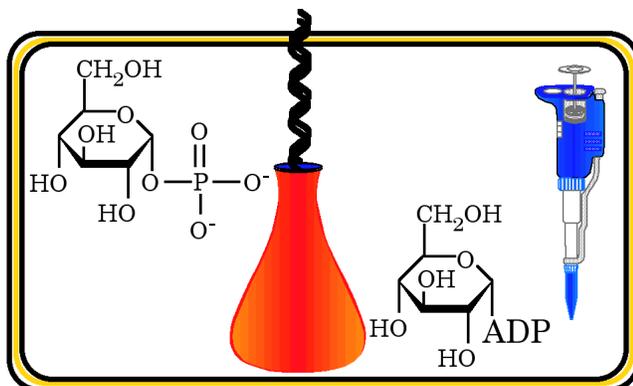




*Department of
Chemistry
and Biochemistry*



Chemistry 472B

BIOTECHNOLOGY

LABORATORY

MANUAL

Mark Brandt, Ph.D.
Second edition
January, 2002

Table of Contents

Introduction to the Laboratory.....	1
General Information: Keeping a laboratory notebook.....	3
General Information: Laboratory reports.....	5
Basic Laboratory Techniques: Pipetting.....	14
Basic Laboratory Techniques: Measuring absorbance.....	16
Basic Laboratory Techniques: Performing dilutions.....	18
Basic Laboratory Techniques: Buffers.....	20
General Information: Molecular biology.....	22
General Information: Cell genotypes.....	24
Methods: Polymerase chain reaction.....	26
Methods: Plasmid preparation.....	29
Procedure for Plasmid miniprep:.....	31
Methods: Ligation.....	32
Methods: Competent cell preparation and Transformation.....	34
Methods: Selection and Screening.....	35
General Information: Restriction digestion.....	38
Electrophoretic Techniques: SDS PAGE.....	39
Electrophoretic Techniques: Agarose gel electrophoresis.....	44
Electrophoretic Techniques: DNA sequencing.....	46
Electrophoretic Techniques: Western blotting.....	49
Procedure for running the SDS PAGE and blotting:.....	50
Western blotting incubations.....	52
Methods: Protein assays.....	56
Methods: Protein purification.....	58
Procedure for pouring a column.....	63
Methods: Cell lysis for protein purification.....	65
Methods: ADP-glucose pyrophosphorylase assay.....	66
ADP-glucose pyrophosphorylase assay procedure.....	68
Methods: Protein crystallography.....	70
Introduction to enzyme kinetics.....	74
Definitions.....	83

Other useful Information:

Biochemistry Stockroom: MH-277

Chemistry & Biochemistry Office: MH-580

Introduction to the Laboratory

This course is intended to introduce you to some of the most widely used experimental procedures in biotechnology, including DNA isolation, manipulation, cloning, and mutagenesis, and protein purification and characterization. You will also gain some familiarity with some of the types of equipment frequently used in biochemistry and molecular biology.

Research is often a collaborative effort in which many people may contribute to different aspects of a given project. Few papers in the scientific literature are written by single authors; the vast majority of papers have at least two authors, and many papers have more than ten contributing people. In part to provide a more authentic experience of actual lab work, experiments will be done in groups of two or three. You may choose partners, or you can ask to be assigned to a group.

The biotechnology laboratory course, like all laboratory courses, is an exploration of procedures. This means that, in order to get full benefit from the course, you will need to read the manual, and you should participate as much as possible in the discussions. You should ask questions in or out of class. You should also try to participate in the actual lab work (and not simply allow your lab partners to do things for you). The more effort you put into the course work, the more you will learn. The class is an opportunity to learn valuable skills; take full advantage of it!

Prior to many of the lab periods, you will need to spend some time reading the Laboratory Manual. This reading will provide background information and an outline of the procedures to be performed. If you do not do this, you will find yourself wasting large amounts of class time, and annoying both your lab partners and your instructor. To encourage your understanding of the material, you will have problem sets that cover material related to the planned experiments.

The biotechnology laboratory is conducted as a “directed” research project. This means that although the general procedures are well established, the overall goal of each experiment is the acquisition of **new** information. Because of the nature of scientific research, predicting the outcome of experiments that have not previously been performed is difficult. It may therefore be necessary to design new experiments based on the results of previous ones, or to repeat experiments that yielded uninterpretable or ambiguous results. If you expect to know what exactly you will be doing weeks in advance, you will be in for a shock. On the other hand, if you approach the course with an open and flexible mindset, you will learn how research is performed in a biotechnology laboratory.

SAFETY: Laboratories contain hazards of various kinds. Everyone is **required** to wear **closed-toe shoes, long pants, goggles with side shields**, and a **lab coat** while performing laboratory work. Students should not work in the laboratory if the instructor is not present.

Some of the chemicals used are toxic, mutagenic, or teratogenic. If you believe that you have a health condition that puts you at exceptional risk, or believe yourself to be pregnant, please see your instructor in private to discuss the issue. If you have questions or concerns about exposure to hazardous chemicals, please consult your instructor or go to the Research and Instructional Safety Office (MH-557).

PHILOSOPHICAL ISSUES: Scientific research involves an exploration of the unknown. In some classes, a question has a single “correct” answer, which is known to the instructor, and imparted to the students. In research, however, the correct answer is rarely known ahead of time, and must instead be inferred from the experimental results. Researchers must therefore become accustomed to some level of uncertainty about the “correct” answer to any experimental question, and must always remain open to experimental evidence that contradicts a hypothesis that has arisen from previous experiments. Your task as a scientist will be to consider **your** data, and to attempt to interpret it. In this context, “wrong” answers are answers that are contradicted by your data or that do not arise logically from the data you have collected.

This uncertainty as to the “correct” answer means that you must be careful when reporting what you did and what you observed, especially if you observe something unexpected. Humans are good at fooling themselves; you need to guard against reporting what you *expect to see* rather than what you actually *did* see. Scientific fraud, in which people intentionally report false data, is considered very serious because it results in a difficult-to-overcome belief in an answer that conflicts with the truth. You will occasionally see retractions, in which a scientist publishes a statement that information in a previously published paper is the result of an artifact, and is not a reflection of the “correct” answer. Avoiding the embarrassment of publishing a retraction is one reason for the care that people take in performing experiments and in interpreting the results.

Another ethical issue is the proper citation of the sources of information you use for any scientific writing. You should always properly reference the authors of papers or books you consult. It also means that you should cite the inventors of methods that you use for your experiments. If you do not, in effect you are claiming credit for work performed by others.

General Information: Keeping a Laboratory Notebook

All students will be **required** to maintain a laboratory notebook. The notebook will be used for the recording of laboratory data and calculations, and will be critically important for writing your lab reports.

The purpose of a laboratory notebook is to allow anyone with some biochemical knowledge to understand **exactly** what you did. You need to record the information in sufficient detail so as to be able to repeat it, and you must be able to understand exactly what your results were. You will need good notes to be able to write your lab reports; in addition, as your understanding of biochemistry improves, your notebook should allow you to figure out why some parts of your experiments did not work as expected.

Companies that perform research require their employees to keep proper notebooks. In many companies, company policy dictates that any work not recorded in the notebook was, by definition, never actually performed. As a result, the work must be repeated, which tends to have deleterious effects on the career opportunities of the employees involved. In cases of disputes as to priority, notebook dates are sometimes used to indicate exactly when an experiment was performed. Ownership of patents (and in some cases large amounts of money) can therefore be critically dependent on keeping a proper notebook. Instruction in keeping laboratory notebooks is therefore a major part of most laboratory courses.

In your notebook, each experiment should begin with a **title**, a **date**, and a statement of the **objective** of the planned work. You should also **record exactly what you did at each step** (being sure to mention anything that you did that differed from the information in the Manual). In addition, you should record any numerical information, such as the weights of reagents used, absorbance readings, enzyme activities, protein concentrations, and buffer concentrations.

Most experiments will extend over several days, and over several pages in your notebook. To allow you to keep track of what you have done, you should include the day's date at the top of each page. Including sub-titles for each page may make it easier to keep track of what you did at each step.

Everything you do should be recorded **directly** into your lab notebook in **pen**. If you make a mistake, draw a line through it, and write the correction next to the mistake. (It may turn out that the original information was correct after all, so do not obliterate the original information by erasing it, or by removing the page from your notebook.) Any calculations performed should be written directly into your book. Any work done on a computer, or printouts from laboratory instruments, should be taped directly into your lab notebook.

Writing important information on scrap paper, and then recording it in your notebook later is **not** acceptable. **If you are writing something while in the laboratory, you should be writing it directly into the notebook.**

At each step in your experiment (after each assay or measurement), in addition to the results, record your thoughts regarding the experiment and how you think it is going. Record your mistakes, and your attempts to rectify them. Record the calculations involved in any type of data analysis, as well as explanations for both what you did and what you think it means. A research project is a journey into the unknown; your laboratory notebook is usually your only guide through the forests of uncertainty.

It is also a good idea to look over your notebook periodically during the semester, and make notes of things that you do not understand, so that you can ask questions before the lab reports are due.

Do not say “well, I will remember what this means”; instead, **write it down!** Do not say “I will remember what I was thinking while I did this experiment”; instead, **write it down!** If you use your lab notebook properly, you will find that writing your lab reports is much easier, and you will be developing good habits for the future.

General Information: Laboratory Reports

The laboratory reports are major written assignments, due at intervals during the semester. The laboratory reports should be written in the form of a scientific paper. To help you learn to write a scientific paper correctly, the laboratory reports will be due in two sections, with the second report building on the first one. The second report should contain all of the information from the previous report, plus all of the new work. You should incorporate the instructor's suggestions, using these comments to guide you in the generation of the new sections. Note that the second laboratory report will be graded more stringently than the first one: you are expected to learn from your mistakes!

All of the laboratory reports are expected to be well-formatted, word-processed documents, written in standard scientific American English. The use of spell-checkers and grammar-checkers is strongly recommended. (Note: the Appendix does not have to be neatly as neatly formatted as the rest of the report, and, if necessary, may be handwritten.)

In scientific research, results are reported to the world in the form of scientific papers published in the peer-reviewed scientific literature. These papers are not only important in disseminating the results of the research, but are critical for essentially all aspects of career advancement for the scientists involved. Learning to write a proper scientific paper is therefore an important part of the education of all scientists.

Scientific papers are expected to be written in a well-defined format. The overall format is generally similar in all journals, although the specific details vary somewhat. In this class, the laboratory reports should be in the form of a paper in the *Journal of Biological Chemistry*. Looking for papers in the *Journal of Biological Chemistry* to use as examples is strongly recommended. (Note that the formatting that you should attempt to emulate applies to **content**; you do **not** need to spend time generating the specific page layout of a *Journal of Biological Chemistry* paper. The preferred page layout for lab report submission has the body of your paper in double-spaced text.) In keeping with this formatting, the report should have the following sections: Title Page, Abstract, Introduction, Materials and Methods, Results, Discussion, and Appendix.

Many scientists have their own preferred ways of writing papers. Most scientists, however, use an iterative process of writing, in which they write the paper, and then rewrite it several times before submitting the paper to the journal for review and (hopefully) publication. In addition, most papers are written in an order that deviates from the final format. A common procedure is to write the Methods section first, followed by the Results section. The Methods section is a simple description of procedures and can be written before the experimental results have been analyzed. The Results section contains the observations that constitute the study to be published. Once these sections are written, most people write an incomplete draft of the Discussion section that explains the results in the context of the paper.

After the Results section is written, and some thought put into interpreting the results, most people write the Introduction. When writing your Introduction, you should think of the Introduction as an episode of “Jeopardy”: the Results are the answers, and now it is necessary to come up with corresponding questions. You do not need to write the “questions” in the form of a question, but you should think about raising questions in the readers’ mind that you will then answer in the Results and Discussion sections.

After writing the Introduction, you should then look at how you have written Introduction, and rewrite the Results section to more clearly answer the questions raised in the Introduction, and then write the Discussion to interpret and clarify the answers. When properly done, each rewrite acts as an impetus for the rewrite of a different section, until all of the sections fit together into a coherent story.

Finally, after all of the other sections have been written, you can write the abstract, by extracting the most important information from each section and combining the information into a single paragraph.

You should keep these general concepts for writing a paper in mind while considering the content of each section. The content of each section of a scientific paper is discussed below. (Remember that you may be well advised not to write the paper in this order.)

Title Page: This should include the title of your report, the author’s name (*i.e.* your name), your lab partner’s name(s), and your address (your e-mail address is sufficient).

Abstract: This should be a *brief* version of the **entire** paper. It therefore should include a brief introduction, methods, results, and discussion, expressed in ~200 words. This truncation is normally achieved in part by greatly abbreviating the methods portion, unless the methods involved are novel or are crucial to understanding the findings presented.

Thousands of papers are published every week. Most literature database search engines include the title and abstract, but do not include the remainder of the paper. In writing the abstract, remember that the vast majority of readers probably will **not** read the paper, because they lack the time. Therefore, in order to present your information to the largest possible audience, you need to have an abstract that is clearly written, that is understandable without having to read the paper, and that contains all of the relevant findings from the paper.

The abstract must include the **overall conclusions** from the paper; once again, this is important because you **want** people to know what you have discovered. Your job/grant funding/promotions/fame and fortune/ability to do more experiments/ability to retire to the exotic locale of your choice may depend on having people understand what you have done. (This applies to the entire paper, but the abstract tends to be at least skimmed by vast numbers of people who will never read the paper.)

Introduction: This section should include background information setting up the scientific problem you are attempting to address and the overall goal of the experiments you performed. What is the hypothesis you are testing? What directly relevant information is necessary to understand this hypothesis and why is it important? What is not known that you hope to address? What are you planning to attempt to accomplish? (Very briefly) How did you accomplish this?

In writing an introduction, you are attempting to orient the readers, so that they will know what to consider as they read the rest of the paper. This means that you should carefully consider whether you are presenting information that is irrelevant or misleading. If you discuss an issue related to your protein in the introduction, the reader will expect you to address that issue in the remainder of your paper. In addition, after having read your introduction, the reader should have an appreciation of the questions you were attempting to address with your experiments and why these questions are important. If someone can read your introduction without wanting to read the rest of your paper to find the answer to the provocative questions that you raised, you have not written your introduction properly!

Methods: This should be a *concise* summary of what you did. It should include enough detail so that any reasonably intelligent biochemist could repeat your work, but not a minute-by-minute recitation of the hours you spent performing the experiment. One common mistake is to include information that belongs in the Results section; the Methods section is for **methods**. For example, a description of a protein assay should describe the procedure used, but generally should **not** include a list of the samples measured in the assay. On the other hand, a common mistake is to fail to include some methods, such as the techniques used to analyze the data obtained during the study.

When most people read a paper, they tend to *skip* the Methods section unless they need to know exactly how an experiment was performed. This means that they will not read the Methods unless they do not believe your description in the Results section, or because they work in the field and want to see if you used a novel technique. Because many people skip the Methods section, the Methods section should only be a description of the methods used. With the possible exception of one-time events such as plasmid constructions, it is rarely a good idea to include results in the Methods section. If you do include results in the Methods section, these results should be at least summarized in the Results section also.

The Methods section should also contain the source of the important reagents and identifying information for any equipment used. Because research reagents of high quality are available from many vendors, the precise source of most reagents is much less important than it once was. It is common practice, however, to state in the Methods section that, for example, “the ADP-glucose pyrophosphorylase expression vector was a generous gift of Dr. C. Meyer”.

Results: This section should be a description of what you did **in words**, illustrated with figures and tables. It is **not** enough merely to have several figures; you need to

explain what each figure means. Try to avoid merely listing results in the text; instead, explain the findings and briefly fit them into the overall context of the paper.

For each set of experiments, you need to consider the following questions: What are you doing? Why and how are you doing it? What was the rationale for the methods you employed? What is the point of the experiment you are about to describe? What strategy are you using to address the experimental question you are asking?

None of your answers to the above questions should be lengthy, but you do need to consider these questions in writing your report. It may be totally obvious to **you** why you performed your brilliant experiment, but unless you explain the purpose and rationale behind the experiment, your flawless reasoning may not be obvious to your readers.

Remember that you are telling a story to people who have not done the experiments. You cannot assume that the reader will know what you are doing and why. In addition, you are telling a story that people will be predisposed to **disbelieve**. You therefore need to present your information as clearly as possible. If you do so, people will (at worst) understand what they are criticizing, and (at best) see that you have put enough thought and effort into your work as to make it likely that you are trustworthy.

What data do you need to report? Do **not** report data merely because it is available. Instead, report data to make a point. You are trying to tell a factual story. This means that you **cannot** lie to your readers. On the other hand, if you perform an irrelevant experiment, reporting the results may be confusing. For example, if you perform five SDS-PAGE electrophoresis experiments that show essentially the same results, you do not need to include the results of each individual gel.

In reporting the results of an experiment that yielded numerical data, it is poor writing technique to simply list in the text the same values listed in a table or shown in a graph. The raw numbers are meaningless unless put into context. In other words, cite in the text only the *important* numbers, and explain *why* these values are important.

For reporting numbers in the text, convert the numbers to reasonable values. A number such as 0.0014567 mg/ μ l is **not** reasonable for two reasons: 1) converting the value to 1.4567 mg/ml results in a number that is much easier to read, and 2) the number of significant figures reported seems excessive (unless you **really** believe that your experiment was accurate to five significant figures).

As an example, you will be writing a description of ADP-glucose pyrophosphorylase purification and enzyme assays in your Results section. You should consider the following in writing this section.

Restriction analysis: why did you digest the DNA with these enzymes? What band sizes did you obtain? What band sizes did you expect to obtain? What does this tell you about whether the plasmid is the correct one?

Purification: Why did you perform the purification? What strategy did you employ for the purification? Why did you use the steps you used and not others? During the purification, what step resulted in the greatest purification? When did you observe the ADP-glucose pyrophosphorylase to elute from the column? Was this expected, unexpected, or did you have no basis for making a prediction? Is there a figure you could generate to clarify your results? (Is a figure necessary to clarify your results?) Based on your data, was your purification successful or unsuccessful? Why? Do you have any data other than fold-purification to indicate whether your purification was successful? How did your purification compare to literature values obtained for similar proteins?

One important feature in scientific papers describing protein purification is a table that presents data concerning the success of the purification procedure. This table usually has two goals: 1) to measure the removal of contaminants during the purification, and 2) to assess the efficiency of each step, and the overall efficiency of the entire purification process.

The typical purification table for an enzyme gives the values listed for each step during the purification. In order to set up a purification table, you will need to know the enzyme activity (corrected to nmol/min of product formed per μl of enzyme solution). For the ADP-glucose pyrophosphorylase assay, the **activity** can be calculated by:

$$\text{Activity} \left(\frac{\text{nmol}}{\text{min} / \mu\text{l}} \right) = \frac{\text{CPM}}{\text{standard CPM} / \text{nmol}} \cdot \frac{1 \text{ ml total}}{0.5 \text{ ml measured}} \cdot \frac{1 \text{ min}}{10 \text{ min}} \cdot \frac{1 \mu\text{l}}{\# \text{ of } \mu\text{l used}}$$

In the equation above, “standard CPM/nmol” is the CPM measured for the radioactive standards on the day of the experiment. Although the total assay volume is 1.0 ml, you will measure the radioactivity in only 0.5 ml. You will usually be using 10 μl of the enzyme; to correct to per μl of enzyme, you therefore must multiply by 0.1. You will usually be running the assay for 10 minutes, and therefore need to multiply by 0.1 to find the activity per minute. If you diluted your enzyme preparation, you will need to multiply by the dilution factor. You will be performing this calculation repeatedly during the course; you should make certain that you understand the purpose of each term within the equation.

Total Enzyme Activity: the activity per μl of enzyme solution multiplied by the total volume of that fraction. (If you have 50 ml of the original homogenate, the volume is 50 ml, even if you only saved 0.5 ml for the assay.)

Total Protein: the protein concentration per ml of solution multiplied by the total volume of that fraction.

Specific activity: Total Activity/Total Protein (or [Activity/ μl]/[Protein concentration in $\mu\text{g}/\mu\text{l}$])

Fold-purification: (Specific activity at a given step)/(Specific activity of starting sample)

%Yield: (Total activity at a given step)/(Total activity of starting sample)*100 (Note: this is intended to track the amount of ADP-glucose pyrophosphorylase present in your sample, and **not** the total amount of all of the proteins in the fraction.

Which purification steps do you think should be included in your table? Why do you think that each of the values listed above is important? (What are the values telling you about your purification procedure?)

Your table should include, at minimum, the columns listed below.

Step	Activity (nmol/min /μl)	Protein conc. (mg/ml)	Total volume (ml)	Total protein (mg)	Total Activity (μmol/min)	Specific activity (μmol/min/ mg)	Fold Purifi- cation	Yield (%)
------	-------------------------------	-----------------------------	-------------------------	--------------------------	---------------------------------	---	---------------------------	--------------

Based on your data, was your purification successful or unsuccessful? Why? Do you have any data other than fold-purification to indicate whether your purification was successful? How did your purification compare to literature values obtained for similar proteins? How did your purification of the mutant compare to the purification of the wild-type?

Enzyme assay: what can you learn from each enzyme assay? (If the answer is “nothing”, is it worth including these results in the paper?) How do you know that the assay results are valid? What assumptions are you making about the enzyme reaction actually occurring in the reaction tube? Are these assumptions likely to be correct for each assay? Are these assumptions likely to be correct for some assays but not for others? What controls did you run to ensure that the results were at least potentially meaningful?

In some cases, the answers to the above questions do not need to be stated explicitly. However, you **always** need to consider the answers before writing the paper. Knowingly incorporating the results of a flawed experiment in a paper is a good way to lose grant funding or become unemployed, and may result in your finding yourself in court defending yourself in a lawsuit or in a criminal trial. This does not mean that experiments that later turn out to be less informative than you would like are useless, but merely means that you should look carefully at your data, and try to understand the validity of each experiment before mentioning it in a written document.

Scientific research involves *intelligent* observation. In other words, you need to look at your data critically, and to attempt to understand everything it is telling you. Once you believe that you understand the data, you need to describe the results of your experiments so that others will be able to appreciate your insights, and be able to decide whether they agree with your conclusions.

Discussion: This section should begin with a brief summary of your results, and an explanation of what they mean. What were you hoping to accomplish? What did you discover as a result of your experiments? Which of your results are interesting? What

can you say about your hypotheses now that you have additional data? What did you expect to see? Did you see what you expected? Did you find surprising results?

At least in part, the Discussion section should be the section in which you answer the questions you raised in the Introduction. Sometimes the answer is that your original hypothesis turned out to be flawed; in this case, you should point out how your data indicate the flaws, and propose a brilliant new hypothesis to account for your observations. Sometimes your original hypothesis is supported by the data, in which case you point out how your original brilliant concept predicted your results.

You should end your discussion section with your conclusions. Did your experiment achieve your goals? How are your results going to change the world?

Figures and figure legends: In writing a paper, figures can be extremely useful. They are rarely, however, self-explanatory. This means that you need to refer to the figure in the text. In addition, you need to include some relevant information in the figure legend, so that people simply glancing through the paper can derive useful information from the figures.

As an example, in a figure of a gel, you should indicate the identity of the samples loaded in the figure legend. If more than one band is present in an important lane, it is often a good idea to highlight the important band in some way. (Note: in doing so, **do not** write on the actual lane; instead, place an arrow or other marker beside the gel, or beside the lane.)

Designing figures requires considerable thought. What point are you trying to make with the figure? Is the point necessary? If the point is a necessary one, how can the figure be used to make the point as clear as possible? Can you design a figure to present more information, or present the information more clearly?

Figure legends can be extremely useful in allowing you to present relevant information that would disrupt the orderly flow of ideas in the text. The figure legends are also necessary in clarifying the information presented in the figure.

References: In any scholarly endeavor, it is customary to give credit to your sources of information. The Reference section allows you to properly credit the originators of the information you are presenting. Where did your introductory information come from? Where did your methods come from? (Note that, unless you invented the method, you should always reference the paper that first described the work.)

Acknowledgments: In scientific papers, it is customary to thank the agency, company, or private foundation that funded the research. In addition, it is polite to acknowledge gifts of reagents or other supplies. Note that, if you purchased the reagent, the source of the reagent should be cited in the Methods section.

Appendix: Finally, the report should contain an appendix that contains your raw data and the calculations that you used to reduce your data to understandable form. In a real paper, Appendix sections are only included for the description of novel

calculations; in this course, the Appendix is included so that your lab instructor can correct your calculation mistakes.

In each section, attempt to organize the information you are presenting logically. Scientific papers are written for intelligent people who have not done the experiments you are describing. If your report is disorganized they may not understand it. If you do not write well, the reader will not believe your conclusions. (In the real world, a poorly written paper will not be published, and you will not get grant funding! In this class, if you instructor does not believe your conclusions, you will not get a good grade.)

The list of questions below is designed to help you write each section of the report correctly.

Criteria for Judging Lab Reports:

General:

- Does it contain the required sections?
- Is it clearly written?
- Does it use scientific terms properly?
- Does it use good grammar?
- Are the words spelled correctly?
- Are the calculations performed correctly?
- Is it unnecessarily long?
- Is the title meaningful?
- Does the title page contain the author's name and address?
- Does the title page contain the name(s) of the author's lab partners?

Abstract:

- Does it introduce the overall topic?
- Does it explain the hypothesis being tested?
- Are the important methods described?
- Does it reach logical conclusions supported by the data?
- Does it flow well? Is it logically written? Is it concise?

Introduction:

- Does it give general background?
- Does it point out poorly understood or unknown factors related to the study?
- Does it raise questions?
- Does it explain the hypothesis being tested?
- Does it discuss the significance of the work?
- Does it flow well? Is it logically written? Is it concise?

Materials and Methods:

- Could the experiments be understood based on the information given?
- Does it include the source of the reagents?
- Does it include information that belongs in the Results section?
- Does it describe all of the methods used?
- Is it excessively long?

Results:

- Does it explain the rationale and strategy for the experiments performed?
- Does it describe, in words, what was done?
- Does it answer the questions raised in the Introduction?
- Does it flow well? Is it logically written? Is it concise?

Discussion:

- Does it summarize the findings obtained in the Results section?
- Does it discuss the expected results?
- Does it discuss the unexpected results?
- Does it answer the questions raised in the Introduction?
- Does it reach conclusions?
- Does it explain why the conclusions are important?
- Does it flow well? Is it logically written? Is it concise?

Figures:

- Are the figures well designed?
- Do the figures include informative legends?
- Do the figures present information useful for understanding the text?

Tables:

- Are the tables well designed?
- Do the tables present information useful for understanding the text?
- Is the information in the tables redundant?

Acknowledgments:

- Are the sources of funding given credit?

References:

- Is the information obtained from published sources properly referenced?

Appendix:

- Are the raw data and the calculations included?

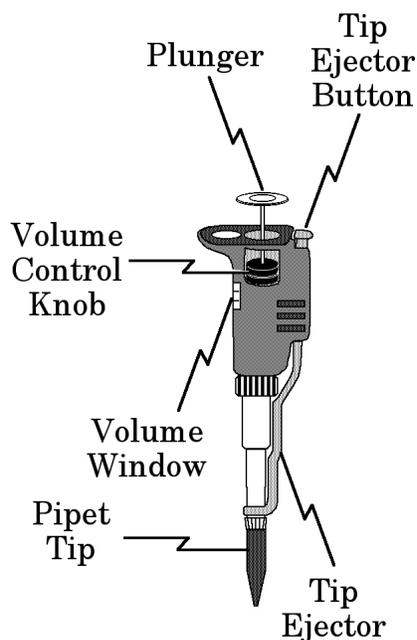
Reading over this list of questions before writing a draft of the report is strongly recommended. Reading these questions after writing your first draft, and using the questions to guide your revisions is also strongly recommended.

Basic Laboratory Techniques: Pipetting

In molecular biology and biochemistry, the ability to accurately and reproducibly measure and transfer small volumes of liquids is critical for obtaining useful results. For volumes less than 1 ml, the most common method for measuring liquid volumes involves the use of a device known as a pipetman. (Note: “Pipetman” is the brand name of the most commonly used of these types of pipets; however, all of these pipetting devices work on similar principles.)

A drawing of a pipetman is shown at right. The devices you use may not look exactly like the one shown. The pipetmen used in this course come in three different types: P1000, P200, and P20.

P1000 are useful for volumes from 200 to 1000 μl . P200 are useful for volumes from 20 to 200 μl . P20 are useful for volumes from 0.5 to 20 μl . Make sure that you are using the correct pipetman for the volume you need. Also, make sure that the pipetman is actually set for the volume you need by looking in the “volume window”, and, if necessary, turning the “volume control knob” until the pipetman displays the correct volume (the pipetmen do *not* read your mind; because several people will use the pipets, they may not always be set as you expect them to be). **Do not attempt to set pipetmen for volumes larger than their maximum, or for volumes less than zero**; doing so will damage the pipetman.



All pipetmen use disposable tips (**do not pipet liquids without using the appropriate tip**, because this will contaminate the pipetman and may damage it). When attaching the tip, make certain that the tip is the correct type for the pipetman you are using, and that the tip is properly seated on the end of the pipetman.

Try depressing the plunger. As the plunger depresses, you will feel a sudden increase in resistance. This is the first “stop”. If you continue pushing, you will find a point where the plunger no longer moves downward (the second “stop”). When using the pipet, depress the plunger to the first “stop”, place the tip into the liquid, and **in a slow, controlled manner**, allow the plunger to move upwards. (Do not simply let the plunger go; doing so will cause the liquid to splatter within the tip, resulting in inaccurate volumes and in contamination of the pipet.)

Now, take the pipetman (carrying the pipetted liquid in the tip) to the container to which you wish to add liquid. Depress the plunger to the first, and then to the second stop. If you watch carefully, you will note that depressing to the second stop expels all of the liquid from the tip. (Actually, this is true for most aqueous solutions. In some cases, however, such as for organic solvents, or for solutions containing large amounts of protein, it is often difficult to get all of the liquid out of the tip. In these

cases, it is best to “wet” the tip, by pipetting the original solution once, expelling it, and then taking up the liquid a second time.)

Although pipetmen are tremendously useful, they have a potential drawback. If used improperly, pipetmen will transfer inaccurate volumes. In addition, pipetmen may lose calibration. If used incautiously, therefore, pipetmen may yield misleading or even totally useless results. Checking the calibration of pipetmen is a simple procedure that can save considerable time, energy, and reagents.

Basic Laboratory Techniques: Measuring Absorbance

A spectrophotometer is an instrument for measuring the absorbance of a solution. Absorbance is a useful quantity. The Beer-Lambert law states that:

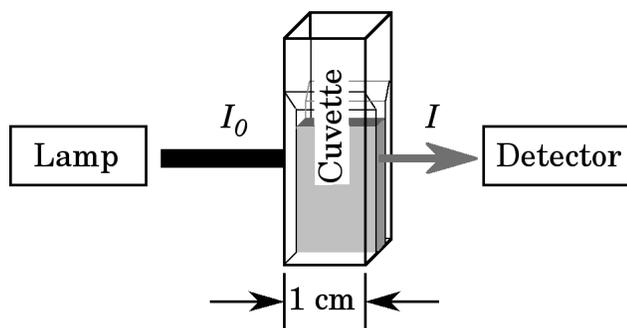
$$A = \epsilon cl$$

where A is the absorbance of the sample at a particular wavelength, ϵ is the extinction coefficient for the compound at that wavelength in $(M \cdot \text{cm})^{-1}$, c is the molar concentration of the absorbing species, and l is the path length of the solution in cm. Thus, if the extinction coefficient of an absorbing species is known, the absorbance of the solution can be used to calculate the concentration of the absorbing species in solution. (This assumes that the species of interest is the only material that absorbs at the wavelength being measured.)

The above is an explanation of *why* we measure absorbance: absorbance allows us to calculate the concentration of compounds in solution. However, it does not explain *what* absorbance is. Another definition of absorbance is:

$$A = \log \frac{I_0}{I}$$

where I_0 is the amount of light entering the sample, and I is the amount of light leaving the sample. Absorbance is therefore a measure of the portion of the light leaving the lamp that actually makes it to the detector. A little thought will reveal that when absorbance = 1, only 10% of the light is reaching the detector; when absorbance = 2, only 1% of the light is reaching the detector.



The typical internal arrangement of a Spectrophotometer

Absorbance values greater than 2 are unreliable, because too little light is reaching the detector to allow accurate measurements. When measuring absorbance, note the values; if the reading is greater than 2, dilute the sample and repeat the measurement.

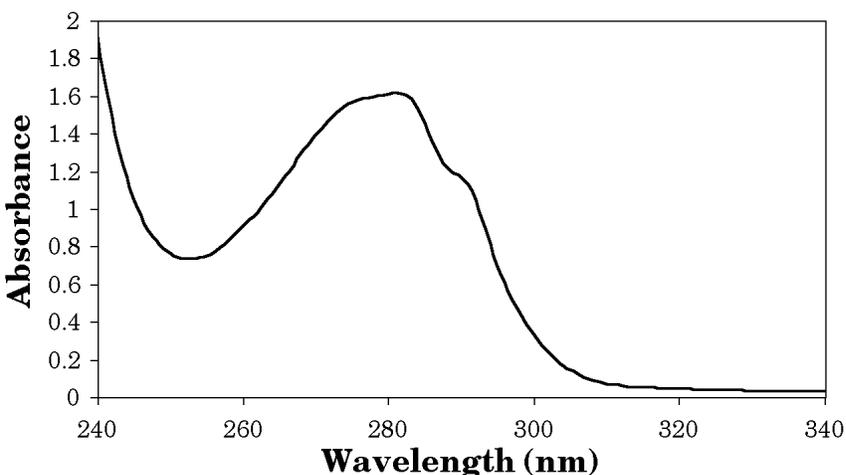
Spectrophotometers measure the decrease in the amount of light reaching the detector. A spectrophotometer will interpret fingerprints on the optical face of the cuvette, or air bubbles, or objects floating in your solution as absorbance; you therefore need to **look carefully at your cuvette** before putting it into the spectrophotometer to make sure that your reading is not subject to these types of artifacts.

Cuvettes are usually square objects 1 cm across (as shown in the above figure). In some cases, the liquid reservoir is not square; in those cases, make sure that the 1 cm dimension is aligned with the light path (note the orientation in the diagram above.)

Some cuvettes are designed for visible light only. When the spectrophotometer is set for ultraviolet wavelengths (wavelengths of 340 nm or less) **make sure that your cuvette does not have a large absorbance when it contains only water.**

The term “spectroscopy” comes from the word “spectrum” which originally referred to the multiple colors of light apparent in an analysis of white light using a prism. “Spectroscopy” therefore implies the use of multiple wavelengths of light. Spectrophotometers have the ability to specifically measure absorbance at specific wavelengths. The most commonly used method to allow this involves a “monochromator”, a device (either a prism, or more commonly, a diffraction grating) that splits the incident light into its component wavelengths, and allows only light of the desired wavelength to reach the sample. The ability to measure absorbance at different wavelengths is very useful, because the **extinction coefficient of a compound varies with wavelength**. In addition, the absorbance spectrum of a compound can vary dramatically depending on the chemical composition of the compound, and depending on the environment (such as the solvent) around the compound.

The graph at right shows the absorbance spectrum of a protein. The protein has a strong absorbance peak near 280 nm, but exhibits very little absorbance at longer wavelengths. For this protein, the only chromophores (chemical groups within a compound that absorb light) are the aromatic amino acids tryptophan and tyrosine. For many proteins, these two residues are the only chromophores; because tryptophan and tyrosine only absorb in the ultraviolet portion of the spectrum, such proteins are colorless molecules. Colored proteins, such as hemoglobin, exhibit their color due to chromophores (heme, in the case of hemoglobin) that absorb in the visible portion of the spectrum.



The extinction coefficient of a molecule at a given wavelength can be calculated using the Beer-Lambert equation from absorbance measurements for solutions of known concentration.

Basic Laboratory Techniques: Performing Dilutions

Many solutions used in biochemistry are prepared by the dilution of a more concentrated stock solution. In preparing to make a dilution (or series of dilutions), you need to consider the goal of the procedure. This means that you need to consider both the desired final concentration and required volume of the diluted material. A simple equation allows the dilution to be calculated readily:

$$C_1V_1 = C_2V_2$$

where C_1 is the concentration of the initial solution; V_1 is the volume of the initial solution available to be used for dilution (this may not be the total volume of the initial solution, and instead may be a small fraction of the initial solution), C_2 is the desired final concentration, and V_2 is the desired final volume.

In most cases, the initial concentration and the final concentration are either known or are chosen in order to work correctly in the experiment being planned. The final volume is usually an amount that is chosen based on the amount required for a given experiment. This means that at least three of the required terms are either known or can be chosen by the experimenter.

Let us consider an example. You are setting up a standard curve. You have a stock solution of 1000 $\mu\text{g/ml}$ BSA, and for one of the points on the curve, you want 200 μl of 20 $\mu\text{g/ml}$. In this case, $C_1 = 1000 \mu\text{g/ml}$; $C_2 = 20 \mu\text{g/ml}$, and $V_2 = 200 \mu\text{l}$. This leaves V_1 as the unknown value (*i.e.* how much of the stock solution must be diluted to 200 μl final volume to yield the desired concentration). Rearranging the dilution equation gives:

$$V_1 = V_2 \frac{C_2}{C_1} \quad \text{and therefore} \quad 4 \mu\text{l} = 200 \mu\text{l} \left(\frac{20 \mu\text{g/ml}}{1000 \mu\text{g/ml}} \right)$$

Thus, you need to dilute 4 μl of the stock solution to a final volume of 200 μl (*i.e.* by adding 196 μl).

If, in the example, you wished to make a solution of 1 $\mu\text{g/ml}$, the same equation would indicate that you need 0.2 μl of the 1000 $\mu\text{g/ml}$ stock solution for 200 μl of the final diluted sample. This is a problem: 0.2 μl is very difficult to measure accurately. You have two choices: change the final volume (*i.e.* if V_2 is larger, then V_1 must also increase), or perform **serial dilutions** (*i.e.* instead of diluting the stock solution by a factor of 1000 in one step, dilute the stock solution, and then make a further dilution of the diluted stock).

In many cases, while the final *concentration* is important, the final *volume* is not (as in the previous paragraph). In these cases, do what was explained in this example: use a **convenient** dilution: a dilution that involves volumes that are easily pipetted. Pipetting 1.3333 μl is usually less accurate than pipetting 4 μl , both because 4 μl is a larger volume, and because it is difficult to set the pipet for 1.3333 μl . In this case, 4 μl is a convenient volume, while 1.3333 μl is not.

In some cases, you may not know the actual starting concentration. If, for example, you need to measure the enzyme activity in a sample, and you find that the activity is too high to measure accurately, you will need to dilute the starting material. Since you don't know the actual starting concentration, all you know is the concentration **ratio** between starting and final solutions. As long as you keep track of the concentration ratio in all of your dilutions, you can easily determine the enzyme activity in the initial solution, even though you cannot measure it directly.

Concentration ratios are frequently of considerable value. For example, you have a stock solution of buffer that contains 450 mM Tris-HCl, 10 mM EDTA, and 500 mM NaCl. You actually wish to use a final concentration of 45 mM Tris-HCl, 1 mM EDTA, and 50 mM NaCl. In each case the concentration of the final buffer is one-tenth that of the original. Simply performing a 1:10 dilution of the stock solution then gives the appropriate final concentration of each component. The stock solution of buffer is typically called a 10x stock, because it is ten-times more concentrated than the final, useful buffer.

Note, in the previous paragraph, the "1:10" dilution. The description uses the chemistry convention for this term, which will be used throughout this course. The 1:10 dilution mentioned is performed by taking one part of the initial solution, and adding nine parts of solvent (usually water). This results in a final concentration that is ten-fold lower than the original.

Basic Laboratory Techniques: Buffers

Proteins, and especially enzymes, are generally quite sensitive to changes in the concentrations of various solution components. A buffer is a solution that is used to control the properties of a process occurring in an experimental aqueous medium. The term “buffer” is related to the ability of these solutions to resist changes in the hydrogen ion concentration, but buffers also contain other molecules, and are used to attempt to influence the ionic strength, the activity of proteases, and other parameters of the experiment in addition to the hydrogen ion concentration.

In any biochemistry experiment, the buffer components must be chosen based on their effect on the experiment. Ideal buffer components control pH and ionic strength without interacting in other ways with the system being studied. For example, while phosphate is a common physiological buffer, it may not be appropriate for some biochemical experiments, especially if phosphate is a substrate or product of the reaction being studied. In addition, some proteins interact poorly with some buffer components (a fact usually discovered by trial and error). As an example, Tris is less than ideal because of its high pK_a and the large change in pK_a that it exhibits upon changes in temperature. However, Tris is inexpensive, most proteins are stable in Tris buffers, and Tris rarely reacts with biological compounds; as a result, Tris is commonly used in biochemistry.

You will have seen the Henderson-Hasselbalch equation in previous courses. This equation is useful for calculating the **theoretical** pH of a solution. It is also useful for predicting whether a particular compound will be useful as a buffer over a given pH range. However, the Henderson-Hasselbalch equation has its drawbacks. Many buffers used in biochemical experiments deviate significantly from ideal Henderson-Hasselbalch behavior.

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \qquad \text{Henderson-Hasselbalch equation}$$

Because of the commonly observed deviations from ideal behavior, buffers are typically prepared by adding the buffer components to a container, adjusting the solution to the desired pH by adding an acid or a base, and then adding sufficient water to reach the expected final volume.

For example, a 1 liter of 50 mM Tris-HCl buffer (pH 7.4) with 200 mM sodium chloride would be prepared by adding 50 mmoles of Tris base and 200 mmoles of sodium chloride to a flask and adding water to about 900 ml. HCl would then be added to reduce the pH to 7.4, using a pH meter to monitor the changing pH, followed by addition of enough water to yield a 1 liter final volume. (If the solution contained 1 liter before addition of the HCl, the final volume would be more than 1 liter, and therefore the buffer would be less concentrated than it should be.)

Note that in order to produce most biochemically useful buffers, several components must be added together. This frequently requires careful consideration of the necessary dilutions for each of the components.

When performing experiments with proteins, it is rarely a good idea to dilute the protein with water, unless denaturation of the protein is not a concern. When denaturation is a concern, such as when performing dilutions for enzyme assays, perform the dilution using a suitable buffer to prevent undesirable alterations in the structure of the protein in solution.

General Information: Molecular Biology

Purifying a protein from cells that normally contain the protein has certain advantages. You know that the protein was synthesized in its natural environment, and you can assume that it was folded correctly, and was subject to any normal post-translational modifications. Purifying proteins from normal cells has been the traditional method for most of the history of biochemistry.

However, using normal cells as a source of protein also has significant drawbacks. The proteins suitable for study are limited to those expressed by readily available sources, and to those expressed in significant quantities. For example, studying human proteins tends to be somewhat difficult due to limitations in availability of starting material. In addition, the experimental modifications that can be introduced into the protein are limited both in type and in specificity.

As a result, expression of genes in heterologous organisms (especially bacteria) has become a frequently used technique. Bacterial protein expression generally allows the protein to be expressed in very large quantities, allows the researcher to choose the form of the protein to be expressed, and allows the researcher to introduce mutations in the protein sequence to examine the roles of individual amino acids in the function of the protein.

Escherichia coli is a bacterium found in the intestinal tracts of a many species of mammals, including humans. It has been used for a wide variety of molecular biological experiments, and a large number of specialized laboratory *E. coli* strains have been produced. Techniques for manipulating DNA in *E. coli* are well established.

Comments on molecular biological techniques

Molecular biology in its modern meaning (*i.e.* referring to genetic manipulation and analysis techniques) is a fairly new science; nearly all of the techniques used were invented after 1970. This means that techniques are still being invented, and many of the procedures have changed (often dramatically) during the last few years.

Molecular biological techniques differ slightly from biochemical techniques. In molecular biology, exact quantitation of the reagents is of variable importance. In some cases (as with some of the buffers) concentrations and pH are critically important; in others, such as ligation reactions, the DNA concentration can vary within a factor of ten (or more) and still allow the procedure to be successful.

On the other hand, some aspects of molecular biology require considerable care. **Humans (and many other organisms) secrete enzymes that degrade DNA and RNA;** it is therefore necessary to be completely aware of what you are doing. Allowing the reagents to contact your skin will cause problems – not necessarily to you, but definitely to your samples. DNase (the non-specific enzyme that degrades DNA) is rapidly denatured by heating at 68°C; for this reason, it is a good idea to heat treat your DNA samples if there is any chance of DNase contamination. Your plasticware (the pipet tips and microfuge tubes) have been autoclaved (heated to 121°C at

elevated pressure) to denature any DNase associated with them; this means that you should not handle the plasticware unnecessarily (e.g., pour out a few microfuge tubes; do **not** reach into the container with your fingers).

Many molecular biological procedures are very commonly used. Some techniques are so commonly used that all of the required reagents are available in kit form. These kits have made many aspects of molecular biology much easier.

General Information: Cell Genotypes

A large variety of *E. coli* strains are used in molecular biological research projects. The strains are usually differentiated by their genetic properties. Researchers choose the strain or strains that they will use for an experiment based on the genetic properties of the strains, and therefore need some understanding of the genes that are frequently modified in laboratory strains of *E. coli*.

A genotype for an *E. coli* strain includes the genes that are known to be modified when compared to the wild-type strain from which the strain was derived. Most laboratory strains are derived from either K12 or B strains, which are considered to be standard “wild-type” strains. Genes not listed in a genotype are thought not to be mutated, although laboratory strains often deviate from wild-type in additional (poorly characterized) ways. In most cases, the genes listed are mutated to the point of inactivity. In a few cases, a “-” (indicating **inactive** or **absent**), “+” (indicating **active**), or “ ” (indicating **deleted**) is added to avoid ambiguity.

The genotype is a factor in choosing a bacterial strain. However, once a strain has been tentatively chosen, it usually has to be tested to verify its usefulness in the experiment.

Genotypes for two commonly used *E. coli* strains are shown below.

Top10 genotype: F⁻, *mcrA*, (*mrr-hsdRMS-mcrBC*) 80*lacZ* M15 *lacX74* *deoR* *recA1* *araD139* (*ara-leu*)7697 *galU* *galK* *rpsL* *endA1* *nupG*

JM109 genotype: F' *traD36 lacI^q lacZ* M15 *proA⁺B⁺/e14⁻* (*McrA⁻*) (*lac-proAB*) *thi* *gyrA96* (*Nal^r*) *endA1* *hsdR17* (*r_K⁻ m_K⁺*) *relA1* *supE44* *recA1*

F' = a low copy number plasmid (also called an episome) that can be transferred from one cell to another. The genes on the F' plasmid are listed immediately after this; genes after the slash are genomic.

F⁻ = cells that lack the F' plasmid. This plasmid was of considerable concern during the early years of recombinant DNA experimentation, because it allowed bacteria to exchange DNA without the assistance of the researcher.

lacZ M15 = A mutated form of *lacZ* (β -galactosidase) missing the first 15 codons. Co-expression of *lacZ α* , a gene for a short peptide (of 41 to ~150 residues) results in restoration of β -galactosidase activity.

lacI^q = a modified form of the *lacI* repressor gene. In this case, the mutation results in over-production of the LacI protein, and therefore greater suppression of lac promoter (and lac-derived promoter) driven expression.

e14 = a DNA element present in K12, but not many derivative strains. The element contains the *McrA* gene, so cells lacking *e14* are also lacking *McrA*.

mcrA, *mcrBC*, *mrr* = genes for methylation-requiring restriction enzymes. DNA with methylation is degraded by these enzymes; DNA molecules isolated from higher organisms and from some non-*E. coli* bacteria (although not from *Drosophila* or *Saccharomyces*), contain methyl groups at sites recognized by these enzymes. (Note that this only applies to DNA isolated directly from the organism; PCR products and other *in vitro* synthesized DNA molecules are not methylated.)

proA⁺B⁺ = contains active forms of two genes required for proline biosynthesis. These genes are sometimes used as a selection mechanism. In JM109 cells, the episome contains the *proAB* genes, while the chromosomal DNA is “*proAB*”. If JM109 cells are grown on a culture medium lacking proline, only those cells harboring the episome will be able to survive.

endA = Endonuclease I; deleting this gene reduces non-specific degradation of DNA.

recA1 = recombination negative, a mutation generally necessary for stable replication of plasmid DNA.

hsdR = deletion of the normal K12 restriction enzyme *EcoKI*. The genotype *hsdR17* (*r_K⁻ m_K⁺*) means that the restriction enzyme is deleted, but that the methylation is retained; this allows DNA replicated in this strain to be used in *hsdR⁺* strains.

hsdRMS = deletion of both *EcoKI* restriction enzyme and the corresponding methylase. This means that DNA produced in this strain may be degraded if transformed into a non-*hsdRMS* strain.

supE = glutamine-inserting amber (UAG) suppressor tRNA.

supF = tyrosine-inserting amber suppressor tRNA. The *supE* and *supF* suppressors are required for growth of some bacteriophages. The presence of *supE* and *supF* in many laboratory strains means that use of TAG stop codons may result in unpredictable results, and therefore, when possible, the other stop codons (TAA or TGA) should be used.

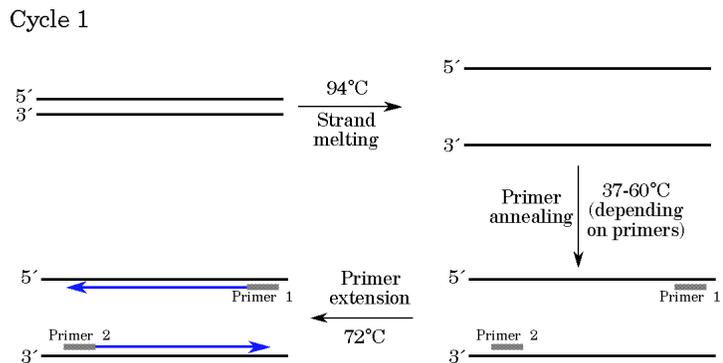
Methods: Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a technique that allows the generation of large amounts of a single DNA sequence from a mixture of sequences; the fragment generated can be designed to contain specific starting and ending positions based on the needs of the experiment.

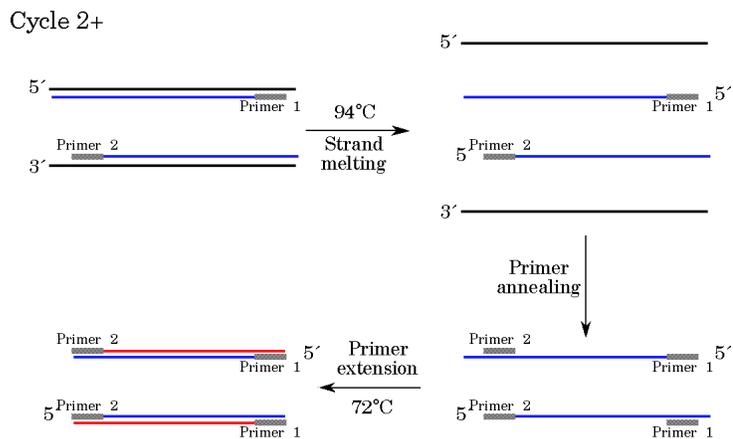
PCR uses a DNA polymerase (an enzyme that synthesizes DNA). DNA polymerases require a “signal” to begin synthesis. The signal is a short fragment of DNA. PCR uses two synthetic oligonucleotides (known as primers) that correspond to (and thus base-pair to) the ends of the sequence of interest and act as synthesis initiation signals.

If the Cycle 1 reaction (the top of the figure at right) were the entire process, PCR would not be very useful. However, an examination of the figure reveals that, at the end of the first cycle, twice as much template DNA exists as was present at the beginning. Therefore, repeating the cycle allows the amount of product DNA to increase geometrically.

In theory, the amount of product will double each cycle. In practice, PCR is not quite that efficient, although it can produce tremendous quantities of DNA. It is literally possible to begin with a single molecule of DNA and generate enough DNA for any molecular biological technique. In addition, the DNA synthesized in the PCR reaction will have specific starting and ending points: the primer sequences define the ends of the fragments.



The two strands of the DNA template are separated by heating (usually to 94°C). The temperature is then decreased to allow the primers to bind to the template DNA. Once the primers have bound, the polymerase is allowed to synthesize new DNA strands (the polymerase most commonly used has a temperature optimum of 72°C.)



Good PCR primers contain approximately 50% G+C and 50% A+T. This reduces problems in inducing template strand separation caused by the high affinity of G for C, and reduces non-specific priming common with high AT content.

In addition, good primers have few regions of complementarity either internal to primer (*i.e.* secondary structure) or between two primers, especially at 3' end of primer, and avoid repeated sequences (*e.g.*, AAAA or GTGTGT). Primers that are "poor" by these criteria can cause artifacts that prevent amplification of the desired sequence. In some cases, these potential problems are difficult to avoid due to constraints imposed by the sequence of interest; in these cases, the use of longer primers (*i.e.* 24 bases) may solve specificity problems.

The primer should be long enough to have a reasonable melting temperature (*i.e.* a melting temperature of 55-70°C). Melting temperature depends on a number of variables. In most cases, an approximation [$(4^{\circ}\text{C}$ for each G or C) + $(2^{\circ}\text{C}$ for each A or T)] will yield a value close enough to design the PCR experiment.

PCR also allows the generation of mutations at the ends of the fragment, because the primer does not need to be an exact match to the template DNA (it needs to be a sufficiently good match to allow primer binding, but it does not need to be a perfect match). In most cases, the introduction of mismatches at the 5' end of the primer has little effect. However, mismatches at the 3' end of the primer may prevent synthesis of the new strand.

The mutations inserted by PCR are frequently used to generate restriction sites to simplify cloning of the PCR fragment. Although PCR can also be used to generate mutations within coding sequences, this is somewhat more complex, because the PCR primers only affect the sequence at the ends of the PCR fragments. However, modified forms of PCR are quite useful for site-directed mutagenesis experiments.

PCR requires the use of a DNA polymerase to make the copies of the DNA sequence used as a template. As noted in the figure above, the PCR method involves heating the sample to ~94°C to separate the chains of the double-stranded DNA. In addition, while the oligonucleotide binding and polymerization reactions can occur at a range of temperatures, oligonucleotide binding is much more specific (*i.e.* it is more likely that the oligonucleotide will bind the correct sequence) at higher temperatures. To prevent problems with low temperature incubations and problems due to denaturation of the polymerase, most PCR experiments employ thermostable DNA polymerases. The DNA polymerase most commonly used for PCR is derived from the bacterium *Thermus aquaticus*. *T. aquaticus* prefers to live at temperature of about 70°C, and therefore its proteins (including its DNA polymerases) are stable at elevated temperatures.

Although the *Taq* polymerase is highly thermostable, it does begin to denature at temperatures above 90°C. Its half-life decreases rapidly with increasing temperature. For this reason, melting times of greater than 1 minute for the chain reaction cycles should be avoided. If the thermal cycler melting temperature drifts above 95°C, the enzyme may be inactivated prior to completion of the program.

The *Taq* polymerase has a primer extension rate of 60-100 bases/second under optimum conditions; thus it may be advantageous to use short (1-10 second) extension times, particularly for short products (*i.e.* below 500 bp) to decrease formation of non-specific products. However, for longer fragments (greater than 1000

bases), optimum primer extension rates are rarely achieved, and longer extension times should be used.

Although the *Taq* polymerase is a popular enzyme due to its ability to catalyze primer extension under a wide variety of conditions, it has some drawbacks. Its worst drawback is its relative lack of fidelity; it has a significant error rate, and lacks proofreading functions. As a result, many experiments employ thermostable DNA polymerases that have lower probabilities of misincorporation. The use of other polymerase may require minor modifications to the PCR procedure described below.

PCR conditions, and especially annealing temperature, must be chosen empirically to optimize PCR product formation. Some primer-template combinations work under a wide variety of conditions; others result in product only in a narrow range of conditions. For some reactions the annealing temperature is a critical parameter, with no specific product formed except in a narrow optimum range. It is usually best to begin testing conditions with an annealing temperature of about 55°C, because the use of annealing temperatures above 50°C often prevents certain types of mismatch artifacts.

The temperature profile shown below is a typical one that works for many different primers and templates:

PCR procedure:

Mix reagents in a PCR tube:
16 µl of mixture of 1.25 mM dNTP
10 µl 10x PCR buffer
5 µl 20 µM 5' primer
5 µl 20 µM 3' primer
1 µl of DNA template mixture
62.5 µl deionized water
0.5 µl *Taq* polymerase
Place the PCR tube in the PCR machine.

PCR Temperature Profile

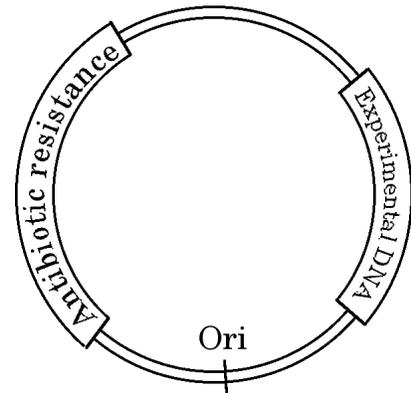
Temp.	Time	Function
94°C	0.5 minutes	Denature DNA
55°C	0.5 minutes	Primer Annealing
72°C	0.75 minutes	Primer Extension

Run the PCR program as shown above (with appropriate modifications if necessary for the specific polymerase, primers, and template being used in the experiment).

Methods: Plasmid Preparation

PCR is one method for generating large amounts of DNA. A second method is to allow bacteria to replicate the DNA and then purify the replicated DNA from the bacteria. In most cases, the DNA of interest for this method is plasmid DNA. A plasmid is a double stranded DNA molecule that will replicate in an organism. A typical plasmid used for molecular biology contains at least four features.

1) The plasmid must be **circular**, because bacteria generally will not replicate linear DNA. 2) The plasmid must contain a sequence that functions as an **origin of replication (ori)**. 3) The plasmid must contain a **selection mechanism** that will force the bacteria to retain the DNA; the most common type of selection mechanism used in bacteria is a gene for resistance to an antibiotic such as ampicillin. 4) The plasmid must contain a region for the insertion of the **experimental DNA**. A generic plasmid exhibiting these features is shown at right.



An **expression plasmid** is a specific type of plasmid used to allow expression of heterologous DNA. An expression plasmid must therefore have, in addition to the features listed above, a strong promoter element that will drive transcription of the foreign DNA¹ in the host organism, and an effective ribosome binding site that will allow efficient translation of the transcribed RNA.

Because plasmids are much smaller than chromosomal DNA (for *E. coli*, a typical plasmid contains 5 to 10 kilobase pairs (kb), while *E. coli* chromosomal DNA contains about 4,800 kb), separating the two types of DNA molecule is relatively straightforward. In addition, most plasmids used in molecular biology are “high copy number plasmids”; in other words, each bacterium contains many copies (usually >100) of the plasmid. Therefore, although each plasmid molecule is much smaller than the chromosome, plasmid DNA often comprises ~10% or more of the total DNA in the bacterium.

Plasmid preparation requires several steps:

1) **Growth of bacteria containing the plasmid of interest.** This involves starting a liquid culture using a single clone (each colony on a plate represents a single clone, so cultures are started by picking one colony and adding it to the culture medium). The bacteria are then grown until they reach stationary phase; stationary phase occurs when the bacteria have either used up all of the available nutrients, or when the bacterial waste products have reached levels that preclude further growth. The commonly used laboratory bacterium *E. coli*, reaches stationary phase after 12-18 hours of growth (typically overnight growth is assumed to result in stationary phase).

¹Note that “foreign DNA” can be from any organism, including the organism used as a host. Thus, *E. coli* can be used to express *E. coli* proteins from expression plasmids. In most cases, however, the DNA being expressed is from a different organism, and is being expressed in *E. coli* for convenience.

(Note that some other bacterial types, and all yeast and mammalian cells in culture grow more slowly than does *E. coli*.)

2) **Separation of the bacteria from the culture medium.** The next step has two purposes, concentrating the cells, and transferring them into a buffer that will facilitate plasmid purification. The usual method is to centrifuge the culture and to discard the spent culture medium. The cells can then be frozen for storage (freezing the cells also increases the vulnerability of the cells to the lysis conditions), or resuspended immediately in a buffer for plasmid preparation.

3) **Lysis of the bacteria** (*i.e.* disruption of the bacteria to release plasmid DNA). Bacterial cells are much tougher than human cells, and lysing bacteria requires some effort. The technique usually used to extract plasmid DNA from bacteria involves lysing the cells with a mixture of SDS and NaOH, which disrupts the cell membranes and cell wall. **Once the SDS/NaOH mixture has been added to the cells, it is imperative that the cells be treated gently**; DNA molecules are long and fragile, and vigorous treatment will readily damage the DNA molecules released from the cells. In addition, when performed properly, the alkaline lysis technique does not extract the chromosomal DNA (unless the chromosomal DNA is fragmented by violent shearing forces). The intention in a plasmid prep is to purify the plasmid DNA while obtaining as little chromosomal DNA as possible.

4) **Neutralization of the NaOH.** Once the cells have been lysed, it is a good idea to lower the pH to near neutral. Neutralization is typically performed by adding an acetate buffer. This has two effects: it prevents degradation of the DNA, and it precipitates some proteins and most of the lipids.

5) **Purification of the plasmid DNA.** A variety of different techniques are used for separating plasmid DNA from the other molecules present in the cell. Most of the commonly used techniques involve binding the plasmid DNA to an insoluble material, washing off unbound material (protein, lipids, and small molecules), and then eluting the DNA.

Procedure for Plasmid miniprep:

The procedure outline given below should be used as a general guide for the alkaline lysis technique. However, because of minor differences in the plasmid prep kits in common use, it is a good idea to consult the manufacturer's instructions for the exact procedure.

1. Grow a 3 ml overnight culture of the cells containing the plasmid, using the appropriate culture medium (usually either LB or TB containing an antibiotic to force the cells to retain the plasmid).
2. Centrifuge 1.5 ml of the culture in a microfuge tube to pellet the cells; discard the supernatant. Depending on the size of the cell pellet, it is frequently preferable to repeat the procedure by adding an additional 1.5 ml from the same culture to the pelleted cells, and recentrifuging.
3. Resuspend the cells in the Resuspension Buffer solution (the amount to be used varies from 200 to 350 μ l depending on the kit). Make sure that the cells

are evenly suspended. (The suspension should be cloudy with no obvious clumps).

4. Lyse the cells using a mixture of SDS and NaOH (usually 1% SDS in 0.2 M NaOH². After adding the SDS and NaOH, be gentle with the solution to avoid disrupting the chromosomal DNA. The solution should become fairly clear.

Neutralize the pH using the acetate solution. You should observe the formation of a precipitate. Centrifuge the sample to pellet the precipitate.

The remainder of the technique depends on the exact plasmid prep kit being used; consult the manufacturer's instructions for the recommended procedure.

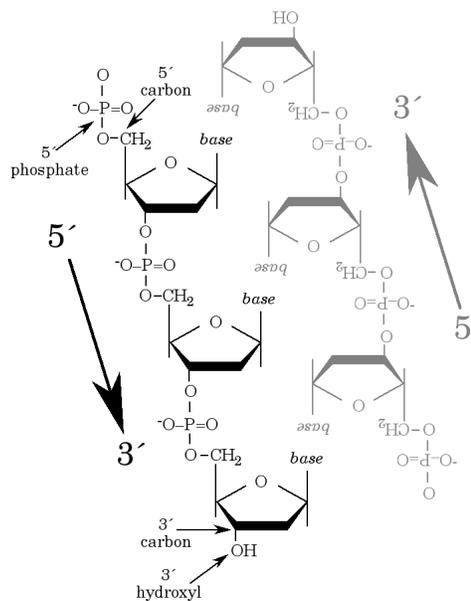
When the procedure is complete, it is usually a good idea to heat the DNA sample to 68°C for 10 minutes to inactivate any contaminating DNase. After this step, store the plasmid DNA at -20°C.

² In preparing the SDS/NaOH solution, addition of the SDS to NaOH at a concentration much higher than the 0.2 M of the final solution will result in precipitation of the SDS. It is therefore necessary to dilute the NaOH prior to addition of the SDS.

Methods: Ligation

The process of connecting two pieces of DNA together is called **ligation**, and is catalyzed by an enzyme called **ligase**. The ligase used most often in molecular biology is derived from the T4 bacteriophage, and uses ATP to supply the energy necessary for the reaction. In addition to ATP, *T4 ligase requires DNA with a 5'-phosphate group and free 3'-hydroxyl groups*.

The drawing at right shows a (very short) region of double stranded DNA. Both strands of the DNA molecule contain 5'-phosphates and free 3'-hydroxyl groups; this DNA molecule is therefore capable of being ligated. This DNA fragment is blunt ended (*i.e.* all of the bases are paired with bases from the opposite strand); note that some restriction enzymes leave blunt ends, while others leave "overhangs", which are short stretches of single stranded DNA at either the 5' or 3' end.



Synthetic oligonucleotides contain free 5'-hydroxyl groups, and therefore must be subjected to phosphorylation prior to ligation. In contrast, most (although not all) restriction enzymes leave 5'-phosphate groups; most restriction fragments can be ligated immediately after digestion.

Ligation also requires compatible ends to the DNA. The drawings below show examples of different types of compatible and incompatible ends (N's imply that any arbitrary sequence could be present).

Compatible Blunt ends		Compatible 4-base, 5'-sticky ends		Compatible 2-base, 3'-sticky ends	
5' -NNNNN	NNNNN	5' -N	CATGN	5' -NNNTG	NNN
3' -NNNNN	NNNNN	3' -NGTAC	N	3' -NNN	ACNNN
Incompatible (blunt/non-blunt ends)		Incompatible 4-base, 5'-overhangs		Incompatible 4-base, 3'-overhangs	
5' -NNNNN	N	5' -N	AGCTN	5' -NCTAG	N
3' -NNNNN	GTACN	3' -NGTAC	N	3' -N	AATTN

The drawing below shows DNA fragments generated from restriction digests that used *Nco*I and *Hind*III; these enzymes both leave 4-base 5'-overhangs. The end generated by digestion with one of these enzymes is compatible with other ends generated by the **same** enzyme, but not by the ends generated by the other enzyme. In attempting to construct a plasmid, scientists typically try to take advantage of the specificity of the restriction enzymes and of the ligase to force the creation a plasmid with the insert in the correct orientation.

Methods: Competent Cell Preparation and Transformation

Replication of a plasmid requires the insertion of the plasmid DNA into the bacteria used for the replication process. In general, bacteria are reluctant to take up DNA from their environment (at least, they typically will not do so without degrading it first). In order to improve the probability that the bacteria will actually internalize the plasmid DNA, it is necessary to make the cells “competent” to absorb the plasmid DNA. One procedure that results in efficient competent cells is given below.

Competent cells are significantly more fragile than normal bacteria. Vortexing the cells, heating the cells above 42°C or to 42°C for prolonged periods, or exposure of the cells to any of a number of other even mildly abusive treatments may kill them.

Competent Cell Preparation:

TSS: 85% LB medium 10% PEG 8000 5% DMSO 50 mM MgCl ₂ pH 6.5 <i>Sterile filter</i> DMSO may degrade into transformation inhibitors; for best results, fresh DMSO should be stored in frozen aliquots.	Grow an overnight culture of the desired <i>E. coli</i> strain under the appropriate conditions (LB + antibiotic, if used). Add 1 ml aliquot of cells to 50 ml LB + 20 mM MgSO ₄ ; grow at ~18-20°C (note: growth at higher temperatures results in somewhat lower efficiencies) to OD ₆₀₀ 0.9 (<i>i.e.</i> mid-log phase). Cool the cells on ice for 20 minutes. Spin down the cells (4000 rpm for 5 minutes) in sterile tubes. Resuspend the cells in 1/10 volume TSS (5 ml for 50 ml culture). Cells may be used immediately or frozen for storage. Quick freeze the cells in 310 µl aliquots in a dry ice:ethanol bath and store at -70°C.
---	---

Transformation:

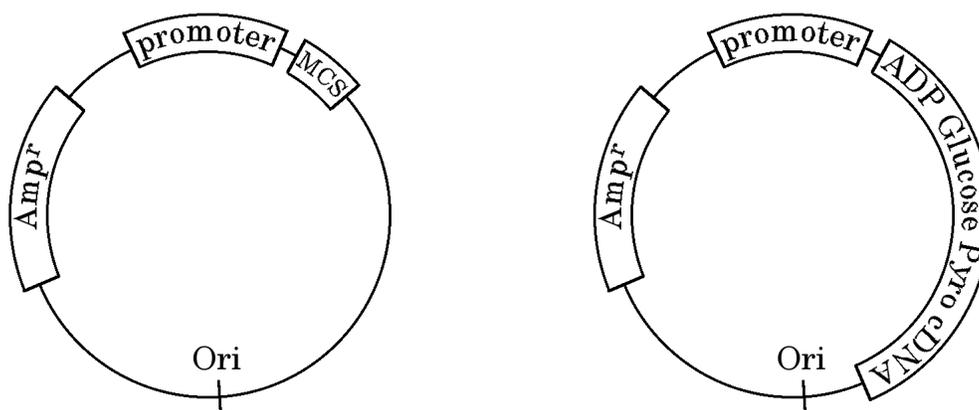
The procedure for allowing competent cells to take up DNA is called transformation. Transformation requires mixing a small volume of DNA (usually 1-3 µl) with 100 µl of competent cells. The mixture is incubated on ice for 30 minutes to allow the cells to take up the DNA. The cells are then heat-shocked at 37-42°C for 1 minute, and cold-shocked on ice for 2 minutes. The heat shock/cold shock reverses the effect of the competent cell process, by, in effect, healing the damage inflicted by the competent cell solution. Adding 0.5 ml of culture medium (usually LB) and incubating the cells for 10 - 30 minutes at 37°C assists in this healing process. This cell/medium mixture is then spread onto bacterial culture plates containing agar, LB, and a selection mechanism.

Methods: Selection and Screening

Transformation is usually an extremely inefficient process; in most cases, only a very small fraction of cells actually take up DNA (usually less than 1 in 10^5 cells). If it were necessary to sort out the few cells that took up DNA from the vast majority that did not, cloning experiments would be very difficult. However, selection mechanisms (*i.e.* conditions where cells grow if they express a gene, but die if they do not) make this fairly straightforward.

The selection mechanism most commonly used is antibiotic resistance. If the *E. coli* strain used is not resistant to the antibiotic, but the plasmid DNA contains a gene coding for resistance, then only the cells that have taken up the plasmid will be able to grow. Most commonly used plasmids contain a gene for resistance to ampicillin; only cells that have taken up the plasmid will be able to grow and form colonies. (Note that, since taking up plasmid DNA is a very rare event, each group of cells visible on a plate represents the offspring of a single cell that internalized a single molecule of plasmid DNA.)

In most cloning procedures, however, only a fraction of the colonies contain the correct plasmid. The selection technique means that only cells containing *some* plasmid DNA have grown; it is still possible for cells to have taken up incorrect plasmids (as long as the plasmid contains an ampicillin resistance gene and origin of replication). The figure below shows the starting plasmid (on the left), and a plasmid you might wish to construct (on the right). These are some (although not all) of the possible plasmids that may have resulted in colonies. The next step, therefore, is a screening procedure: an attempt to find cells containing the plasmid of interest.



Screening is more labor intensive than selection; selection techniques only allow cells that *might* be of interest to grow, while screening requires you to actually do something to find cells that contain the correct plasmid. A large number of screening techniques have been developed; this brief manual cannot cover all of them.

All screening techniques depend on finding a difference between plasmids of interest and other likely plasmids that might be found in cells on the same culture plate. These techniques often involve directly testing for the presence of specific DNA. For example, performing PCR on the colonies to look for the presence of the specific insert

DNA or hybridization screening of the cells for the presence of the DNA looks directly at the DNA within the cells. Purifying the plasmid DNA and then testing the plasmid by restriction analysis or DNA sequencing also looks directly at the plasmid DNA. Alternatively, the screening technique can analyze the DNA indirectly, by looking for production of a heterologous protein or for an activity that must be due to a protein encoded by the plasmid of interest.

The screening technique used depends largely on the features of the DNA. Purifying plasmid DNA is much more labor intensive (and more expensive) than most techniques that assay the cells directly. Thus, in many cases, PCR screening and hybridization screening (which are performed on cells derived directly from the culture plate) are preferable. However, these techniques may not yield unambiguous results for small differences in DNA. Site-directed mutagenesis, for example, usually involves single base changes, which are difficult to detect by PCR. Instead, PCR screening is best suited for detecting the presence or absence of relatively large DNA fragments.

PCR Screening procedure

The procedure is to pick the colony with a sterile pipet tip, streak the cells on a new plate (so that, if the colony turns out to be correct, you will still have viable cells to replicate), and then soak the pipet tip in 10 μl of water. This water (actually the few cells that drop off into the water) can then act as a template for a PCR. It is obviously necessary to use PCR primers that are specific for the DNA of the insert. Colonies that result in PCR products of the correct size (based on running the samples on an agarose gel) can tentatively be assumed to contain the correct plasmid.

A PCR screening reaction does not require a large reaction volume. Instead, a common method is to prepare a master solution of all of the required reagents except template, followed by adding the same volume of master mix to each PCR tube. The last step should be the addition of the template to the PCR tube (adding the template last reduces the chance of cross-contamination of the PCR samples).

A master mix that allows a final reaction volume of 25 μl and a template volume of 5 μl can be prepared as follows:

Per sample to be run:

4 μl 1.25 mM dNTP mixture

2.5 μl 10x PCR buffer

1.25 μl 20 μM 5' primer

1.25 μl 20 μM 3' primer

0.125 μl Thermostable DNA polymerase

11 μl autoclaved water

In screening procedures (as in most experimental procedures in general), it is a good idea to set up a positive control: an experimental condition that **should** work. In PCR screening, the positive control is intended to verify that the reaction mixture was prepared properly, and that impurities introduced by using whole cells as templates are not inhibiting the polymerase.

General Information: Restriction Digestion

Restriction enzymes are crucial reagents in molecular biology. Restriction enzymes cleave DNA only at specific sequences. Molecular biology would be extremely difficult without restriction enzymes. Fortunately, different bacteria have evolved a large variety of restriction enzymes, and enzymes with specificity for large numbers of sequences are now commercially available.

In order to decide which restriction enzymes to use for a particular process, you will need to know the sequence of the DNA molecules of interest. A number of computer programs that automate the process of analyzing DNA sequences for the presence of restriction sites are available.

Restriction digestion is typically used for two purposes: restriction mapping and specific DNA cleavage for production of new constructs.

Plasmid preparation procedures are non-specific: they can be used to purify any plasmid present within the bacteria. This is a major advantage, because it means that the protocol does not need to be changed for different plasmids. However, it also means that it is possible to purify the wrong plasmid if the cells used did not contain the intended plasmid DNA. (All *E. coli* look alike; it is possible to take the wrong plate out of the refrigerator and inadvertently use the wrong cells.) A simple method for checking the identity of plasmids is a technique called **restriction mapping**, in which the plasmid is subjected to digestion with different restriction endonucleases. Because restriction enzymes cleave DNA at specific sequences, the size of the resulting fragments can be predicted. If the observed restriction fragment sizes match the predicted sizes, it is likely that the plasmid is indeed correct.

In performing plasmid construction, cleavage of DNA molecules is usually necessary to create the fragments to be ligated together. In some cases, restriction sites will be present in appropriate places in the parent DNA molecules. On the other hand, in many cases, no restriction sites will be available; in these cases, one important feature of PCR is the ability to readily engineer restriction sites at the ends of the PCR products.

Restriction enzymes must cleave both strands of the double stranded DNA. They can do this in a number of ways. Some restriction enzymes cleave both strands at the same location, resulting in a “blunt” end. Other restriction enzymes cleave at different locations on the different strands, leaving short stretches of single stranded DNA. For commercially available restriction enzymes, these short stretches of single stranded DNA are typically either two bases or four bases long (see the comments on compatible enzymes in the “Ligation” notes).

One potential problem with restriction enzymes is that they are most active in specific buffers; the best buffer for one enzyme may not be the best one for another enzyme. In some cases, it is necessary to run one reaction, perform a DNA clean-up procedure, and then run the other reaction. In other cases, a compromise buffer can be used, in which the enzymes will both digest the DNA with somewhat reduced efficiency.

Electrophoretic Techniques: SDS PAGE

SDS polyacrylamide gel electrophoresis (PAGE) allows the separation of proteins based on their molecular weight. This technique can be used to determine whether a given protein is present in a sample, and for assessment of purity of the preparation, for estimation of the approximate quantity of the protein, and for measurement of the size of the protein.

Electrophoresis is a process in which molecules are exposed to an electric field and separated on the basis of their differential mobilities in that field. The observed differential mobility is a result of different charge magnitudes on different molecules, and the result of different resistance to movement through the medium. For molecules with similar shapes, the mobility is proportional to the charge-to-mass ratio of the molecule. For molecules of similar shapes and similar charge-to-mass ratios, the motion through the medium will be proportional to the size of the molecule, because friction increases as a function of size.

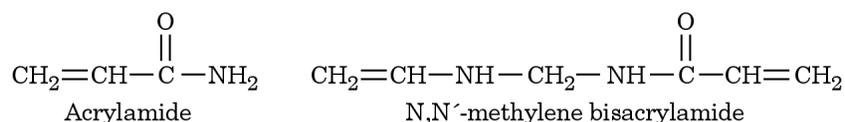
The velocity of a charged molecule is given by:

$$v = \frac{qE}{f}$$

where q is the charge on the molecule, E is the electrical potential gradient, and f is the frictional coefficient of the medium for the molecule. For similarly shaped molecules, f is proportional to size of the molecule. For molecules in which charge increases in proportion to size, larger molecules move more slowly than small ones, because f increases faster than charge.

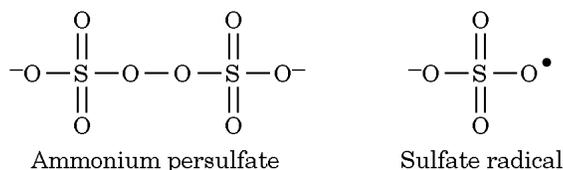
Gel electrophoresis uses a matrix of large uncharged molecules to provide the required friction. The matrix also serves to inhibit diffusion, and therefore to prevent degradation of the separation that is achieved. The separation of proteins usually involves the use of polyacrylamide as the matrix.

Polyacrylamide is formed by polymerization of acrylamide monomers in the presence of N,N' -methylene bis-acrylamide.



The bis-acrylamide contains two double bonds, which allow the compound to act as a cross-linker between polyacrylamide chains. The presence of the cross-linking agent results in formation of a gel matrix rather than a simple linear polymer.

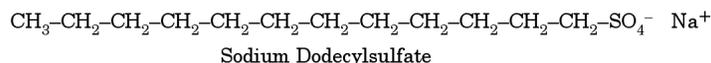
The polymerization reaction is a serial reaction using a free radical mechanism. The formation of the free radicals is initiated by the unstable compound ammonium persulfate. The sulfate radicals formed then react with tetramethylethylenediamine (TEMED), forming TEMED radicals which react with acrylamide to begin the actual polymerization reaction.



Varying the amount of acrylamide monomer and bis-acrylamide cross-linker present controls the formation of the matrix. The use of larger amounts of these components results in a denser matrix. Denser matrices are used for separating smaller proteins; larger proteins may find the pores in a dense matrix too small to enter, and may therefore not enter the gel at all. The table below lists approximate useful ranges for different gel densities.

Percent acrylamide	Useful Molecular weight range
7	30,000 to 200,000
10	20,000 to 150,000
12	10,000 to 100,000
15	5,000 to 70,000

Real proteins have different proportions of charged side-chains. As a result, real proteins do not have constant charge-to-mass ratios. Real proteins also have varying three-dimensional shapes. In order to measure molecular weight, it is necessary to induce the formation of a similar shape and charge-to-mass ratio. Boiling the protein in the presence of the detergent sodium dodecyl sulfate (SDS) and the reducing agent β -mercaptoethanol (which reduces disulfide bonds) results in disruption of the three dimensional structure of the protein. In addition, large amounts of SDS bind to the protein (approximately one molecule of SDS for every two amino acid residues). Given the fact that each SDS molecule has a negative charge at the pH used for electrophoresis, the use of SDS results in a large negative charge that overwhelms any intrinsic charge present in the protein.

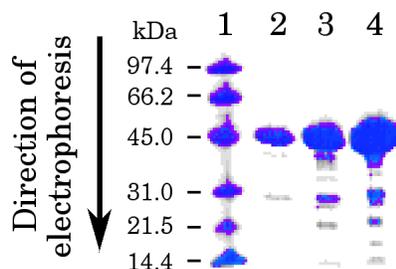


Note that treatment with SDS and β -mercaptoethanol will result in the formation of denatured protein monomers; it is these protein monomers that are separated on the SDS PAGE.

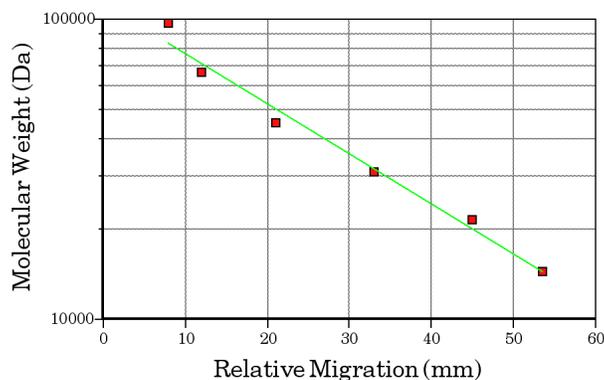
After the electrophoresis has been performed, the protein must be detected. The most commonly used method for detecting protein is Coomassie Blue R-250. Coomassie blue is a dye that binds proteins. Staining is performed by placing the gel in a solution of Coomassie blue in acetic acid and methanol. The function of the acetic acid and methanol is to cross-link the proteins into the gel so that they do not diffuse. Following staining of the proteins, the gel is placed in a destaining solution of acetic acid and methanol, which results in removal of the excess Coomassie blue.

An example of an SDS gel is shown at right. Lane 1 in the example is comprised of proteins of known molecular weight called **molecular weight standards**. Lanes 2, 3,

and 4 contain increasing concentrations of an experimental protein sample. Loading increasing amounts of protein makes it possible to see minor impurities, which are difficult to see in lane 2 and fairly obvious in lane 4.



The molecular weight standards can be used to calibrate the migration of proteins of differing sizes on the gel. For any given gel, the migration will be inversely proportional to the log of the molecular weight. Thus, a plot of **log molecular weight versus migration distance** for proteins of known size can act as a standard curve to allow measurement of molecular weights of unknown proteins. An example of an SDS PAGE standard curve is shown below.

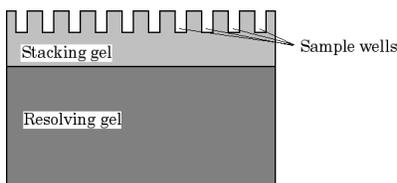


Note that proteins within an SDS polyacrylamide gel are denatured; the molecular weight determined will be that of the individual monomers of multimeric proteins.

A commonly used set of molecular weight standards for SDS PAGE experiments is shown in the table below.

Standard protein	Molecular weight (Da)
Rabbit muscle phosphorylase b	97400
Bovine serum albumin	66200
Chicken ovalbumin	45000
Bovine carbonic anhydrase	31000
Soybean trypsin inhibitor	21500
Chicken lysozyme	14400

A picture of an SDS PAGE gel is shown below.

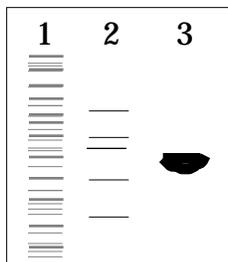


The gel consists of two layers: the top layer is the **stacking gel** and the bottom layer is the **resolving gel**. The purpose of the two sections differs: the stacking gel concentrates all of the protein in a narrow region, while the resolving gel performs the actual separation of the proteins by molecular weight. The stacking gel also contains the wells into which the samples are loaded.

The stacking gel is prepared at a lower pH (6.8) and lower acrylamide percentage (6%). At this low pH the glycine of the electrophoresis tank buffer is in the neutral zwitterionic form and is not an effective carrier of current. The chloride ions also present are highly charged and migrate rapidly toward the anode. The SDS-coated protein molecules and the dye, which have charge-to-mass ratios greater than that of the glycine but less than that of Cl^- , must migrate behind the Cl^- and ahead of the glycine. This has the effect of concentrating the proteins in a thin band sandwiched between the Cl^- ions and the glycine molecules. In addition, because the acrylamide concentration of the stacking gel is very low most proteins are not retarded and move freely through the gel matrix. When the sample reaches the end of the stacking gel it should appear as a thin blue band.

The resolving gel is at pH 8.8 and has the desired acrylamide concentration for separation of proteins in the appropriate size range (probably 10% for your experiment). When the stacked samples enter the resolving gel the higher pH results in negatively charged glycine molecules which then migrate with the Cl^- ions. The protein samples lag behind and are separated by the sieving effect of the gel. The tracking dye (bromophenol blue) will migrate faster than the proteins. When the tracking dye reaches the end of the gel, the electrophoresis should be terminated so that the proteins do not run off the gel.

Gels have limits in terms of the amount of protein that can be loaded. Loading too much protein results in “overloading” in which the sample runs unevenly. Loading too little protein makes the protein bands difficult to detect. In choosing the amount of protein to load, you need to consider the number of proteins in your sample; if your sample contains many proteins, you can load more total protein.



In the gel drawn at left, the same mass of total protein was loaded in each lane. In the first lane, there were many proteins, and therefore none resulted in an intense band. You could have loaded more protein for this sample. In the second lane, there are five proteins visible; the amount loaded was appropriate for this sample. The third lane is overloaded; there is too much of the one protein present, and it did not run cleanly. In fact, this band probably ran at a molecular weight lower than the actual molecular weight of the protein, and it may have distorted the protein migration in adjacent lanes.

Chemicals required for SDS PAGE:

10% Resolving gel

1.5 ml 1.5 M Tris-HCl, pH 8.8
 2 ml 30% acrylamide
 0.06 ml 10% SDS
 2.44 ml water

Add **last** to initiate reaction:
 30 μ l 10 % ammonium persulfate
 5 μ l TEMED

5x Sample Buffer

60 mM Tris-HCl, pH 6.8
 25% glycerol
 2% SDS
 14.4 mM β -mercaptoethanol
 1% bromophenol blue

Electrophoresis tank buffer

25 mM Tris
 192 mM glycine, pH 8.8
 0.1% SDS

Coomassie Staining Solution (10% Acetic acid, 25% Methanol, 0.05% Coomassie R-250 or Bio-Safe Coomassie Blue solution)

Stacking gel

0.875 ml 1.0 M Tris-HCl, pH 6.8
 0.583 ml 30% acrylamide
 0.035 ml 10% SDS
 2.007 ml water

Add **last** to initiate reaction:
 30 μ l 10 % ammonium persulfate
 5 μ l TEMED

Hardware required:

Electrophoresis apparatus
 Gel loading tips
 Weigh boats for gel staining
 100°C water bath
 Microcentrifuge tubes

Electrophoretic Techniques: Agarose Gel Electrophoresis

DNA is a negatively charged molecule. Since all DNA sequences have the same phosphate backbone, fragments of all sizes have the same charge-to-mass ratio. As with proteins, it is often useful to run DNA on gels. DNA gels are slightly different from protein gels. Because DNA has an intrinsically constant charge-to-mass ratio, addition of SDS is unnecessary. In addition, DNA fragments are generally much larger than protein molecules; in most cases, DNA is run on agarose gels instead of polyacrylamide gels, because agarose forms a lower density matrix more suited to running larger molecules. Various concentrations of agarose can be used to separate different sized DNA: 0.75% for DNA > 3 kb, 2% or 3% for 50 – 400 bp, and 1% for DNA between these ranges.

Agarose gels allow estimation of DNA fragment size. This can be useful for verifying that a fragment of the correct size was produced in a PCR reaction, or to assess the size of restriction enzyme digestion fragments. It is also useful to run plasmid DNA on a gel in order to assess both the concentration of the DNA and the quality of the preparation (*i.e.* to look for contaminating nucleic acids or excessive fragmentation of the plasmid DNA).

Agarose gels are usually run in either TBE (Tris-borate-EDTA) or TAE (Tris-acetate-EDTA) buffer. Because the agarose comprises a very small percentage of the gel while the remainder is the buffer, agarose gels can be run as “submarine gels”, in which the gel is merely submerged in the buffer, with the electrical current running through both the buffer and the gel.

Agarose Gel Procedure

Carefully seal the ends of casting tray. (The diagram at right shows the ends sealed with tape; the apparatus you will probably be using is slightly different, and has its own built-in sealing mechanism.)

Place the comb in position on the casting tray.

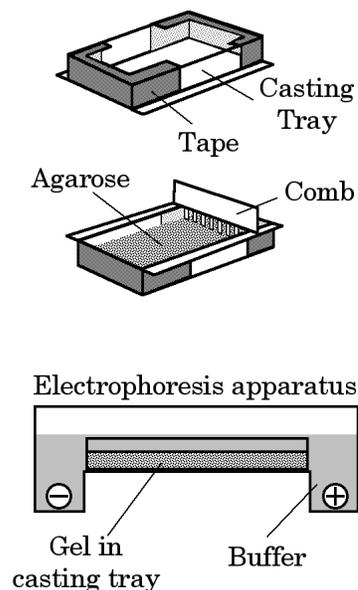
Mix the agarose and TBE at a ratio of 1 g agarose per 100 ml TBE, and heat the solution in microwave oven until the agarose melts and is evenly distributed in the solution.

Add 1 μ l of 10 mg/ml ethidium bromide for every 25 ml of melted agarose.

Allow the agarose to cool slightly (to avoid warping the plastic) and pour the agarose onto the casting tray.

After the gel cools (20 minutes), remove the sealing tape and the comb. Place the casting tray with the gel into the electrophoresis apparatus, and add 1x TBE to just above the level of the gel.

Load DNA samples containing 10% Loading Buffer. Run the gel under constant voltage conditions at 100 volts.



In the presence of DNA, when excited by ultraviolet light, ethidium bromide fluoresces bright orange. This DNA is usually visualized by placing the gel on a UV transilluminator. A Wratten 22A orange filter, which filters out the other wavelengths emitted by the transilluminator, can then be used to photograph the gel to create a permanent record.

Agarose gel Loading Buffer:

42 mg Bromophenol Blue
6.7 g sucrose
to 10 ml in H₂O

1x TBE:

0.09 M Tris-HCl, pH 8.0
0.09 M boric acid
2 mM EDTA

4x TBE: (4 L)

64 ml 0.5 M EDTA
174.4 g Tris base
89.0 g Boric acid
QS to 4 L with ddH₂O

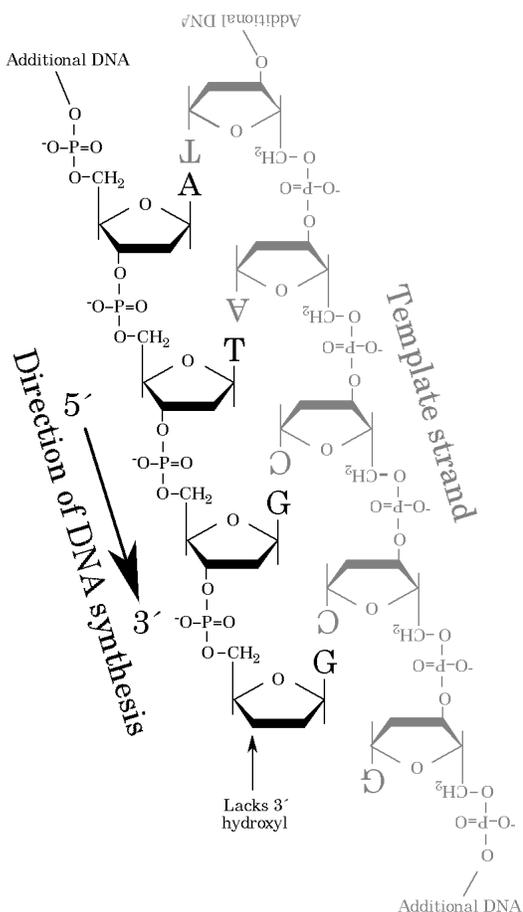
Electrophoretic Techniques: DNA Sequencing

The last step in constructing a plasmid, and especially in constructing mutant forms of a plasmid, is to verify that the correct plasmid has actually been constructed. Many of the procedures involved in plasmid construction (especially PCR) are associated with significant error rates; on the other hand, mutagenesis techniques do not always result in the intended mutation (and in some cases, do not result in any mutation at all). The best method for verifying the presence of intended mutations and the absence of unintended ones is to sequence the DNA.

DNA sequencing is a well-established procedure; it is far easier to sequence DNA than to sequence either RNA or proteins. Like PCR, DNA sequencing involves a modification of a normal cellular DNA replication process. As with most DNA synthetic reactions, sequencing reactions require an oligonucleotide to use as a primer, a template to use for synthesizing a complementary strand, and dNTPs to use as substrates for synthesis. However, the DNA sequencing reaction also requires a way of identifying each base added to the growing DNA strand.

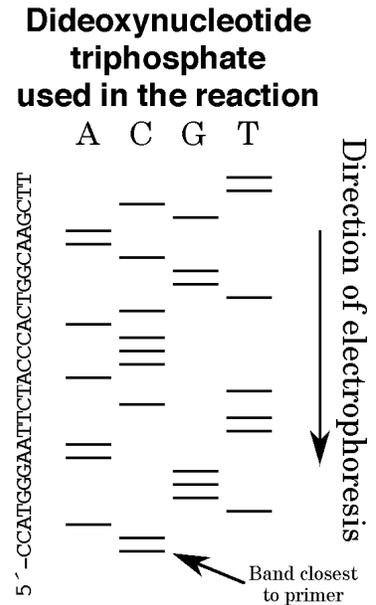
The commonly used methods for DNA sequencing all involve the use of “chain terminators”, which are modified nucleotides that can be incorporated into the new DNA strand, but do not permit continued synthesis of DNA. The chain-terminators lack a 3'-hydroxyl group (in the example at right, dideoxyGTP was inserted instead of dGTP). The absence of the 3' hydroxyl group prevents addition of the next base, and therefore terminates DNA synthesis. If the sequencing reaction is set up such that the identity of the chain terminator nucleotide added is known, it is possible to identify the nucleotide at each position.

One method, being used with increasing frequency, involves the use of fluorescently labeled chain terminators. If each type (*i.e.* A, C, G, and T) of nucleotide has a different fluorescent label, the sequence can be determined very readily using an automated system. Variations on this method are being used in the genome sequencing projects currently in progress. This method is also used by sequencing services (you send your DNA sample to a lab, and they tell you what the sequence is without you having to actually do the work). The fluorescent-tag system, however, requires expensive equipment.



Laboratories that perform small scale sequencing usually use a somewhat older method in which four separate reactions are run, one for each of the possible chain terminator nucleotides; in each case, the newly synthesized DNA is radioactively labeled so that it can be detected (usually by exposing the gel to film). After the reaction is complete, the DNA is run on a polyacrylamide gel; sequencing gels can separate DNA fragments that differ in size by a single nucleotide.

A representation of a sequencing gel autoradiograph (*i.e.* the developed film after an incubation of the gel with the film) is shown at right. Four reactions were run using the same template DNA, but with only one dideoxynucleotide. The resulting DNA fragments were then separated based on their size; as usual, the smallest bands migrated at the bottom. The sequence can be read directly from the gel; simply start at the bottom, and for each band, note the lane in which the band occurred (the sequence derived from this “experiment” is shown on the left side of the figure).



The semi-automated sequencing machines use a generally similar method; the difference is that, instead of stopping the gel part-way through the separation (as is shown here), the bands are detected (based on the fluorescent tags on the ddNTPs) as each one runs off the end of the gel.

Depending on how well the reaction runs, on the homogeneity of the gel, on the quality of the template DNA, and on the exact method used, a single sequencing reaction can yield anywhere from 100 to 1000 bases of information. A cDNA of 1000 bp would usually require 2 to 4 reactions (using different primer positions) to read the entire sequence. The human genome (3×10^9 bp of unique sequence) would require at least 3×10^6 sequencing reactions; in practice, it is going to require many more sequencing reactions than that, in order to sequence each part of the genome more than once.

The “gel” shown above is an idealized example. Some sequencing gels actually look similar to this; most, however, have artifacts of various types, which make reading the sequence somewhat more challenging. Artifacts can be due to inhomogeneities in the gel (if some lanes run faster than others, it can be difficult to read the sequence). If some copies of the DNA template have holes (or other problems that force the polymerase to stop synthesis), the gel may show bands in all of the lanes (making it difficult to decide which is the correct base). In addition, the sequence itself may cause problems: the GC base-pair binds more tightly than the AT base-pair; sequences with high GC content are especially subject to artifacts where the polymerase has trouble synthesizing the new strand, or where the DNA forms secondary structure while running on the gel which tends to alter its speed of migration through the gel. While experienced researchers can usually compensate for these problems, determining the sequence of unknown DNA samples can be quite difficult. Verifying a known or expected sequence (for example, when checking for the presence of point mutations

following a site-directed mutagenesis experiment) is usually much less difficult, but can still be subject to errors.

Procedure:

Sequencing reactions require a DNA template. In most cases, the DNA template is obtained by performing plasmid preps on some of the colonies previously identified as positives for a sequence of interest.

As with most molecular biological techniques, DNA sequencing is typically performed using a kit. Consult the manufacturer's instructions for the precise protocol to use. The procedure will involve mixing an oligonucleotide primer with the template DNA and allowing a polymerase to synthesize a new strand of DNA. The procedure is thus a variation on the PCR method, with two modifications: 1) only one oligonucleotide is necessary, and 2) four separate reactions must be run, with each reaction containing a single dideoxynucleotide for specific chain termination (in addition to the four normal deoxynucleotides).

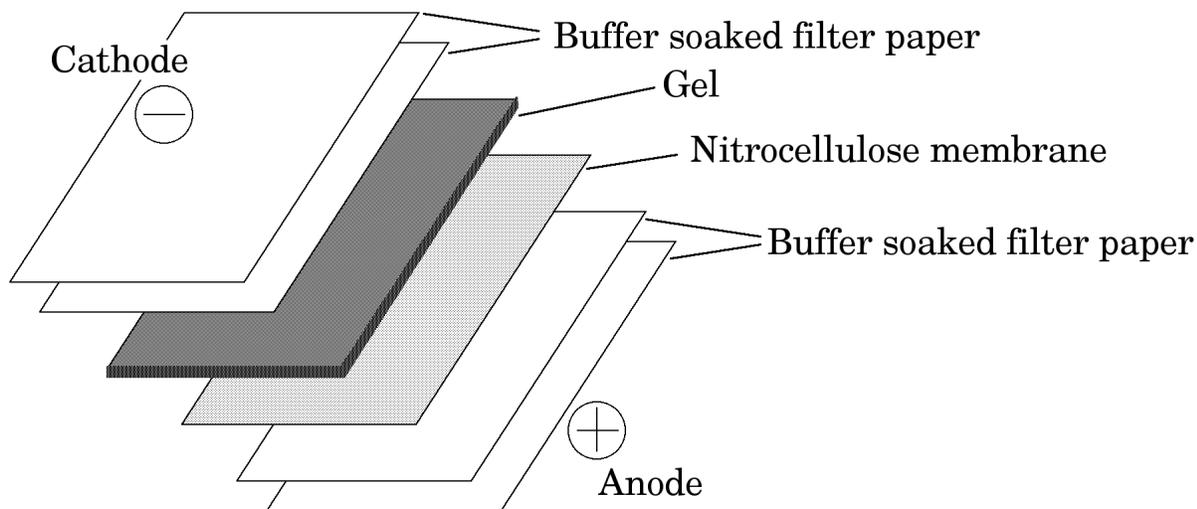
Electrophoretic Techniques: Western Blotting

SDS PAGE allows the separation of proteins based on their size. However, it does not allow the unambiguous identification of any given protein, because the Coomassie blue dye used for protein detection binds to all proteins. Unambiguous identification requires a more specific detection method. Western blotting is one technique that allows the identification of the location of one particular protein from an electrophoretically separated mixture of proteins. In Western blotting, proteins are separated on a gel, are transferred to a solid support, and then are probed for using reagents specific for a particular amino acid sequence. For Western blots, the most commonly used probe is an antiserum raised against the protein of interest. However, protein-antibody interactions alone will not reveal the location of the protein on the solid support, since no visible change occurs. This primary interaction must therefore be coupled with another interaction that produces a detectable product. Detectable products include colored precipitates, emitted light, or radioactive compounds that specifically bind the antibody.

Transfer of the gel

The first step of western blotting involves transferring the protein bands from the gel to a solid support. Various membranes are available, with polyvinylidene difluoride (PVDF) and nitrocellulose being the most commonly used. PVDF membranes are favored because of their high mechanical strength, chemical stability, and enhanced protein binding; on the other hand, nitrocellulose is somewhat less expensive. The transfer of protein bands is accomplished by electrophoretic migration of the proteins from the gel to the membrane.

The most frequently used procedure involves the semi-dry method of gel transfer. The set up is shown below.



Procedure for running the SDS PAGE and blotting:

Chemicals required:

Transfer Buffer

192 mM Glycine
25 mM Tris pH 8.3
1.3 mM SDS
20% methanol

Blocking solution (1% gelatin or 3% BSA in TBST + 0.02% sodium azide)

Tris Buffered Saline (TBS)

50 mM Tris-HCl, pH 7.5
150 mM NaCl

TBST: TBS with 0.1% Tween-20

Hardware required:

Electrophoresis apparatus
100°C Water bath
Gel loading tips
Weigh boats for gel staining
Hoefer Semi-Dry Transfer apparatus
Filter paper
Transfer membrane

10% SDS polyacrylamide gels
Electrophoresis tank buffer
5x Sample buffer
Coomassie Staining Solution

1. Set up the gel apparatus

Mark the location of wells before adding tank buffer. After adding the tank buffer, remove the bubbles from the bottom of the gel.

2. Prepare samples

In most cases, it is necessary to run at least one SDS-PAGE gel to determine the appropriate concentration of protein to load for the Western blot. Ideally, Western blotting involves concentrations of protein that yield sharp, easily detectable, individual bands. Once the correct protein concentration is determined, enough of each sample should be prepared to allow two lanes of each sample to be run.

3. Loading samples

A common method of running a Western blot is to cut the SDS PAGE in half after electrophoresis, with one half stained with Coomassie blue and the other transferred for Western blotting. Therefore, load to gel to yield two identical halves.

Record how the samples were loaded!

4. Run the gel

Attach the electrodes to the gel apparatus, and turn on the power supply. Run the gel at constant current (15 mA per gel) until the tracking dye reaches the end of the gel. Be careful: the high current levels used during gel electrophoresis can be dangerous!

5. Disassemble the gel

Turn off the power supply and remove the electrodes. Remove the gel sandwich. Remove one spacer, and insert it in one corner between the glass plates. Use the spacer to gently pry off one glass plate. Remove the stacking gel and discard it.

6. Separate the gel

Cut the resolving gel in half. Place half of the gel (the half with the unstained protein standards) in a weigh boat with Coomassie blue stain and incubate overnight. The other half of the gel (the one with the stained protein markers) will be transferred to the membrane.

7. Transfer the protein

To prepare the “transphor sandwich” cut 4 pieces of filter paper and 1 piece of membrane to the same size as the gel. Briefly soak the filter paper and the membrane in transfer buffer. Place 2 pieces of the soaked filter paper on the anode (positive electrode), followed by the membrane. Then place the gel on top of the membrane and put two more pieces of the soaked filter paper on top. Make sure no air bubbles are trapped underneath; this can be accomplished by rolling a glass pipet or test tube over the transphor sandwich to remove any trapped air bubbles.

Transfer will occur for 30 minutes at the appropriate amperage. In most cases, an appropriate amperage can be calculated based on the size of the gel, in order to obtain 0.8 mA/cm^2 of gel. For gels run using the Bio-Rad gel apparatus, is approximately 30 mA.

8. Block the non-specific protein binding sites

When the transfer is complete, place the membrane in blocking solution in a weigh boat on the shaker platform until next laboratory period.

Western blotting incubations

Chemicals required:

Tris Buffered Saline (TBS)

50 mM Tris-HCl, pH 7.5
150 mM NaCl

TBST:

TBS with 0.1% Tween-20

Antibody Buffer:

1% gelatin in TBST

Primary Antibody

Secondary Antibody (Biotin-conjugated Goat anti-rabbit immunoglobulin 10 μ l/30 ml antibody buffer)

Avidin-alkaline phosphatase

(AvidinAP) (10 μ l/30 ml antibody buffer)

Development Solution

200 μ l BCIP (15 mg/ml in DMF)
200 μ l 30 mg/ml NBT in 70% DMF
20 ml 100 mM Tris-HCl, 200 mM NaCl,
5 mM MgCl₂, pH 8.0

Procedure

Although Western blots can be accomplished in a single day, in most cases, the process is a two-day procedure. The first day is used to run the gel, to transfer the proteins to the membrane, and to block the non-specific protein binding sites using the blocking solution. The second day is used to perform the antibody incubations and to carry out the detection process.

The second part of the Western blot procedure involves incubation of the membrane with the antibody and the detection reagents. An overview of the Western blot procedure is as follows:

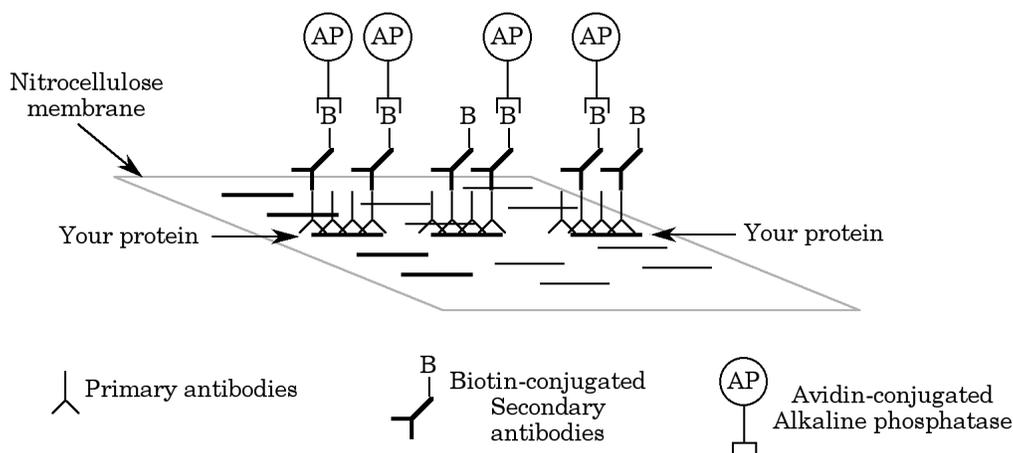
1. Block unbound membrane protein binding sites
2. Incubate with primary antibody
3. Wash with TBST
4. Incubate with secondary antibody
5. Wash with TBST
6. Incubate in color development reagents

Incubations

Blocking. The first step (which should be complete) is to incubate the membrane in a blocking solution (a solution which contains high concentration of protein). The blocking solution is usually prepared with either bovine serum albumin or milk proteins. This step is referred to “blocking” because the protein in the blocking solution binds to non-specific protein binding sites on the membrane, and therefore prevents antibodies (which are proteins) from binding to these sites.

Primary antibody. After blocking, the membrane is incubated in a solution containing the primary antibody. The antibody typically used is an antiserum from rabbits inoculated with either the specific protein of interest, or a closely related protein. The antibody should bind to any of its antigen molecules present on the membrane. Most antibodies have very high binding affinities for their antigens, and the complex formed should remain intact during the wash steps. Excess antibody is removed by washing in a Tris-buffered saline solution that contains small amounts of the detergent Tween-20 (TBST buffer).

Secondary antibody. To detect the antibody-antigen interaction a secondary antibody is added. An example of a secondary antibody that can be used is a goat-anti-rabbit immunoglobulin antibody. This is an antibody raised in goats that recognizes rabbit immunoglobulin. Incubation of the membrane will create an antigen:primary-antibody:secondary-antibody ternary complex as shown below. Excess secondary antibody is removed by washing in TBST buffer as was done for the primary antibody.



Color detection. One method used for color detection is to take advantage of the strong affinity between the protein avidin and the small molecule biotin. In this approach, biotin is covalently attached to the secondary antibody (creating the “biotin-conjugated goat anti-rabbit immunoglobulin antibody” that is frequently used in these types of experiments). After incubation of the membrane with the biotin-conjugated antibody, the membrane is incubated with an enzyme covalently attached to avidin. The membrane containing the multiprotein complex (see the figure above) can then be incubated with substrates for the enzyme reaction that ultimately produces a detectable product. One commonly used enzyme for this process is

alkaline phosphatase, which catalyzes reactions that look like this: $\text{ROPO}_3^{2-} + \text{H}_2\text{O} \rightarrow \text{ROH} + \text{HPO}_4^-$. The substrate 5-bromo-4-chloro-3-indolyl phosphate (BCIP) is dephosphorylated by the alkaline phosphatase; the dephosphorylated product can then react with nitro blue tetrazolium (NBT) to form a purple/black precipitate. This reaction can detect 25-50 pg of alkaline phosphatase on a Western blot. The sensitivity may be increased by increasing the incubation time with the substrates and allow more product to form.

Remember to wear gloves when handling the membrane to prevent finger proteins and oils from interfering in the experiment. Please take **only** the amount of reagent needed.

After the membrane has been incubated in blocking solution for a minimum of two hours, it is possible to begin the antibody incubations.

Note: The successive washes are time consuming. Work efficiently and be prepared for the next step!

1. Prepare an appropriate dilution of primary antibody in antibody buffer.
2. Transfer the blocked membrane to the primary antibody solution (made up in step 1). Incubate with shaking at room temperature for approximately 30 minutes.
3. Wash the membrane three times (5 minutes each wash) with 30 ml (each wash) of TBST buffer.
4. Prepare the secondary antibody dilution.
5. Transfer the blocked membrane to the secondary antibody solution (made up in step 4). Incubate with shaking at room temperature for approximately 30 minutes.
6. Wash the membrane three times (5 minutes each wash) with 30 ml (each wash) of TBST buffer.
7. Transfer the membrane to 30 ml of Avidin-AP solution and incubate at room temperature with shaking for approximately 45 minutes.
8. Wash the membrane three times (5 minutes each wash) with 30 ml (each wash) of TBST buffer.
9. Transfer the membrane to freshly prepared Development Solution. Let color develop sufficiently (color develops best in relatively low light). When color has developed, stop the reaction by immersing and rinsing the membrane in deionized water for approximately 10 minutes.

10. Photograph the developed membrane to create a permanent record of the experiment.

Methods: Protein Assays

In biochemistry, it is frequently useful to know the total protein concentration in a solution. A number of methods have been developed for measuring total protein concentration. Perhaps the most commonly used method is called the Bradford Assay, after the name of the scientist who developed the method. The method was first described in a paper written more than 20 years ago³.

The Bradford reagent binds proteins; when it does so, its extinction coefficient at 595 nm increases. Because the magnitude of the extinction coefficient change is somewhat dependent on the conditions, it is necessary to calibrate the change in absorbance induced by different amounts of protein. This calibration procedure is called “generating a standard curve”.

Ideally, the absorbance readings obtained in a standard curve should be linearly proportional to the protein concentration (which means that there are limits as to how much protein that can be used, because too much protein will give absorbance readings too high to be meaningful). Time constraints mean that it will probably be able possible to run each sample only once. In a normal research laboratory project would require running replicates for both the standard curve and the unknown samples, so as to be able to assess the uncertainty in the data.

Supplies

Bradford Reagent
1 mg/ml Bovine serum albumin (BSA)

Cuvettes

Procedure

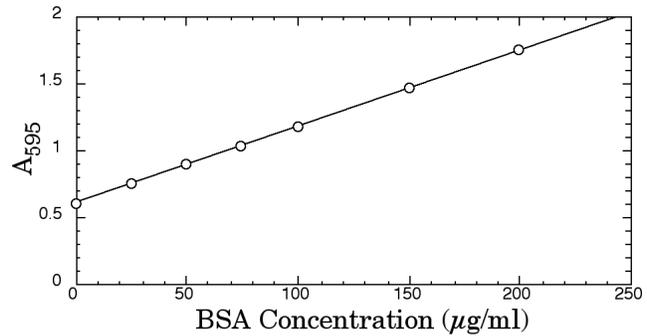
Standard Curve: Most laboratories have stock solutions of standard proteins (usually bovine serum albumin (BSA) or γ -globulin). The following discussion assumes the use of a stock solution of 1 mg/ml BSA. Note that 1 mg/ml = 1000 μ g/ml. In order to set up a standard curve, it is necessary to make up several dilutions of the BSA stock, and then use a constant volume from each protein dilution in the actual assay tubes. **One useful set of standards making 25 μ g/ml, 50 μ g/ml, 75 μ g/ml, 100 μ g/ml, 150 μ g/ml, and 200 μ g/ml solutions of BSA.** Since the assay requires 100 μ l of the protein solution, it is probably best to make convenient dilutions with final volumes of 150 to 400 μ l. **When planning to perform a Bradford assay, decide how to make up these dilutions *before* coming to class.**

Mix **800 μ l of H₂O, 200 μ l of the Bradford Coomassie Blue solution and 100 μ l of the BSA** standard dilution together, allow them to equilibrate for a few minutes, and measure the absorbance at 595 nm. In addition to BSA protein solutions, run one sample without protein (*i.e.* which contains the same assay mixture, using buffer

³Bradford, M.M. (1976) *Anal. Biochem.* **72**: 248-254

instead of protein). This zero protein data point is a useful control and acts as a part of the standard curve.

Your plot of the absorbance at 595 nm *versus* protein concentration should look similar to the one shown at right (your data may not fit the line as closely as these do, although they will if you are careful with your dilutions).



You will probably also need to dilute your unknown protein solutions. You will prefer to use small volumes of your precious protein solutions (you worked hard to prepare those protein solutions, and you will need them again for later experiments). You will find that many of the samples you have will be too concentrated to run undiluted in the assay. If you perform 1:3 serial dilutions of your protein, you will eventually obtain samples that fall within the standard curve. (Reminder: **serial dilutions** are performed by diluting a sample, then taking an aliquot from the diluted sample and diluting it further, and then taking an aliquot from the second dilution and diluting it further, and) You will need to make sure that at least one of your diluted protein samples has an absorbance that falls within the standard curve.

Once you have measured the absorbance values for the standard curve and for your unknown protein samples, you will need to use the standard curve to determine the protein concentrations for your unknowns. The best method is to plot the values obtained for the standard curve, determine the slope of the best-fit line, and then use the equation of the line to give you the protein concentrations for your unknown samples. (Don't forget to correct the value obtained in this method for the dilution you performed: your goal is to calculate the protein concentration in the initial sample.)

Methods: The Art and Science of Protein Purification

When attempting to understand how a protein works, it is usually necessary to isolate the protein from other proteins that are present in the tissue. This allows you to study the protein with some assurance that the results reflect the protein of interest and are not due to other molecules that were originally present in the tissue. Protein purification is therefore a commonly used biochemical technique.

Most proteins are fairly large molecules. They are smaller than DNA molecules, but they are tremendously large when compared to the molecules typical organic chemists are concerned with. The three-dimensional structure of most proteins is a consequence of many relatively weak non-covalent interactions. Disrupting this three-dimensional structure, on which the function of the protein depends, is therefore a relatively easy process. Conversely, preventing the loss of the non-covalent structure (and sometimes the covalent structure) is frequently difficult.

Disrupting cellular structure is required to release the proteins from the cell. However, the process has two side effects that may damage proteins: 1) cell disruption typically involves shearing forces and heat, both of which can damage proteins, and 2) cells normally contain proteases (enzymes that hydrolyze other proteins). In most cells, proteases are carefully controlled; however, disruption of the cell usually also releases the proteases from their control systems, and may allow the cleavage of the protein of interest.

Purification of proteins involves taking advantage of differences between the protein of interest and the remaining proteins present in the mixture. Because proteins are all polymers of the same twenty amino acids, the differences in properties tend to be fairly small.

In most cases, current understanding of protein structural properties is insufficient to allow a purification method to be generated theoretically. The “Art” in the title of this section reflects the fact that development of most protein purification procedures is a matter of trial and error. The table below lists some of the general properties of proteins that can be useful for protein purification, and some of the methods that take advantage of these properties. Each of these general methods will be discussed in some detail below. Note that for any given protein, only some of these methods will be useful, and therefore protein purification schemes vary widely.

Property	Technique
Solubility	Ammonium sulfate precipitation
Charge	Ion-exchange chromatography
Hydrophobicity	Hydrophobic interaction chromatography
Size	Gel-filtration chromatography
Function	Affinity chromatography
Stability	Heat-treatment, pH treatment

Ammonium sulfate precipitation

In many cases, cell lysates can be loaded directly onto chromatography columns.

However, in some cases other molecules present in the lysate interfere with binding of the protein to the resin. In addition, some resins (especially affinity resins and sepharose-based resins) are fairly expensive; loading crude cell lysates on these columns may result in binding of cellular material (e.g. lipids and DNA) that are difficult to remove, and which may damage the column. As a result, purification methods often begin with one of several possible simple techniques that remove at least some of these unwanted materials prior to using an expensive column.

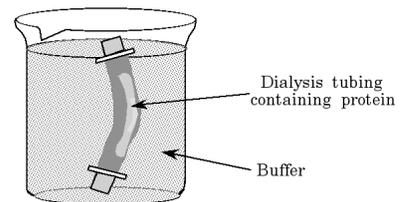
One of the most commonly used crude purification techniques involves the use of differential **solubility**. Proteins precipitate with increasing ammonium sulfate concentrations, with most proteins precipitating somewhere between 10% and 60% ammonium sulfate. (The percentages are relative to a saturated solution, which has a concentration of about 4 M; thus most proteins precipitate between 0.4 M and 2.4 M.) This can allow a simple, partial, purification of a protein; if the protein of interest precipitates at 40% ammonium sulfate, many other proteins will remain in solution, as will many other non-protein molecules.

Most proteins are not damaged by ammonium sulfate precipitation, and can be resuspended in a small volume of buffer. Ammonium sulfate precipitation results in a high salt concentration in the protein solution; this may be advantageous (if the intended next step is hydrophobic interaction chromatography), or deleterious (if the next step is ion exchange chromatography).

When necessary, two methods are frequently used to remove the salt. One method is gel filtration chromatography (discussed briefly below). Another frequently used method is dialysis.

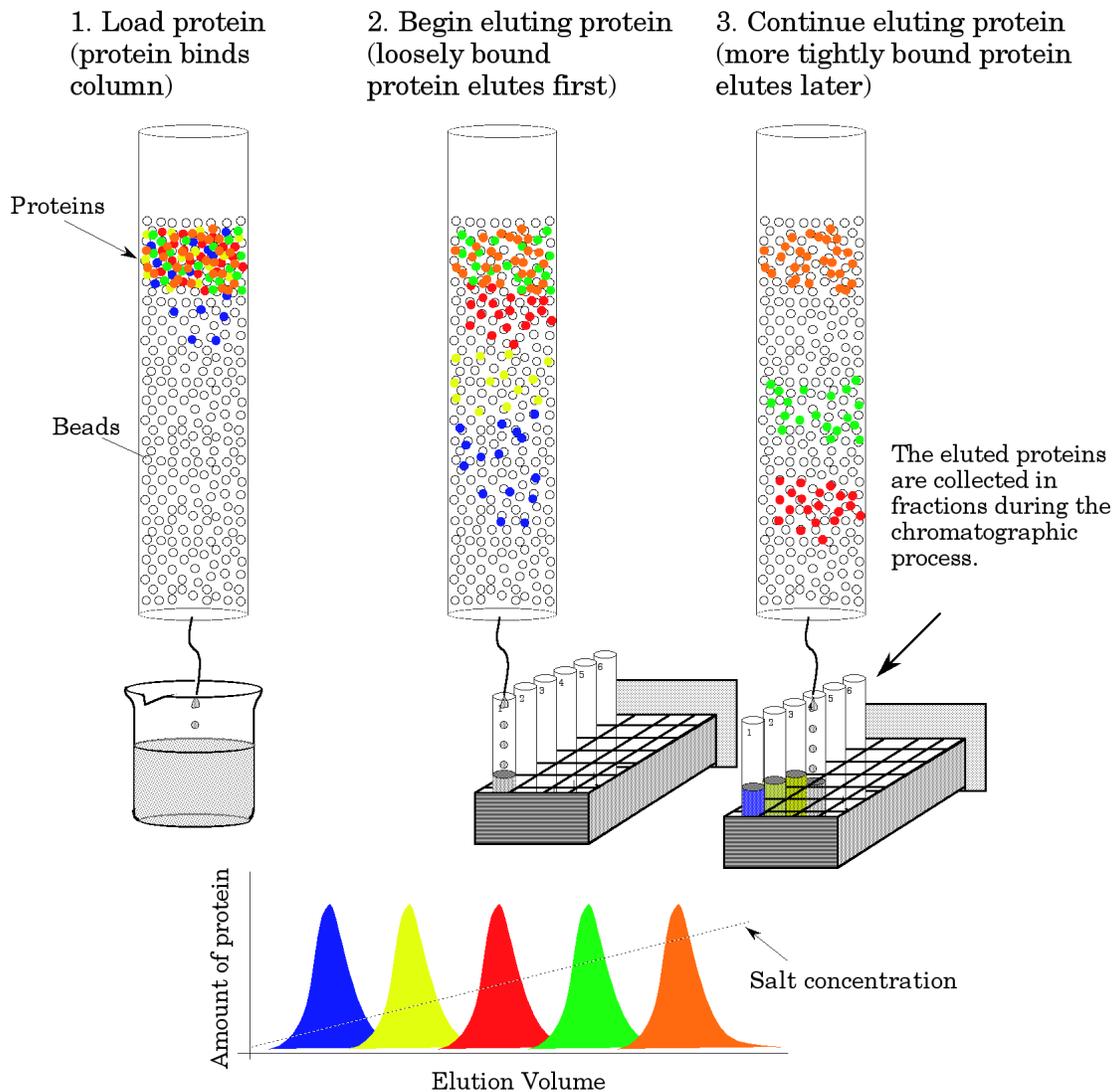
Dialysis

Dialysis involves placing the protein solution in a semi-permeable membrane, and placing the membrane in a large container of buffer. Small molecules (such as salt ions) pass through the dialysis membrane (moving from high concentration to low concentration), while large molecules are unable to cross the membrane. Dialysis membranes come in a variety of pore sizes, and are therefore useful for removing a variety of different sized solutes. In principle, dialysis could allow separation of large proteins from small ones; in practice, however, the pores in the tubing are insufficiently uniform to allow this technique to be used effectively.



Chromatographic methods

Most purification methods involve chromatography. Chromatographic methods involve a column of an insoluble material that can bind molecules based on specific properties common to proteins. The solution containing the mixture of proteins is then allowed to pass through the column; the protein of interest may bind (depending on its properties), while at least some impurities remain in solution and leave the column. The procedure is completed by eluting (*i.e.* “removing”) the proteins that have bound to the column.



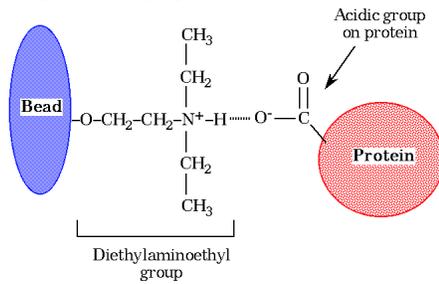
An illustration of a chromatographic run is shown above. The initial sample contains five different proteins (the differently colored filled circles). These proteins are bound to the column fairly tightly. Once elution begins, the proteins begin leaving the column. The graph at the bottom of the diagram shows proteins eluting with increasing salt concentration, in the manner that would occur with an ion exchange column; otherwise, this diagram applies to essentially any type of chromatographic method. Note: most columns do **not** run this neatly, especially in the beginning of a purification procedure.

Ion exchange chromatography

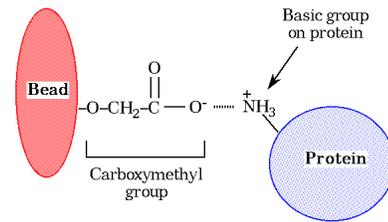
Proteins are charged molecules. They will therefore bind to other molecules of opposite charge. Ion exchange columns are produced by covalently attaching charged molecules such as diethyl-aminoethyl groups to insoluble carbohydrate resins. In many cases, small differences in charge can result in significant separations on ion exchange columns. Ion exchange columns are typically loaded at low ionic strength, and the protein removed by raising the ionic strength.

Types of ion exchange chromatography

Anion exchange chromatography



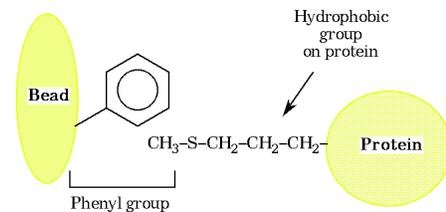
Cation exchange chromatography



Hydrophobic interaction chromatography

Proteins contain hydrophobic amino acid side chains, some of which are exposed at the surface of the protein. Proteins will therefore often bind to other hydrophobic molecules. Hydrophobic interaction columns are produced by covalently attaching hydrophobic molecules such as acyl chains or phenyl groups to insoluble carbohydrate resins.

Hydrophobic interaction chromatography

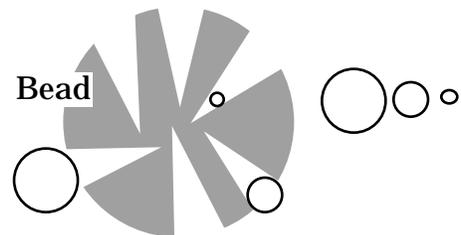


The hydrophobic effect is strongest under high ionic strength conditions; hydrophobic interaction columns are therefore typically loaded at high ionic strength, and the protein removed by lowering the ionic strength (thus, these columns are the opposite of ion exchange columns).

Gel filtration chromatography

In gel filtration chromatography (also known as size exclusion, gel permeation, or molecular sieve chromatography), molecules are separated based on size. As with SDS PAGE, a gel filtration chromatography column can be calibrated by using molecules of known size; the gel filtration column can then be used to determine the size of an unknown protein. In contrast to SDS PAGE, gel filtration measures the size of the native protein (as opposed to denatured proteins).

Gel filtration columns are made of porous beads packed into a column. The beads can be polymers of dextran (Sephadex), agarose (Sephacrose and Superose), agarose cross-linked to dextran (Superdex), polyacrylamide (Sephacryl), or related compounds. Different types of beads have somewhat different physical properties that may make them more appropriate for different proteins.



In contrast to ion exchange or hydrophobic interaction chromatography, in gel filtration chromatography the protein sample should be applied to the column in the smallest reasonable volume to ensure good separation. The molecules do not form tight associations with the column, and if loaded in a large volume, will elute in a large volume with little separation between large and small molecules. The maximum

volume that can be applied to a column varies somewhat with bead type (e.g., about 2 to 3% of the bed volume for Sephadex, and about 0.5 % of the bed volume for Superdex).

As a solution containing molecules of varying sizes passes through the column, the molecules distribute between the inside and outside of the pores depending on their size. Molecules too big for the pores are totally excluded, and elute from the column first. Smaller molecules fit in the pores, and therefore elute later. The elution volume for a molecule is thus inversely related to the size of the molecule.

Gel filtration beads are characterized by their composition. They are also characterized by their exclusion limit, which is the size of the smallest molecule incapable of entering the beads. Beads with a variety of compositions and pore sizes are available. For example, a Sephadex G-100 bead is made of a dextran polymer and should be able to separate molecules of up to 100 kDa; molecules bigger than 100 kDa will be completely excluded and will run at the void volume.

The “**void volume**” is the elution volume of a molecule so large that it is totally unable to enter any of the pores. The void volume therefore represents the minimum possible elution volume for any molecule. Another useful parameter for a gel filtration column is “**column volume**”, which is the volume of an empty cylinder the size of the column being used. This represents the maximum elution volume for a very small molecule.

Certain problems are inherent in the process of gel filtration. First, even if the sample is applied as a narrow band to the column, the turbulence associated with passing the mobile phase through the column results in a broadening of the bands of the applied molecules as they travel through the column. The broadening of the bands works against the purification of the individual components of the sample, since the broad peaks can partially overlap one another. Thus, it is advisable to apply the sample as a narrow band on the top of the column. This can limit the usefulness of the method, because the sample may in some cases be difficult to concentrate sufficiently to make the method optimal. Another problem with gel filtration is that proteins may interact with the column matrix. This interaction will retard the elution of the protein and the apparent molecular weight will be lower than expected. To prevent interaction of the protein with the matrix the gel filtration should be run under moderate ionic strength to prevent ionic interactions between the proteins and the column matrix.

Separation on gel filtration columns increases in proportion to the square root of the column length. Longer columns therefore result in higher resolution. Unfortunately, both the cost of the column and the time the column takes to run increase in direct proportion to the length of the column.

Separation is inversely proportional to flow rate; thus, improved separation typically requires longer chromatography times. The maximum flow rate for a particular separation on a particular column can be estimated from theoretical considerations, but usually has to be determined empirically (*i.e.* try a flow rate, and see if the results are acceptable).

Affinity chromatography

Many proteins exhibit specific interactions with other molecules (called ligands); for example, enzymes must have the ability to bind to their substrates, and antibodies exhibit high affinity interactions with their antigens. In principle, it is possible to covalently attach the ligand to an insoluble resin. A column produced from this resin is called an affinity column.

Affinity chromatography is somewhat less commonly used than the forms of chromatography discussed above. In many cases, the covalent attachment of the ligand to the column results in steric clashes that prevent the protein from binding. In some cases, although the protein will bind to the affinity resin, the resin is so expensive that other purification methods are used instead. However, affinity chromatography can be extremely useful for purifying difficult-to-isolate proteins. While most proteins contain charged groups, the specificity of protein-ligand interactions means that only a very small fraction of the proteins in a cell will bind to any given affinity resin. Affinity chromatography can therefore be an extremely useful purification technique; in some cases, a single affinity chromatography step may be the only step necessary to completely purify a protein.

Procedure for pouring a column

HPLC columns are packed by machines, and tend to cost several hundred dollars each. Gravity columns, however, are generally prepared by the user, and are usually much less expensive.

Preparing a gravity column requires preparing the beads. Some types of beads, (especially gel filtration beads) need to be “swollen”. The swelling of the beads is a consequence of hydration of the beads, and usually takes 24 to 72 hours, depending on the bead type.

1. Make sure that the beads and buffer that you will use are at the temperature you will use for pouring the column. If you are pouring the column at room temperature using cold buffer, air bubbles will form as the buffer warms.
2. Gently suspend the beads in buffer using a stir rod. You want an even suspension, but you do not want to damage or crush the beads in the process. (Do not use a magnetic stirrer!) Make sure that a reasonable amount of buffer is present (too little buffer will make the suspension too viscous to pour evenly, while too much will require a long time to pour. A volume of buffer equivalent to about half the bead volume is usually reasonable.
3. Pour the bead suspension into the column. Use a stir rod to help run the suspension down the side of the column. Alternatively, tilt the column to make pouring easier. If air bubbles form, you will need to re-pour the column.
4. Allow the buffer to begin draining from the column. As the buffer drains, the beads will become more concentrated. In addition, the beads will begin to settle toward the bottom of the column. Periodically add more bead suspension to the column. Do not

allow the bead suspension to completely settle before adding more beads; this will create a discontinuity in the column that will adversely affect resolution. In addition, when pouring, try to avoid disturbing the bed of previously settled beads. You want the bed to be as even as possible.

5. After all of the beads have been added to the column, or when you believe that the column bed is sufficiently high, place the top on the column, and begin equilibrating the column with buffer.

Make sure that any column you are running does not run dry at any point. If an HPLC column runs dry, it may be irreparably damaged. If a gravity column runs dry, you will need to re-pour it.

Protein Purification Strategies

Developing a scheme for purifying a protein remains an empirical process. However, in purifying a new protein, it is sometimes possible to adapt methods used for purifying similar proteins. In addition, planning the procedure before simply trying different methods can be extremely useful. Examples of this include using an ammonium sulfate precipitation step before a hydrophobic interaction chromatography step, because the high concentration of ammonium sulfate that results from the precipitation will allow the precipitated protein (or the non-precipitated protein remaining in solution) to be loaded directly onto the column. In contrast, an ammonium sulfate precipitated protein must be dialyzed (or otherwise desalted) prior to loading on an ion exchange column. Another frequently used scheme involves an inexpensive technique such as ammonium sulfate precipitation prior to remove bulk contaminants prior to running a higher resolution but more expensive technique such as affinity chromatography.

As with most scientific procedures, the more you know about the protein, and the more you know about protein purification, the more likely it is that you will be able to design a successful purification procedure.

Methods: Cell Lysis for Protein Purification

Cell lysis procedure:

1. Pellet the cells in a microfuge tube: add 1.5 ml culture, centrifuge for 15 seconds at maximum speed, discard the supernatant, and then repeat the procedure. This will result in the cells from 3 ml of culture in a single tube.
2. Resuspend the cells in 1 ml Lysis buffer.
3. Take 300 μ l aliquot for Bradford protein assay.
4. Add 50 μ l lysozyme, and mix gently.
5. After 10-20 minutes at room temperature, microfuge for 10 minutes to pellet cell debris.
6. Use supernatant for LDH assay.

Required materials:

Lysis Buffer

50 mM Tris-HCl, pH 8.0

10 mM EDTA

2 mM Dithiothreitol

Lysozyme

50 mg/ml lysozyme in Lysis buffer

LDH assay reagents

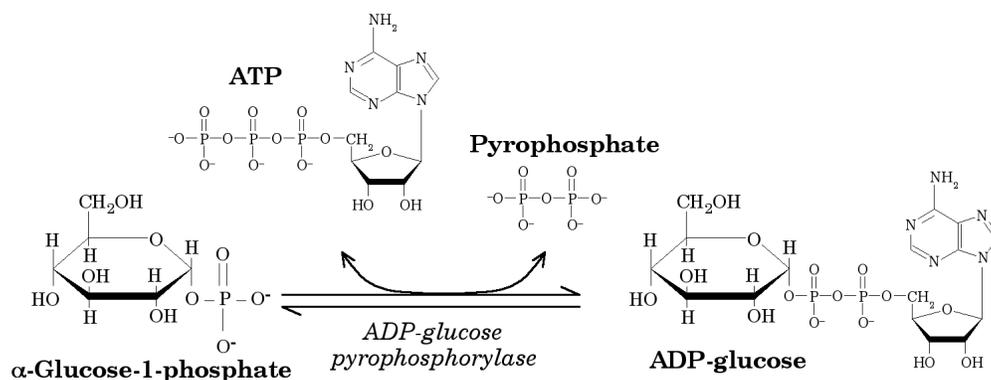
Bradford assay reagents.

Microfuge

1.5 ml microfuge tubes

Methods: ADP-Glucose Pyrophosphorylase Assay

ADP-glucose pyrophosphorylase catalyzes the rate-limiting reaction in starch and glycogen biosynthesis in plants and bacteria. The reaction it catalyzes is shown below:



The reaction as drawn is readily reversible. In the cell, pyrophosphatase hydrolyzes the pyrophosphate to release inorganic phosphate; the low pyrophosphate concentration resulting from its cleavage prevents the reverse reaction from occurring.

The enzyme can be assayed in either direction. In crude preparations, it is usually necessary to follow conversion of ADP-glucose + pyrophosphate \rightarrow ATP + glucose-1-phosphate. This is because crude preparations contain other enzymes that use glucose-1-phosphate and ATP as substrates. While the crude preparations may also contain pyrophosphatase, this enzyme can readily be inhibited under assay conditions.

Enzyme assays require the ability to specifically measure either the disappearance of substrate or the appearance of product. In most cases, looking for product formation is preferable, because in most enzyme assays, substrate concentration is fairly high, and looking for a small decrease in the amount of a compound present in high concentration is difficult. In the case of the ADP-glucose pyrophosphorylase assay you will be performing, you will be looking for formation of radioactive ATP from radioactive pyrophosphate.

You need to be able to specifically measure the radioactivity in the ATP; this means that you need to physically separate the ATP from the unreacted pyrophosphate still present in the assay mixture. To do this, you can take advantage of the fact that ATP is an organic compound; like most organic compounds, it adsorbs to charcoal. Therefore, if you add charcoal to the reaction mixture, and then centrifuge the mixture, the ATP will end up in the pellet, while the pyrophosphate will remain in solution.

Enzymatic reactions are time dependent. This means that time is a variable that must be carefully controlled. The ADP-glucose pyrophosphorylase assay you will run involves allowing the reaction to proceed for 10 minutes **exactly**. You will not be able to start and stop the reaction for all tubes simultaneously; therefore you must

stagger both the initiation and termination of the reactions. The commonly used procedure involves using 30 second intervals:

Time (seconds)	Event
0	Addition of enzyme to tube 1
30	Addition of enzyme to tube 2
60	Addition of enzyme to tube 3
90	Addition of enzyme to tube 4
.	.
.	.
.	.
600	Termination of reaction in tube 1
630	Termination of reaction in tube 2
660	Termination of reaction in tube 3
690	Termination of reaction in tube 4
.	.
.	.
.	.

Termination of the reactions involves addition of 5% Trichloroacetic acid (TCA). This has two purposes: it dilutes the reagents, and, more importantly, it denatures the enzymes in the mixture; the combination of these two processes prevents further production of radioactive ATP.

ADP-glucose pyrophosphorylase assay procedure

Label enough 16x150 mm glass tubes for each sample to be run.

Add 240 μ l of assay solution to each tube. (This may be done for you.)

Initiate reaction by adding 10 μ l of the enzyme solution to the glass tube.

Vortex the tube to mix the contents, and place the tube in the 37°C water bath.

Repeat for each tube at 30 second intervals until all tubes are in the water bath.

Terminate reaction by adding 3 ml 5% TCA to the tube, vortexing, and placing the tube on ice. Note that this needs to be after exactly 600 seconds (10 minutes) for each tube.

Add 200 μ l 15% Norit A charcoal and 100 μ l 100 mM unlabeled pyrophosphate to each tube and vortex.

Centrifuge 1 minute to pellet charcoal.

Use the vacuum suction device to aspirate off the supernatant (be careful to avoid disturbing the pellet).

Add 3 ml 5% TCA and vortex.

Centrifuge 1 minute to pellet charcoal.

Aspirate supernatant.

Add 3 ml 5% TCA and vortex.

Centrifuge 1 minute to pellet charcoal.

Aspirate supernatant.

Add 1 ml 1 M HCl and vortex.

Heat the tube in a boiling water bath for 10 minutes (place marbles on top of tubes to prevent evaporation of the water in your samples).

Cool on ice.

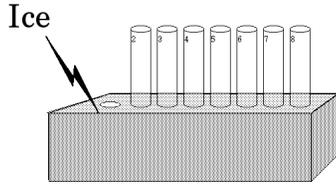
Centrifuge 1 minute to pellet charcoal.

Pipet 0.5 ml of HCl supernatant into scintillation vial. Add 5 ml scintillation fluid, and mix.

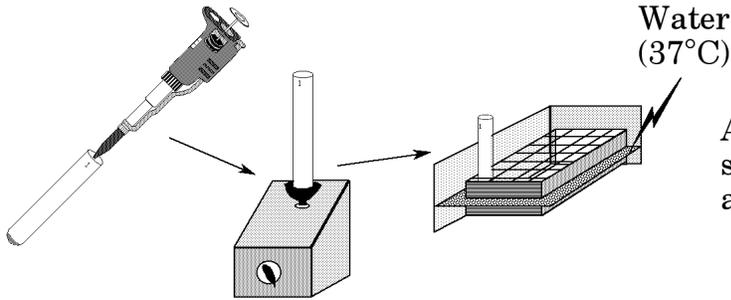
Measure radioactivity using scintillation counter; count standards to allow quantitation of radioactive ATP produced.

The standards will allow you to calculate the number of nmol of product produced during the assay. Typically, you will need to convert your CPM data to nmol/min/ μ l of enzyme (and eventually, to nmol/min/mg enzyme). This is an important calculation to understand.

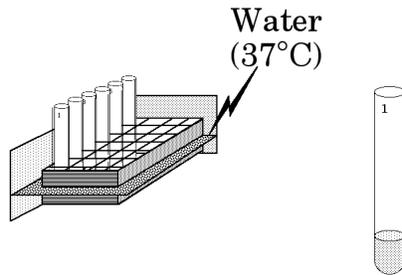
The assay procedure is summarized in the figure on the next page.



Add 240 μl of assay solution to each tube and place on ice.



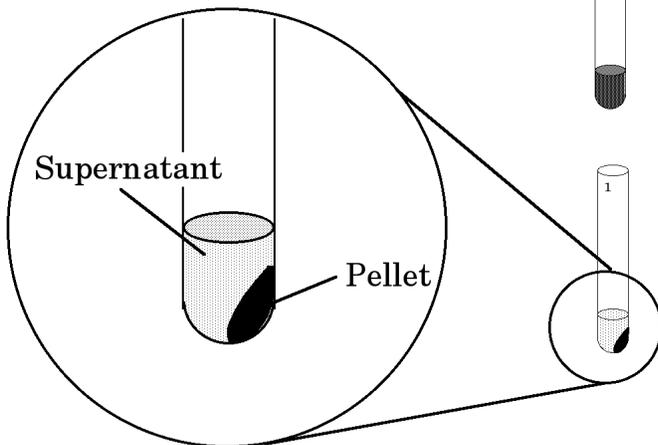
At **timed intervals**, add enzyme sample to corresponding tube, mix, and place in 37°C water bath.



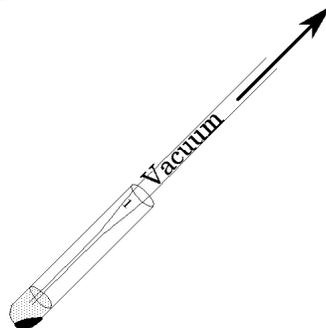
After **exactly** 10 minutes, add 3 ml of 5% TCA and mix. Place on ice while waiting for the remaining reactions to complete.



Add 200 μl of Norit A charcoal and 100 μl of 100 mM unlabelled pyrophosphate and mix.



Centrifuge for 1 minute to pellet charcoal.



Carefully aspirate supernatant (*i.e.* remove using a pasteur pipet attached to a vacuum source). **Avoid aspirating the pellet.**

Methods: Protein Crystallography

Proteins, and especially enzymes, have functions because of the precise placement of their amino acid side-chains in three-dimensional space. In order to fully understand any protein, it is usually necessary to understand the three-dimensional structure of the protein.

The most frequently used method for determining molecular three-dimensional structures involves the use of X-ray crystallographic analysis. In this technique, a crystal of the molecule is irradiated with short wavelength photons (typical X-rays have a wavelength of $\sim 1.5 \text{ \AA}$). These high energy photons are diffracted by the molecular structure, and the diffracted photons can be collected and analyzed.

In order for X-ray diffraction to work, crystals of the molecule of interest are necessary. This is true because individual molecules do not diffract X-rays in sufficient quantities for analysis. In good crystals, all of the molecules are arranged in a regular array, and therefore the diffraction pattern from the crystal is equivalent to that from a single molecule. Crystals of reasonable size provide enough diffracted photons to allow measurement of the diffraction pattern.

For small molecules, crystallization is relatively simple. However, proteins are large molecules, and crystallization can be correspondingly more difficult.

In order for a protein to crystallize, its tendency to remain in solution must be lower than its concentration in the solution. However, if this disparity is too great, the protein will aggregate and precipitate in a non-ordered fashion. It is therefore necessary to find conditions where the protein will leave the solution in an orderly manner.

If it were possible to predict the conditions that would allow the protein to crystallize, X-ray crystallography would be considerably simpler. However, in general, it is impossible to predict conditions that will allow crystallization of any given protein. Instead, the conditions must be determined empirically (*i.e.* by trying many different conditions until a condition under which crystals form is found).

The current method for finding crystallization conditions involves a screening procedure. The protein is tested with many different solutions that have previously been shown to result in crystals when used for other proteins. These solutions all contain **precipitants**, which are molecules that alter the structure of water in a way that induces proteins to leave the solution.

While the precipitant will induce the protein to leave the solution, the protein must leave the solution slowly. If the protein leaves the solution too rapidly, it will tend to aggregate rather than crystallize. This means that the concentration of precipitant must change slowly from a concentration too low to induce precipitation to a concentration above the precipitation threshold.

The usual method for raising the precipitant concentration slowly involves a technique called **vapor diffusion**. If two separate solutions are confined in a small

sealed container, solvent from the more dilute solution will tend to evaporate and condense in the more concentrated solution until both solutions have the same solute concentration. This is especially true if the more concentrated solution has a much larger volume. If the protein is added to the more dilute solution, it will experience a slow increase in solute concentration during this process. If the protein is soluble in the dilute solution, and less soluble in the more concentrated solution, it will tend to leave the solution in a relative slow fashion, and therefore have the opportunity to crystallize.

One version of the vapor diffusion technique is called the **hanging drop method**. In this method, a small volume the protein is mixed with a known ratio of the precipitant solution (usually involving a 1:2 dilution of the precipitant). This small volume of protein/precipitant is then inverted and suspended above a larger, undiluted solution of the precipitant. Assuming that the chamber is sealed, the concentration of precipitant in the protein solution will begin to rise. The time required for equilibration varies depending on precipitant type, and chamber and solution sizes, but typically involves 24-48 hours.

Finding the correct conditions for crystallization frequently involves setting up several thousand of these hanging drops using different types of precipitants, with different types of ions present, and different types of buffers present, while varying the pH and the concentration of all of these components. Precipitants commonly used for the technique include detergents, ammonium sulfate, polyethylene glycols of varying size, and small organic alcohols such as ethanol and isopropanol.

Crystallography projects tend to be relatively lengthy undertakings. The first step is obtaining purified protein (either by purifying it yourself, or by talking someone into purifying it for you). Because large-scale screening of crystallization conditions is usually necessary, it is rarely worth beginning a crystallography project unless at least 100 mg of purified protein are available. The second step is obtaining crystals, the step you are undertaking today. The third step is obtaining diffraction quality crystals. The crystals obtained from screening are usually too small, and are sometimes insufficiently well ordered to allow diffraction analysis. Once crystallization conditions are found, the conditions usually need to be modified to obtain suitable crystals. In general, reasonably large crystals are most useful; for proteins “large” crystals are about 200 μm in their smallest dimension.

The first three steps of a crystallography project involve fairly standard biochemistry. The remaining steps (collecting the diffraction data, analyzing the data, and solving the structure) are more complex, and are most commonly performed by individuals with at least some specialized training in the field of X-ray diffraction analysis.

Reagents and Supplies

24 well plates

Glass cover slips

5 ml syringes

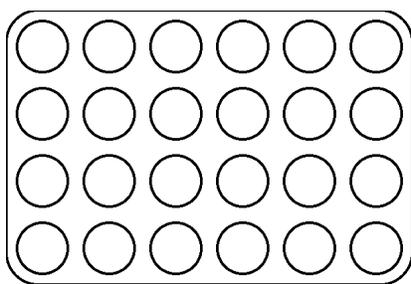
Immersion oil

Solutions for crystallography, and proteins to test for crystal formation.

Setting up a “crystal box”

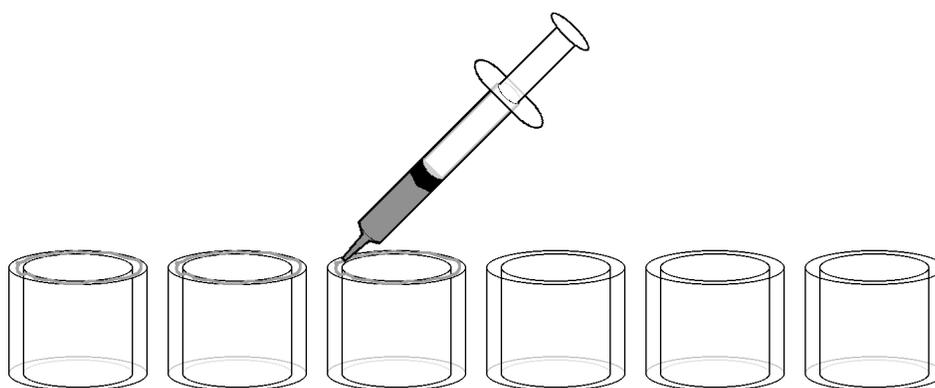
Because crystal condition screening requires testing a variety of conditions, the screening is frequently performed in 24-well culture plates. These allow a variety of conditions to be tested simultaneously, and are well suited to the hanging drop method.

The box is arranged in a six by four array. When setting up multiple conditions, this array is frequently used to allow a range of concentrations to be run in series across the box.



Procedure

Use a syringe to put a narrow band of oil around each of the six wells in the first row of the plate.

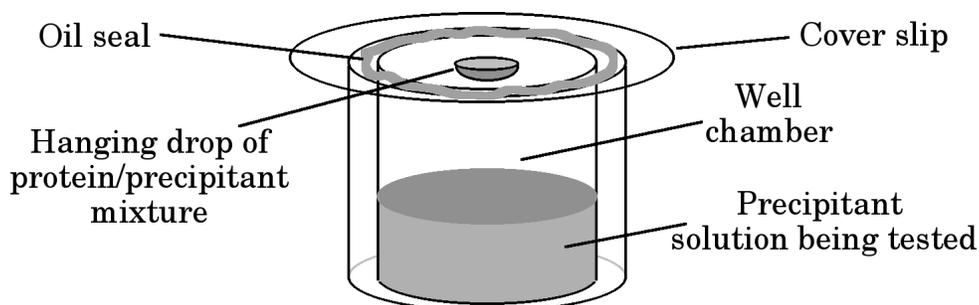


Add 1 ml of the solution to be tested in that well to each of the wells.

Lay out six glass cover slips on a small piece of paper. Avoid touching the surface of the cover slips, and try to keep the cover slips as clean as possible. Place 2 μ l of protein in the center of each of the six cover slips.

Add 2 μl of the corresponding well solution to the protein drop.

Pick up the cover slip, and place it, inverted, on the well, making sure that the oil forms a seal around the entire perimeter of the well.



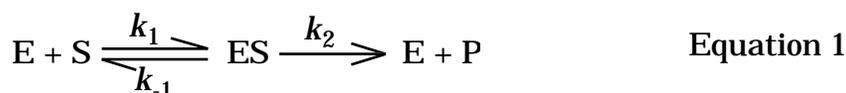
After all of the wells are set up, check the drops under a microscope for the immediate formation of precipitate, and for the presence of dirt, dust, and other objects that may appear as crystalline artifacts. After checking the drops, place the box in a cabinet where it will be protected from traffic.

Periodically (once every day or so) examine the drops under a microscope for the presence of crystals.

Introduction to Enzyme Kinetics

ADP-glucose pyrophosphorylase is an enzyme. In biochemistry and physiology, enzymes are critically important molecules, and understanding them requires understanding their interactions with their substrates. This is especially true for enzymes that you are planning to mutate; understanding the effect of the mutation requires characterization of both the wild-type and mutant forms of the enzyme.

Enzymes are catalysts: they alter the rate of a reaction. When characterizing an enzyme, it is therefore necessary to study the reaction kinetics. A general scheme for an enzyme-catalyzed reaction is:⁴



Contemplation of this reaction scheme reveals that product formation is a first order function of the concentration of ES complex. This first order rate of product formation, also termed velocity, is usually expressed as:

$$\text{velocity} = \frac{d[P]}{dt} = k_2[ES] \quad \text{Equation 2}$$

Standard chemical kinetic considerations state that concentration of ES complex depends on its rate of formation and its rate of disappearance. Allowing variations in ES complex concentration greatly increases the difficulty of studying the enzymatic reaction. Fortunately, it is usually possible to set up conditions in which ES complex variability is minimized. In most cases, the rate-limiting step is the formation of product. This means that, following initiation of the reaction, $E + S \rightleftharpoons ES$ equilibration occurs before significant substrate is converted to product. Thereafter, as long as the substrate concentration does not vary significantly during the period of interest, then the amount of ES complex will be essentially constant. The assumption that the amount of ES complex is invariant is known as the **steady state assumption**, and is expressed mathematically as:⁵

$$\frac{d[ES]}{dt} = k_1[E]_f[S] - k_{-1}[ES] - k_2[ES] = 0 \quad \text{Equation 3}$$

⁴Note that this scheme is applicable to reversible enzymes; it merely assumes that the initial concentration of product is zero, and therefore the reverse reaction does not occur. This situation is somewhat unusual in a physiological system, but is readily achieved in experiments performed using purified enzymes.

⁵Equation 3 assumes that the reverse reaction can be ignored (a good assumption if product concentration is initially zero). Note the “[E]_f”; this is the concentration of enzyme **not** bound to substrate; in most cases, [E]_f will be significantly less than the total enzyme concentration (usually abbreviated [E]_t). In contrast, steady state conditions assume that [S]_t [S]_f (i.e. [ES] is a negligible fraction of [S]_t because [S]_f >> [E]_t), and therefore it is not necessary to correct [S]_t for the amount of S present in the ES complex.

In order for steady state conditions to be achieved, the enzyme concentration must be a very small fraction of the substrate concentration. This is necessary in order to prevent conversion of substrate into product from resulting in significant changes in substrate concentration. If substrate concentration decreases significantly, ES complex concentration will also decrease, and Equation 3 will become a poor approximation of the true situation.

In studying an enzyme, it is usually best to measure product formation at multiple time points. In a well characterized system, it is possible to measure product formation at a single time point, under the assumption that product formation is a linear function of time. The major potential problem with performing a single point assay is that non-linear product formation may not be readily apparent.

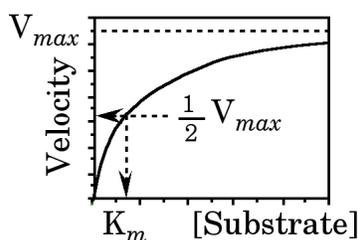
Using the steady state assumption, and some simple kinetics calculations, we can derive an equation for the rate of product formation as a function of substrate concentration:

$$v = \frac{k_2[E]_t[S]}{\frac{k_{-1} + k_2}{k_1} + [S]} \quad \text{Equation 4}$$

If the amount of enzyme used ($[E]_t$) is the same in each assay, then $k_2[E]_t$ will be a constant, usually termed V_{max} . Combining the rate constants in the denominator into a new constant, K_m , results in an equation you have seen before, the Michaelis-Menten equation:

$$v = \frac{V_{max}[S]}{K_m + [S]} \quad \text{Equation 5}$$

Contemplation of Equation 4 or Equation 5 reveals that velocity is a linear function of the amount of enzyme present (assuming that steady state conditions hold): for any substrate concentration, increasing the concentration of enzyme will increase the velocity in direct proportion. However, at any enzyme concentration, the velocity depends *non-linearly* on the amount of substrate present. The equation for velocity as a function of substrate concentration is that of a rectangular hyperbola.



The non-linear nature of the Michaelis-Menten equation has caused some difficulty, because determining the parameters for non-linear equations from real experimental data by standard analytical methods is impossible. The most common method used for determining the K_m and V_{max} for an enzyme has historically been the use of one of three linear transformations of the Michaelis-Menten equation:

$$\frac{1}{v} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}} \quad \text{Equation 6}$$

$$v = -K_m \frac{v}{[S]} + V_{max} \quad \text{Equation 7}$$

$$\frac{[S]}{v} = \frac{K_m}{V_{max}} + \frac{[S]}{V_{max}} \quad \text{Equation 8}$$

The linear transformations of the Michaelis-Menten equation allow the use of least-squares linear regression to determine the K_m and V_{max} for an enzyme from experimental velocity *versus* $[S]$ data. Linear forms of the Michaelis-Menten equation are useful, because many calculators and most graph-generating computer programs contain linear regression algorithms (such as the “trend-line function” in Excel). In addition, it is much easier to interpret changes for linear plots than for non-linear ones.

Unfortunately, the use of these linear transformations of the Michaelis-Menten equation implies that the enzyme assay data contain **no experimental errors** (a clearly inaccurate assumption), because including an error term makes the mathematical transformation non-linear as well. In addition, the linear equations frequently result in inaccurate values of the parameters K_m and V_{max} , because they tend to be heavily affected by the error in some data points. This is particularly true for Equation 6 (the double-reciprocal transformation often called the Lineweaver-Burk equation), in which **the data point that affects the parameters K_m and V_{max} most strongly is the one the most likely to be inaccurate** (the point corresponding to the lowest substrate concentration).

One partial solution to the problem of using linear transformations is to use more than one of them. If your data fit a rectangular hyperbola closely, all of the linear transformations will yield similar values for the kinetic parameters. On the other hand, if your data deviate significantly from hyperbolic behavior, linear regression fits to the different equations will result in markedly different values for K_m and V_{max} .

A much better method for determining the K_m and V_{max} values involves the use of non-linear regression. Until the advent of computers, non-linear regression was very difficult, because it is an iterative procedure. While most fairly sophisticated plotting programs include non-linear regression routines that allow data to be directly fit to the hyperbolic Michaelis-Menten equation, most students, and many scientists, are unfamiliar with the technique. In addition, non-scientific programs, such as Excel, lack an algorithm for performing non-linear regression.

Multisubstrate Enzymes

The enzyme kinetics concepts briefly outlined above apply primarily to relatively simple enzymes. Enzymes with more than one substrate, or more than one active site may not exhibit Michaelis-Menten kinetics. ADP-glucose pyrophosphorylase has two substrates, and the tetrameric form of the enzyme thought to exist in solution

should have a total of four active sites for each substrate. In addition, ADP-glucose pyrophosphorylase is subject to allosteric regulation.

Dealing with multiple substrates is *relatively* straightforward: if the concentration of one substrate is varied while the other one is held constant at a high value, the kinetics will exhibit pseudo-first order behavior. Thus, if pyrophosphate concentration is high, and ADP-glucose concentration is varied, the K_m and V_{max} values for ADP-glucose can be calculated from the observed velocity data. Note that the V_{max} value obtained will be somewhat lower than the true value: the true V_{max} value is an asymptote in the relevant equation, and its determination requires extrapolation to infinite concentrations of *both* substrates.

Enzymes with more than one active site may exhibit cooperativity, in which binding of one substrate alters either the binding affinity for other substrate molecules or the catalytic rate constant for the reaction. Experimentally, differentiating between deviations from hyperbolic behavior due to cooperativity or due to poor experimental technique is often a non-trivial task.

In addition to substrate-binding altering the binding of other substrates, some enzymes (including ADP-glucose pyrophosphorylase) are subject to allosteric regulation. An allosteric enzyme has a regulator binding site that is distinct from the catalytic site; the binding of allosteric effectors alters the kinetic parameters of the enzyme. The allosteric effector can alter the substrate K_m , the V_{max} , or both.

ADP-glucose pyrophosphorylase enzymes from different species are subject to different types and different degrees of allosteric regulation. Changing the nature and degree of allosteric regulation is a major goal of biotechnology efforts for ADP-glucose pyrophosphorylase.

Side note: Catalytic Rate Constants

The Michaelis-Menten equation (Equation 5) contains two parameters: K_m and V_{max} . While K_m is a constant for each enzyme, V_{max} is not really a constant, because it depends on the enzyme concentration. Enzyme concentrations can be chosen arbitrarily; different experiments may intentionally or inadvertently use different enzyme concentrations. Factoring out the enzyme concentration from the V_{max} leaves a rate constant that is, like the K_m , a constant physical property of the enzyme.⁶ Comparing Equations 4 and 5 suggests that this rate constant is k_2 ; this is correct only for enzyme described by the simple reaction scheme shown in Equation 1. A more general term for this is the catalytic rate constant, k_{cat} (also known as the *turnover number*); k_{cat} is the rate constant that characterizes the slowest step of the enzymatic process under investigation. The k_{cat} is the number of substrate molecules converted to product by a single enzyme molecule per unit time.

⁶Actually, both K_m and k_{cat} are only constant under a given set of conditions. For Michaelis-Menten enzymes, these values are independent of enzyme concentration, but are altered by changes in temperature, pH, and other factors. Small changes in these parameters probably accounts for some of the observed experimental errors.

For any enzyme, k_{cat} is a far more useful value than V_{max} , because k_{cat} is independent of enzyme concentration. Calculation of k_{cat} is straightforward:

$$k_{cat} = \frac{V_{max}}{[E]_t} \quad \text{Equation 9}$$

Obviously, the use of equation 9 means that you need to have determined the concentration of the enzyme in some way. Obviously also, the enzyme concentration in Equation 9 is the concentration in the reaction, not the concentration in your stock solution.

If you have measured the amount of protein present in a reasonably purified ADP-glucose pyrophosphorylase fraction, and if you assume that this protein concentration (in mg/ml) represents the ADP-glucose pyrophosphorylase concentration (in mg/ml), you can use this protein concentration and the molecular weight of ADP-glucose pyrophosphorylase to determine the enzyme concentration value required to estimate k_{cat} .

Side Note: Linear Regression Analysis

In dealing with experimentally obtained numerical data, it is frequently necessary to propose a mathematical model that describes the system. In order to be useful, this model should have some basis in reality: the parameters in the model should describe properties of the system being studied. The above discussion summarizes a mathematical model (the Michaelis-Menten equation) that is commonly used to describe simple enzyme-catalyzed reactions. The mathematical model that describes enzyme-catalyzed product formation is a rectangular hyperbola, which is a non-linear equation.

Because linear equations are much easier to solve, and somewhat easier to understand, the hyperbolic function is usually transformed into one of several possible linear equations.

A linear equation is an equation of the form:

$$y = mx + b \quad \text{Equation 10}$$

A linear equation states that, if x varies, y varies in direct proportion; m is the constant of proportionality (known as the slope of the line), and b is the value of y when $x = 0$ (known as the y -intercept).

When working with experimental data, all of the measured data points will not fit an equation of the line exactly. It is therefore necessary to estimate the slope and y -intercept that best fit the experimental data. For linear equations these parameters can be calculated analytically from the experimental data. The slope and y -intercept can be calculated using Equation 11 and Equation 12, respectively. Equation 11 and

12 are *least-squares linear regression equations*: the slope and y-intercept are chosen such that the values give the smallest possible average errors for each data point.⁷

$$m = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x} \bar{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} \quad \text{Equation 11}$$

$$b = \bar{y} - m\bar{x} \quad \text{Equation 12}^8$$

Equation 11 is moderately complex, and therefore it is usually desirable to use a computer program (or calculator) that deals with this equation automatically.

A useful measure of how well the data points fit the line is the correlation coefficient, usually abbreviated R. This is calculated by:

$$R = m \frac{\sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}}{\sqrt{\frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2}} \quad \text{Equation 13}$$

A second, related, way of assessing the fit is R^2 (the square of the correlation coefficient); R^2 is the fraction of the variation in y that can be accounted for by the variation in x. If $R^2 = 1$, the data points all fit the line perfectly; if R^2 is less than 1, some of the variation in y is due to other factors (either to errors, or to non-linear behavior of the system). The smaller the R^2 , the poorer the fit of the data to the line.

The major advantage of linearized forms of the Michaelis-Menten equation is that the linear regression can be solved exactly (non-linear regressions cannot be solved exactly).⁹ The linear forms of the Michaelis-Menten equation have some

⁷“Least-squares” refers to the common technique of squaring the error (*i.e.* the difference between the average value and the actual value) for each measurement. While the error can be positive or negative, the square of the error is always a positive number. The parameter that yields the smallest overall error (the “least” error) is the one chosen. The sum of the squared errors

= $\sum_{i=1}^n (x_i - x_t)^2$, where x_i is each data point, and x_t is the theoretical value for that data point based on the fit to the data.

⁸In Equation 12, \bar{x} is the mean of the x values, and \bar{y} is the mean of the y values.

⁹Least-squares non-linear regression algorithms do exist; all require some variant of guessing a value, calculating the sum of the squared errors, guessing a new value based on the previous value, calculating the sum of the errors, and repeating. The algorithm can only approximate the

disadvantages. One problem (a problem common to all linear regressions) is that the slope depends heavily on the points at the end of the lines, especially if those points are well separated from points near the center of the line, while points near the center of the line have little effect on the slope. The other major problem, mentioned above, is that the linear forms of the Michaelis-Menten equation markedly distort the errors in the data points.

Both linear and non-linear regressions involve attempting to fit a set of data to a defined equation. In some cases, the equation will be inappropriate for the actual data; this usually requires the investigator (*i.e.* **you**) to examine the plot, and decide whether the deviation from the curve is due to experimental errors, or to a systematic deviation from ideal behavior. As with any experimental data analytical technique, *computer-based curve fitting is **not** a substitute for a thoughtful appraisal of the data collected.*

Procedural Notes

Enzyme catalyzed reactions, like all chemical processes, have temperature-dependent rates. This means that, in order to obtain meaningful results, you need to perform all of the measurements at the same temperature (ADP-glucose pyrophosphorylase assays are typically performed at 37°C). This means that you need to control the temperature, and the way the temperature changes, during the assay.

Enzyme catalyzed reactions are time-dependent. This means that you need to carefully control the amount of time the reaction is allowed to run. This is especially important for single-point measurements, such as those that you will perform for ADP-glucose pyrophosphorylase.

As noted above, the rate of product formation in an enzymatic reaction is dependent on the concentration of the ES complex (see Equation 2). Under steady state conditions, the concentration of ES complex is essentially constant, and therefore velocity should also be constant. Variation in ES complex concentration can also be a consequence of varying enzyme concentration. Most experiments assume that enzyme concentration is a constant. Protein solutions are somewhat difficult to pipet accurately, you will need to exercise care in pipetting your enzyme solutions to prevent variations in enzyme concentration that will greatly complicate your analyses.

A related aspect of single-point measurements is that you must be sure that you are actually working with steady state conditions. Whether you have steady state conditions is usually determined by measuring the activity at more than one enzyme concentration. If the apparent velocity is a linear function of enzyme concentration, you can assume that you have steady state conditions. However, if the apparent velocity is not a linear function of enzyme concentration, it is likely that some (or all) of your samples have too much enzyme to allow steady state kinetics.

true value of the parameters by repeating the guessing procedure a number of times. This tedious procedure is best left to a computer.

Performing Linear Regression Analyses on Real Data

Using the linear forms of the Michaelis-Menten equation requires figuring out which of the terms should be plotted as x and y, and figuring out how to determine the K_m and V_{max} values from the regression analysis. Table 1 shows the relevant information for each of the three most common linear plots used for analyzing enzyme kinetic data. The names given reflect the authors of the early papers that used each plot. Depending on the source, you may find other names associated with some of these plots (especially the plot for Equation 6).

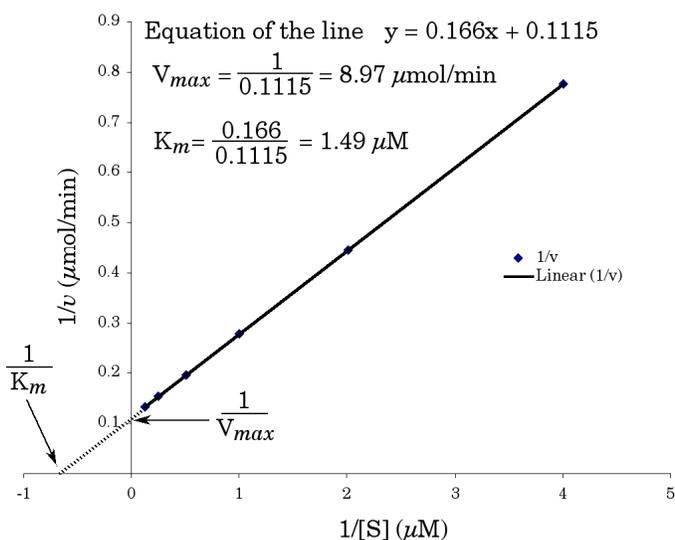
Setting up one of these plots requires obtaining a series of velocities that correspond to substrate concentrations. These data are then transformed to the value to be plotted according to the relevant equation, and plotted on a graph. The transformed data are then subjected to linear regression analysis (using a function such as the trend-line algorithm of Excel), and the slope and y-intercept values then used to determine the K_m and V_{max} values.

Table 1

Plot	x	y	Slope	y-intercept
Equation 6 (Lineweaver-Burk)	$\frac{1}{[S]}$	$\frac{1}{v}$	$\frac{K_m}{V_{max}}$	$\frac{1}{V_{max}}$
Equation 7 (Eadie-Hofstee)	$\frac{v}{[S]}$	v	$-K_m$	V_{max}
Equation 8 (Hanes-Woolf)	[S]	$\frac{[S]}{v}$	$\frac{1}{V_{max}}$	$\frac{K_m}{V_{max}}$

The example below shows this procedure for a set of data using the Lineweaver-Burk plot.

[S] (μM)	v ($\mu\text{mol}/\text{min}$)	1/[S]	1/v
0.25	1.29	4	0.7752
0.5	2.25	2	0.4444
1	3.6	1	0.2778
2	5.14	0.5	0.1946
4	6.55	0.25	0.1527
8	7.58	0.125	0.1319



The equation of the line ($y = 0.166x + 0.1115$) includes both the slope and y-intercept information. In the case of the Lineweaver-Burk plot, $V_{max} = 1/(\text{y-intercept})$, and $K_m = (\text{slope})/(\text{y-intercept})$. These relationships were derived from the information in Table 1.

Note that the values for K_m and V_{max} shown on the graph include units. These units are critically important for comparison purposes. You should **always** include the units for these parameters any time that you report results from a kinetic analysis of an enzyme.

The data given in the example fit a rectangular hyperbola fairly closely, and as a result can be readily analyzed using the Lineweaver-Burk plot. However, most real experiments do not yield data this well behaved. You therefore need to look at the data and at the resulting graphs, and not merely accept (and report) the values calculated by Excel. Once again, **there is no substitute for having an intelligent human look at the results** of a kinetic (or any other kind of) experiment.

References

All biochemistry textbooks have sections on enzyme kinetics. In addition, a number of books have been written on enzyme kinetics. The classic textbook is:

Segel, I.H. *Enzyme Kinetics* John Wiley & Sons, New York, 1975.

Other important literature includes:

Fischer, E. (1894) *Berichte* **27**, 2985.

Henri, V. (1902) *Acad. Sci., Paris* **135**, 916.

Michaelis, L. and Menten, M. (1913) *Biochem. Z.* **49**, 333.

Haldane, J.B.S. *Enzymes* Longmans, Green & Comp. 1930, reprinted in 1965 by MIT Press.

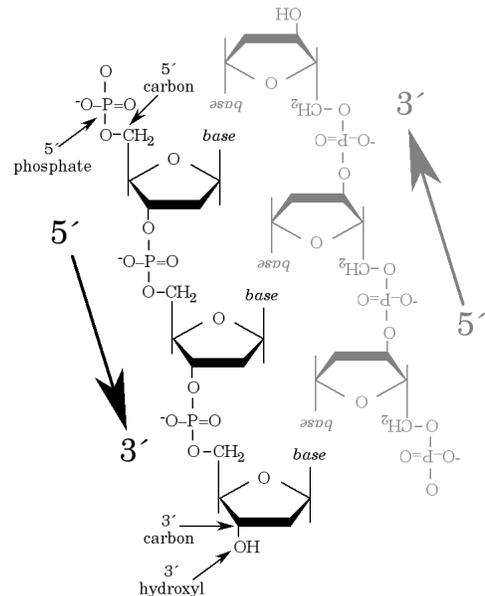
Lineweaver, H. and Burk, D. (1934) *J. Am. Chem. Soc.* **157**, 427.

Eisenthal, R. and Cornish-Bowden, A. (1974) *Biochem. J.* **139**, 715.

Definitions

3' ("three prime"): the 3-carbon of the second ring in a structure. In molecular biology, the ribose ring of the nucleotide is considered to be the second ring. The 3'-end of a DNA strand is the site of new synthesis by DNA polymerases.

5' ("five prime"): the 5-carbon of the second ring in a structure. In molecular biology, the ribose ring of the nucleotide is considered to be the second ring. Because DNA synthesis and protein translation occur 5' to 3', the 5' end is usually considered to be the starting position; a feature that is 5' relative to a particular site is considered to be upstream, while a feature that is 3' is considered to be downstream.



Affinity chromatography: a technique for separating a protein from a mixture on the basis of a property specific to that particular protein. For example, one substrate for lactate dehydrogenase is NADH; a column with NADH or a structurally related molecule may bind lactate dehydrogenase (and probably other dehydrogenases) with high affinity, while not binding the vast majority of other proteins.

Amino acid: strictly, any organic compound containing an acidic function and an amino group; in biochemistry, this term is often used to refer to any of the nineteen amino acid and one imino acid compounds typically used in biological protein synthesis.

Ammonium sulfate: a salt, $(\text{NH}_4)_2\text{SO}_4$, which has the property of reducing the solubility of proteins in the same solution, usually without causing structural alterations in the protein. Because different proteins exhibit differential solubility in ammonium sulfate, this salt can be used to separate proteins based on gross physical differences. Ammonium sulfate concentration is often given in percent. This "percent" is slightly unusual, in that it refers to "percent of saturation", rather than grams of solid per 100 ml of solution. The percent ammonium sulfate varies somewhat with temperature.

Anion exchange chromatography: a type of ion exchange chromatography in which the resin is derivatized using positively charged compounds such as DEAE or quaternary ethyl amino groups. The positively charged resin then allows the exchange of anions: it exchanges negatively charged proteins with counterions from the buffer.

Antibiotic: a compound that either inhibits growth of, or is toxic to, bacteria, even when used systemically. Thus, compounds such as penicillin, which can be used to treat bacterial infections, are considered to be antibiotics, while compounds such as ethanol (which kills bacteria when in direct contact but not if taken systemically) are not considered to be antibiotics.

Antibiotic resistance: the ability to grow in the presence of an antibiotic. In most cases this trait is the result of a gene coding for an enzyme that degrades the antibiotic (for example, the gene for β -lactamase confers resistance to β -lactam antibiotics such as penicillin and ampicillin, because the enzyme inactivates the antibiotic by cleaving part of its structure).

Bacteriophage: a virus that infects bacteria (also called simply “phage”). Bacteriophages, like all viruses, take over cellular machinery as part of their replication process. Engineered bacteriophages are useful for propagating DNA for a number of molecular biological processes.

Base-pair: *n.* a set of nucleotides, one from each strand of a double-stranded nucleic acid, that form a hydrogen-bonded complex with one another. *v.* to form a series of such sets of nucleotides. The average molecular weight of a base pair is about 650 Da (assuming sodium as the counter ion).

-ME (β -mercaptoethanol): a commonly used reducing agent. Mercaptoethanol and DTT are used to maintain the cysteine residues in the free sulfhydryl form.

Cation exchange chromatography: a type of ion exchange chromatography in which the resin is derivatized using negatively charged compounds such as carboxymethyl groups. The negatively charged resin then allows the exchange of anions: it exchanges positively charged proteins with counterions from the buffer.

cDNA: a DNA sequence complementary to another nucleic acid sequence. The term cDNA is usually used to refer to DNA generated by reverse transcribing an mRNA. As such, cDNAs represent actively transcribed genomic DNA but do not contain introns.

cDNA library: a mixture of cDNA fragments comprised of copies of most of the mRNAs expressed within the source tissue. In most cases, the cDNAs that comprise the library are inserted into a modified form of the *E. coli* bacteriophage λ , which allows both the propagation of the DNA and, for some cloning techniques, the isolation of individual cDNAs.

Chromophore: a chemical functional group within a molecule that absorbs electromagnetic radiation. While the term chromophore applies to groups that absorb radiation of any wavelength, it applies especially to groups that absorb within the visible portion of the spectrum, because these groups add color (“chromo” is derived from the Greek word for color) to a molecule.

Codon: a sequence of three bases that can be translated into an amino acid. In order to be considered a codon, the DNA (or RNA) sequence must be part of a coding sequence, and must be in the correct reading frame.

Cohesive end: the segment of single-stranded DNA extending 5' or 3' from a double stranded DNA fragment resulting from digestion by a restriction enzyme, which is capable of base-pairing to a compatible end of another DNA fragment (or the opposite end of the same fragment). Cohesive ends are typically referred to as “sticky ends” except in formal writing.

Column: a cylindrical apparatus containing chromatography resins used for chromatographic processes. Columns typically have an inlet, which allows loading of samples and addition of running buffer, and an outlet, which allows collection of the material that is not bound to the column.

Compatible ends: termini of linear DNA fragments that are capable of being ligated. Compatible ends can be blunt, or can be comprised of sticky ends, where the protruding single stranded DNA sequence of one end can base-pair to the other. (Note: it is possible for the ends of a single DNA fragment to be compatible, in which case, the fragment will tend to circularize if ligated.)

Competent cells: bacteria treated with a solution that greatly increases their likelihood of taking up DNA from their surroundings. Competent cells are significantly more fragile than normal bacteria, and are easily killed by violent treatment.

Complementary: in molecular biology, having a sequence that will base-pair to a sequence of interest. The sequence 5'-GGACTG is complementary to the sequence 5'-CAGTCC.

DEAE: diethyl-aminoethyl, a positively charged functional group frequently attached to resins used for anion exchange chromatography.

Deoxynucleotide: a compound containing a purine or pyrimidine base attached to ribose phosphate, in which the ribose is missing one of the hydroxyl groups normally present. Unless specified, the hydroxyl is missing from the 2'-position. Deoxynucleotides are the monomer units for DNA.

Dideoxynucleotide: a modified deoxynucleotide, in which both the 2'- and 3'-hydroxyl groups are missing. Dideoxynucleotides are used as chain terminators for DNA sequencing. DNA polymerases normally add the next nucleotide to the 3'-hydroxyl of the previous nucleotide; if the previous nucleotide lacks a 3'-hydroxyl, adding another nucleotide is impossible.

DNA (deoxyribonucleic acid): the genetic material of some viruses and all known non-viral organisms. DNA is a deoxyribonucleotide polymer comprised of four types of bases (adenine (A), cytosine (C), guanine (G), and thymine (T)). The specific base sequence, in the presence of cellular structures, determines the role of the DNA (*i.e.* coding regions, non-coding control regions, regions of other, less well defined functions, or regions with no known function).

DNase: any of a number of enzymes capable of hydrolyzing DNA into small fragments. Unlike restriction enzymes, DNase does not exhibit any sequence specificity, and will cleave essentially any DNA strand into smaller fragments. Humans secrete DNase; it is therefore necessary to avoid contact between human skin and any valuable DNA samples. DNase is rapidly inactivated by heating to 68°C.

DTT (dithiothreitol): a commonly used reducing agent. Mercaptoethanol and DTT are used to maintain the cysteine residues in the free sulfhydryl form, although DTT is somewhat more effective and somewhat more stable in aqueous solution.

EDTA (ethylenediamine tetraacetic acid): a chelating agent used in many buffers to sequester metal ions that may affect biochemical systems. EDTA inhibits calcium-dependent proteases by reducing the free calcium concentration.

Elution: the process of allowing protein to dissociate from a column resin. Elution usually involves altering the running buffer to decrease the strength of the interaction between the protein and the resin.

Exon: a DNA sequence that becomes part of the mature mRNA. Exons may include both coding and non-coding sequences.

Expression: in molecular biology, synthesis of RNA or (usually) protein from a DNA coding sequence.

Expression vector: a plasmid designed for expression of foreign proteins in a particular host cell (often *E. coli*). In addition to the normal features of a plasmid, expression vectors contain a strong promoter and ribosome-binding site upstream of unique restriction sites intended to allow the insertion of foreign DNA.

Extinction coefficient (ϵ): the Beer-Lambert law ($A = \epsilon cl$) proportionality constant that relates absorbance to concentration for a given molecule at a given wavelength in a cuvette of a given pathlength. The extinction coefficient is dependent on the probability that the molecule will absorb light at the applicable wavelength.

Frame: short for “Reading frame” (see below).

Gel filtration chromatography: a technique for separating molecules on the basis of size. Gel filtration resins contain small pores; small molecules enter the pores, while larger molecules cannot. Thus, large molecules experience a smaller total volume in the column, and elute first, followed by other molecules in order of decreasing size. Gel filtration chromatography can be used in purification techniques, or can be used analytically to measure the apparent molecular weight of molecules and folded proteins in solution. (Note: this technique is also called gel permeation or size exclusion chromatography.)

Genetic code: the algorithm that cells use to translate nucleic sequences into protein sequences. Exon sequences are essentially a substitution code; each group of three bases (*i.e.* each codon) defines an amino acid. Since organisms require only 20 amino acids and a stop signal, while the code includes 64 possible codons (4^3), the code contains some redundancy (for example, the genetic code contains 6 codons for the amino acid serine).

Although the control elements (such as promoters) may differ markedly between species, nearly all organisms translate nucleic acid sequences into proteins using the same code, and therefore, foreign DNA expressed in an organism nearly

always results in the same protein sequence as is found in the parent organism. However, all organisms do not use all codons with the same frequency. Some prokaryotes use tRNA availability as one method for regulating protein synthesis rates. Some foreign proteins are poorly expressed in *E. coli* due to large numbers of rare codons (*i.e.* having corresponding tRNAs that are produced in relatively small amounts). A rare codon frequency of 15% or less usually results in high expression, unless the rare codons are in close proximity to one another in the coding sequence.

Genetic Code

First Position	Second Position								Third Position
	T		C		A		G		
T	TTT	Phe	<i>TCT</i>	<i>Ser</i>	TAT	Tyr	TGT	Cys	T
	<i>TTC</i>	<i>Phe</i>	<i>TCC</i>	<i>Ser</i>	<i>TAC</i>	<i>Tyr</i>	TGC	Cys	C
	TTA	Leu	TCA	Ser	TAA	Stop	TGA	Stop	A
	TTG	Leu	TCG	Ser	TAG	Stop	TGG	Trp	G
C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T
	CTC	Leu	CCC	Pro	<i>CAC</i>	<i>His</i>	<i>CGC</i>	<i>Arg</i>	C
	CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	<i>CTG</i>	<i>Leu</i>	CCG	Pro	<i>CAG</i>	<i>Gln</i>	CGG	Arg	G
A	ATT	Ile	<i>ACT</i>	<i>Thr</i>	AAT	Asn	AGT	Ser	T
	<i>ATC</i>	<i>Ile</i>	<i>ACC</i>	<i>Thr</i>	AAC	Asn	AGC	Ser	C
	ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	<i>GTT</i>	<i>Val</i>	<i>GCT</i>	<i>Ala</i>	GAT	Asp	<i>GGT</i>	<i>Gly</i>	T
	GTC	Val	GCC	Ala	GAC	Asp	<i>GGC</i>	<i>Gly</i>	C
	<i>GTA</i>	<i>Val</i>	<i>GCA</i>	<i>Ala</i>	<i>GAA</i>	<i>Glu</i>	GGA	Gly	A
	GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

Italics indicates preferred tRNA in *E. coli*. **Bold** indicates minor tRNA in *E. coli*.

Adapted from: "Biased Codon Usage: An exploration of its role in optimization of translation." In: *Maximizing Gene Expression*, pp. 225-285, 1986.

Genotype: the genetic makeup of an organism. In most cases, the genotype of a given organism is assumed to be identical to that of the wild-type unless explicitly stated to be mutated. A large number of *E. coli* strains with known mutations are available; many of these mutations result in characteristics useful for various molecular biological applications.

Gradient: in column chromatography, a gradual change in the concentration of some component of the running buffer. Gradients can be smooth (typically linear, unless computer controlled pumps are used) or “step”. A smooth gradient involves a constantly changing concentration of the component(s) of the running buffer. A step gradient uses a constant concentration of the component, followed one or more times by the use of a new solution containing a different concentration the component.

HPLC (high performance liquid chromatography): a form of chromatographic technique that uses sophisticated high-pressure pumps to move the liquid phase through the column. HPLC pumps are capable of generating pressure of 50 megaPascals or more (over 7000 pounds per square inch). HPLC columns are designed to withstand high pressures, although most HPLC columns will be damaged by the maximum pressure output of the pump.

Hydrophobic interaction chromatography: a technique for separating molecules on the basis of their ability to interact with hydrophobic functional groups covalently attached to a resin.

Incubation: storage under defined conditions, especially at a controlled temperature.

Induction: in molecular biology, to cause initiation of transcription of a gene by some external intervention. For example, IPTG is used to induce expression of genes under control of several *lac*-derived promoters used in expression vectors.

Ion exchange chromatography: a technique for separating molecules on the basis of charge. Ion exchange resins contain charged groups; proteins containing amino acid residues of opposite charge will bind to the charged groups. Raising the ionic strength (usually by raising the salt concentration) of the running buffer causes proteins to elute from the column.

Ionic strength: a measure of the total amount of charged species present in a solution. Mathematically, ionic strength = $\frac{1}{2} \sum_i C_i Z_i^2$, where C_i is the concentration of the i^{th} species present, and Z_i is the charge on that species.

IPTG (isopropylthio- β -D-galactoside): a non-hydrolyzable carbohydrate derivative. IPTG binds to the *lac* repressor, causing its dissociation from *lac* promoter elements, and consequent activation of transcription from the promoter. Most promoter elements used to drive expression of engineered genes are derived from the *lac* promoter; in most strains of *E. coli*, transcription of the engineered gene only occurs in the presence of lactose or synthetic analogs such as IPTG.

Intron: a DNA sequence that is transcribed and then removed during mRNA maturation. Mammalian genes frequently contain hundreds of kilobases of DNA; after the introns are removed the residual mRNA is usually less than ten kilobases long. Because mammalian DNA contains introns, and because prokaryotes lack the ability to remove introns, mammalian genomic DNA cannot be used directly as a source of genetic material for expression of mammalian

proteins in bacteria. Instead, it is necessary to use cDNA (*i.e.* reverse transcribed mature mRNA) as the source of genetic material.

K-12: a wild-type strain of *E. coli*, from which many laboratory *E. coli* strains have been derived.

Kinase: an enzyme that phosphorylates its substrate, generally using ATP as the phosphate donor. Most kinases are specific for certain types of substrates (for example, Protein Kinase C phosphorylates specific proteins on specific serine and threonine residues). Polynucleotide kinase phosphorylates free 5'-hydroxyl groups of DNA, and is therefore frequently used to prepare DNA lacking a 5'-phosphate for ligation reactions.

Lag-phase: the period during which bacteria grow slowly after being taken from an environment in which nutrients are limiting to an environment in which nutrients are plentiful (for normal *E. coli*, this period is usually 1-2 hours).

LB (Luria-Bertani broth): a commonly used "rich" bacterial growth medium. Rich media contain all of the nutrients required for cell growth. (This is in contrast to minimal media, which contain only a few minerals, a carbon source, and a nitrogen source, and therefore require the cells to synthesize all of the required metabolites.)

LDH (lactate dehydrogenase): a ubiquitous nicotinamide coenzyme-dependent oxidoreductase that interconverts pyruvate and lactate. LDH is expressed in relatively large amounts in some tissues, and is an easily purified, stable protein.

Ligase: an enzyme that catalyzes the formation of a covalent bond between a 5'-phosphorylated and a 3'-free-hydroxyl end of a DNA strand. The function of a ligase is thus to connect two DNA fragments. The bacteriophage T4-derived ligase commonly used in molecular biology uses ATP as a co-substrate. Ligase is inhibited by a number of impurities that may be present in DNA samples.

Ligation: in molecular biology, a reaction in which two fragments of (usually) double-stranded DNA are connected together.

Load: in column chromatography, the process of allowing a protein-containing solution to enter a column, usually under conditions in which the protein of interest would be expected to bind to the resin. In gel electrophoresis, the process of placing a sample within a well on the gel prior to the application of a potential gradient to separate the components of the sample.

Log-phase: the process of rapidly growing in an environment rich in nutrients. Bacteria in log-phase divide every 20-40 minutes (in contrast, human cells divide roughly every 24 hours). During log-phase growth, bacteria express a group of genes somewhat different from those expressed during stationary phase.

MCS (multiple cloning site): a region of a plasmid containing a number of unique restriction sites that is intended as the insertion site for foreign DNA. In expression vectors, the MCS is located in close proximity to the promoter and

other signal sequences that drive transcription and translation of the inserted gene.

Mismatch: in molecular biology, a base, or series of bases, which do not form base pairs with the corresponding bases on the opposite strand. The presence of mismatches implies that one strand has undergone a mutation. Synthetic oligonucleotides frequently contain a few mismatches to create desired mutations (usually to create restriction sites, or to create modified protein sequences); note that the presence of too many mismatches will prevent the oligonucleotide from binding to its intended complementary strand.

mRNA (messenger RNA): a single-stranded RNA that contains a protein coding sequence and 5'- and 3'-untranslated regions which contain control elements. Most mRNAs in multicellular organisms contain a poly-A tail (a stretch of multiple adenosine residues at the 3'-end), which allows the isolation of mRNA; this is not true in bacteria, and therefore bacterial mRNA usually cannot be separated from other bacterial RNAs.

Nucleotide: a compound containing a purine or pyrimidine base attached to ribose phosphate. Nucleotides are the monomer units for RNA; nucleotides are also used for metabolic reactions and for intracellular signaling.

Oligonucleotide: a short, usually single-stranded sequence of (usually) DNA. Most oligonucleotides are synthesized using chemical methods; oligonucleotides are also the product of extensive digestion of nucleic acids with hydrolytic enzymes.

Open reading frame (ORF): a sequence of DNA that begins with ATG and ends with an in-frame stop codon, often used to refer to DNA sequences not certainly identified as being actively expressed (*i.e.* not known to be genes). In most cases, to be identified as an ORF the region must be long enough (>200 bases) to code for a peptide of reasonable size (few coding regions smaller than ~200 bases have been observed).

Ori (origin of replication): a DNA sequence in a plasmid required for replication of DNA in bacteria.

Palindrome: a sequence of characters that reads identically in both forward and reverse direction. In molecular biology, a palindromic sequence is one in which the one strand has the same sequence as the complementary strand. For example, AAGCTT is a palindrome (the complementary strand also reads 5'-AAGCTT). Most (although not all) restriction enzyme recognition sequences are palindromic (AAGCTT is a *HindIII* site).

Phenyl sepharose: a hydrophobic interaction chromatography resin, in which phenyl groups are covalently attached to sepharose (a cross-linked carbohydrate derivative).

Plasmid: a circular double-stranded non-chromosomal DNA molecule that bacteria will replicate. Most plasmids contain a gene for antibiotic resistance, an origin of replication, and one or more genes of interest to the researcher.

Plasmid prep: a procedure for purifying plasmid DNA from bacteria. In most cases, the cells are lysed with detergent (usually SDS) and high pH, followed by precipitation of chromosomal DNA; the procedure then uses one of a variety of methods of separating the plasmid DNA from residual contaminants (soluble proteins, carbohydrates, lipids, and other molecules).

PMSF (phenylmethylsulfonyl fluoride): a commonly used protease inhibitor. PMSF is an irreversible inhibitor of serine proteases. It has limited solubility and limited stability in aqueous solutions, and is toxic; alternative inhibitors of serine proteases have been developed that lack these drawbacks, but that are considerably more expensive.

Polymerase chain reaction (PCR): a technique for producing large amounts of a specific DNA fragment from a small amount of mixed DNA sequences. Briefly, a sample of DNA is heated to ~94°C (to separate the strands), the temperature is lowered to 37-60°C to allow binding of specific oligonucleotides to the ends of the sequence of interest, and a polymerase is used to replicate the sequence 3' to the oligonucleotides; this procedure is then repeated 20-50 times. In most cases, the polymerase used is heat-stable (the polymerase is usually derived from a bacterium that prefers living at 70°-100°C), and therefore the polymerase only needs to be added at the beginning of the process. The “chain reaction” occurs because each cycle results in an increased amount (roughly a doubling) of DNA that can act as a template for further DNA synthesis.

Primer: a short oligonucleotide sequence complementary to a sequence of interest. Most DNA polymerases cannot begin synthesizing nucleic acids without a template and at least a short region of double stranded nucleic acid to act as a starting place. Primers act as the necessary starting place for nucleic acid synthesis for a variety of molecular biological techniques, including PCR and DNA sequencing.

Promoter: a DNA sequence recognized by the transcription machinery (*i.e.* proteins involved in the synthesis of RNA from DNA). Promoters act as signals for initiation of RNA synthesis. Expression vectors typically contain strong promoters, such as the *trc* promoter, that are used to initiate mRNA synthesis using the inserted foreign gene as a template.

Reading frame: each codon has three bases; if the sequence is read beginning with one base, the translated protein will have a different sequence from a protein translated beginning with the following base. For example, the sequence ATGTGGTAA codes for Met-Trp-Stop if read from the first base, x-Cys-Gly-x (the “x” refer to the partial codons) if read beginning with the second base, and x-Val-Val-x) if read beginning with the third base. In this example, the TAA stop codon is in-frame with the methionine codon, but not with the cysteine or valine codons in the other reading frames.

Recombinant DNA: genetic material that has been engineered in some fashion. Most commonly, the term recombinant DNA refers to coding sequences taken from one organism and placed in another organism to allow expression of the foreign gene in the new environment.

Replication: the process of synthesizing a new DNA strand using the preexisting strand as a template. In normal cells, the result of replication is a doubling of the total amount of DNA, and only occurs immediately prior to cell division, with each daughter cell receiving one complete set of DNA molecules.

Repressor: a protein that prevents transcription from a promoter element. In most cases, repressors release from DNA in the presence of cellular stimuli. For example the *lac* repressor binds DNA in the absence, but not the presence of lactose; as a result, only when lactose is present in its environment does *E. coli* expend energy synthesizing the enzymes necessary to metabolize lactose. Repressor/promoter pairs are used in molecular biology to allow protein expression only under desired conditions.

Resin: an insoluble material, usually a modified carbohydrate polymer, used to form the matrix of a column. More generally, the term resin applies both to the insoluble polymer, and to the polymer that has been derivatized with functional groups that allow separation of proteins. Thus, DEAE-cellulose is an anion-exchange resin.

Restriction enzyme: an enzyme that cleaves specific sequences of double-stranded DNA. For example, *Nco* I cleaves CCATGG between the two “C”; because it cleaves both strands the same way; digestion with *Nco* I leaves a four base stretch of single-stranded DNA extending from the 5′ end. This four-base overhang is called a “sticky end”. (Restriction enzymes are one mechanism that bacteria use to degrade foreign DNA. Each wild-type strain contains a restriction enzyme and a methylase. The methylase tags the host cell DNA with methyl groups on A or C residues of specific sequences; this modification prevents degradation of the host cell’s own DNA by the endogenous restriction enzyme. Laboratory strains typically have the restriction enzyme system inactivated to prevent the degradation of introduced plasmid DNA (for example, in K-12-derived *E. coli* strains, *hsdRMS* mutants have both restriction enzyme *Eco*KI and the corresponding methylase genes deleted, while *hsdR17* ($r_K^- m_K^+$) mutants have the *Eco*KI gene inactivated, but retain the methylase). Note, however, that some foreign (e.g., commercially available) restriction enzymes will not cleave DNA when the host cell methylation patterns alter the bases in their recognition sequence.)

Reverse transcriptase: a specialized DNA polymerase, usually derived from a retrovirus, capable of using RNA as a template for DNA synthesis. The normal paradigm for information flow within a cell is from DNA to RNA to protein; reverse transcriptases were given their name to reflect the fact that these enzymes alter this standard direction of information flow.

RNA (ribonucleic acid): a polymer of nucleotides normally containing four types of bases (adenine (A), cytosine (C), guanine (G), and uracil (U)), although some forms of RNA include additional types of nucleotide residues. RNA molecules have varying functions, most of these functions being involved with protein synthesis. In some viruses, RNA acts as the sole genetic material.

RNase: any of a number of enzymes capable of hydrolyzing RNA into small fragments. Unlike DNase, most isozymes of RNase are very stable enzymes that are extraordinarily resistant to heat inactivation. Because humans (and most other species) secrete RNase, and because RNase is much more difficult than DNase to inactivate, working with RNA is somewhat more challenging than working with DNA.

Running buffer: the solution (for proteins, almost exclusively an aqueous solution) used for chromatography. Running buffers usually contain a pH-buffering species as well as salts and other molecules designed to either enhance or prevent binding of proteins to column resins. Alternatively, the term running buffer is sometimes used to describe the electrophoresis tank buffer used for running electrophoretic gels.
Screening: a method for locating separating the desired from undesirable cells based on some property that requires testing by the investigator.

Scintillation counter: an instrument for measuring radioactivity. In scintillation counting, radioactive decay excites organic molecules (scintillants); the molecules emit the energy in the form of light that is detected by the counter.

Scintillation fluid: a solution that aids in the detection and quantitation of radioactivity. The solution contains scintillants, which are molecules that emit absorbed energy (in this case, from radioactive decay) in the form of light.

Screening: a method for finding desirable cells in a mixture of cells by a process that requires testing by the investigator.

Selection: a method for separating desired from undesirable cells based on ability to survive and/or grow. One common selection technique is to grow cells in the presence of an antibiotic.

Start codon: a sequence that signals initiation of translation. Most genes use the sequence AUG (frequently referred to as ATG, because the AUG is derived from the ATG sequence found in the DNA); a few genes use GUG.

Stationary phase: in column chromatography, the solid resin support material that allows the molecules of interest to separate.

Stationary phase: in molecular biology, the period of little or no growth that occurs when the nutrients in an environment have been consumed, or when waste products have reached toxic levels. Stationary phase involves an adaptive response, in which the cells alter the genes being expressed to allow survival under limiting conditions.

Sticky end: the segment of single-stranded DNA extending 5' or 3' from a double stranded DNA fragment following digestion by a restriction enzyme that is capable of base-pairing to a compatible end of another DNA fragment (or the opposite end of the same fragment). Note: "sticky end" is a slang term; the technical term is "cohesive end"; however, few people use the term "cohesive end" except when writing formal papers.

Stop codon: a nucleotide sequence that signals the termination of translation. Three stop codons are commonly used: UAA, UGA, and UAG (frequently referred to as TAA, TGA, and TAG). Many laboratory strains of *E. coli* contain a suppressor tRNA for TAG stop codons; it is therefore preferable to avoid TAG as a stop codon in engineered DNA sequences intended for use in *E. coli*.

Suppressor tRNA: a tRNA that binds what is ordinarily a stop codon, but allows protein synthesis to continue by inserting an amino acid instead of terminating translation. In effect, the suppressor tRNA converts the stop codon into a codon for the amino acid. Many laboratory strains of *E. coli* contain TAG suppressor codons (especially supE, which inserts glutamate, or supF, which inserts phenylalanine).

Transcription: the process of synthesizing RNA from a DNA template. (Derived from the standard English term for converting information from one form to another: e.g., verbal English is transcribed into written English; nucleic acid information is converted to a different type of nucleic acid information.)

Transformation: the process of inducing cells to take up DNA from their environment. Transforming bacteria involves using a salt solution to make “competent” cells.

Translation: the process of synthesizing protein from an RNA template. (Derived from the standard English term for converting information from one language to another; information in the form of nucleic acid is converted to a different “language”: protein.)

Tris (tris-(hydroxymethyl) aminomethane): a buffer commonly used for biochemical experiments . Tris rarely interferes in biochemical reactions, and is inexpensive. However, Tris has a relatively high pK_a (8.1 at 25°C). In addition, the pK_a value for Tris changes by -0.031 pH units per °C, resulting in a large temperature-dependent pH variation in Tris buffers.

tRNA (transfer RNA): a small RNA molecule that mediates the incorporation of amino acid residues into a growing protein chain. Each tRNA is specific for one type of amino acid, and contains a sequence complementary to the corresponding codon.

Vector: an entity or mechanism for transmitting biological information. In molecular biology, this term is often used to refer to plasmid DNA, especially if the plasmid DNA contains a foreign gene and elements to drive transcription and translation of that gene.