is communicating your results to other scientists, either at scientific meetings or through a publication in a journal. When you submit a paper to refereed journals, it is read critically by other scientists in your field, and your methods, results, and conclusions are scrutinized. If any errors are discovered, they are corrected before your results are communicated to the scientific community at large.

**Poster sessions** are an excellent way to share preliminary findings with your colleagues. The emphasis in poster presentations is on the methods and the results. The informal atmosphere promotes the exchange of ideas among scientists with common interests. See Chapter 7 on how to prepare a poster.

**Oral presentations** are different from both journal articles and poster sessions, because the speaker's delivery plays a critical role in the success of the communication. See Chapter 8 for tips on preparing and delivering an effective oral presentation.

Chapter 2

# Developing a Literature Search Strategy

The development of library research skills is an essential part of your training as a biology student. A vast body of literature is available on just about every topic. Finding exactly what you need is the hard part.

In biology, sources are divided broadly into primary and secondary references. **Primary references** are the research articles, dissertations, technical reports, or conference papers in which a scientist describes his or her original work. Primary references are written for fellow scientists—in other words, for a specialized audience. The objective of a primary reference is to present the essence of a scientist's work in a way that permits readers to duplicate the work for their own purposes and to refute or build on that work.

**Secondary references** include encyclopedias, textbooks, articles in popular magazines, and information posted on the websites of professional societies, government agencies, and other scientific organizations. Secondary references are based on primary references, but they address a wider, less-specialized audience. In secondary references, there is less emphasis on the methodology and presentation of data. Instead, the results and their implications are described in general terms for the benefit of non-specialist readers.

You will delve into the biological literature when you write laboratory reports, research papers, and other assignments. Although secondary references provide a good starting point for your work, it is important to be able to locate the primary sources on which the secondary sources are based. Only the primary literature provides you with a description of the methodology and the actual experimental results. With this information, you can draw your own conclusions from the author's data.

Although initially it may be difficult to read primary literature, it will become easier with practice, and the rewards are well worth it. One benefit of *reading* research articles is that you will become a better *writer*. Through reading, you become familiar with the writing style and overall structure of research articles, so that you have a model when you write your own lab