ELEMENTARY BIOPHYSICS
SELECTED TOPICS

H. T. EPSTEIN

PRINCIPLES OF BIOLOGY SERIES

ADDISON-WESLEY PUBLISHING COMPANY

ELEMENTARY BIOPHYSICS
SELECTED TOPICS

HERMAN T. EPSTEIN, *Brandeis University*

ADDISON-WESLEY PUBLISHING COMPANY, INC.
READING, MASSACHUSETTS · PALO ALTO · LONDON

This book is in the
ADDISON-WESLEY
SERIES IN
THE LIFE SCIENCES

Library of Congress Catalog No. 63–16566

For Becky, Karen, Erika and David

Preface

This monograph has an explicit purpose: to facilitate the introduction of biophysics into the elementary biology course.

The reason for wanting to introduce biophysics is that the obvious trend in biology is toward physical and chemical biology. By the time the present generation of students has reached scientific maturity, twenty years will have passed, and our students will be scientifically obsolete unless they have been taught to think in physico-chemical terms from their earliest courses.

At present, practically all biophysics courses are given on the senior-graduate level. In teaching my own course on that level, I have been struck many times by the fact that a substantial fraction of the topics in biophysics utilizes only elementary mathematics, physics, and chemistry. There arose the possibility of designing an introductory biophysics course using only those elementary courses as prerequisites. However, it is unlikely at present that any department, including my own, would install such a course, because of the pressure of covering the broad scope of science that is included in present-day biology. Another means of achieving the goal is to give, say, a three- or four-week introduction to biophysics within the introductory course as just another of the facets of biology that are surveyed. The economy of paperback monographs has permitted the writing of a short introduction to biophysics without writing an entire textbook.

The majority of introductory biology courses appears to be given in the freshman year. Thus, even by placing the "month of biophysics" at the end of the course, the students will have had only the year of biology and probably a year of chemistry or physics, along with high school mathematics and science. Accordingly, one of the severe problems has been to choose topics suitable for a very elementary presentation. Therefore there have been important omissions, aside from the fact that this short coverage is intended for beginning students. For example, there is no discussion of membrane properties, nerve physiology, enzyme kinetics, information theory, etc. Some items are omitted for the reason that the complexity is too great, others because I judge them not sufficiently developed at the present time, and still others because a monograph is necessarily rather short. Even though I wanted to offer teachers a selection among possible topics, some have been omitted because they are likely to be covered in other introductory courses. What remains, then, is a selection that seems reasonable to me.

The chief writing problem has been to achieve simplification without oversimplification so that statements have a reasonable range of validity rather than no validity at all. There are tremendous simplifications in the topics presented. Some simplifications shock *me,* let alone others who would have simplified in an entirely different way. I offer no apology for what has been done, for my judgment is that I have chose a reasonably good way of getting ideas and attitudes across to beginning students. I would prefer this way to an appreciably more involved one required for more accurate presentations.

The quantitative, experimental point of view of the physicist and chemist has been, in my judgment, their most important contribution to biology. I have presented topics which seem to me to communicate this quantitative, experimental point of view. Biometry has been included because I think that an introduction is desirable early in the game and that, for most purposes, *no more than* an introduction is needed by most students. To say this bluntly, I think that statistical methods are highly overrated, and too often serve as excuses for not doing better thought-out experiments. To be fair, I must also state that professional statisticians know the uses and limitations of their subject; it is the statistically ill-informed scientist who tends to overrate and misuse statistics. Therefore I have presented the topic from the point of view of the problems faced in evaluating measurements, rather than as an exercise in application of mathematics to experimentally obtained numbers.

The chapter on physical forces and chemical bonds is placed next because I believe that the molecular biology of today is only a pale beginning of molecular studies in biology. Molecular differentiation, molecular neurophysiology, etc. are on the horizon.

The physics of vision, hearing, and muscles is going to develop increasingly, and in 20 years may well be the first real meeting place of biology and psychology.

Biophysical methods will continue to serve as basic equipment in the experimental baggage of biological scientists. An introductory survey has been written for those who would like just a quick look at the methods. For those who want more depth, I have chosen to illustrate with the ones which seem to me most likely to be generally applicable.

Finally, I would like to express my gratitude to the National Science Foundation for their award of a Senior Faculty Fellowship during the year 1959–60. This fellowship gave me the leisure to think about biophysics, and I developed the point of view which has now led to this monograph.

Waltham, Mass. H. T. E.
June 1963

Contents

CHAPTER 1 *The Mathematical Treatment of Data*

INTRODUCTION

With a given set of data, the problem is always to extract the maximum amount of information. This in no sense relieves the investigator of the problem of designing better and more extensive and inclusive experiments. But the existing data surely deserve to be analyzed as fully as warranted. The word *warranted* is the problem word, and it is to this problem that the mathematical analysis of data is directed. We would like to know the information that can be gotten from some data including the most likely values of the various quantities measured and, equally importantly, the probable uncertainty of the values thus obtained.

A measurement of any property of a system is, by itself, almost without significance, because we do not know the uncertainty in the measurement. A statement as extreme as this needs some justification. We feel intuitively that the purpose of a measurement is to know something accurate about a property. If we are greatly uncertain about the result, then we have learned little. For we then have to say that the measurement might be some particular number but that it also might be, say, 1000 times greater or less than that number. Thus, not even knowing the uncertainty in the measurement is equivalent to changing the number 1000 to any number you may wish to insert. The whole purpose of statistical analysis is to show us how to maximize the relevance of the measurements we make.

MEASUREMENTS AND THEIR VARIATION

As a simple first example, let us take a series of n individual measurements of something: $x_1, x_2, \ldots, x_n$. Why do we take more than one measurement? What have we gained by taking, as we usually do, the arithmetic average of these n measurements?

If we consider that the sources of uncertainties in measurements act randomly, then in a set of measurements we are as likely to get an individual result higher than the "true" value as to get one equally much lower than the "true" value. Indeed, unless measurements somehow

cluster around the "true" value, measurements will avail us nothing. Thus we are led to the assumption that the value around which the results cluster is near to the "true" value; indeed, if we think about the problem carefully, we come to the realization that we have no way of ever knowing the "true" value. So, we *define* the "true" value as the value around which measurements do, in fact, cluster. Of course, we now need to specify a means of extracting the true value (we henceforth omit the quotation marks, which, nevertheless, are still there so far as knowing the "truth" is concerned). If we could make an infinite number of measurements, the value which recurred most frequently would be taken as the true value. What do we do about the fact that we normally take only a few measurements and essentially never take a number even remotely resembling a very large number? How do we extract an estimate of the true value from a small, finite set of n measurements? The answer involves an analysis of kinds of errors.

In principle there are two kinds of errors in measurements. First, there are the so-called *systematic* errors—such as those resulting from the use of an inaccurate scale. This kind of error source is eliminated only through painstaking examination of the measurements themselves and analysis of the machines and procedures involved. We have no way of being sure that all such errors have been eliminated; we hope for the best.

The second kind is the *chance* error with which biometry is concerned. We assume that these chance errors are just as likely to yield measurements that are higher than the true value, as they are to yield measurements that are lower than the true value. On this assumption, the deviations, $d_i (= x_i - a)$, of the individual measurements, x_i, from the true value, a, should total zero. This condition can be stated algebraically and we now show how this enables us to determine the value of a from a given set of measurements.

$$\sum_i d_i = d_1 + d_2 + d_3 + \cdots + d_n = \sum_i (x_i - a)$$
$$= x_1 + x_2 + x_3 + \cdots + x_n - na.$$

Setting the sum of the deviations equal to zero gives

$$0 = x_1 + x_2 + x_3 + \cdots + x_n - na$$

or

$$a = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_i x_i}{n}.$$

This condition leads directly to the choice of the arithmetic mean as the desired average. With this choice, the average deviation is zero, since there must be as much positive deviation as negative deviation. To get

information about the deviations, we cannot use the average deviation, since a set of measurements with large deviations from the average will appear equivalent to a set with small deviations, in that both have an average deviation of zero. It is plausible that we are really interested in the absolute values of the deviations as a measure of the accuracy and reproducibility of our measurements. Now, it is not simple mathematically to work with absolute values. Therefore statisticians have used the least complicated way which is almost equivalent to using the absolute values: they use the squares of the values of the deviations of the individual measurements from the average. Since these are always positive, the sum of these squared deviations is also always positive, and the quality of the set of measurements can be inferred from the smallness of the sum of the squared deviations.

Now, however, we must ask what average must be taken to minimize the sum of the squared deviations; this value of a will be the one around which the individual measurements cluster according to our proposed criterion of least-squared deviations.

For those who cannot use the differential calculus, the result is given a few lines below. For those who can use the calculus, the derivation follows.

We wish to minimize the expression for the total variation:

$$\sum_i d_i^2 = \sum_i (x_i - a)^2 = (x_1 - a)^2 + (x_2 - a)^2 + \cdots + (x_n - a)^2.$$

To minimize this expression we differentiate with respect to a, keeping everything else constant, and set the result equal to zero:

$$2(x_1 - a)(-1) + 2(x_2 - a)(-1) + \cdots + 2(x_n - a)(-1) = 2\sum_i (x_i - a)(-1) = 0.$$

Dividing by 2 and collecting terms, we find

$$-(x_1 + x_2 + \cdots + x_n) + na = 0.$$

Thus, finally, solving for a, we obtain

$$a = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_i x_i}{n}.$$

We have thereby shown that the value of a which minimizes the sum of the squared deviations is again just the arithmetic mean.

We need to introduce a few new terms at this point. First, V, the *average variation* is simply the total variation divided by the number of independent measurements taken:

$$V = \frac{\sum_i di^2}{n} = \frac{\sum_i (x_i - a)^2}{n}.$$

This *average variation* is called the *variance,* and we shall use the term, even though it seems unnecessary to lose the directly given description inherent in the longer name.

Now, we could specify the results of our set of n measurements by saying they yield the estimate of the true value to be $a \pm V$. This seems somewhat silly, for it would mean, for instance, that we would give the length of something as 5.7 cm $\pm$ 2.3 cm^2, and it is silly to talk of squared centimeters as expressing the variation in the measurement of centimeters. It would appear more reasonable to take the square root of the variance, so that the average and its variation would then have the same dimensions. Thus, in the illustration above, we would say 5.7 cm $\pm$ 1.5 cm, and this statement has a sensible ring to it.

Adopting this change, we need to name the new quantity—the square root of the variance—and it is called the *standard deviation.* It is usually denoted by s, or by the equivalent Greek letter σ (sigma). So the results of a set of measurements are summarized as $a \pm s$, where

$$a = \frac{\sum_i x_i}{n},$$

and

$$s = \sqrt{\frac{\sum_i (x_i - a)^2}{n}}.$$

Now that we have this formidable looking apparatus, we have to ask what it means. Of what use is the averaging procedure? What significance attaches to this standard deviation? The procedures of statistics supply answers to these questions, and the answer is based upon finding that sets of n measurements do not exhibit many different kinds of clustering around the average a but that, surprisingly, only two kinds of clustering are actually obtained. To be concrete, if we plot n_i, the number of times we get the measurement x_i, we obtain only two different curves, as sketched in Fig. 1. In practice, we divide n_i by the total number of measurements to obtain the fraction of measurements which yield the result x_i. This fraction is called the frequency, f_i, of a particular measurement, and it is f_i which is indicated in the figure.

Note that these curves differ in that one is symmetrical, the other asymmetrical. Now, it has been possible to derive mathematical expressions for frequency curves under assumptions to be set forth later. It turns out that there are basically two types of curves, and the two types, when graphed, look very much like the two empirically obtained curves. There is a famous quotation to the effect that mathematicians believe their theoretical curves are relevant because they have been shown to exist in nature, while experimental scientists believe the experi-

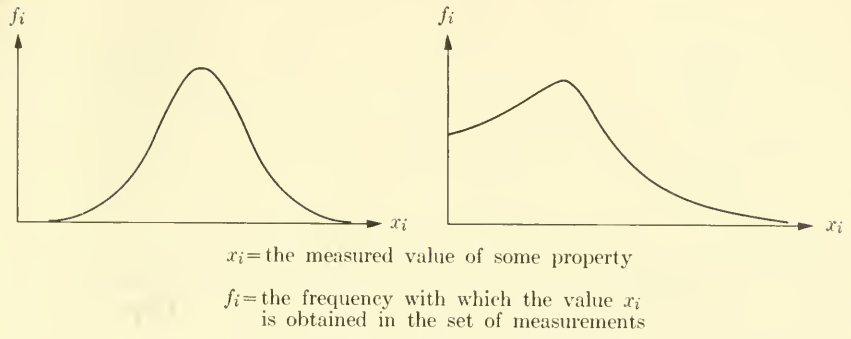


FIG. 1. A schematic representation of the two kinds of sets of measurements usually obtained experimentally.

mental curves are reasonable because their relevance has been proven by the mathematicians. In fact, of course, there is an inseparable mutual interaction between the two approaches.

To derive the symmetrical curve, the mathematicians assume that there exists a large number of small, randomly occurring, positive and negative contributions of error to the individual measurements. Thereby they derive the following expression for the frequency, f_i, of deviations from the average value:

$$f_i = \frac{1}{s\sqrt{2\pi}} e^{-(x_i-a)^2/2s^2}.$$

Here the pi and 2's result from the requirement that the individual fractional frequencies add up to unity. That means that the fractions of cases with all possible deviations from the average must include all cases, so that the sum is unity (or 100%, if expressed in percentage terms).

This distribution of deviation frequencies is called, variously, the Normal Error Curve, the Normal Distribution, or the Gaussian Distribution, after the great 19th-century scientist Gauss. The bell-shaped distribution is plotted in Fig. 2, which also indicates the frequencies for deviations equal to zero, $\pm s$, $\pm 2s$, and $\pm 3s$.

We can now begin to answer some of the many questions we have been accumulating. First, what fraction of the deviations lies within $\pm s$ of the average value a? This amounts to finding the fraction of cases in the shaded area of Fig. 3. (For those who understand the calculus, we can say that the fraction is found mathematically by integrating the Gaussian Distribution from $a - s$ to $a + s$.) The result is that about $\frac{2}{3}$ of the deviations fall in this range; conversely, of course, $\frac{1}{3}$ fall outside this range. This means that *if the average value of a measurement is a, then by chance alone one will have a measurement deviating from a by*

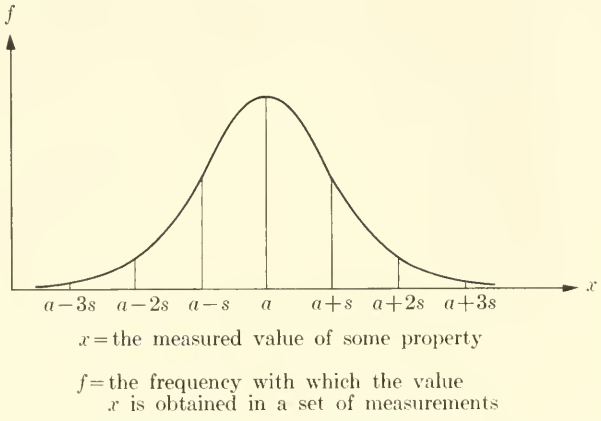


FIG. 2. Gaussian distribution of measurements, showing the average value a and the frequency of measurements equaling a or deviating from a by one, two, or three standard deviations.

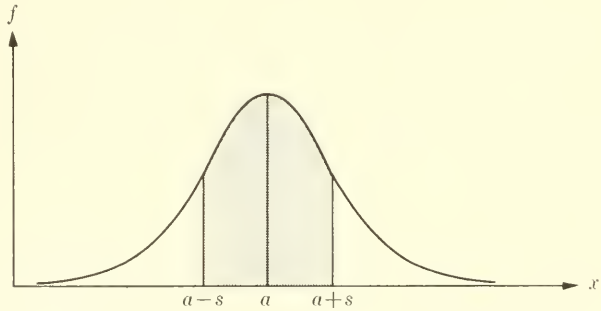


FIG. 3. The shaded area is proportional to the fraction of all measurements lying between $a - s$ and $a + s$.

at least $\pm s$ *in one of every three measurements.* By calculating in an entirely similar way, the following table is obtained:

A deviation of:	*Occurs in fraction of measurements:*
$\pm s$ or more	1/3
$\pm 2s$ or more	1/20
$\pm 2.3s$ or more	1/100
$\pm 3s$ or more	1/400

Note that a deviation in only *one* direction will occur in half the given fraction of measurements. That is, for example, a deviation of $3s$ or more will occur, by chance alone, once in 800 measurements. Similarly,

a deviation of $-2s$ or more (in the negative direction) will occur once in 40 measurements.

This table will provide the basis for much of the analysis we will present. With its presentation, we have answered one of the two questions we set out to answer about the standard deviation: what is its significance?

The other question concerns the utility of the averaging procedure employed in obtaining the arithmetic average. We already know that the clustering of measurements around this value is greatest, but still more can be said. If we ask the mathematicians to compute the clustering of the averages themselves, we are asking what would be the distribution of averages if we made a set of n measurements many, many times. Another way of asking the same question is to ask for the standard deviation of the average itself. The mathematician's reply to the latter question is

$$s_a = \frac{s_x}{\sqrt{n}}, \qquad \text{or} \qquad V_a = \frac{V_x}{n}.$$

That is, the average variation, or variance, of the average measurement is only one-nth of the variance of the individual measurements. This is the result we want. It says that the reason for taking the average is that the average is roughly $\sqrt{n}$ times more likely to be the true value than is any individual measurement. s_a is called the *standard error* by statisticians.

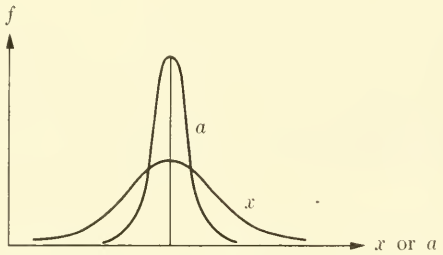


FIG. 4. A schematic representation of a set of individual measurements, x, and of a set of averages, a, obtained by making many repeated sets of measurements. The relative narrowness of the curve for a shows that the average obtained in any single set of measurements is very likely to be close to the "true" value.

Still another way of indicating the same information is shown in the schematic drawing in Fig. 4, which shows the distribution of individual measurements and also the distribution of averages. The latter is so narrow that it is very unlikely that we will ever get a much different result, *even by chance.*

One problem with our analysis thus far is that we have asked the mathematicians to tell us the distributions when we have many, many sets of measurements. Suppose we don't have so many measurements. After all, who feels like making more than, say, a half dozen, or maybe ten, measurements, let alone an essentially infinite number? The answer to our suppositional question is a surprisingly simple one. The mathematicians tell us that the difference comes in two places: in computing the standard deviation, and in computing the percentages given in our reference table above. The computation of s changes only by substituting $n - 1$ for n. This has the effect of increasing s, and is increasingly important the smaller n is. For instance, if we make only two measurements, the variance would double as a result of this correction. Next, if we ask what the table percentages become, we have to realize that the answer will depend on n—there will be a different answer for each value of n. Thus we would need not just the simple table above, but a set of tables. We will not present such a set, but tell you that you can find them in any standard statistics book. But, in the spirit of the present exposition of the subject, we can tell you that these tables aren't really so necessary. In practice, if n is as large as 10, one makes little error in using our own little reference table; if n is 20 or more, there is really no error in our table. So, just by modifying the expression for s to

$$s = \sqrt{\frac{(x_i - a)^2}{n - 1}}$$

we obtain most of the correction needed because of the small samples.

Up to this point, we have been dealing with the symmetrical curve for the distribution of deviations from the average. We have still to discuss the asymmetrical distribution before giving some applications of the analytical procedures which have been developed.

THE ASYMMETRICAL DISTRIBUTION

The Gaussian Distribution of deviations dealt with the case of a large number of small sources of error. The asymmetrical distribution deals with the case of having basically two alternative situations, one of which is very unlikely. Indeed, the one which is unlikely is so unlikely that it would be irrelevant if it were not for the fact that there were so many cases in question that the total possibility is not negligible. Take, for example, the case of distributing samples, each 1/100 ml, of a bacterial suspension containing 10,000,000 cells per milliliter. The probability of finding any individual bacterium in a particular sample is very small: 1/100 of 10,000,000, or 1/100,000. Yet, because of the large number of bacteria in the sample, the probability of finding *some* bacteria in the

sample is actually reasonably large. This situation has been analyzed and the distribution which governs it has been given the name of the French mathematician Poisson. The Poisson Distribution is

$$P_n(a) = \frac{e^{-a}a^n}{n!},$$

where e is the base of natural logarithms, and equals 2.71828. This represents the probability of finding precisely n of something if the average is a. If we were talking the language of deviations, we would be talking about the quantity $n - a$. Here, however, the least we can have is a equal to zero, so that we do not have very many negative deviations: if the average were 2, then we would have negative deviations of only -1 and -2. Thus this distribution is basically asymmetrical.

As an illustration of the use of this distribution, suppose we distribute 100 spores into 100 test tubes at random. The probability of finding any given spore in any given test tube has an average value of 1/100, but the probability of finding *some* spore is quite appreciable. Indeed, the average number of spores per tube is 100/100, or 1. The Poisson formula says that if the average is 1, then, for example, the probability of finding precisely 2 spores in a tube is

$$P_2(1) = \frac{e^{-1} \times 1^2}{2!} = 0.18,$$

as we can determine from tables or from direct computation using the numerical value of e.

One of the most important uses of the formula has to do with finding the zero class—the probability of finding none if the average is a. We know intuitively that if we tossed 100 spores randomly into 100 test tubes, there would be a substantial number of tubes having no spores. The formula gives the zero class as

$$P_0(a) = e^{-a}.$$

In the illustration above, e^{-1} equals 0.37. In 37% of the tubes there will be no spores at all.

The zero class formula is just as useful in the inverse way. If we measure the zero class, we can use the formula to obtain the average. If, for example, we found 2% of the test tubes without spores, we would solve the equation

$$0.02 = e^{-a}.$$

A look at some tables gives the result for a as a little less than 4 spores

per cell. The following table for the zero class may be of some use:

$a =$	1	2	3	4	5
$P_0(a) =$	0.37	0.14	0.05	0.018	0.0067

The Poisson distribution turns out to have another interesting and useful property. If the standard deviation of things satisfying this distribution is calculated, it is found that

$$s^2 = a.$$

This remarkable result permits obtaining the standard deviation from the average alone. Since the average can be obtained from the zero class, it becomes an exceedingly simple matter to obtain the standard deviation. In the illustration above, where there were 2% of the tubes unoccupied by spores, we not only know from the little table that the average occupancy is about 4, but also that the standard deviation is close to 2.

This property is frequently used in the design of experiments. Suppose we wish to determine the number of spores to an accuracy of, say 5%. We have only to solve the relation

$$\frac{s}{a} = \frac{\sqrt{a}}{a} = 0.05.$$

The result is a equals about 400. So we know that we must count until a total of four hundred spores have been tallied in order to know the number 400 to an uncertainty of 5%.

COMPUTATION OF ERRORS IN COMPOUND QUANTITIES

It frequently happens that the desired quantity cannot be measured directly, but can be determined from two or more other measurements. For example, the molecular weight of a particle can be determined by a combination of sedimentation and diffusion data according to the following formula:

$$M = \frac{s}{D} \frac{RT}{1 - (d_l/d_p)},$$

where M is the molecular weight of the particle of density d_p, suspended in a liquid of density d_l, at a temperature T, and R is the gas constant. Respectively, D and s are the coefficients of diffusion and sedimentation, and are obtained experimentally. For our purposes, the desired quantity, M, is given in terms of the complex ratio of 5 measured quantities. If each of these has its own standard deviation, what is the standard deviation of M?

We consider the problem in two simplified special cases. First, the case where some quantity M equals the product of two other measurements:

$$M = AB.$$

Then, using s_M,s_A, and s_B for the standard deviations of the three quantities, we have

$$M \pm s_M = (A \pm s_A)(B \pm s_B) = AB \pm Bs_A \pm As_B \pm s_As_B$$

or, neglecting the term s_As_B as being usually much smaller than the terms retained,

$$s_M = \pm Bs_A \pm As_B.$$

Dividing by M and its equivalent AB yields

$$\frac{s_M}{M} = \pm \frac{s_A}{A} \pm \frac{s_B}{B}.$$

This tells us that the *fractional* uncertainty in M is compounded of the *fractional* uncertainties in the two factors A and B. Because of the plus or minus signs, we don't know what to do at this point. Statisticians have been able to show that the most likely value in this case is given by the square root of the sum of the squares of the fractional uncertainties in A and B:

$$\frac{s_M}{M} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2}$$

It is a relatively simple matter to extend this to the product of more than two factors. And since a quotient A/C is equivalent to a product $A(1/C)$, an entirely similar formula holds for quotients. The general results can then be indicated as

$$M = \frac{A \cdot B}{C},$$

$$\frac{s_M}{M} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2 + \left(\frac{s_C}{C}\right)^2}.$$

The second part of the problem is the case where M equals the sum (or difference) of two quantities:

$$M = A + B,$$
$$M \pm s_M = A \pm s_A + B \pm s_B,$$
$$s_M = \pm s_A \pm s_B.$$

Thus, when the relationship is addition (or subtraction), the *actual* uncertainty, *not* the *fractional* uncertainty, is compounded of the actual

uncertainties of the factors making up M. As before, the most likely value is the root-mean-square one:

$$s_M = \sqrt{s_A^2 + s_B^2}$$

Now, if one has a more complicated relationship, such as

$$M = \frac{AB}{C + D},$$

the standard deviation may be written as

$$\frac{s_M}{M} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2 + \left(\frac{s_{C+D}}{C + D}\right)^2}.$$

As we have just shown,

$$s_{C+D}^2 = s_C^2 + s_D^2.$$

Thus, finally,

$$\frac{s_M}{M} = \sqrt{\left(\frac{s_A}{A}\right)^2 + \left(\frac{s_B}{B}\right)^2 + \frac{(s_C^2 + s_D^2)}{(C + D)^2}}.$$

APPLICATIONS OF STATISTICAL ANALYSIS

The general idea in these analyses is to compute the probability that a particular situation could have arisen by chance. If it could have occurred by chance once in a million times, everyone would agree that chance is not the explanation; there must be an underlying mechanism for the deviation. Suppose, however, that by chance alone the situation could have arisen once in 10 times—what this means is that a deviation *this great or greater* will occur that often. Do we then say that a 1/10 chance is too unlikely, or that it is so likely to happen that the deviation is not to be regarded as significant?

Statisticians have found through experience that an event that can occur once in 10 times by chance alone is not sufficiently unlikely to warrant undue interest. They have arrived at the following statements:

A probability of 0.05 is significant.
A probability of 0.01 is highly significant.

In the applications we shall present, these two levels of significance will be utilized repeatedly.

There are three main kinds of statistical computations that we shall set forth. These involve:

1. the ratio of a given deviation to the standard deviation;
2. the ratio of one variance to another variance;
3. the chi-squared test.

1. *The ratio of a given deviation to the standard deviation*

As an illustration, suppose we are studying the effect of adding various biochemicals to a nutrient medium and find that, after a given time, there is an average of 183 organisms per test tube, with a standard deviation of 9 organisms. One test tube (that has *one particular biochemical* additive) contains 212 organisms. Is this increase significant?

The deviation is 29 organisms, which is slightly more than $3s$. From our previous reference table we find that a deviation of this much or more will occur by chance alone about one time in 800. Therefore we say that the deviation is highly significant and it is therefore highly likely that the particular biochemical had a real, positive effect.

If there had been another tube with 198 organisms, the deviation would be 15 organisms, and the ratio of this deviation to the standard deviation is almost 2. Thus the probability of this happening by chance alone is almost 0.025, and is therefore considered almost significant. Thus we don't quite know whether to consider that the particular additive had a real effect. We have to do more or better experiments to decide.

The ratio of the observed deviation to the standard deviation should really be evaluated by using tables which depend on the number of independent measurements, as explained previously. This computation was done many years ago by a man publishing under the pseudonym *Student*. He called the ratio t, and you will therefore find this test in the textbooks under the name Student's t test.

The evaluation of n is not entirely trivial. When we choose a particular average a, the variability in the set of measurements is reduced by the choice. Indeed, given $n - 1$ measurements and a, one can reconstruct the nth measurement. Thus, knowing a, there are only $n - 1$ independent measurements. Similarly, every time we fix a parameter like a or s or the total number of things counted, we impose a condition which allows us to reconstruct the entire series of measurements with one fewer measurement.

In a way entirely similar to that for individual measurements, we can compare deviations of averages with standard deviations of averages. If two laboratories measure the concentration of viruses in a vaccine preparation, the deviation between their average values should be compared with the standard deviation of the averages (the standard error) to see whether there is agreement or systematic deviation between their results.

2. *The ratio of one variance to another variance*

It happens with surprising frequency that there is need to compare two variances. The analysis of the problem ends up with an estimate of whether the ratio of the two variances is sufficiently different from

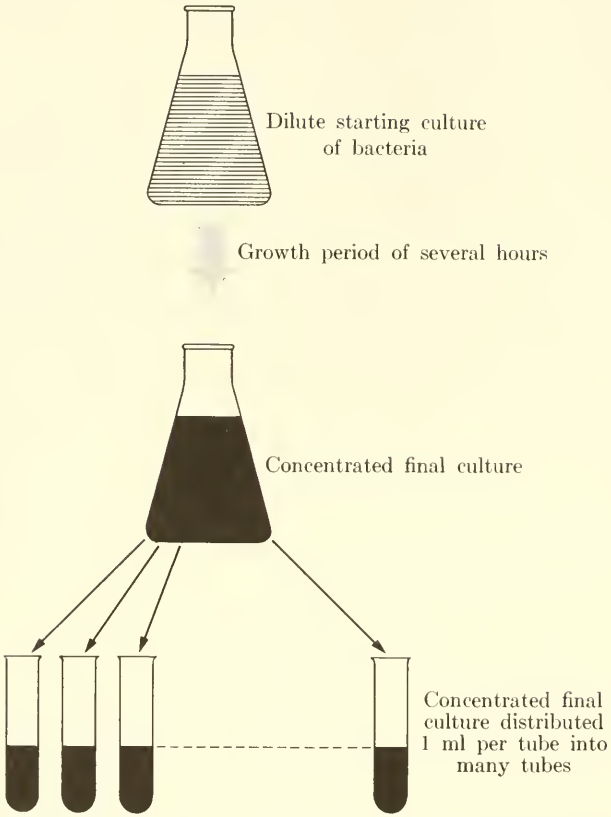


FIG. 5. Bacteria are grown for a given period of time from a dilute starting culture to a concentrated final culture. This final culture is distributed in 1 ml aliquots into many tubes. These tubes are then sampled to determine the number of mutant bacteria in each tube.

unity to warrant the deduction that there are different mechanisms involved in producing the fluctuations which result in the deviations. As with the deviation ratio test, the actual conclusion is that by chance alone a ratio this large or larger will occur in a certain fraction of cases with the given number of independent measurements in each of the variances. And, once again, the question of significance is decided on the basis of one chance in 20 being significant and one chance in 100 being highly significant.

As an example, we take an investigation which was historically a turning point in the understanding of the mechanism of mutations. The appearance of a mutant cell is detected by placing the cells in an environment in which the mutant property can be expressed. An illustration from bacterial genetics would be the appearance of resistance to a given anti-

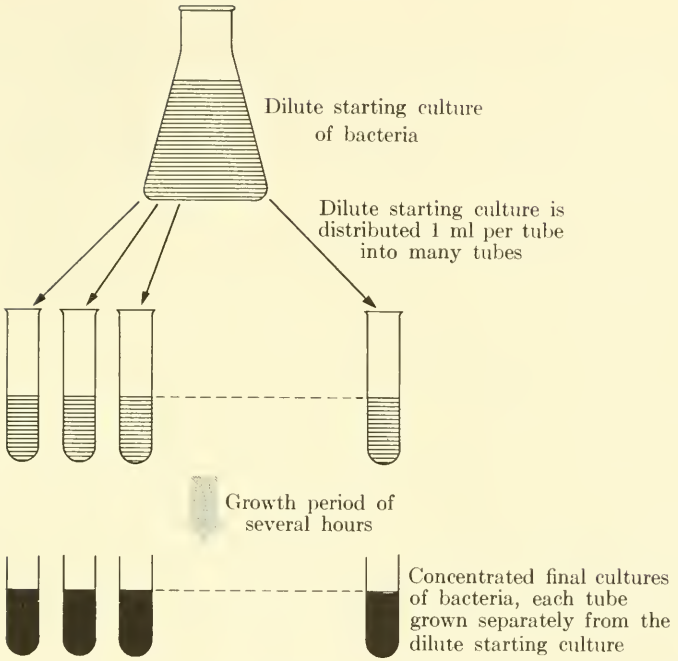


FIG. 6. This experiment differs from that of Fig. 5 in that the bacteria are distributed into separate tubes *before* the growth occurs. The final tubes are then sampled to determine the number of mutant bacteria in each tube.

biotic. One asks whether the *testing* with the drug *induced* the resistance in a small fraction of the cells or whether the property had occurred previously, by chance (*spontaneously*), in that small fraction of the cells. Thus the question was posed as to whether mutations are spontaneous or induced. The answer to the question involved an anlysis of variance.

Consider, as in Fig. 5, a flask containing 100 ml of medium, inoculated with some cells which then grow up to some final concentration. If the 100 ml are distributed in 100 tubes and analyzed for the number of mutants per tube, we will find an average value for the mutant concentration and a variance calculated from the deviations from the average.

Next, consider a parallel experiment in which the 100 ml are distributed into the 100 tubes as soon as the flask is inoculated. The bacteria then grow up to the same final concentration as before (see Fig. 6). These 100 tubes are now analyzed for the number of mutants per tube and, as before, an average and a variance are obtained.

If the mutation mechanism is by *induction* of drug resistance, then the results of the two experiments should be entirely similar, since the drug is acting on the same number of cells in the same number of tubes.

If the mutants arise spontaneously, we will expect the same average number of mutants, since the same total number of cells is being tested in each case. But in the second experiment the deviations should be greater, since in some tubes a mutant may arise early during growth and give rise to a large percentage of mutant cells; alternatively, in some tubes mutants may chance to arise very late, so that the percentage of mutant cells in these tubes will be very small. Therefore there is a clear difference in prediction by the two models. The induction model predicts the same variance in both experiments; the spontaneous arisal model predicts a variance in the second experiment which is much greater than the variance in the first experiment.

In several control experiments, in which the number of resistant bacteria was determined in different samples from the same culture flask, it was found that the mean number of mutants equaled the variance with reasonable fidelity. For example, where the mean was 3.3, the variance was 3.8; where the mean was 51.4, the variance was 27. Now, in the second part of the experiment, the number of resistant bacteria was determined in the set of similar samples from different culture tubes. Here the means were very different from the variances. For example, where the mean was 3.8, the variance was 40.8; where the mean was 48.2, the variance was 1171.

Using tables of the variance ratio (with the proper number of independent measurements for each variance), it was shown that the experimentally found ratio was far outside any chance fluctuation of the ratio away from unity. Thus the mechanism of spontaneous, noninduced mutation was first demonstrated.

3. *The chi-squared test*

Chi squared (χ^2) is defined as

$$\chi^2 = \frac{\sum_i (x_i - a)^2}{s^2_{\text{th}}}.$$

Now, for a large number, n, of independent measurements, this ratio should be approximately n, since the numerator is n times the experimental value of the variance and the denominator is the theoretical value of the variance obtained by using an essentially infinite number of measurements. So the tables of chi squared show the number of independent measurements as a parameter. Up to this point, chi squared is just another variance ratio test, with the denominator being a theoretical variance (approximated by taking an inordinately large number of measurements or estimated by theoretical methods beyond the scope of our treatment).

The chief use of chi squared is in determining the ratio for a whole set of different points with different theoretical values. For instance, if we wish to compare an experimental set of values with any theoretical curve, we use chi squared. In this instance, the variance in the denominator is replaced by the mean value predicted by the curve, since the Poisson Distribution afforded us the insight that the mean is equal to the variance. The numerator remains n times the experimental variance even though in application of chi squared there is usually only one measurement made at each point, so that, usually, $n = 1$. Next, we sum the expression for all the points of the curve at which experimental values were taken. This sum over the various points of the curve is indicated by a subscript p to distinguish it from the sum of deviations at individual points. This sum of chi squareds is then written as

$$\chi^2 = \sum_p \frac{(x_p - a_p)^2}{a_p}.$$

Since, as we have remarked, the means are good estimates of the theoretical values at each point, the expression is more usually written as

$$\chi^2 = \sum_p \frac{(x_{\text{exp}} - x_{\text{th}})^2}{x_{\text{th}}}.$$

This is the formula given in most textbooks.

Since the value of chi squared at each point should be unity, on the average, the sum should total simply to the number of independent measurements. Indeed, if we look at tables for chi squared, we find that the values are very closely equal to the number of independent measurements at a probability of occurrence equal to $\frac{1}{3}$. The following table will assist us in demonstrating the use of chi squared.

χ^2			
n	$p = 1/3$	$p = 0.05$	$p = 0.01$
1	1.09	3.8	6.6
2	2.2	6.0	9.2
3	3.3	7.8	11.3
4	4.3	9.5	13.3
5	5.4	11.7	15.1
6	6.5	12.6	16.8
10	10.6	18.3	23.2
20	20.5	31.4	37.6

As an illustration, we compare some experimental data with a straight-line formula to find whether it can represent the data adequately.

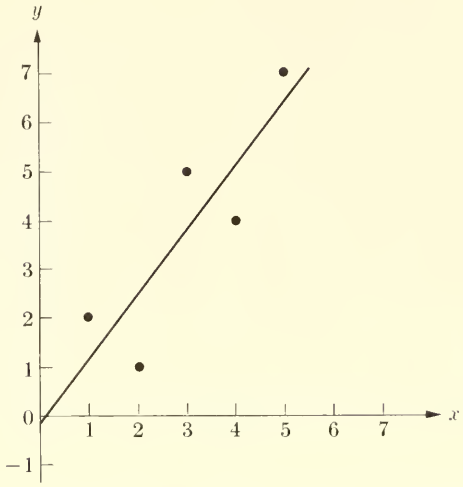


FIG. 7. A plot of some experimental data and a line which is to be tested by the chi-squared criterion for its goodness of fit to the data.

The plot of the data and the theoretical curve are shown in Fig. 7. The theoretical curve was derived by methods to be discussed in a later section. The equation is $y_{\text{th}} = -0.1 + 1.3x$. The values of y_{exp} and y_{th} at the experimental values of x are

$$\begin{aligned} x &= 1 \quad 2 \quad 3 \quad 4 \quad 5 \\ y_{\text{th}} &= 1.2 \quad 2.5 \quad 3.8 \quad 5.1 \quad 6.4 \\ y_{\text{exp}} &= 2 \quad 1 \quad 5 \quad 4 \quad 7 \end{aligned}$$

The deviations are, for the straight line,

$$(2 - 1.2),\ (1 - 2.5),\ (5 - 3.8),\ (4 - 5.1), \quad \text{and} \quad (7 - 6.4).$$

Their chi squared is then given by

$$\begin{aligned} \chi^2 &= \frac{0.64}{1.2} + \frac{2.25}{2.5} + \frac{1.44}{3.8} + \frac{1.21}{5.1} + \frac{0.36}{6.4} \\ &= 0.533 + 0.900 + 0.379 + 0.237 + 0.564 \\ &= 2.613. \end{aligned}$$

There are five measurements, and we have chosen the two parameters of the straight line (the intercept on the y-axis and the slope of the line) to make a best fit, so there are only three independent measurements.

Looking at the chi-squared table, we see that for three independent measurements we could get a chi squared of 3.3 or more in one of each

three sets of measurements we might make. Therefore the deviations from the straight line are not significantly different from what would be expected by chance alone. Thus the straight line is an adequate representation of the data.

In genetics, chi squared is usually used as in the following example. The problem is whether there is a correlation between two properties: hair color and eye color. It was found that, of 80 individuals, there were 36 with blue eyes and light hair, 11 with brown eyes and light hair, 9 with blue eyes and dark hair, and 24 with brown eyes and dark hair. Are the deviations from chance great enough to warrant the conclusion that hair color and eye color tend to vary together?

Note that there are 47 light-haired individuals (and therefore 33 with dark hair). There are 45 blue-eyed people (and therefore 35 with brown eyes). If the light-haired people had the same proportion of blue and brown eyes as the entire population studied, then 45/80 of them would have blue eyes. This is $45/80 \times 47$ or 26.43 people.

It is not necessary to compute any of the other classes of individuals because there must be a total of 47 light-haired people, so that theoretically there would be $47 - 26.43 = 20.57$ light-haired people with brown eyes if the eye color distributes as for the entire group of people. Further, since there are 45 blue-eyed people in all, and 26.43 have light hair, the remaining $45 - 26.43 = 18.57$ people should have dark hair. Finally, the number of dark-haired, brown-eyed people should be 14.43.

Chi squared is then computed as the sum of the four terms:

$$\chi^2 = \frac{(36 - 26.43)^2}{26.43} + \frac{(11 - 20.57)^2}{20.57} + \frac{(9 - 18.57)^2}{18.57} + \frac{(24 - 14.43)^2}{14.43}$$

$$= 19.19.$$

The number of independent measurements is unity. We saw this directly, since as soon as we had computed one theoretical number (26.43 blue-eyed, light-haired people), the other three numbers were fixed. We can also fairly readily see this by counting the number of quantities we have fixed, since we fixed the total number of people to be 80, the total number of blue-eyed people to be 45 and the total number of light-haired people to be 47. Thus we have only one of the four measurements which can be arbitrarily made.

Looking at our table, we see that for $n = 1$, the probability of getting a value of 6.6 for chi squared is already 0.01. To get a value of 19.19 would correspond to a probability much smaller than this. Therefore the deviations from theoretical values can by no means be accounted for by chance fluctuations. Thus we conclude that the hypothesis must be that eye color and hair color do not vary randomly, but are associated.

TREATMENT OF EXPERIMENTAL DATA

1. *Data-fitting using the least-squares principle*

The least-squares principle was used to obtain an index of the variation in data. It has other uses, one of which is the subject of this section. The problem of finding a theoretical expression to fit data utilizes this principle to choose the parameters of the curve so as to minimize the variation around the curve. We illustrate the approach by finding the expressions for the parameters of a straight line which is to be fitted to the data.

The equation of a straight line is

$$y = a + bx.$$

Consider the difference between the theoretically and the experimentally determined y:

$$y_{\text{th}} - y_{\text{exp}} = a + bx - y_{\text{exp}}.$$

These differences are sketched in Fig. 8.

Our principle of least squares tells us to square this difference, add all the squared differences for all the points, and then to choose a and b so as to minimize the sum. For those with mathematical facility, the derivation of a and b will be presented. For other readers, the results are given below.

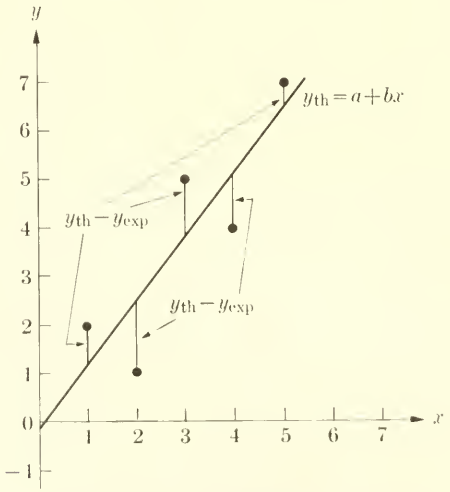


FIG. 8. A plot of some experimental data relating y and x and a theoretical line whose slope and intercept are to be chosen to minimize the total deviation of the theoretical values of y from the experimental values of y. The short vertical lines are the deviations themselves: $y_{\text{th}} - y_{\text{exp}}$.

The mathematical problem is to minimize

$$\sum(y - y_{\text{exp}})^2 = \sum(a + bx - y_{\text{exp}})^2.$$

We take the derivative with respect to a and b in turn and set the derivative equal to zero, obtaining

$$0 = \sum 2(a + bx - y_{\text{exp}}),$$

$$0 = \sum 2(a + bx - y_{\text{exp}})\,(x).$$

Factoring out the 2's, and multiplying as indicated in the second equation, we obtain

$$\sum a + b\sum x - \sum y_{\text{exp}} = 0 = na + b(\sum x) - \sum y_{\text{exp}},$$

$$\sum ax + b\sum x^2 - \sum xy_{\text{exp}} = 0 = a(\sum x) + b(\sum x^2) - \sum xy_{\text{exp}}.$$

We can solve these two equations for a and b to give us the values of a and b which minimize the deviations from the theoretical line:

$$a = \frac{(\sum x^2)(\sum y_{\text{exp}}) - (\sum xy_{\text{exp}})(\sum x)}{n(\sum x^2) - (\sum x)^2},$$

$$b = \frac{n(\sum xy_{\text{exp}}) - (\sum x)(\sum y_{\text{exp}})}{n(\sum x^2) - (\sum x)^2}.$$

To show the use of these expressions, and also to demonstrate that the expressions are much more fearful-looking than actually harmful, we illustrate by obtaining the straight line used in the previous section on chi squared. Because we will need the sums, we tabulate the data again, along with the operations needed for this application.

$x =$	1	2	3	4	5;	$\sum x = 1 + 2 + 3 + 4 + 5 = 15$
$y_{\text{exp}} =$	2	1	5	4	7;	$\sum y_{\text{exp}} = 19$
$x^2 =$	1	4	9	16	25;	$\sum x^2 = 55$
$xy_{\text{exp}} =$	2	2	15	16	35;	$\sum xy_{\text{exp}} = 70$

We substitute the various numbers in the expressions for a and b, obtaining

$$a = \frac{(55)(19) - (70)(15)}{(5)(55) - (15)^2} = \frac{-5}{50} = -0.1,$$

$$b = \frac{(5)(70) - (15)(19)}{5(55) - (15)^2} = \frac{65}{50} = 1.3.$$

The least-squares straight line for these data then is

$$y = -0.1 + 1.3x.$$

If we had wanted to fit a more complicated curve, the procedure would have been entirely similar. For example, suppose we had decided to try something quite complicated, such as

$$y = a \sin bx + c \log x.$$

Aside from the purely mechanical problem of carrying out the calculus operations, the procedure is entirely similar. We find expressions for a, b, and c so as to minimize the sum of the squares of the deviations from this expression, thereby obtaining three equations for the three parameters. We could, in fact, go as high as five parameters with these data, and then we would be able to fit the data exactly *at the experimental points,* although the curve might fluctuate wildly between those points.

There are other methods of obtaining curves to fit data, but the curves obtained by the least-squares principle are the generally accepted ones. Further, there are many subtleties in the least-squares approach which the student can find in the standard texts.

2. Correlation

We can say a few things about the concept of correlation. Here we ask the question of the extent to which variations in one variable are accompanied by a parallel variation in another variable. It must be strongly emphasized that there is *not the slightest implication* of a direct connection between the two variables. The pedagogically favored illustration of this point is the finding of an almost perfect proportionality between teachers' salaries and liquor consumption. The pedagogue then declares that it is manifestly ridiculous to deduce the validity of the causative relation—the teachers are drinking all that liquor. If we study the details of the situation, we discover that teachers' salaries rise during prosperous times and during such times the entire population has extra money, some of which is spent on liquor.

In the way in which we will introduce correlation, we look at the extent to which points cluster about the presumed line expressing the possible correlation, and compare it with the clustering around the mean value of all the points. Consider Fig. 9.

If we call $\overline{y}$ the average value of y, then the variation around the average is $\sum(y - \overline{y})^2$.

The variation around a proposed line (to be found theoretically) is

$$\sum(y - y_{th})^2.$$

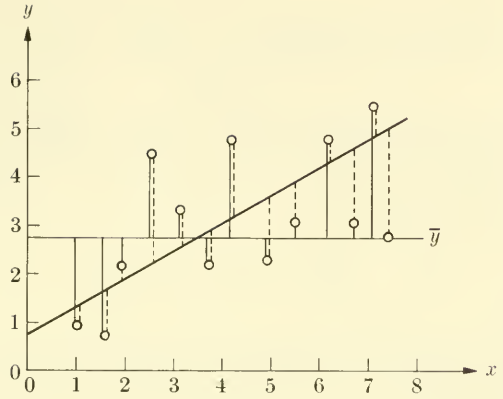


FIG. 9. A plot of some experimental data (○); their deviations (----) from the mean, $\overline{y}$; and their deviations (----) from a line proposed to fit the data. The extent to which x and y are said to be correlated is given by the extent to which the line is a better representation of the data than is the value $\overline{y}$. The criterion for better representation is the ratio of the sum of the squares of the deviation from the line to the sum of the squares of the deviations from the mean, $\overline{y}$. The quantitative use of the criterion is given in the text.

Consider the ratio of these two variations:

$$\frac{\sum(y - y_{\text{th}})^2}{\sum(y - \overline{y})^2}.$$

If there were a perfect correlation between y and x—that is to say, if the line fit the data perfectly—the numerator would be zero. If the line is less than a perfect fit, the numerator becomes increasingly large until, finally, if there is absolutely no connection between y and x, there will be as much variation around the line as around the mean value of y. Then the ratio will be unity. Thus the values of this ratio vary from zero for a perfect correlation to unity for completely unrelated things. Since we wish to speak of the extent to which things are "co-related," the statisticians have chosen to use one minus this ratio as the index of correlation, R. Then perfect relation is unity and complete lack of relation is zero.

$$R = 1 - \frac{\sum(y - y_{\text{th}})^2}{\sum(y - \overline{y})^2}.$$

There is one more complication to be taken care of. The way we have chosen the relationship index, it is always positive. This presents a problem, because it is possible for things to be negatively connected, in that an increase in one variable correlates with a decrease in the other.

This problem has been solved by using the square root of the relationship index with the proper sign. That is,

$$r = \pm\sqrt{R},$$

where r is what is technically defined as the correlation coefficient. It can be shown algebraically that this definition of r is entirely equivalent to the one used in standard texts:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \cdot \sum(y - \bar{y})^2}}.$$

Here the numerator is to be understood as showing the extent to which x and y vary together. For a large number of pairs of values of x and y, if they were unrelated, then for each particular plus value of $(y - \bar{y})$, there would be both a positive and a negative value of $(x - \bar{x})$, so that in the sum these terms would cancel out. If there were actually a correlation, then when the x difference was positive the y difference would also tend to be positive and there would be no cancellation. The purpose of the denominator is to make the coefficient dimensionless and to give it the values zero and unity for, respectively, unrelated and perfectly related variables. To prove the equality of the two definitions, use must be made of the least-squares line relating y and x.

By using our definition based on clustering, those of us with mathematical facility may use any type of theoretical expression for y and can see how to compute the correlation even for curved lines. The rest of us have to take the expressions for r as given.

A point which is very much understressed in many standard treatments is that the correlation index—the square of the correlation coefficient—is the direct measure of the extent to which variables are correlated. Thus, a correlation coefficient of 0.6 looks respectably large, but we note that its square is 0.36. This tells us that only about $\frac{1}{3}$ of the variation in y is correlated with variations in x, the remainder being associated with factors which do not affect x. From this point of view we may say that one-third of the factors determining x and y are shared, the others being different for the two variables.

CHAPTER 2 *Physical Forces and Chemical Bonds*

INTRODUCTION

As the current pace of advances in molecular biology and biochemistry accelerates, a detailed understanding of the structure, interactions, and functions of biologically important molecules becomes increasingly necessary. This means that the forces involved in chemical reactions must be increasingly taken into account, and this requires detailed information about the nature of the various chemical bonds. Consequently, in this chapter we present the basic physical forces and try to give an idea of how they operate to create the various chemical bonds. Fundamentally, the study of such matters involves comprehension of the complex mathematical apparatus of modern theoretical physics and theoretical chemistry, and it is impossible to get around this requirement. Nevertheless, almost all scientists have geometrical and mechanical images and models in their minds which guide them in their use of such a mathematical apparatus to the point that these images and models serve as the bases of their "intuition" when considering new problems. These images and models are admittedly limited in scope and accuracy, and each scientist may have his own peculiar set. This writer believes that it is useful for beginning students to acquire some intuitive feelings about such matters even before they acquire the detailed apparatus for understanding the matters. Indeed, in some respects it is only to the extent that today's complexity becomes incorporated into the intellectual equipment of the youth that youth has the capacity to transcend the elders. In this spirit, then, is the following discussion of forces and bonds presented.

PHYSICAL FORCES

The difference between Newtonian and quantum physics is of a nature which far transcends their procedural differences. But, for our purposes, we find it useful at this point to look at an aspect of the procedural differences: the emphasis of Newtonian physics on the concept of force; and the emphasis of quantum physics on the concept of energy. Indeed, it is in great part this shift which permitted physics to provide the basis for

its deep contributions to chemistry, which has always worked on the level of energy.

Newton started from the Galilean principle of inertia (known also as Newton's first law of motion) which asserts that bodies move with constant velocity unless a force acts on them. Thus, the observation of a nonconstant velocity (an acceleration) is taken as evidence for the existence of a force on the body. Then Newton's second law provides a recipe for quantitative work by asserting that the acceleration produced by the force is directly proportional to the force. It is also inversely proportional to the mass of the body. These two assertions are combined in the usual statement of the law:

$$a = \frac{f}{m}.$$

When an acceleration is then observed, it is studied in as many different situations as needed to give us an idea of its general behavior: its dependence on time, distance between bodies, dependence on masses, velocities, etc. By such a study, Newton was able to deduce the expression for the first force ever studied as such: the gravitational force. He deduced its formula as

$$f_{\text{grav}} = \frac{m_1 m_2}{Gr^2},$$

where m_1 and m_2 are the masses of the two bodies exerting gravitational forces on each other, r is their separation, and G is the gravitational coefficient which puts f in the proper units.

The procedure for using this result in a case where a gravitational force is known to be acting is to insert the expression for f, with the proper values of the masses, into the first equation, solving for the acceleration, a. From a, by standard mathematical procedures, the position and velocity of the body in question can be determined at all times. For physics, the problem is solved.

Are there any other forces in nature? How could we recognize them? In a situation in which we can show the gravitational force not to be effective, the observation of an acceleration is evidence for a new force.

Peasants from the province of Magnesia (Greece) long ago observed that some stones could attract or repel other stones, and two aspects of the situation permit us to deduce the existence of a new force. First, there is both an attraction and a repulsion, whereas with gravitation there is only attraction. Second, these stones exert their effect along a table top, at right angles to the predominant gravitational attraction of the earth. Thus the existence of what we call *magnetic force* was deduced.

Next, when a piece of fur is stroked with a piece of amber, both the amber and the fur acquire the capacity to attract pieces of lint, paper,

and each other. Stones from Magnesia, however, have no effect. The mutual gravitational forces of the amber and fur are inappreciable. Therefore, there is a third force in Nature, called the *electrical force.*

Except for the nuclear force (which does not concern us here), these are the only forces now known in Nature, and the explanations of natural phenomena must necessarily involve these three forces in various combinations. With them, science has explained the consequences of a ball's hitting a bat, light being absorbed by leaves, the trajectories of stars, and the properties of molecules interacting with each other.

For completeness, it is worth mentioning that Einstein showed that the electric and magnetic forces are really only two special aspects of a single, unified, electromagnetic force.

There are other situations in which we speak of forces acting. For example, when particles move through a solution, or an airplane moves through the air, we speak of the drag or viscous forces acting on the particle or airplane. In fact, on the molecular level there is no such thing as a viscous force. There is something properly called viscous energy, perhaps, but even this terminology would be unacceptable to some scientists.

When a particle tries to move through a solution, it meets a resistance deriving from the fact that in the liquid state the solvent molecules attract each other strongly (generally by the van der Waals "force" to be explained later). Thus it takes work for the particle to move between the solvent molecules, since it must separate these molecules. Therefore, when the particle has pushed the solvent molecules apart, and then moves farther on, the molecules come back together with sufficient energy (imparted by the particle and thus lost by that particle) that they bounce off each other. This oscillational energy is communicated by collision to the other molecules of the solvent, with the final result that the average oscillation (thermal vibrational) energy is increased somewhat. In this fashion, the energy required to move a particle through the solvent produces heat, which our usual manner of speaking ascribes to the work done against frictional forces.

Similar explanations are available for other apparent forces.

CHEMICAL BONDS

The accent on forces lasted more than 200 years, until quantum physics shifted the emphasis to energies. In the meantime, chemists had become increasingly occupied with energies, as thermodynamics developed and was able to make assertions about energies involved in reactions while saying nothing directly about forces. In this spirit, the chemists talked about the energies of the bonds holding atoms and molecules to-

gether, rather than about the forces. This has turned out to be more than just a verbal difference, because all the bonds have turned out to be electrical (or, rather, electromagnetic) in origin, so that the study of chemistry using the force concept would have led to impossibly difficult models.

The picture that we have of atoms is that of miniature solar systems with a central positively charged nucleus and a number of planetary electrons whose total charge is normally equal to the net positive charge of the nucleus. If we consider a crystal of salt, sodium chloride, and ask what holds it together, we have to know something of the conditions within the crystal. If we toss some metallic sodium into water, the chemists tell us that there will be a violent reaction, perhaps an explosion and a fire. If we separately bubble some chlorine gas through water, some of the gas seems to dissolve, the rest comes bubbling out. Breathing the chlorine gas or drinking the chlorine water is harmful. Yet when we toss some salt into water, nothing happens and no one is hurt by drinking the water. Clearly, *the sodium and chlorine cannot be in their native states* while in the salt crystal or in the dissolved state. X-ray analysis of the salt crystal shows us that the centers of the sodium and chlorine atoms are clearly separated from each other. The *only difference in state* that we can imagine is that sodium and chlorine may have done some swapping of their planetary electrons. Could such an apparently simple swap have accomplished such profound changes in the properties of the substances? The answer is a resounding yes, on purely experimental grounds as sketched above.

In following up this kind of analysis, physics and chemistry have shown that certain numbers of electrons are much more stable than all others. Further, it was soon noticed that atoms which naturally had one of these stable numbers were quite thoroughly inactive chemically. Thus, the first attempt to explain chemistry from atomic properties utilized the closeness of the atoms concerned to one of these stable states. In our case, sodium turns out to have one electron too many; chlorine has one too few. What could be more natural than for sodium to give its unwanted electron to the chlorine, thereby transforming both atoms into stable configurations? If we make this hypothesis, we see that we have simultaneously created a reason for the combination of sodium and chlorine. For the former now has a net plus charge and the latter a net minus charge. Electrically, the two atoms are simply pulled together.

Is it possible actually to explain the bonding by a simple electrostatic interaction? The mutual energy of two charges depends in a very simple way on the charges and their separation:

$$E_{el} = \frac{q_1 q_2}{r},$$

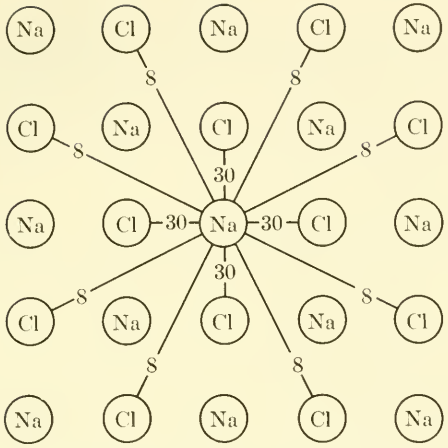


FIG. 10. A two-dimensional illustration of the interaction of the central Na^+ atom with its Cl^- neighbors. Most of the 180 kcal/mol binding energy comes from interactions with the four nearest Cl^- neighbors; each of these bonds is shown to be about 30 kcal/mol, so that these four bonds add up to 120 kcal/mol. The interactions with the eight next nearest Cl^- neighbors are much weaker because the separations are much greater. In this illustration, each bond is about 8 kcal/mol, so that these eight bonds total to 64 kcal/mol. In practice, the situation is really three-dimensional, and the repulsion of the nearest Na^+ atoms must also be taken into account.

where E_{el} is the electrical energy of the two charges of magnitudes q_1 and q_2 separated by a distance r.

The value of r can be deduced in several ways (including x-ray analysis) and the calculated mutual energy turns out to be 120 kcal/mol (Fig. 10). The experimental binding energy is 180 kcal/mol. Thus we see that the sodium chloride binding may be understood fairly well on the basis of what is called a simple *ionic bond* between the two constituent atoms. Electrical attraction of the ions with still other neighboring ions accounts for the other 60 kcal/mol.

Is it possible to explain all chemical bonds in these ionic terms? As we have remarked above, if the electrical forces didn't exist, there would be no atoms and no chemistry. But the simple picture of atoms giving or receiving electrons and thereby being electrically bonded is too simple. Consider the hydrogen molecule. It is composed of two originally neutral hydrogen atoms. There is no reason to suspect that one of the atoms has given its electron to the other. Perhaps it is reasonable to propose a model in which the two atoms share each other's electron. The state of atoms having two electrons had been found to be one of the stable states described above. So, by sharing each other's electron, *both* hydrogen atoms can attain this stable state. It is necessary to use quantum physics

to do the calculation properly, but the sharing idea has proved to be a good one, in that excellent agreement of theory and experiment is obtained for the binding energy of the hydrogen molecule. Since the electrons involved are the valence electrons of the atoms and since they are shared in a cooperative way, the shared electron bond has been given the name *covalent bond.*

For other molecular systems, an obvious extension of these notions has proved useful. If the atoms share the electrons unequally—so that the electrons spend somewhat more time with one atom than with another—then there is a combination of sharing and ionic bonds. These situations are usually spoken of as exhibiting some *partial ionic bonding.* To the extent that the electrons are shared, we calculate the binding energy as for a covalent bond. To the extent that electrons spend more time with one atom than with another, the first is more negatively charged; this requires that we calculate the electrical interaction as though each atom had something less than a whole electron's worth of electric charge.

In this fashion, then, the electric bond between atoms has been decomposed into ionic bonds, covalent bonds, and intermediate cases of covalent bonds with partial ionic character. Do these models permit us to picture everything that takes place in the whirling clouds of electrons? Unfortunately, the answer is no. There are still three other bonds which are thought of as involving quite different electron configurations. And there is a slight extension of the covalent bond that we present first.

It is possible for electrons to be shared not just by pairs of atoms but also by all the constituents of a piece of matter. Essentially, the electrons, or rather, some of them, roam freely throughout the piece of matter. This extreme case is that of metals; metals are defined as substances in which electrons are completely free to move. We will not deal further with this type of interaction.

The next effect to be set forth involves the situation in which alternative electron configurations are possible. We return to the hydrogen molecule already mentioned as being bonded by a sharing principle which we did not elucidate at all. If we now look a little more at this case, we can find a feature which Newtonian physics would consider irrelevant. At any one time, the electron which belonged to one of the atoms might be near that nucleus while the electron which arrived with the other atom would be near its nucleus. But the electrons might also be with the opposite nuclei. Now, it would seem to make no difference which electron is with which nucleus, since in any event only one electron will be at one place at one time. But (simplifying the physics enormously) the procedures of quantum physics force us to count *both* possibilities, even though our intuition tells us this is counting things twice. The *exchange* of electrons shouldn't make any difference in Newton's physics, but it

makes a difference in quantum physics. The ultimate test rests on the experimental value of the binding energy, and the result is unequivocally in favor of quantum physics. *Exchange energy* does, indeed, exist.

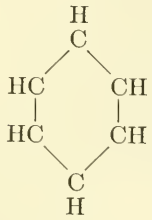


FIG. 11. A schematic representation of the benzene molecule: C_6H_6.

For biology, the exchange frequently appears in another form which is called by another name: *resonance*. To illustrate this, consider the benzene molecule shown in Fig. 11. Each carbon atom has four valence electrons and therefore makes four covalent bonds. Each carbon atom is bonded to one hydrogen atom. Thus, of the 24 bonding electrons initially present, 18 are left for the intercarbon bonds, making three bonds per atom. Figure 12 shows that there are two different ways of achieving this bonding, both using the idea of forming double bonds between half the pairs of atoms. As before, quantum physics tells us to count both possibilities in computing the binding energy, and experiment confirms this prediction. In this situation, there is a model for the exchange which results in the name of "resonance" energy; the molecule is said to resonate between the two forms. One can think that the electrons forming the double bond spend half the time on one side of each carbon atom, and half the time on the other side. By racing back and forth between these two configurations, the electrons tend to pull the carbon atoms closer

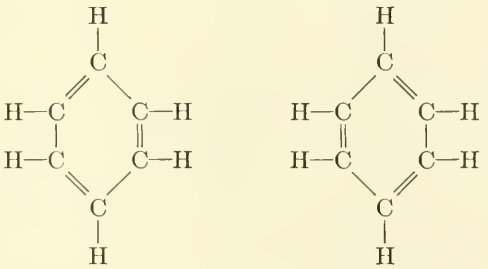


FIG. 12. Two equivalent representations of the benzene molecule, with all the bonds indicated. There are three bonds remaining per carbon atom after the bond to hydrogen has been established. The three double bonds can be placed in the two ways shown, and chemists speak of the structures as resonating from one form to the other.

FIG. 13. A particle is represented with its charges distributed so that the negative charges tend to be concentrated near one end and the positive charges near the other. The particle is thus equivalent electrically to the dipole indicated on the right.

together than with single bonds. Indeed, the measured separation (1.39 A) between the carbon atoms in benzene is actually intermediate between the 1.54 A for a single bond and 1.34 A for a double bond between carbon atoms.

The next kind of bond to be discussed is that called variously the molecular energy or the van der Waals energy. This is a bond between electrically neutral atoms or molecules, or groups of atoms or molecules. Electric neutrality does not necessarily imply zero electric charge; it may also arise from the equality of plus and minus charges. If the particle we are now discussing happens to have some of the negative charges separated from an equal amount of its positive charges, it is still neutral as a whole, but forms what is called a dipole, as in Fig. 13. Here the separation of the charges is indicated schematically by the circled plus and minus signs. Such an entity, composed of plus and minus charges, is called an electric dipole, by analogy with the similar situation with magnetic poles. Now, dipoles can interact electrically, too, although the attraction of one charge is offset to quite an extent by the repulsion exerted by the other charge of the dipole. Thus dipole forces are usually quite weak. We shall not enter into a discussion of these dipole forces other than to say that there are several varieties of them, all contributing to the van der Waals energy, and all decreasing much more rapidly with particle separation than by the regular interaction described at the beginning of this chapter. Because these dipoles are involved in the dispersion of light, the energy involved is also frequently called *dispersion energy*.

The final bond to be discussed is the *hydrogen bond*. Consider, for the moment, the OH^- ion. We recall that oxygen picks up two electrons to reach the nearest stable atomic configuration, and that hydrogen nor-

mally gives up one electron. Thus the OH^- ion should really be thought of as $O^=$ coupled to H^+. This is an ionic bond. When we examine the NH bond, a similar situation is found. For example, in ammonia, NH_3, the nitrogen would be triply charged negatively to reach the stable configuration. Now if the geometrical structures of molecules were such that an NH_3 group should be near an oxygen atom (such as one forming part of a C=O group) one of the plus-charged hydrogens from the NH_3 group could oscillate back and forth between its group and the oxygen atom, making essentially electrical bonds and thereby binding the two groups together. Since the hydrogen nucleus is nearly 2000 times as massive as the electron, the oscillations are much slower than for electron bonds, and the resultant bond is correspondingly weak. But since large molecules of biological importance contain many such groups, there can be many hydrogen bonds which add up to a respectably large total binding energy. For example, the nucleic acids' characteristic double-stranded structure appears to be held together entirely by hydrogen bonds.

To give an idea of the magnitudes of these bonds, the following table shows typical values in each of the five categories of bonds:

ionic	100 kcal/mol
covalent	100 kcal/mol
partial ionic	50 kcal/mol
resonance	40 kcal/mol
hydrogen	2–10 kcal/mol

One final feature to be discussed is called *saturation.* In the ionic bond, the ions interact electrically with all other electrically charged groupings in their neighborhood; the ionic bond is therefore said to be unsaturated, since other ions can join in. The covalent bond we have pictured is due to a sharing of electrons, and the result would be that the shared electrons would lie primarily between the two atoms. Therefore no other atoms could normally become involved—without breaking this covalent bond. Thus the covalent bond is said to be saturated, since it normally involves just one pair of atoms. Such considerations become important in estimating the effects of the entry of other substances in the neighborhood of the existing bonded substances.

MOLECULES IN SOLUTION

Practically all the chemistry occurring in biological systems takes place in solutions, almost all of which are water solutions. Thus it is important to look at the effects of water on the forces and bonds we have outlined.

Consider a charged particle, an ion, in a solution which also contains many small ions (e.g., Na^+, Cl^-, SO_4, etc.). The minus-charged ion on, say, a protein molecule will attract some of the cations in solution by a simple electrostatic force. As shown in Fig. 14, these cations will tend to cluster around the negatively charged protein, while the negative ions will be repelled. The situation would look as in the figure except that the plus ions also repel each other and attract the anions. Therefore the real picture will have the cations pushed out quite a bit, intermingled to some extent with the negative ions, so that it is only statistically that the protein has more plus charges nearby than minus charges. This picture has been worked out mathematically by Debye and Hückel, and the actual distribution of cations and anions around the protein can be shown to be just the kind of swirling, interpenetrating clouds that our simple approach yields. At some distance from the protein, of course, the electric force of the protein is vanishingly small, so that for the solution as a whole there are Debye-Hückel clouds around each charged molecule.

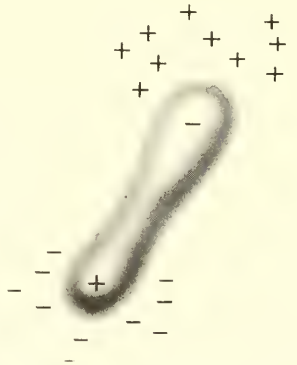


FIG. 14. The clustering of plus and minus solution ions near the respective minus and plus charged ends of a "dipolar" particle.

These ionic clouds affect reactions appreciably, for reacting molecules must drag along their sycophantic clouds, thereby reducing their mobility in solution, and therefore the reactivity. Also, because of the existence of these clouds of ions, molecules held together by ionic bonds or hydrogen bonds will have their binding appreciably reduced. It is possible that this effect is important biologically, especially since hydrogen bonds have been found to be responsible for the twisting and folding of such biologically important molecules as proteins and nucleic acids. Indeed, since the latter are held together chiefly, if not entirely, by hydrogen bonds, changes in the ionic concentration near nucleic acids could well be associated with their ability to separate into component strands during replication. The ease of separation of the strands has been verified as being greatly augmented by lowering the salt concentration while heat-

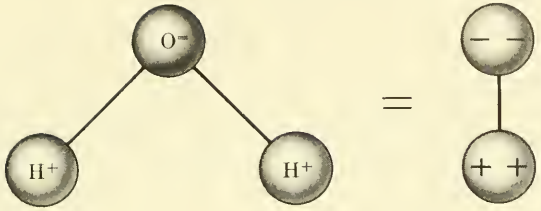


FIG. 15. A sketch of the water molecule and its electrically equivalent dipole.

ing nucleic acid solutions; this effect is, however, primarily associated with increase of the repulsion due to the negatively charged phosphate groups.

A second aspect derives from the fact that the water molecules are themselves dipoles; their positive and negative charges are permanently separated from each other, so that the molecule as a whole is electrically neutral, but still can effect electrical interactions. Thus water can act to split molecules bonded appreciably by electric forces. As an example, salt (sodium chloride) is found to be entirely separated into constituent ions when salt is dissolved in water.

Some investigations have yielded results suggesting that, in water, the hydrogen bond strength may be less than 1000 cal/mol. In any event, the strengths of ionic and hydrogen bonds, due to electric forces, are subject to important modifications when the molecules are dissolved in water.

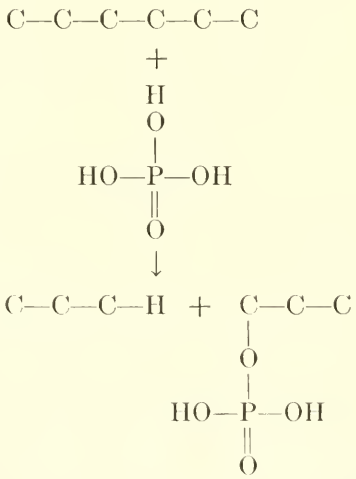
BONDS AND ENERGY EXCHANGE

Given the existence of the various chemical bonds, it remains to explain the energy exchanges involved in chemical reactions. In order to stick two atoms together in, say, a covalent bonding, we have to bring them close together. As long as they are many atomic diameters apart, there is no force acting between the atoms, but as they approach each other, the outer electrons of each exert a repulsive force on the other, so that we have to do work to push them still closer together. It is not until we have pushed them so close that an electron from one atom experiences the attractive force of the nucleus of the other atom that the electron swap which makes the two atoms exert a net attraction on each other can take place. It turns out that the net attraction is not very great, but as long as it exists, it holds the atoms together. If we now supply the energy to move the electron back where it was in the first place, the repulsion between the atoms is enough to force them apart and give them an appreciable velocity. This velocity corresponds to an energy of motion, which is the form in which we get back the energy we initially expended in forcing the atoms together. Thus, as long as the atoms are bound to each other, they have this latent energy. In this fashion, com-

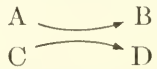
plex molecules which contain great amounts of latent energy may be synthesized.

Now, if such an energy-rich molecule collides, by Brownian motion, with another molecule of the right kind (and evolution has so arranged matters that there is a reasonable likelihood of a molecule of the right kind being in the neighborhood), a rearrangement of both molecules may occur. If the first molecule had a carbon-carbon bond, it could be broken by interacting with a molecule of phosphoric acid so that one carbon receives a hydrogen from the acid, and the other atom is coupled with the remaining phosphate group. This carbon-phosphate bond also has appreciable energy, *derived in this case from the carbon-carbon energy which now produces very little of the high velocity of separation that a simple splitting would have produced.*

The way this could happen would be for the phosphoric acid to approach the covalent bond so closely that the electron orbits were thereby disturbed. There would then be a certain probability that the covalent bond would be established with the phosphate group by displacing a hydrogen atom; the latter would then fall into place with the remaining carbon atom. Schematically, the reaction may be written as follows:



The generalization then is that energy is exchanged by coupling a dissociation reaction with an association reaction, with little energy loss resulting. This coupling of reactions is a basic feature of biochemical transformations, and in texts you find this written in general symbols as



Here the symbolism means that the energy-releasing reaction $A \rightarrow B$ is coupled with the energy-requiring reaction $C \rightarrow D$.

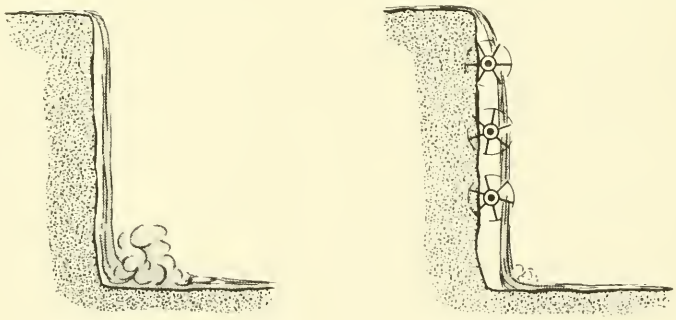
ENERGY AND BIOLOGICAL PROCESSES

Energy exchanges provide the key to understanding biological processes. Such a statement either must be obviously true (since the laws of physics and chemistry must be obeyed) or else some vitalistic factors must operate. Since none of the latter has been found necessary, as yet, to explain biological processes, science continues to work along physicochemical lines.

The way in which energy is used in biological processes is of some interest, and will be explained with the analogy of a waterfall, as in Fig. 16.

If, as in the left sketch, the water falls straight down to the ground, it bounces wildly because it has accumulated an enormous speed as it falls from the top. The energy in the water at the top (its ability to do work) is converted to the energy of motion of the water, and when the water strikes the ground level, there is a glorious display of hydropyrotechnics, after which the previously swift water molecules settle down to their new (energy-wise) impoverished state of being like all other water molecules. The energy of the waterfall has been dissipated into heat, for the water at the bottom is now somewhat warmer than it was before.

In the right-hand sketch, the water falls only a small distance before striking the vanes of a water wheel, which thereby is caused to turn. If the water wheel is connected to something, it can turn that something



(a) (b)

FIG. 16. In (a), the water falls freely to the bottom, bounces high because of the large velocity each molecule acquires while falling, and finally settles down to a quiet flow to the right. In (b), the water falls freely only to the first water wheel. The small velocity acquired is imparted to the vanes of the wheel, thereby turning the wheel and causing useful work to be done by whatever is connected with the wheel. The water then falls freely again for a short distance, picking up more velocity, which is imparted to the next wheel. Finally, the water arrives at the bottom with very little energy to dissipate, so there is little wild bouncing, and the water quickly settles down to its quiet flow to the right.

and cause it to do some mechanical work. Thus the water energy is partly converted into usable mechanical energy. We say "partly" because it took some energy to turn the wheel itself, some energy is lost in friction, and some in the bouncing off the wheel. Then the water falls a little more before going through a second water wheel. The water must fall somewhat, or it will have no motion to impart to the wheel; each time it hits a wheel, some of the energy is lost in the three processes mentioned. Thus, in a series of interactions with a set of wheels, the initial energy of the water can be fairly efficiently converted to usable energy.

One way of storing the energy extracted by each water wheel is to have it turn a magnet to create electric energy which could then, if desired, be stored in a battery for later use.

The same process, in spirit, exists in biological events. If a complex molecule exists, there is energy stored in its bonds, as described previously. The molecule exists because work was done in creating it, and this energy is now to be extracted by taking the molecule apart. If the molecule were taken apart all at once, the energy would be primarily released as heat, for it is difficult to couple an explosion efficiently to an energy storage device. Instead, the complex molecule is taken apart one atom (or a very few atoms) at a time, and some of the energy released is used to build up a small molecule. These small molecules play the role of storage batteries, and may serve subsequently to provide the energy for needs of the cell, such as movement, replication, etc.

In this stepwise fashion, the energy originally stored in the complex molecules can be used, with reasonable efficiency, to create energy storage molecules which then are of general use to effect the biochemical reactions. In practice, for example, the complex molecule could be a sugar, and the small battery-like molecule could be ATP. The device of storing energy in special molecules like ATP avoids the problem which would otherwise arise in coupling energy-yielding dissociation of complex molecules to energy-requiring reactions. There would otherwise be problems of different locales of the two processes, as well as problems in timing.

CHAPTER 3 *Physical Aspects of Vision*

Some of the physical aspects of vision are accessible to those who have some knowledge of statistical analysis. There are many kinds of studies that have been made, and we shall deal with several of these.

SENSITIVITY OF THE EYE

One important question about vision has to do with the number of photons required for seeing. The question has to be refined, because an immediate problem arises in deciding where these photons must be absorbed. Do they all have to be absorbed in the same rod (cones are neglected, since we know they are used for high light-intensity vision)? Can a single rod effectively absorb more than one photon in any arbitrarily short time interval? Do rods or groups of rods cooperate in vision? Do they cooperate in experiments to determine the *minimum* number of photons that may be detected?

Studies of the sensitivity of the eye to light are carried out by exposing the eye to flashes of dim light and measuring the fraction of times (the probability) that the flash is seen. If light of any intensity could be seen, the experiment would be uninteresting. But the results of these experiments are more than normally interesting.

There are two ways of doing the experiment. The most direct way is to shine progressively dimmer light flashes into the eye and to determine the smallest intensity that is normally seen. The actual mechanics of this experiment is of interest. The subject sits in a dark box with his teeth gripping a plaster cast specially made to fit the subject's teeth. This, plus some clamps on the subject's head, permits him to gaze directly ahead at a very dim red light which also helps to fix the subject's eyes. The rods are quite insensitive to red, so this light doesn't disturb the experiment. Dark-adaptation was achieved by having the subject sit in total darkness for 45 minutes. An aperture which can be illuminated with light of arbitrary intensity and wavelength is situated at an angle of about 20 degrees away from the line of sight. The size of the aperture is chosen so that the image formed on the retina covers only

about 500 rods. The amount of light needed for vision can be estimated from the fact that stars of the 8th magnitude are visible, corresponding to about 10^{-9} erg/sec. In the laboratory setup, even less energy was needed for detection.

The subjects were able to detect light of wavelength 510 mμ whose intensity corresponded to about 10^{-10} energy units (ergs) incident on the eye. This corresponds to about 100 photons. This energy has to be corrected for the various reflections and absorptions experienced by photons before they reach the retina and are absorbed. At the cornea about 5% is reflected, leaving 95 photons to continue. Half of these are absorbed in the humors of the eye, leaving about 45 photons. Eighty percent of these are absorbed in the nonsensitive portions of the retinal structure, leaving about 10 photons to be absorbed for this barely detectable flash.

In the actual experimental situation, it was found that light of a particular intensity could always be seen. But when a slightly lower intensity was used, the subject was always able to see the flash in a certain percentage of cases. That is, for an intensity which can always be seen, one cannot be found just slightly lower which is never seen. Therefore, for each subject, that intensity which was seen 60% of the time was arbitrarily taken as the threshold for that particular run. There were variations from individual to individual, and even the same person exhibited a day-to-day variation. In the experiments being discussed, seven subjects were studied, and the range for threshold seeing was 2.1 to 5.7×10^{-10} erg. One particular subject had a range of 4.83 to 5.68×10^{-10} erg.

At a wavelength of 510 mμ, the energy of a quantum is 3.89×10^{-12} erg. Thus, the actual range in photons is from

$$\frac{2.1 \times 10^{-10}}{3.89 \times 10^{-12}} \quad \text{to} \quad \frac{5.7 \times 10^{-10}}{3.89 \times 10^{-12}},$$

which is 54 to 158 photons. Since about 10% of the photons get through to the retina, the number of photons needed for threshold sensitivity is 5 to 16.

We have 5 to 16 photons falling on 500 rods. On each rod, therefore, there is an average (taking the 5-photon number for calculating purposes) of 1 photon per 100 rods, or a chance of 0.01 that any one rod received a photon. The chance that 2 photons will fall on the same rod is then approximately 0.01×0.01 or 0.0001—one chance in 10,000. Since a total of only 5 photons reaches the rods, it is highly improbable that two photons per rod are needed to produce a response. Thus, one photon must react with one rod, affecting one molecule of the visual pigment,

and in turn this reaction initiates a sequence of other chemical reactions leading finally to stimulation of a single nerve fiber. It must then be true that the requirement for 5 photons is exerted at the level of the nerves. The nerve excitation resulting from summing the effects of 5 individual nerve fibers is what is finally sufficient to trigger the optical nerve into transmitting an impulse to the brain.

On quite general grounds, a result greater than one photon is to be expected. If one photon were the threshold, then a single firing of a nerve fiber would result in vision. A single nerve fiber can be excited randomly by heat, fluctuations in chemical reactions in the retina, pressure, etc., so that we should be seeing flashes all the time. Therefore, we would guess that more than one photon would be needed, so as to reduce the random triggering of vision by these agents when it is irrelevant to the seeing processes. Of course, psychologists might maintain the entirely possible position that the triggerings occur but are suppressed by the mind so as to free the eye for purposeful vision.

A number such as 5 photons is interesting because the fluctuations in this number (from the Poisson Distribution) yield a standard deviation of $\sqrt{5}$, which is about 2.3. Two standard deviations ($2s$) equal about 5. Therefore, in some 5% of the cases where an average of 5 photons was delivered to the retina, the fluctuations should be great enough to result in *no photons* getting through and therefore the subject would not see the flash that the experimenter thought he was giving.

This leads us to another way of approaching the visual threshold problem. Consider that, for a subject to detect a flash, a threshold of n photons must be equaled or exceeded on the retina in the area being stimulated. When the number of photons arriving is less than n, the subject will see no flash at all; when it is n or greater, the flash will be seen each time. Thus, if this simple model were correct, the frequency of seeing flashes of increasing intensity would rise abruptly from zero to 100% at the threshold value of n. However, there are two sources of variation in such an experiment: biological variability and physical variability. We can't do anything about biological variability, for the term includes factors which we don't understand or cannot control. So we focus attention on physical variability. In this situation it means that when one shines n photons, on the average, on a retina, there will be fluctuations of this number. Indeed, we know that the standard deviation of the number is $\sqrt{n}$. That is, in one third of the cases, the number will differ from n by $\sqrt{n}$ or more. Now, when the average is n, the fraction of times there will be n or more is not unity, as in the simple nonfluctuating example first mentioned in the paragraph, but is something less than unity. For any given intensity of light, the flash will contain

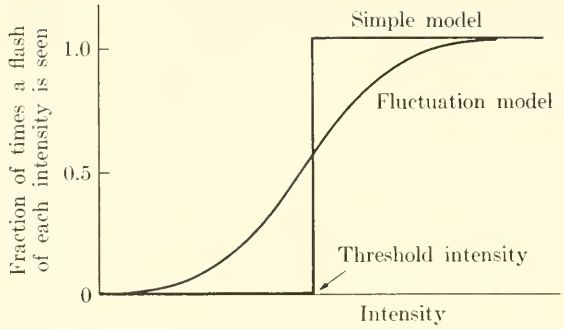


FIG. 17. A schematic representation of the frequency with which a flash of each intensity is seen, for two models. The simple model assumes that each flash has the same number of photons. The fluctuation model takes into account the fact that a given average number of photons (plotted on the abscissa) will have a fluctuating number of photons in each individual flash, for experimental reasons.

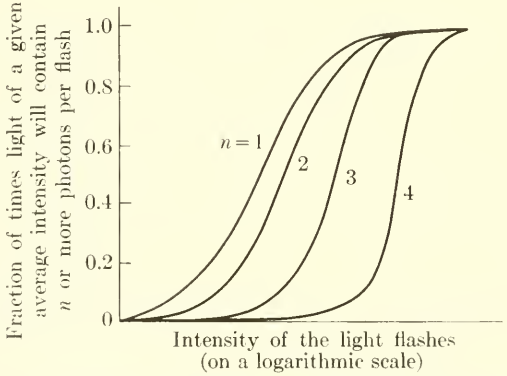


FIG. 18. Calculated curves of the fraction of times a light flash will contain at least n photons for the average intensity plotted on the abscissa.

a fluctuating number of photons, and only flashes containing at least the threshold number will be seen. As the intensity increases, a steadily increasing proportion of flashes will exceed the threshold number. Finally, intensities will be reached for which, despite the fluctuations, all flashes will exceed the threshold number of photons.

The simple model and the fluctuation model are presented schematically in Fig. 17. A flash of a given intensity is presented repeatedly to the subject, and the fraction of times he sees the flash is recorded. Then, in succeeding experiments, using a series of intensities, similar data are recorded. The data are then plotted to show the fraction of times the flash is seen for each incident intensity.

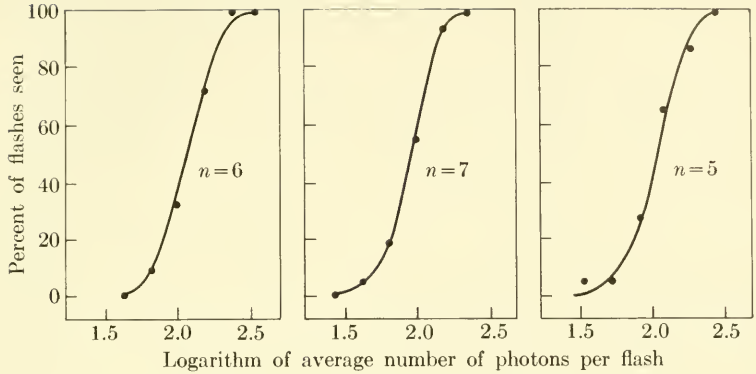


FIG. 19. This figure presents the experimental findings on three different subjects. The lines are the theoretical curves given in the previous figure. As may be appreciated by studying the theoretical curves, the choice of a curve with n greater or smaller by unity would give an appreciably poorer fit to the experimental points. (From Hecht, Shlaer, and Pirenne, *J. Gen. Physiol.* **25**, 819, 1942; courtesy the authors and *J. Gen. Physiol.*, Rockefeller Institute Press, New York.)

The theoretical analysis utilizes the Poisson formula. For any average number of incident photons, we can figure out the fraction of times there will be a minimum of, say, 1 photon hitting the retina. This curve is plotted in Fig. 18 with a 1 beside it. Similarly, we can figure out the fraction of times there will be at least 2 photons hitting the retina for any average intensity. And a similar calculation can be carried out for any threshold number of photons. Figure 18 shows these curves for various values of this threshold number. Next we try to fit the calculated curves of Fig. 18 to the actual data obtained. Since the shape of the curve depends strongly on the value of the threshold number, it is possible to distinguish quite well between possible curves. In this way, by trying successively 3, 4, 5, 6, 7, etc. photons for the threshold, it is possible to show that the best number is about 6 photons. This means that at low intensities there were few times that the flash was seen because there were few flashes which chanced to contain 6 photons, since the average number of photons was less than 6. When an intensity was reached such that an average of 6 photons reached the retinal absorbers, not all the flashes were seen because some of the flashes contained 0, 1, 2, 3, 4, and 5 photons, none of which could be seen. When very high intensities were used, all flashes had at least 6 photons, and so all were seen. In fact, as shown in Fig. 19, the data for three different subjects gave 5, 6, and 7 photons as the number which best fitted the data. Note that this experiment could have been done simply by fitting the curves to the data—it

is not necessary to measure the absolute intensities, so long as we know the ratio of the intensities.

It is of interest to point out that this experiment could have been done in precisely this way 150 or more years ago. If someone had had the inspiration to fit the results with the already known Poisson Distribution, the quantum nature of light could have been discovered that long ago.

STRUCTURE OF THE RETINA

One of the problems which perturb biophysicists is the reason for the size and shape of biological structures. In this category comes the question of why we have a concave retina. In the early 1800's, one Johannes Müller was trying to figure out the answer to this question, and in speculating on the problem he remarked on some interesting aspects. Consider a flat, plane retina, as in Fig. 20(a). An object such as the arrow to the left will send light into many of the rods shown schematically in the retina. As a result, only the presence or absence of an object will be detected by such an eye. If the retina is made concave, as in part (b) of the figure, some of the rods will still see many parts of the arrow, so that little improvement in vision will result. But consider the convex retina in part (c). Here, if the angular opening of each rod is sufficiently small, only a very few rods see the various portions of the arrow, so that what amounts to an erect image of the arrow is formed on this retina.

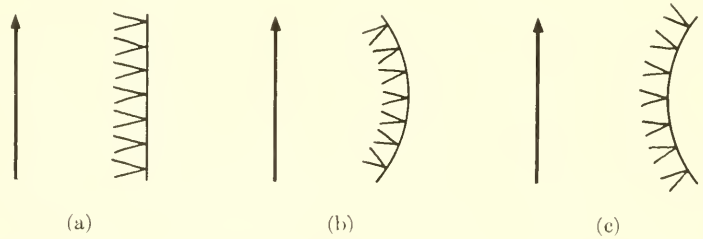


FIG. 20. A schematic representation of the problem of detecting the shape of an arrow as seen by retinas which are, respectively, flat, concave, and convex to the arrow.

Up until the time of Müller no such retinal structures were known, and it is to his credit that he took the trouble to go to Nature to discover that this is essentially the insect eye. The rods are not really rods, but are conelike structures called *ommatidia,* whose angular opening is about one degree. It is therefore not surprising that the resolution of this insect eye is about one degree. This information is gathered by making a grid of black and white bars of such a thickness that at the eye of the insect

the individual bars subtend, in various experiments, 0.1 degree, 0.5 degree, 1 degree, 2 degrees, etc. In response to a movement of the 2-degree bars, the insect responds by moving away. When the 0.5-degree bars were used, the insect made no response, thus showing that the fineness of the bars was such that the images overlapped and looked grey to him; he could perceive no motion in these circumstances.

The human eye can be studied in a similar way, and the resolution of 0.01 degree corresponds well with the angle subtended at the lens by neighboring retinal elements. To give an idea of what these numbers mean, a finger at arm's length subtends about one degree. Thus an insect can just make out the individual fingers of a man one yard away; a human can make them out about 100 yards away.

It should be realized that the experiments described in this chapter were done 20 years ago and that much work of a far more sophisticated nature has been done in the intervening years. Also during that time, enormous success has been achieved in studying the biochemistry of vision. Some biophysicists feel that studies of vision will provide the most direct route to an understanding of the nerve interplay which goes by the name of neurophysiology. Accordingly, the neurophysiology of vision is emerging as one of the most important and most exciting areas of research today.

CHAPTER 4 *Physical Aspects of Hearing*

INTRODUCTION

Physics has established that the thing we call sound travels through matter in a manner similar to the way waves travel. Sinusoidal oscillation of the particles of matter causes one particle to collide with the next, thereby transmitting its energy to the second particle. We therefore speak of *sound waves,* and ascribe to sound the various properties of waves, such as wavelength, amplitude, intensity, interference, diffraction, etc. The energy in the wave resides in the amplitude of the particle motion and in the number of particles executing the motion. Thus the expression for the intensity of the sound (energy passing through unit area in unit time) has been shown to be proportional to the density of the matter, the square of the sound frequency, and the square of the amplitude.

These aspects of sound are usually taken up in an introductory course in physics, and will therefore not be dealt with here. We will focus our attention on the following questions. How does our brain get information about the amplitude, frequency, and intensity of sound waves? To make one of these questions more concrete, how does the frequency information get to the brain? Is the frequency faithfully reproduced in the auditory nerve? Or is some particular nerve fiber sensitive to a particular frequency, so that all that goes to the brain is the information that particular nerve fibers have been fired, meaning that certain frequencies have been received? Such questions clearly involve knowledge of properties of nerves and nerve fibers, so that we shall have to introduce this study of sound with some aspects of nerve physiology.

The mechanism of transmission of an impulse down a nerve is believed to be associated with an alteration of the electrical state of the nerve membrane. The speeds of these impulses are of the order of 10 m/sec. Once the impulse has passed each point, the recovery processes set in. Upon the length of time required for these recovery processes depends the frequency of impulses that the nerve can transmit. Another factor which affects the frequency of transmission by a single nerve fiber is the time taken to transmit the impulse across a synapse and the recovery time at

the synapse. The events at the synapse by themselves show some of the limitations involved. The law of diffusion is, approximately,

$$x^2 \sim Dt,$$

where x is the distance diffused by a substance in time t if D is the diffusion coefficient. Putting in typical numbers for the acetylcholine molecule (shown to be involved in many synaptic transmissions) t turns out to be about one millisecond. Thus if a second impulse were to arrive at the synapse during this time, the impulses would be blended together. Therefore we can estimate an upper limit of 1000 cycles/sec for the transmission frequency of nerves. Experimental measurements have shown that this is indeed a good estimate, and that most nerve fibers can transmit only up to a few hundred cycles per second. If this limit holds for the auditory nerve, then it is clear that the nerve cannot transmit input frequencies of up to 15,000 cycles/sec, which is roughly the highest frequency that adults can hear. Accordingly, the early studies of the hearing mechanism focused on the finding of individual nerve fibers which responded characteristically to single frequencies. Several of the models proposed will now be examined briefly.

MODELS FOR SPECIFIC NERVES FOR SPECIFIC FREQUENCIES

There are two kinds of models which have been envisaged. The first proposes that certain nerves respond to certain frequencies, due to some as yet unknown mechanism. The second proposes that the nerves which respond to certain frequencies do so because they are connected to other elements which themselves respond to the frequencies; the nerves themselves could transmit any frequency up to their maximum. In the latter category are all the so-called resonance models, which propose, for example, that there are resonating structural fibers in the ear which respond only to characteristic frequencies. When these structural fibers oscillate, they trigger the nerve fibers to which they are connected. Proponents of the resonance theories have then had to demonstrate the existence of such resonating structures in the ear.

To proceed further, we need a sketch of the relevant portions of the ear. Figure 21(a) is a diagram of the snail-like cochlea. Part (b) is a section of the cochlea, showing its division into three so-called canals and the place of insertion of the nerve fiber into the basilar membrane.

The eardrum is connected by means of three auditory bones (malleus, incus, and stapes) to the vestibular canal. The approximate area of the stapes connection is also indicated in part (b) of the figure. Thus vibrations of the eardrum are communicated to the bones of the inner ear,

which finally transmit the vibrations, and the associated pressures, to the cochlea fluid filling the canals.

There are two kinds of resonance theories possible at this point: one based on resonances in the fluid, the other on resonances in the membranes. It can be shown that fluid resonances are physically excluded for a variety of reasons. Consider the schematic representation of the cochlea in Fig. 21(c). Here the cochlea has been shown as a straight tube with its two parts separated by a membrane (we neglect the cochlear duct here—it simply makes the dividing membrane more complicated). The aperture (helicotrema) connecting the two parts is also shown. The actual cochlea would differ from the schematic one chiefly in being twisted around into the snail-like shape.

When the stapes moves in response to an impulse entering the ear, the fluid is pushed. If the canals were filled with a gas or a compressible liquid, waves could travel along the canal and bounce off the other end,

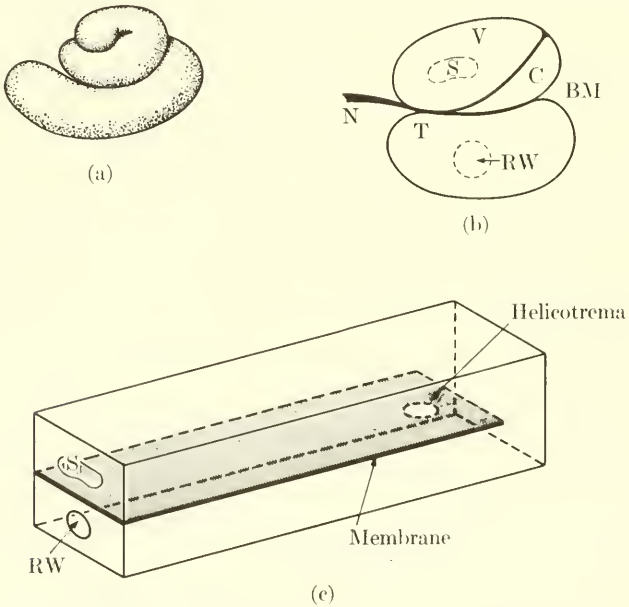


FIG. 21. Sketch of the cochlea (a) and a cross section of the cochlea (b). Part (b) shows the separation of the snail-like tube into Vestibular (V), Median or Cochlear (C), and Tympanic (T) canals. The stapes insertion at the end of the vestibular canal is indicated by the dotted line containing the symbol S, and the round window (RW) at the end of the tympanic canal is similarly indicated. BM is the location of the basilar membrane. N is the auditory nerve as it emerges from the cochlea. Part (c) shows a schematized version of the cochlea unwound from its actual snail-like configuration. The basilar membrane divides the two parts of the cochlea; the cochlear canal is omitted for simplicity. The helicotrema allows passage of fluid from one part of the cochlea to the other.

thus establishing standing waves, as in an ordinary musical instrument. The positions of the maximum pressures in these waves would activate pressure-sensitive nerve endings in the membrane. Thus the localization of points of maximum pressure would activate different nerves, depending on the frequency of the incoming waves. Unfortunately, this simple picture is entirely out of the question. It would require the existence of a cochlea with dimensions similar to those of musical instruments, and this is obviously not the case. In addition, the cochlear fluid is essentially incompressible, so that a more refined analysis is required. (We shall return to this aspect toward the end of this chapter.) Further, because of the viscosity of the fluid, it is hard to imagine that the frequencies could be so sharply separated from each other in what are called standing waves that there could be good frequency (pitch) discrimination. Therefore we confine our attention to *membrane* vibrations.

At first sight, standing waves in the membrane would seem to be ruled out for the same reason as were waves in the fluid. There exists, however, a phenomenon called, among other things, neural sharpening. We can illustrate this with an experiment you yourself can perform. Touch the tip of a finger with a reasonably sharp pencil point; it is best to do this with your eyes closed. Then touch approximately the same place with the tip of a ball-point pen whose point is retracted. The statement of neural sharpening is that you cannot tell the difference. Schematically, the statement is that your neural structure is such that you cannot discriminate between a point (·) and an object (◯) centered around the point—provided the touch is not too strong. The stimulation of nerves around the periphery of the circle is interpreted by the individual as a stimulation of the center point itself.

A more readily appreciable experiment (though less directly illustrative) that exhibits the neural sharpening phenomenon involves touching the skin at the same time with two reasonably sharp pencils. If the points touch the skin less than about 2 cm apart, only one stimulus is felt, whereas at about 3-cm spacing one is aware of the two separate pencil points.

The application to standing waves in the membrane comes in observing that even if a relatively broad area of the membrane is set into vibration by the incoming wave, neural sharpening results in the sensation that only the central point of the vibrating area has been affected. Thus only the particular frequency at that point will be communicated as having been heard.

According to this kind of theory, the *place* which is stimulated in the membrane is different for each frequency and is therefore the mechanical agency of sorting out the frequencies. There are other varieties of the place theory which will not be discussed.

An alternative resonance theory requires the existence of fibers which are tuned (like the strings in a musical instrument) according to their length and tension to produce a single note or frequency. Then, an incoming sound would cause the corresponding fibers to vibrate and the vibrations would stimulate the nerves connected to them. But there arises the purely experimental question of whether such fibers exist, and if fibers can be found, there is the additional question of whether they have sufficiently different lengths and tensions to respond to the range of frequencies of human hearing (30 to 15,000 cycles/sec).

Studies of the membrane have shown that there are indeed structural fibers in the membrane and that there are also hair cells in an organ lying directly above the membrane, but that neither set of possible resonators has a sufficient range of frequencies to be responsible for human hearing.

Both kinds of theories discussed up to this point are included in the general category of *place theories*, so called because the excitation of a particular place in the cochlea is supposed to inform the brain of the arrival of a particular frequency. It was a great disappointment that the place theories didn't work, because there existed impressive experimental evidence for their validity.

The kinds of evidence for the validity of place theories came from direct examination of the membranes. The ears of individuals known to be deaf to certain frequencies were examined after death, and lesions were found in certain places along the membrane; the places found in a study of many ears presented a consistent picture for the places of various frequencies. It was also found that deafness to certain frequencies could be induced in animals by subjecting an animal to such a great intensity of a particular frequency that the ear was injured and the animal subsequently no longer responded to the frequency. Examination of the resultant lesions in the membranes of such animals gave a place representation of frequencies along the membrane which coincided quite well with that derived from naturally occurring lesions. Thus there would seem to be no question that there is at least a significant amount of validity to the place theories, despite the inability of researchers to find a suitable model.

The resonance theories discussed to this point stemmed from the proposals of the German scientist Helmholtz, who began his studies in the middle of the 19th century. You recall that we began our examination of place and resonance theories because of the finding that individual nerve fibers cannot transmit impulses more often than at the rate of 1000 cycles/sec. Yet in recent years, it has become possible to measure the impulse frequencies directly on the auditory nerve, and faithful

replication of the incoming frequencies was found up to several thousand cycles per second!

The explanation of the apparent discrepancy lies in the fact that although individual nerve fibers cannot transmit above 1000 cycles/sec, the groups of fibers making up the auditory nerve may cooperate to transmit higher frequencies. In this so-called *volley principle* explanation, nerve fiber #1 fires, say, every 0.001 second, neuron #2 fires every 0.001 second, but starting 0.00033 second later. Neuron #3 fires every 0.001 second, but starting after 0.00066 second. The combination of the three neurons then fires three times every one-thousandth of a second, thereby reproducing a frequency of 3000 cycles/sec. It is not known where this "volley" action takes place, although the originator of the theory, E. G. Wever, has proposed that it occurs in the neurons which are connected to the inner ear membrane. According to Wever, hearing takes place by a combination of frequency representation (using the volley principle) and place representation (with neural sharpening).

The kind of place theory still has to be specified. We return to the previously mentioned fact that the canals are filled with an incompressible fluid. The properties of the fluid are such that waves travel much faster through the fluid than through the membrane, which is composed of reasonably stiff elements not very tightly coupled to each other. When pressure is exerted by the stapes, the liquid wave reaches the helicotrema and in part goes through to travel down to the round window long before any appreciable part of the membrane can have begun to move. Because the pressures above and below the membrane are not precisely in phase, there is a pressure difference across the membrane which displaces the membrane, producing a bulge. The place where this bulge occurs depends on the frequency of the incoming sound; the higher the frequency, the closer it is to the stapes. The positioning of the bulges seems to depend on a surprising property of the waves which can be predicted by the complicated mathematical theories now available. The property is that almost all of the wave is confined to a thin region of the fluid at the periphery, where the fluid is in contact with the canal material and with the membrane. Thus the situation is somewhat similar to that of surface waves coming in to a beach. Depending on the frequency and amplitude of the waves, they will "break" at various distances from the shore.

Many of these aspects of the rapidly developing field of hearing research are due to von Bekesy, who was awarded a 1961 Nobel Prize.

CHAPTER 5 *Light Absorption Effects*

INTRODUCTION

Quantum physics teaches us that we may talk about radiant energy as traveling either as waves in space or as discrete packages called either photons or quanta. When radiation impinges on matter, it has a probability of being absorbed which depends on (1) the energy in the photon (or its equivalents in wave terms: the wavelength or the frequency of the light), and (2) the nature of the matter. From the photon view, the photon may be absorbed if and only if there are two energy states (roughly two electron orbits) whose energy difference corresponds exactly to the photon energy, as sketched in Fig. 22. And, of course, there must be an electron in the lower of the two states, for there must be some particle to absorb the energy. If the structure of the matter is such that there is no energy difference equal to that of the impinging photon, the photon may impart its energy to the electron, but the electron will rapidly return to its original state and reradiate the energy. Thus, in effect, this kind of matter doesn't absorb the energy at all; because of the time taken for the absorption and re-emission, however, the reradiated photon is delayed relative to its initial path, and we say that its *phase* has been shifted.

Substances (such as glass) in which the energy gap between the electrons and the next higher energy state is so great that virtually none of the incident energy of visible light is absorbed, make suitable insulators. Electrical conductors, on the other hand, are composed of substances in which, by definition, the electrons are free to move, which means that there are so many permissible energy states that almost every incident photon would be absorbed. In advanced physics texts it is shown why the photons re-emitted by electrons in metals are virtually all in the backward direction. The result is that even a thin sheet of metal (e.g., household aluminum foil) is quite opaque, because the incident light is reflected backward.

When light is absorbed, the energy does not remain very long in the absorbing substances. The fact that the colors of substances do not change during illumination may be taken to mean that the absorbing substances quickly revert to their original state, so that they are once again ready to absorb the same wavelengths. There are five main fates

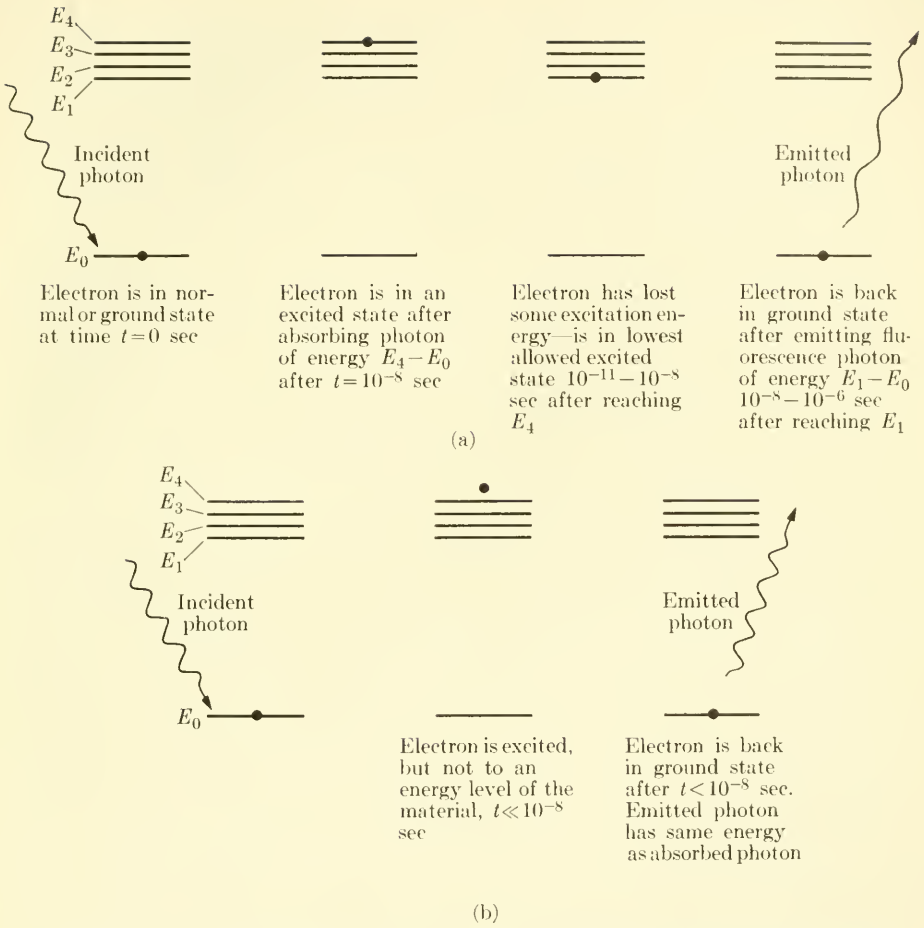


FIG. 22. (a) Light absorption and emission if photon energy equals an energy difference in the material. (b) Light absorption and emission if photon energy does *not* equal an energy difference in the material.

for the bulk of the light energy absorbed: (a) chemical reactions, (b) fluorescence, (c) phosphorescence, (d) energy transfer, and (e) internal conversion.

(a) Photochemistry—the study of the *chemical reactions* resulting from the absorption of light—is among the more fascinating provinces of current biochemical research. One of its most significant aspects is photosynthesis. Since the latter is dealt with in standard introductory biology courses, it will not be covered here. The more esoteric chemical aspects are beyond the scope of these writings. Thus the only segment to be covered in this monograph is that of action spectra, to be dealt with in a later section.

(b) *Fluorescence* is the simplest kind of result of light absorption: the re-emission of most of the light. If the light is emitted very soon after absorption (in no more than one microsecond), the re-emission is termed fluorescence. Because the system cannot emit more energy than it absorbed, the fluorescence is quite generally of a longer wavelength (more toward the red end of the spectrum) than the incident absorbed light. Because of rapid rearrangement of internal electron orbits [these are the changes shown going from (b) to (c) in Fig. 22], fluorescence is *normally* due to a transition from a *particular excited* electron orbit to a *particular lower* orbit very near the normal electron orbit; thus the light is usually monochromatic. If a substance absorbs in several regions of the spectrum (e.g., chlorophyll absorbs strongly both in the blue and in the red) the internal rearrangements referred to yield the result that the fluorescence is as though only the longer wavelength is absorbed (chlorophyll fluoresces only in the red).

(c) *Phosphorescence* is the emission of light considerably later than fluorescence emission—phosphorescences lasting a few seconds are not at all uncommon. The light emitted is primarily the same as in fluorescence. The reason for the delayed light emission in phosphorescence is that the electron is trapped in an orbit (perhaps E_2) from which, according to the rules of quantum physics, it cannot readily jump down to the normal orbit. After a time, which may be as long as a few seconds, the electron manages to get to the particular excited orbit from which it can jump with the emission of the normal fluorescent light. Because the molecules exhibiting appreciable phosphorescence are, *de facto*, in high energy states for considerable periods, they tend to be very reactive chemically.

(d) *Energy transfer* refers to the transfer of the absorbed light energy from the receiving molecule to another molecule. It will occur if the molecules are sufficiently close to each other and if the energy of the excited electron in the absorbing molecule chances to be matched to a possible excited energy state of the other molecule.

(e) *Internal conversion* covers a number of experimental situations. When electrons are excited by absorption of photons, they usually do not reach the next highest orbit but an orbit somewhat higher, so that they execute a vibration around the new orbit (the vibration, naturally, must be one allowed by the rules of quantum physics). This extra vibrational energy is dissipated by being communicated to the surrounding medium or to vibrations of the lattice if the atom is part of a larger structure. The electron thus reaches the lowest state of vibration of the new orbit, and this is the particular excited electron orbit referred to above in the discussion of fluorescence. The extra vibrational energy may be the source of a number of other effects ranging from rearrangements of the atoms composing a molecule to the dissociation of molecules.

The general picture, then, is that photons absorbed by electrons are usually re-emitted backward by free electrons, as in metals, and may be re-emitted primarily in the incident direction if the photon energy does not match any of the energy states available to the electrons.

These possibilities are sufficient to enable us to learn many things through the use of light. The phenomena which we shall discuss include (1) light absorption, (2) birefringence, (3) dichroism, and (4) action spectra.

1. *Light absorption*

If photons impinge on matter, each atom and molecule near which the photons pass has a probability of absorbing the photon. If we call E the probability per molecule of absorbing a photon of a particular energy, then the total probability of a photon's being absorbed is proportional to E and to the number of molecules along the photon path. The number of molecules along the path of a photon is the product of the number per centimeter and the path length in centimeters. Since the number per centimeter is proportional to the concentration of molecules, we can now write the average number of absorptions in a given solution of these molecules as Enx, where n is the concentration in molecules per cubic centimeter and x is the path length in centimeters.

A more usual form of this expression is obtained by multiplying and dividing by Avogadro's number, N_A, the number of molecules per mole:

$$(EN_A)\frac{n}{N_A}x.$$

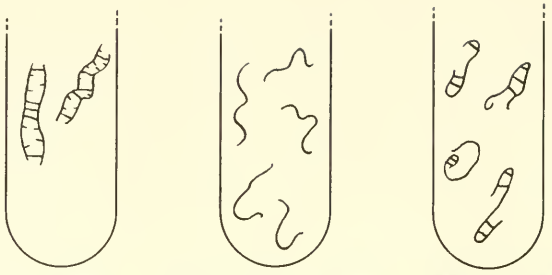
This expression, whose magnitude has been changed not at all by our operations, is now the product of the probability of absorption per mole, the molar concentration, and the path length. This is written as E_mcx.

This expression, as stated above, is the average number of absorptions per photon in the particular experimental situation. We may find the fraction of photons which is *not* absorbed by using the Poisson formula for the fraction of zero cases where the average, a, is the expression just above:

$$P_0(a) = e^{-a} = e^{-E_mcx}.$$

This fraction of photons not absorbed then represents the fraction of light, measured in intensity units, which passes through the solution. We write this as I_x/I_0 (the fraction of the incident intensity, I_0, which gets through a thickness x):

$$\frac{I_x}{I_0} = e^{-E_mcx}.$$



DNA solution heated to some temperature lower than denaturation temperature

DNA solution heated above denaturation temperature to produce strand separation. This is called melted DNA

Melted DNA after sudden chilling

FIG. 23. A schematic representation of the partial and complete separation of DNA strands by breaking some or all of the hydrogen bonds. If the completely separated strands are suddenly chilled, the strands form intra-strand bonds before any inter-strand bonds can be formed. The opposite would have been the case if the solution had been chilled slowly.

It is worth pointing out that machines which measure light absorption —colorimeters, spectrophotometers, etc.—express the result in terms of what is called the *optical density*, which is defined as

$$\text{O.D.} = \log I_0/I_x,$$

since this equals $E_m cx$ and is therefore directly proportional to the concentration of molecules which absorb, which is the quantity usually being sought experimentally.

This expression has been obtained for a single wavelength; the optical density will vary with wavelength if E or E_m varies with wavelength. By measuring the O.D. at various wavelengths for a given setup (keeping c and x constant), we can determine the dependence of E on wavelength. The plot of E versus wavelength is the *intrinsic* absorption spectrum of the molecule.

The absorption and scattering of light by a substance can depend very strongly on its state of chemical combination and on the light-scattering properties of the solvent in which it is suspended. We can see this from an examination of two experimental situations which have proved of interest and significance to biology.

(a) *The hyperchromic effect.* DNA is a long double-stranded molecule, held together by hydrogen bonds along the length of the molecule. As a solution of DNA is heated, the thermal agitation becomes great enough to break more and more of these hydrogen bonds. At any temperature, there will be equilibrium between the breakage of these bonds and their re-formation, so that at any time when there are still some intact hy-

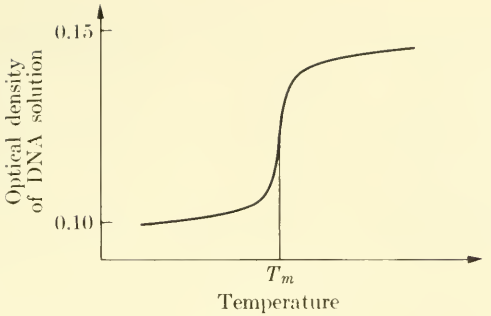


FIG. 24. A sketch of the optical density (O.D.) of a DNA solution as the temperature is slowly raised from room temperature to about 100°C. T_m is the *melting temperature,* defined as the temperature at which half the O.D. change is found. The increase in O.D. is known as the *hyperchromic effect.*

drogen bonds, the strands will be held together. It is possible to raise the temperature so high that all the bonds will be broken at once. If, at this temperature, the solution is suddenly chilled (by placing the tube of DNA solution in an ice bath), the chances are greater that the individual strands will fold up and make internal hydrogen bonds than that they will have time to find the complementary strands to re-form the DNA. This is sketched in Fig. 23.

If the light absorption depends on the chemical state of the substance, it would be possible for the solution to absorb different amounts of light in each of the situations shown in the figure. In fact, when DNA is "melted" there is a 40% increase in light absorption at the peak absorption wavelength (260 mμ) for nucleic acids. Figure 24 shows the light absorption of a DNA solution at various temperatures. The DNA is said to have a melting temperature T_m at the temperature halfway between the native and melted (or denatured) states.

It turns out that the melting temperature T_m is quite characteristic of the DNA of various organisms and that related organisms have melting temperatures which are very close to each other, so that this light absorption measurement has phylogenetic significance. The increase of light absorption is known as the *hyperchromic effect,* and is useful in a number of experimental situations. Since RNA and single-stranded DNA have very small hyperchromic effects, it is possible to deduce their presence or absence by light absorption measurements.

(b) *Matching of refractive indices.* Light is bent in direction when it goes from one medium to another of different optical properties (you know this from the classical case of long objects appearing broken when half in and half out of water). The light-bending property of a substance is known technically as its refractive index. The contents of a cell will

have some particular refractive index, and the cell will be visible if suspended in a medium of another refractive index. This result may be used in the opposite way. By varying the refractive index of the suspending medium, perhaps by varying the sugar content of the medium, it is possible to find a situation of just the right refractive index so that the various parts of the cell are no longer visible. Thereby, the refractive index of these parts of the cell is determined. Since the refractive index is proportional to the concentration of a substance, one can draw inferences about the concentrations of various substances in different parts of cells.

2. *Birefringence*

As hinted at in the introduction to this section, light which is not absorbed is re-emitted in all directions, producing what is called scattering of the light; we mentioned the two extreme cases of total reflection and total transmission. Most cases are intermediate. If electron oscillations are equally possible in all directions, the incident unpolarized light beam will emerge essentially as it entered the suspension—except for the phase shift already mentioned. If, however, the particles or molecules in suspension are not isotropic, i.e., the electrons can oscillate more readily in one direction than in another, an incident unpolarized light beam will be split, because waves oscillating in one direction will have their phases shifted more than those oscillating in another direction. The net result is that so-called anisotropic molecules produce *two plane polarized* emergent beams. Since these emerge in somewhat different directions, the phenomenon is called *double refraction* or *birefringence*.

Clearly, a measurement showing birefringence tells us at once that the material being studied is anisotropic, and since the optical anisotropy is due to anisotropy in the permissible movements of electrons, we can infer something about the structure of the material.

The chief use of this phenomenon in biology presently is in the inverse fashion. We shine plane polarized light on the substance of unknown structure and see what happens to the plane of polarization. Further, if we line up the molecules of a substance (as can be done in several ways), then the refraction (bending of the light) measured along and at right angles to the direction of alignment will tell us even more about the electronic structure of the individual molecules.

Birefringent structure may arise in several ways, of which a few will now be briefly presented.

(a) *Intrinsic birefringence.* The chemical nature and structure of individual molecules may contribute birefringence if there is anisotropy in the movability of the electrons of the molecules. Chemical bonds themselves may be highly anisotropic, the C—C bond being a good example.

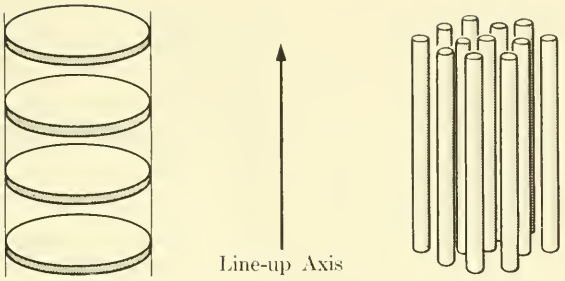


FIG. 25. Disc-shaped and rod-shaped molecules lined up so that their birefringence may be studied. (From G. E. Oster, "Birefringence and Dichroism," in *Physical Techniques in Biological Research*, Vol. I, p. 439, 1955; courtesy Academic Press, Inc., New York.)

This bond, being a covalent bond due to shared electrons lying primarily between the atoms involved, is difficult to distort at right angles to the bond. On the other hand, the triple bond C ≡ C has the atoms bound about three times as strongly, so that the distortion of electron positions which can be effected is mainly at right angles to the bond direction.

(b) *Form (or organizational) birefringence.* As a result of grouping molecules in a regular array (or in a composite structure with an inherent regularity, e.g. nucleic acids) there may well occur directions of easier and harder electron movability in the structure as a whole.

(c) *Flow birefringence.* If a solution is caused to flow, the viscous forces will tend to orient particles with their long axes in the direction of the flow. Spherical particles may even be distorted by these forces. Thus, during the flow, there will be organizational birefringence whose magnitude will depend on the speed of the flow.

Consider the two separate cases of a collection of discs and a collection of rods, as sketched in Fig. 25.

In the case of the discs, electrons can be moved more readily along the planes of the discs than out of the plane; therefore polarized light incident at right angles to the lineup axis will be affected more than polarized light incident along the axis. The case of the rods is precisely opposite, since electrons can more readily be moved along the lineup axis. Thus, the measurement of the birefringence for lined-up molecules tells us whether the molecules have a rodlike or a dislike shape and how they are oriented with respect to the lineup axis.

There are several important instances of the use of birefringence which permitted significant deductions about structures in biology. The case of muscle is the first one, in which the measurement of the birefringence permitted the conclusion that there are both isotropic and anisotropic regions in striated muscle—the bands called *I* and *A* are actually named

for these properties. In the anisotropic region, the result of the birefringence measurement led to the conclusion that the actomyosin fibers must be oriented along the muscle fiber axis.

The second case is that of chloroplasts, whose birefringence showed them to be composed of a stack of discs long before the electron microscope made the structure visible.

The third case is that of DNA. By slowly drying a solution of DNA stretched over a hole by capillary forces, a fiber can be produced. The fiber birefringence shows that the individual elements of the DNA are lined up in the disc configuration, even though it was known from viscosity studies that the individual DNA molecules must be long and thin. Thus it was possible to conclude that the elements composing the DNA —the nucleotides—must be arranged as flat discs perpendicular to the long axis of the molecule.

3. *Dichroism*

The section on birefringence has dealt with the transmission and scattering of light by substances which do not absorb the light. By examining the light scattered in two mutually perpendicular directions, we were able to make inferences about the orientation of the subunits of the substances.

In a very similar manner, we could examine the light transmitted by substances which *absorb* appreciable amounts of the incident plane polarized light. The results of the analysis are of the same type as for birefringence. For example, DNA has been studied by shining plane polarized light of wavelength 260 mμ, the wavelength of maximum absorption by nucleic acids. There is twice as much absorption perpendicular to the fiber axis as there is parallel to it. Thus one concludes that the absorbing elements themselves lie chiefly perpendicular to the fiber axis, as already deduced from birefringence studies.

If more light is absorbed at right angles to the fiber than parallel to it, the fiber is said to exhibit *negative dichroism;* if the converse obtains for the absorption, the fiber exhibits *positive dichroism.*

The energies involved in molecular vibrations and oscillations are such that the wavelengths of light absorbed by these movements lie in the infrared region of the spectrum. Polarized infrared sources now exist, and many substances have been studied by this method. Such studies have been of considerable use because of the characteristic absorption by atomic groupings. For example, the stretching of the N—H bond involves light of 3.0 μ wavelength, C—O stretching is at 6.0 μ, and the bending of the N—H bond is at 6.5 μ. By measuring the dichroism at each of these frequencies, the relative orientations of the groups may be inferred. For

simple molecules, the information may be enough to permit a fairly good preliminary model of the three-dimensional structure.

Finally, it should be noted that when light is absorbed by dichroic structures, the re-emitted light (the so-called *fluorescence*) will be polarized, provided only that the structures (or their constituent parts) do not rotate appreciably during the time (about 10^{-8} sec) required for the re-emission of the light. The study of the fluorescent light then provides additional information about the structures involved.

4. *Action spectra*

For light to produce an effect it must be absorbed, and for absorption to take place there must be energy level differences equal to the energy of the incident photon. Once the photon is absorbed, the pigment (the substance absorbing the light) may itself produce the effect being observed or it may pass the light energy on to another substance by one of several mechanisms. For example, it may happen that a pair of energy levels of the pigment chance to coincide with a pair of levels of an adjacent molecule. If so, the energy may be directly transferred, with the result that the electron of the pigment molecule ends up in its lowest energy state, while an electron of the adjacent pigment is raised to the excited energy state. If this adjacent molecule can produce the effect being studied, the energy transfer will permit the utilization of light not directly absorbed by the effective molecule, thereby expanding the efficiency of the system.

As an example, if light is incident on chlorophyll, the wavelengths absorbed are found to be in the red and blue regions of the spectrum, so that the chlorophyll color is that of the unabsorbed green light; the result of the absorption is the production of photosynthetic activity. Carotenoids absorb in the blue-green, and it is found that such light is effective in photosynthesis, thereby suggesting that the carotenoids and chlorophyll are energetically linked. To prove the linkage, we would have to show that the carotenoids themselves effect no photosynthesis. This will be shown later.

To obtain this information, we measure the effectiveness of light of various wavelengths in producing the effect (in our example of photosynthesis, the effect could be oxygen production, CO_2 fixation, etc.). This is done by irradiating an organism at such an intensity that the photosynthetic activity is not saturated, which means that we should choose an intensity such that the amount of photosynthesis would double if the intensity were doubled. Thereby we obtain the amount of photosynthesis per incident photon of the given wavelength. Then the entire experiment is repeated with light of various wavelengths, so that finally

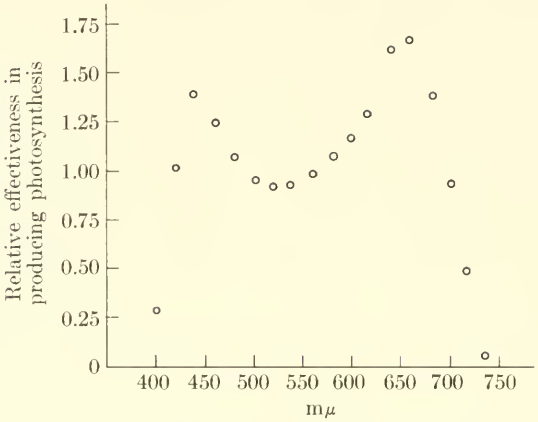


Fig. 26. The effectiveness of various wavelengths of light in producing photosynthetic carbon dioxide fixation. Two clear peaks are in evidence. The fact that the effectiveness does not go down to zero between the peaks indicates the existence of at least one more peak in the wavelengths between the two.

we could plot, as in Fig. 26, the relative effectiveness of the various wavelengths. Of course, to obtain these data we need to measure the incident light intensity and the fraction of the light absorbed by the system. This implies that we must use a solvent medium (photosynthetic algae would be swimming in some specially designed medium) which is entirely nonabsorbing in the spectral region studied, or that the fraction of the light absorbed by the medium must be separately measured and corrected for.

The shape of the effectiveness spectrum or, as it is usually called, the *action spectrum*, is then studied to see whether it resembles any known pigment or collection of pigments. In the figure above, we see basically three peaks. By comparing this spectrum with those in Fig. 27, we see that two of the peaks are similar to those of chlorophyll, the third to that of carotenoids. Thus we deduce that both pigments are involved.

To know that carotenoids are transferring their absorbed photon energy to chlorophyll requires additional experimentation involving fluorescence measurements. When a wavelength absorbed only by chlorophyll is used to illuminate the organisms, the typical red chlorophyll fluorescence is found. When organisms are irradiated with wavelengths absorbed only by carotenoids, no typical carotenoid fluorescence results; only the chlorophyll fluorescence is observed. Thus we may deduce that the energy transfer is only from carotenoids to chlorophyll. Since the efficiency of production of photosynthesis by carotenoid absorption is nearly as high as that of photons absorbed by chlorophyll, we conclude

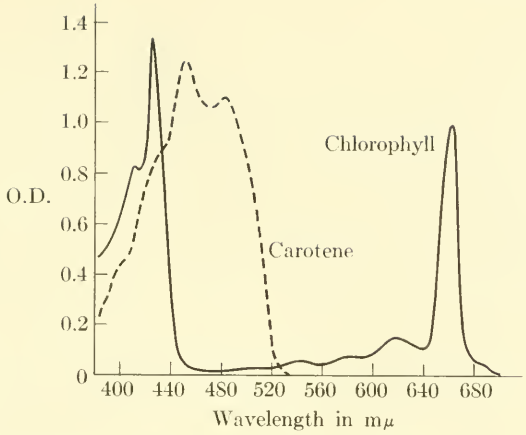


FIG. 27. The absorption spectra of chlorophyll and carotene. The chlorophyll peaks at about 430 and 665 mμ and the carotene peaks near 450 and 490 mμ are identified as being responsible for the action spectrum peaks in the preceding figure.

that virtually all of the energy absorbed by carotenoids is transferred to chlorophyll.

Action spectra then may be used to establish the absorption spectra of compounds involved directly (or indirectly, by energy transfer) in any given light-mediated process. Indeed, in many important investigations the action spectrum has given the first information about the nature of the molecular species involved. The student should recall that the comparison is between the action spectrum and the plot of the extinction coefficient, E, against wavelength. Some additional examples of action spectra utilization include the fitting of the action spectrum for human vision to the absorption spectrum of rhodopsin, the action spectrum of phototaxis with the absorption spectrum of carotenoids, and the action spectrum for flowering of certain plants with the absorption spectrum for a molecule of what is now known as phytochrome.

Up to this point we have considered situations in which the absorbed light *promotes* a given effect. There is a second type of action spectrum in which the absorbed light inhibits or inactivates some process. A trivial example is that of flowering inhibition by far-red light which reverses the action of the phytochrome—this action spectrum has turned out to be that of an excited state of the phytochrome molecule itself.

The more general case of inhibition results from the use of light of such great energy as to damage the pigment. Ultraviolet light is the usual agent of this kind of experimentation, since its photons are sufficiently energetic to break chemical bonds, thereby altering the chemical structure of the absorbing structures. Viruses, bacteria, and many kinds

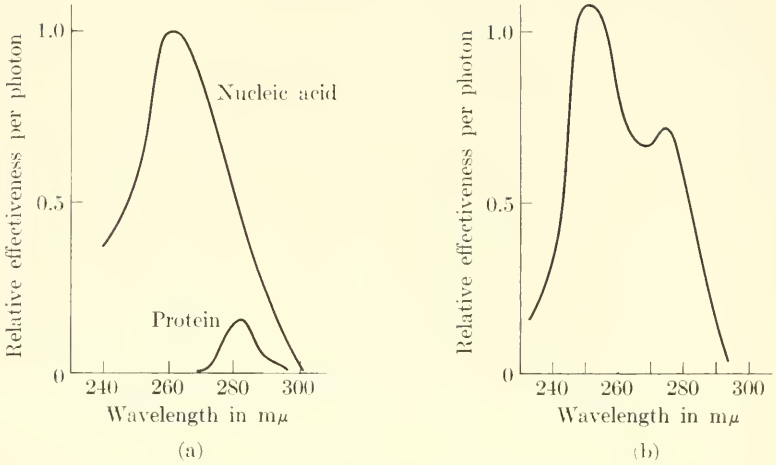


FIG. 28. Part (a) shows the absorption spectra of equal concentrations of nucleic acid and protein. Part (b) shows the action spectrum for inactivating the capacity of Euglena cells to form photosynthetic progeny. Its peaks at 260 and 280 mμ show that both nucleic acids and proteins are involved in this experiment, in which irradiated cells give rise to white, nonphotosynthetic colonies in place of the normal green colonies from unirradiated cells. (After H. Lyman, H. T. Epstein, and J. A. Schiff, *Biochem. Biophys. Acta* **50**, 301, 1961.)

of single cells are inactivated by exposure to ultraviolet light to the extent that they are actually killed.

In these instances, we measure the inactivation spectrum from about 220 mμ to about 350 mμ. In this region the nucleic acids and proteins are the chief pigments, and one usually learns only whether one or both are implicated in the inactivation. In the case of enzyme inactivation, only a spectrum similar to protein absorption is obtained. In virus inactivation usually a spectrum resembling nucleic acid absorption is found, although there are instances of effectiveness peaks corresponding to both nucleic acid and protein. The absorption spectra for nucleic acids and proteins are shown in Fig. 28, along with an example of an inactivation spectrum.

It must be carefully noted, however, that the process being studied need *not* be viability. To emphasize this point, we present the following list of inactivation spectra possible just for simple organisms such as viruses; more complicated organisms would have an even longer list. One can inactivate viruses with respect to (a) their ability to form viable progeny, (b) their ability to kill cells on which they normally grow, (c) their ability to adsorb to cells, (d) their ability to initiate the production of various metabolic processes, (e) their ability to participate in genetic recombination.

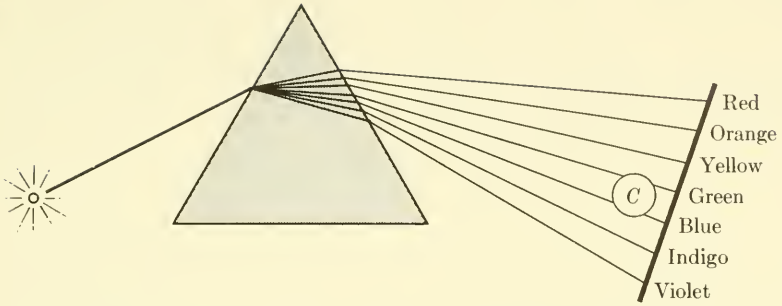


Fig. 29. Light passing through a prism is broken up into the spectrum as shown. A container C of finite size will necessarily intercept a whole range of wavelengths.

For each of these functions an action spectrum can be obtained, and there is no reason to expect the same spectrum for all effects; in fact, different spectra are found. Therefore it is possible to learn some structural and functional lessons from action spectra of this kind, too.

The obtaining of an action spectrum is not as simple as it may seem, due entirely to practical considerations. We have spoken as though it is possible to irradiate with monochromatic light, i.e., light of a single frequency. In practice, the source of light is some high-intensity lamp whose light is used in one of two ways. As indicated in Fig. 29, the light may be passed through a diffraction grating or through a prism to spread the light out into a spectrum. If the container C being irradiated is relatively large, or is placed relatively close to the spectrum-producing device, it will intercept light of many wavelengths, thereby producing an effect composed of all the wavelengths absorbed. If the container is placed so far away from the source that it intercepts only a narrow range of wavelengths, the total energy received becomes very low, because the intensity falls off inversely with the distance. Thus there is a practical limit to the minimum range of wavelengths being absorbed.

A second device is a so-called interference filter which, as its name suggests, uses an interference method to let through only a very small range of wavelengths. This filter may be placed very close to an intense light source to provide a high intensity. However, the band of wavelengths transmitted by such a filter is great enough to give trouble with interpretation in terms of a single wavelength. Thus, in practice, action spectra taken with a finite number of wavelengths are only roughly accurate, so that only broad classes of compounds can be designated as being involved in the action being studied. Happily, this is frequently all the information that is desired.

CHAPTER 6 *Physical Aspects of Muscles*

INTRODUCTION

The study of muscle action provides a good illustration of the different points of view of the biophysicist and physicist. Consider a trapeze performer hanging from a trapeze, supporting a (usually beautiful) lady trapeze artist. The physicist's definition of work is the product of a force on a body and the distance d that the body is displaced in the direction of the force:

$$W = Fd.$$

From this point of view, the artist holding the lady is surely exerting a force, but since there is no motion in the (upward) direction of the force, the physicist claims that no work is being done. This is patent nonsense, even to the physicist, as he discovers by trying to do this job himself. Of course, the reason the physicist insists on talking this nonsense is that he is talking about mechanical work being done against the gravitational pull downward. The work we have in mind—the work that results in sweat and fatigue of the performers—is biophysical and biochemical work. It takes energy to keep the muscles in their tensed state, and the biophysicist (and biochemist) is primarily concerned with this aspect of the performance.

Today, the physics and chemistry of muscle action constitute one of the more interesting research areas of biology. In describing the action of muscles, we will first set forth the over-all aspects of the situation in terms of forces exerted and the contraction and relaxation of the muscles. Then we will inquire into the substructure of the muscles which make possible the observed actions. Finally, we will indicate the molecular aspects of the substructure, and will sketch the possible connection with known biochemical facts of muscle action.

1. *The over-all action*

Contraction of a muscle, or the attempt to contract it, results in a force. Contraction is readily studied since the length of the muscle before and after the contraction can actually be measured and (essen-

tially by hooking the muscle to a very strong spring balance) the force being exerted can be determined. The use of these data depends on the questions being asked. One could ask for the total energy being expended, in which case a simple multiplication of the force and the distance through which it moved (the contraction) gives the external work done by the muscle. This doesn't answer the entire question, because there are quite obviously three energies to be reckoned with:

(a) the external work done,
(b) the work done in maintaining the readiness of the muscle, and
(c) the energy dissipated, as in heat, for there is always some inefficiency in any machine, even the human one.

There are still other questions. An important one derives from reasoning that the total external work done isn't always the important thing—some situations might be better described in terms of the *rate* at which work is being done, for that measures the extent of the immediate availability of the energy. In this case, as shown in an introductory physics course, the rate can be computed from the product of the (assumed constant) force and the velocity of contraction of the muscle. From this rate of doing work the energy expended is computed by first multiplying each rate by the duration of its persistence, and then adding the products, in the same way that an electric company computes the electrical energy we have used by adding the products obtained by multiplying the various rates by the duration of use of each. The formula can be stated as

$$R = \frac{E}{t} = Pv,$$

where R is the rate of energy E expended in a time t if a pull (or force) P is accomplished with a contraction velocity v. (These symbols are chosen to conform to practice in muscle research laboratories.)

By studying a number of muscle preparations from different sources, it has been found, empirically, by A. V. Hill (one of the pioneers of biophysics and of muscle research) that a generalization can be made:

$$(P + a)v = \text{constant}.$$

Here a is a factor whose existence is readily rationalized. It is the internal force expended to make the muscle contract, and therefore the product av is the rate at which work is done to contract the muscle itself. The magnitude of a can be determined from heat measurement, to be discussed later. This term is entirely similar to the corresponding correction to the ideal gas law.

$$PV = \text{constant}$$

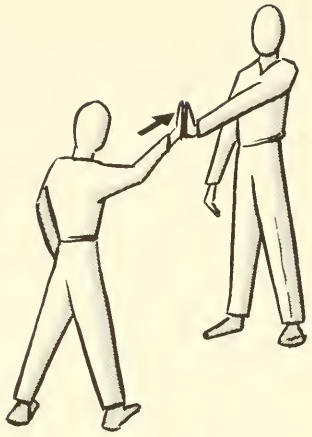


FIG. 30. An experiment to show that with a given energy you can either slowly exert a large force through a small distance or rapidly exert a small force through a large distance. This is demonstrated by having your partner offer either a large or small resistance to your push.

is corrected to

$$(P + a/V^2)V = \text{constant},$$

where the term with a is included to account for the work the gas molecules do to pull apart from each other in the process of expanding to exert a pressure P throughout the surface of the volume V of the container.

There is a second correction to the perfect gas law which takes account of the finite volume of the molecules by subtracting this volume, called b, from the total volume, since the container cannot be made smaller than the volume occupied by the molecules. In the case of muscle, there is a similar correction for the contraction rate of the muscle itself. Thus *Hill's characteristic equation* for muscle action,

$$(P + a)(V + b) = \text{constant},$$

can, by suitable choice of the two arbitrary parameters, be made to fit an extremely wide range of data on muscle contraction. The constant value can be readily determined, since when maximum force is being exerted the muscle will have its smallest contraction rate, which is b. Thus the constant is simply $P_{\max}b$.

If the statement above seems implausible to you, you may familiarize yourself with the point by a simple experiment. Have a friend hold his hand facing you with palm vertical, as in Fig. 30. Now ask him to do the following when you push on his hand. First, have him offer little

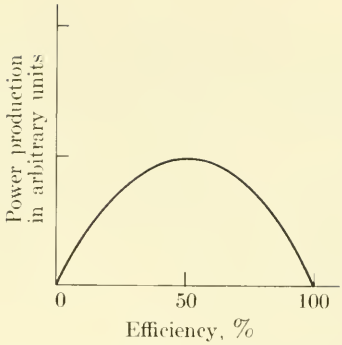


FIG. 31. A representation of the power produced as a function of the efficiency of using the force being applied. The efficiency of power production is equal to 50% when the power delivered is at its maximum value. (From Odum and Pinkerton, *American Scientist*, **43**, 331, 1955; courtesy the authors and *American Scientist*, Princeton, N. J.)

resistance. You will see that his hand is rapidly displaced, but that you are exerting very little force. Second, have him offer his maximal resistance. You will see that his hand is displaced at a minimal rate, but that you are now exerting your own maximum force.

One interesting point seems worth an appreciable digression here. This concerns the design of biological structure from the point of view of energy expenditure. Are biological systems designed to give maximum efficiency (least energy loss) or maximum power deliverable? Clearly, we could conceive that some systems producing energy could be designed for one purpose and some for the other, and perhaps some for other purposes, too. What differences are involved in making this decision?

In an interesting paper on this topic, Odum and Pinkerton point out that the work done depends on the forces and rates. When the rates of displacement are very small, there is maximum force exerted. The power delivered is the product of these two factors so that, since v is very small, the power delivered is small. On the other hand, when the force is very small, the value of v is large, but the product is again small. It is then plausible, and they actually prove, that at intermediate forces and displacement rates, the power delivered is greater than at the two extreme values we have mentioned. This is sketched in Fig. 31. In the cases they discuss, Odum and Pinkerton were able to deduce the exact form of the curve. There is an optimal force to be expected for maximum power delivered. At this value of the force, the efficiency proved to be only 50% for ideal, loss-free systems; for real systems it is less than 50%. That is, by exerting either a very large force at a slow rate or a very small force at a great rate, the physical efficiency is greatest, ap-

proaching 100%. But, if the energy-producing system is to produce the most actual power, the system must be designed to work near 50% efficiency. You may find further discussion of this point in relation to biological systems in the paper whose reference is given at the end of this section.

This digression allows us to conceive of a classification of muscles into two types: those which exert large forces through small distances, and others which exert small forces through large distances. Roughly speaking, these types correspond to what is actually found in nature. Typical striated muscles, such as facial, arm and leg, and heart muscles, contract very little, but exert large forces. Smooth muscles, such as many involuntary muscles, e.g., bladder muscles, contract appreciably, and exert correspondingly smaller forces over longer time periods.

Other than for measuring the external work done, the variables in Hill's equation do not afford much scope for investigation of muscle action problems. However, estimates of the energies involved in the various processes may be obtained from heat measurements. The sensitivity of the temperature measurements involved is fairly great but is well within the scope of modern technology, which has no trouble measuring temperature changes of millionths of a degree. When the temperature changes in muscles are determined before, during, and after a muscle contraction, it is found that there are three main stages of heat production.

(a) *Resting heat.* This is the energy expended in just keeping the muscles ready to perform. The metabolism within the muscle has the function of keeping the metabolic machinery in readiness for action in the extended position so that it can contract when the stimulus reaches it.

(b) *Initial heat.* This energy appears to be associated with the contraction process, being composed of two chief parts: the heat of developing the tension in the muscle, and the heat associated with the actual contraction itself.

(c) *Recovery heat.* This, as the name implies, is associated with the return of the muscle to its pre-contracted state, after which the resting heat keeps the muscles in readiness. The recovery heat includes the buildup of the molecules responsible for the quick availability of energy which characterizes the muscle action. In this case, much of the quick-energy activity is effected through the molecules of ATP and of creatine phosphate, the former of which is a general energy storage battery for the metabolic machinery of cells, the latter of which is an energy storer quite specific for the needs of muscles.

Some of the points mentioned above may be directly observed. It was first shown by Szent-Gyorgy that if an excised muscle is placed in glycerol, many small molecules are dissolved out of the muscle, leaving

chiefly proteins called actin and myosin bound together in a complex protein called actomyosin. Purified actin and myosin mixtures, which can be made to form fibers, were used in later experiments. If ATP molecules are added to the glycerinated muscle preparations or to the synthetic fibers, the result is a readily measured (albeit slow) contraction. The ATP molecules are used up in this reaction, so that part of the recovery heat must be associated with the resynthesis of ATP. Since in living muscle the ATP is present during the resting state of the muscle, it is clear that the mere presence of ATP molecules is not, by itself, the trigger for the contraction.

It is of interest to point out that a similar role of contractile proteins exists in an organism very different from those we ordinarily think about as having muscle action. Bacterial viruses attach to the cells they infect by means of appendages which have been called tails, though there is no reason to think of these long slim appendages as tails any more than as heads or necks. It has been shown that these viruses inject their nucleic acids into the cells they attack, and that this injection is accompanied by the contraction of the tails. Further, it has been shown that there are about 100 molecules of ATP in the tail, and that the ATP is used up during contraction in the same way as in the muscle contraction we have been discussing.

2. *Muscle structure*

The light and electron microscopes have afforded us an excellent picture of the substructure of muscle which should eventually lead to the elucidation of the entire contractile mechanism. Unfortunately, up to this point in time, the considerable technical problems in preparing materials for electron microscopy have prevented the detailed substructure from being determined.

The aspects of muscle visible to the eye resulted in classification of muscles into striated and smooth types. Smooth muscles, by dissection and examination in the light microscope, are found to be composed of short (roughly 20 microns) cells of optically isotropic material. Striated muscle, on the other hand, is found to be composed of bands of materials, the bands being distinguishable initially by being optically different. There is a band of optically isotropic material, called the I-band, and a band of optically anisotropic material, called the A-band. There are two subdivisions in each band. The I-band is divided into two equal parts by a dark-staining thin band called the Z-band (from the German word *Zwischen,* implying a "between" band). The A-band is likewise divided into equal parts by a slightly thicker clear band, called the H-band (from the German word for clear—*hell*). These are shown dia-

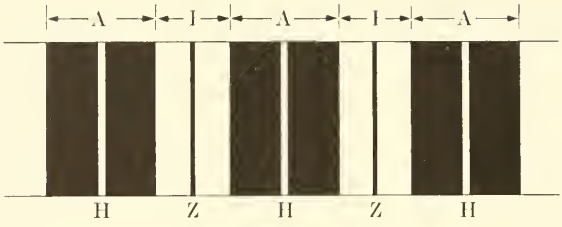


FIG. 32. A schematic representation of a muscle showing the A, I, Z, and H bands. The A-band is optically anisotropic. The I-band is optically isotropic.

grammatically in Fig. 32. More recent work has shown that the H-band contains a thin central M-band, and that the Z-band is similarly constructed. These refinements do not concern us here.

When a muscle is stimulated, there is a contraction, followed by a relaxation and return to the original state; the entire process is called a *twitch*. Observation with the light microscope reveals that during the contractile phase of a twitch, the I-band and the H-band shorten by about the same amount; the A-band width does not change. Since the dark part of the A-band increases in size by the amount that the I-band decreases in size, it is plausible that the material of the I-band is sliding into the A-band, and that this is the mechanism of contraction. Of course, the equality of change in the H-band and I-band could be mere coincidence or have an entirely different interpretation. A once popular theory ascribed the changes to alterations in configuration of the long protein molecules making up the fibers; if the physico-chemical environment changed, the proteins would change in about the same amount. This theory has foundered on the rock of measurement, for x-ray studies have demonstrated that there is no such configuration change in the proteins during muscle contraction. Indeed, this last result teaches us that the muscle shortens while its constituent filaments do not; this is a result which may be restated as showing that there are no contractile proteins in muscle but that there are contractable configurations which leave the filamentous proteins unaltered in length. It is further worth noting that when muscles are stretched, the H- and I-bands both increase by about the same amount, thereby supporting the notion that the two changes are coupled, since the contractions and extensions occur together.

This deduction about material from the I-band sliding into the A-band cannot be the whole story. For if the I-band were to disappear entirely, by sliding into the A-band, the shortening would have a maximum value of about 40%, since the I-band is roughly 40% of the total length of the A- plus I-bands. Yet measurements of intact muscles reveal a shortening

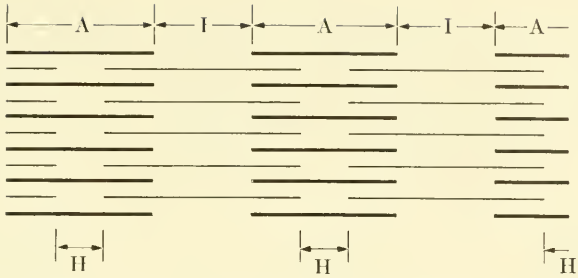


FIG. 33. A diagram of the apparent arrangement of actin and myosin molecules in a striated muscle. The thin lines represent the actin molecules; the thick lines the myosin molecules. When the actin molecules come together end-to-end, the muscle is presumed to shorten maximally, with the disappearance of the H-band. The maximum shortening on this model would be 40%.

of up to 80%. Further, this shortening of 80% may also be observed in smooth, unstriated muscles which presumably are very differently constructed. Furthermore, the synthetic actomyosin filaments, when incubated with ATP, have been observed to contract more than 40%, and in this case there is no possibility of a "sliding" structure.

It is therefore of some weird interest that electron microscopy of muscles has revealed an internal arrangement which dovetails neatly with the sliding structure model. The structure is represented diagrammatically as in Fig. 33. What we see are thick myosin filaments and thin actin filaments. These long protein molecules are placed so that the dark parts of the A-band contain both kinds of molecules, the clear H-band contains only myosin molecules, and the I-band contains only actin molecules. As the sliding structure model would have it, the actin molecules of the I-band slide past the myosin molecules until, as sketched in the next figure, they have gone as far as they can. Thus, the I-band has shortened by as much as the H-band. The evidence for these diagrams derives from studies of cross sections in the I-, A-, and H-regions. As shown in Fig. 34, there are only thick filamentous elements in the H-band. In the dark part of the A-band, there is the same number of thick elements, and twice as many thin elements. In the I-band there are only this doubled number of thin elements. Some clever chemistry has demonstrated that the thick elements are myosin, and that the thin elements are mainly actin.

As may be deduced from the discussion thus far, the structural basis of muscle contraction is simply not known, despite a wealth of detailed information about the nature and arrangement of the materials making up the muscle. Either some vital information is lacking or else there is some misinformation in what we think we already have established.

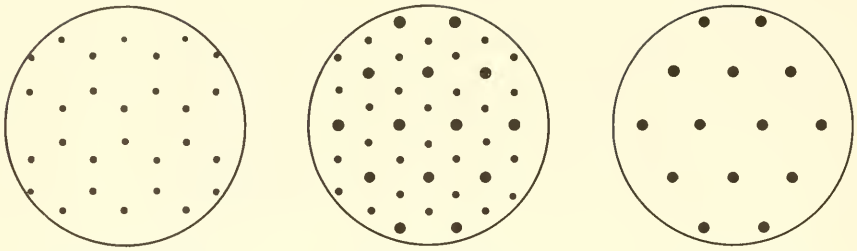


FIG. 34. A diagrammatic representation of the results of electron microscope studies of sections through various parts of a striated muscle. From these sections, the model given in Fig. 33 was deduced.

Some of those working on muscle already feel that there is likely to be something incorrect about the x-ray results, which showed no change of configuration of the muscle proteins during a twitch. Clearly, much remains to be done in this field. It is of particular interest to biophysicists because, since the structural and chemical aspects have been reasonably established, it is likely that the essential missing features of muscles will turn out to be encoded in some esoteric aspects of the physics of molecules and their aggregates. These parts of biophysics belong to a more advanced presentation than we can give here.

CHAPTER 7 *Methods for Determining Molecular Size and Shape*

INTRODUCTION

1. *A survey*

One of the important aspects of biological studies is the determination of the size, shape, and molecular weight of molecules associated with various biological functions. Until the advent of the electron microscope, these properties could be determined directly only if the molecules were larger than, say, 0.4 micron, so that they could be studied in the light microscope. As a result, indirect methods of determining particle size and shape were developed. Even now that the electron microscope has made it possible to obtain visual images of particles as small as 1 to 5 mμ, there are still many instances in which the less direct methods are more practical than electron microscopy. And there are still other instances in which the indirect methods are the only ones which can give the desired information. Furthermore, some of these indirect methods operate on a fineness of structure still inaccessible to the electron microscope. For all these reasons, therefore, it is advisable and worth while to present a survey of such biophysical methods. After presenting these surveys, several of the methods will be presented in more detail.

(a) *Kinetic methods*

Diffusion. If a particle moves through a liquid, its movement will depend on its size, shape, and mass. Therefore, by following the movement, deductions may be made about these properties. If there is no net force acting on the particle, it moves only by thermal agitation in a random fashion. The bigger it is, the less mobile it will be. Therefore the average displacement of such particles will be a measure of the size and shape. If the shape chances to be spherical, it can be shown that the actual diameter can be deduced. If the shape is other than spherical, the observation of the displacement must be coupled with some other measurement to enable the investigator to deduce the actual size and shape.

For a spherical particle, the relationship that is used is

$$\overline{x^2} = \frac{4tkT}{6\pi\eta r},$$

where $\overline{x^2}$ is the average of the square of the displacement of the particles from their starting point, k is the molecular gas constant, t is the time in seconds at which the displacement is x, T is the absolute temperature, η is the viscosity of the medium (to be measured or looked up in handbooks), and r is the radius of the particle.

Sedimentation. If a gravitational or centrifugal force is applied to a particle in solution, the particle will accelerate until its velocity is such that the accelerating force is just balanced by the frictional force. A measurement of this terminal velocity then gives a relationship between the mass, size, and shape of the particle and the frictional force. If the shape is spherical, it is possible to obtain a direct measure of the diameter and the mass of the particle. If the shape is not spherical, one obtains only a relation between the mass and the diameter, and information must be added from other methods to permit the knowledge of the mass m and diameter separately. The relationship for spherical particles is

$$\frac{mv}{F} = \frac{m(1 - d_L/d_p)}{6\pi\eta r},$$

where the symbols are the same as in the relation for the diffusion method, F is the magnitude of the applied force, d_L is the density of the liquid solution, d_p is the density of the particle, and v is the velocity of the particle in cm/sec.

In both the diffusion and sedimentation methods, it is necessary to measure the position of the particle at various times. Since the particles of interest are normally invisible even in the light microscope, it is necessary to use some clever method for determining these properties. This is normally done for diffusion by carefully putting two solutions together at a sharp, flat interface. One solution contains the particles of interest, the other is identical except that these particles are missing. Under these conditions, as the particles move they advance like the front line of a parade of soldiers, as indicated in Fig. 35. By optical methods which will not be discussed here, this boundary can be made visible and its progress measured, thereby giving an average displacement for the particles.

Electrophoresis. If an electric force is applied to a particle in solution, the particle will accelerate until it reaches a constant velocity, as in sedimentation. A measurement of this terminal velocity then gives a relationship between the size and shape of the particle, the electric driving force, and the resisting force. Since the electric force is proportional to the electric charge on the particle, and since that charge is unknown, this relationship does not permit deductions concerning the size and shape of

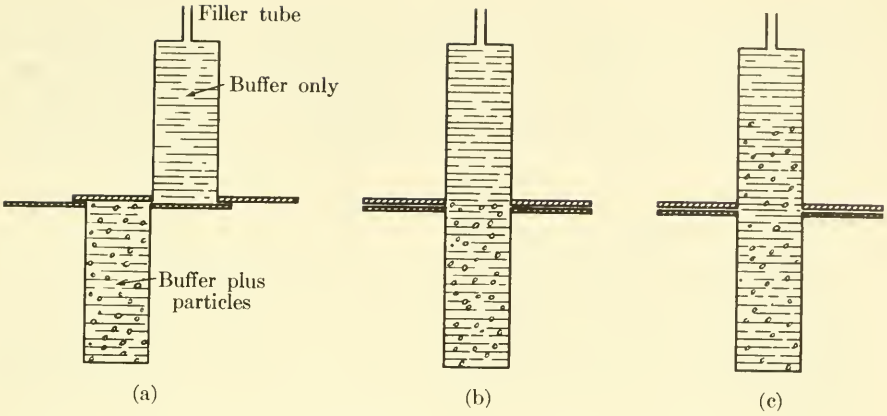


FIG. 35. The setting up of a diffusion experiment. In (a) the lower tube has been filled with a solution containing the particles whose movement is being studied. The tube has flanged ground-glass ends. A similar tube is inverted while empty onto the lower tube in a position such that the flange of one tube covers the chamber portion of the other tube. The top tube is then filled with the solution minus the particles. As shown in (b), the experiment is started by sliding the top tube until the chambers are coincident. Particle diffusion can then begin, and the particles can be found farther and farther up the tube as time progresses. By placing a light-sensitive film behind the joined tubes, and by shining light absorbed by the particles only from the front side, an absorption image of the situation can be obtained at any time.

the particle. Nevertheless, the electrically driven movement, the so-called *electrophoretic mobility,* is characteristic for each particle in a given solution, and can therefore be used to ascertain the presence or absence of the particles. Further, one can find the number of electrophoretically different constituents of unknown solutions, and such information frequently permits important deductions to be drawn. Still further, electrophoretic separations may be used to prepare pure solutions of these constituents. An important example concerns the study of blood. When blood is subjected to an electric force, a number of constituents are found in an entirely reproducible fashion. The constituents are made visible by the boundary detection method described for diffusion and sedimentation. The fastest moving substance has been isolated, and turns out to be serum albumin. The next three fastest substances all turn out to be globular proteins. In the order of speed, these have been called alpha, beta, and gamma. The ignorance of particle size and shape has not prevented investigators from finding that the entire antibody content of blood is in the gamma globulin fraction. The electrophoretic method is used to isolate this fraction for clinical use and for clinical and laboratory research.

Recent developments of the electrophoretic technique include placing the solution on long strips of special filter papers in a moist atmosphere, and applying the electric force across the ends of the paper strip. The separation of the constituents is then achieved simply by using a scissors, and the constituents may then be eluted from the paper in suitable buffers.

(b) *Static methods*

If the particles are stationary, then the physical agent for extracting information has to move into and out of (or at least away from) these particles. Thus, we are led to examine the irradiation of particles by visible light, ultraviolet and x-radiation, and various charged elementary particles.

Visible light scattering. If a particle is very tiny, light waves scattered off its front and rear will be so nearly in the same phase that they scarcely interfere with each other in the sense of waves reinforcing or cancelling each other. If, as indicated in Fig. 36, the particle is comparable in size to the wavelength of light being used, then the wave scattered

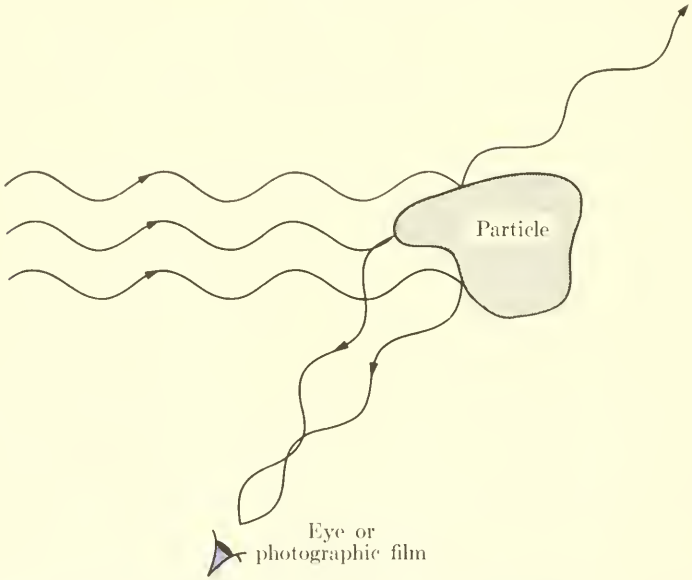


FIG. 36. A light beam whose wavelength is comparable to the dimensions of the scattering particle is reflected in various directions by the particle. The bottom wave travels about a half wavelength more than the middle wave before being reflected, so that it is possible to find a direction, as indicated, in which the eye or a light-sensitive film will record that the reflected beams arrive out of phase, so that there is a cancellation of beams. Other directions could be found in which the beams arrive so that they reinforce each other.

from the front and rear can produce appreciable, and measurable, interference effects whose magnitudes depend on the angle A from which the scattered light is viewed. By measuring the interference for all angles of viewing, it is possible to make significant deductions about the size and shape of the particles even though they may be too small to be visible, even in the light microscope.

A question that has been studied by light scattering is whether heating of DNA separates the two Watson-Crick strands from each other. In one set of measurements it was found that the volume of the scattering molecules was unaltered by heating to temperatures previously believed to be high enough to separate the strands completely. If the separation had occurred, the volume of the individual scattering molecules would have been halved.

Light absorption. Chemical bonds, such as C—H, C—O, C=O, etc., have specific absorption regions in the infrared. Series of atoms with alternating double and single bonds, such as —C=C—C=, absorb at wavelengths which increase progressively with the length of the series. The shorter series absorb in the ultraviolet. Among the materials which absorb in that region of the spectrum, the biologically important ones are proteins and nucleic acids, which contain molecules with short series of these alternating bonds.

Light of the infrared and ultraviolet regions mentioned above, when absorbed by a solution of molecules, will be absorbed in proportion to the concentration of the molecules. Thus instruments designed to measure this light absorption, called *colorimeters* or *spectrophotometers,* can give information as to concentrations, but not as to size and shape.

X-ray absorption. In contradistinction to light absorption, the absorption of x-rays can be used for studies of size and shape. When an x-ray photon passes through a molecule, the properties are such that there is little likelihood that it will actually interact with any molecules. Therefore the chance of interacting is strictly proportional to the number of opportunities for interaction, and therefore to the thickness of the molecule. In addition, the greater the cross-sectional area of the molecule, the greater the probability of the photon's passing through it. Therefore the total probability P of interaction to produce a biological effect is proportional to both factors, i.e., to the thickness t and the cross section A of the molecule:

$$P \propto t \cdot A = V,$$

where $V = tA$ = the volume of the molecule.

Thus, by measurement of the rate at which biological entities are affected by x-rays, it has been possible to make an estimate of the volume of the molecules involved.

2. *Diffusion*

It is well known that thermal motion effects a homogeneous distribution of molecules in solution if enough time elapses for the equilibrium state to be reached. The Brownian motion is the visible result of thermal motion, being due to numerous impacts by molecules on a visible particle; the actual motion observed is the resultant of the collisions occurring practically simultaneously. It should be plausible that a molecule's motion will be inversely related to its size. Since thermal equilibrium means equal average molecular energies, the bigger molecules will have smaller velocities. Thus, if we could follow the individual molecules well enough to measure their average velocities, we could compute their mass from the average thermal energy, using the formula

$$\text{average thermal energy} = \tfrac{1}{2}\ \text{mass} \times (\text{velocity})^2$$

or

$$E_{\text{therm}} = \tfrac{1}{2} Mv^2.$$

In practice, however, this approach is unworkable, because the molecules of interest are not large enough to be seen in the light microscope. Indeed, if they could be seen, their masses could be estimated from their volumes and some reasonable estimate or measurement of their densities. Thus, when we wish to obtain information about masses of invisible molecules, we have to resort to various stratagems. In the section on centrifugation, it is shown how the information can be obtained by using high speed rotors. Here we indicate how similar information can be obtained from measurement of the thermal velocity.

Since we cannot make visible the motion of a single molecule, we use the device of measuring the motion of a large number of molecules. In Fig. 37 we picture a tube on its side set up with the molecules of interest in only one of the two compartments. When the compartments are slid so that molecular interchange can occur, there is an average motion of the molecules of interest, and it is this advancing boundary that gives us a handle on the measurement. In the figure we see schematically what happens when the two compartments are slid together.

In the left-hand column of sketches, we represent the concentration of solute molecules. In the right-hand column, we present the plot of concentration at various positions in the connected compartments. After a while, the sharp concentration difference becomes blurred, as molecules move towards the right. In the second column of sketches, we further indicate the positions where the concentrations are $\frac{3}{4}$ of the original concentration and also $\frac{1}{4}$ of the original concentration. As can be seen, as time goes on these positions separate more and more from each other. By measuring the actual concentration curve by optical absorption means

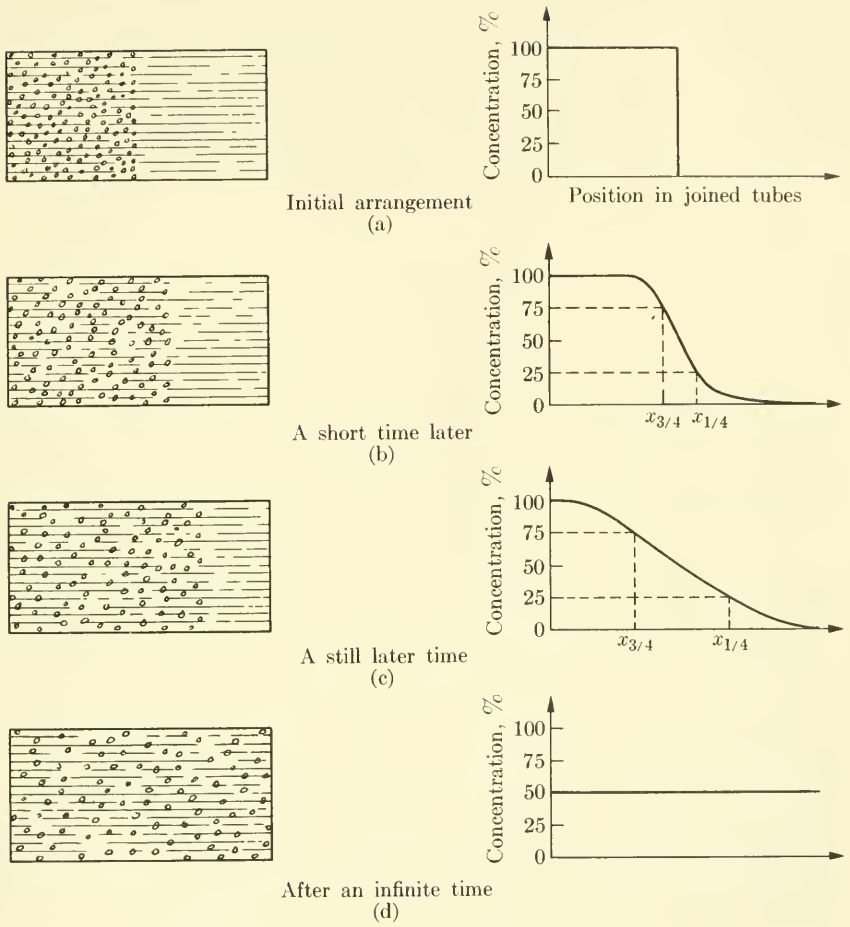


FIG. 37. The initial arrangement has all the particles in one half of the setup. The concentration, plotted on the right, is 100% up to the division point, then falls abruptly to zero. After a short time there is a net flow of particles to the right, so that the concentration now drops more gradually to zero. After more time for diffusion, the particles have reached almost to the right-hand end of the tube, and the concentration falls almost linearly to zero. After an infinite time the particles are evenly distributed through the tube which is twice as large as the starting tube, so the concentration throughout is half the starting concentration. The positions in the tube at which the concentration is 75% or 25% of the starting value are indicated. The distance between these two points is a direct measure of the time elapsed since diffusion commenced.

which we shall not discuss here, we obtain curves like those in the right-hand column. The smaller the molecules involved, the faster the curves fall to the even distribution situation in the bottom sketch. The movement of the molecules can be computed from the actual distribution of molecular velocities, which is known from physicochemical theory.

The rate of spreading of the boundary can be related to the so-called *diffusion coefficient, D,* by the approximate formula

$$x^2 = 4Dt,$$

where x is the distance between the ¾ and ¼ concentration points at time t seconds after starting.

This formula is a very handy one for biophysicists. For example, if we introduce a few enzyme molecules into a solution and these are taken up by a 10-micron cell, can these enzyme molecules possibly reach the substrates on which they act, simply by diffusing? In the table below we give typical values of D for spherical molecules of various sizes.

	Molecular weight	Diffusion coefficient, D
Glycine	75	95×10^{-7} cm^2/sec
Cytochrome C	1.3×10^4	10×10^{-7}
Hemoglobin	6.8×10^4	6.2×10^{-7}
Urease	4.8×10^5	3.5×10^{-7}
Southern bean mosaic virus	6,000,000	1.34×10^{-7}
Bacteriophage T2	200,000,000	3×10^{-8}

A typical enzyme might have a molecular weight of 100,000, and we see that its diffusion coefficient D is therefore about 5×10^{-7} cm^2/sec. Putting this in the formula, we obtain the value of t, the time taken to diffuse just once across the cell:

$$t = \frac{x^2}{4D} = \frac{(10^{-3})^2}{5 \times 10^{-7}} \times 4 = 0.5 \text{ sec.}$$

Thus, this enzyme molecule will diffuse in a waterlike medium so fast that it could travel back and forth across this large cell two times per second, and therefore could well be able to catalyze reactions anywhere in the cell where the substrate happened to be.

To return to the problem of estimating molecular weights, it is necessary to have a connection between the diffusion coefficient and the mass of the molecule. Advanced physical theory gives us the result that

$$D = \frac{kT}{6\pi\eta r},$$

where k is the gas constant per molecule, T is the absolute temperature, η is the viscosity of the solution, and r is the radius of the molecule.

The temperature and viscosity of the solution are obtained by standard measuring instruments, so that the radius of the molecule can be found if D is known. This formula holds only for spherical molecules. For nonspherical molecules, as indicated in the section on centrifugation, we combine data on sedimentation and diffusion to obtain the desired information.

For spherical molecules, we replace r in the formula by its equivalent in terms of molecular weight, using the relationship

$$M = d_p V = \tfrac{4}{3}\pi r^3 d_p.$$

Thus the diffusion coefficient varies inversely as the cube root of the molecular weight. This means that diffusion measurements are quite insensitive to variations in molecular weight, so that other methods should be used for more than approximate values of the molecular weight. That this conclusion is borne out experimentally may be seen by looking at the table given above. There we see that the diffusion coefficient changes from 95×10^{-7} to 0.3×10^{-7}, while the molecular weight changes from 75 to 2×10^8. That is, a 300-fold change in diffusion coefficient corresponds to a 3 million-fold change in molecular weight.

3. *Centrifugation*

There is a radially directed force on a particle which moves along a curved path. If the particle is within a container filled with a liquid which is less dense than the particle, the particle will move radially outward; if the liquid is more dense than the particle, the particle will float inward. As indicated in Fig. 38, the important physical parameters are

r—the radius of circulation of the particle p,
ω—the frequency of rotation of the centrifuge tube, in radians per second,
d_p—the density of the particle, and
d_L—the density of the liquid.

A centrifuge is a device for spinning tubes containing particles dissolved in a liquid. Centrifugation is used currently in five different ways in biology.

(a) By centrifuging a suspension of particles for a sufficiently long time in a liquid that is less dense than the particles, the particles are packed into a pellet at the bottom of the tube. They are then collected by decanting the supernatant fluid. We say that the particles have been *pelleted* or *sedimented*.

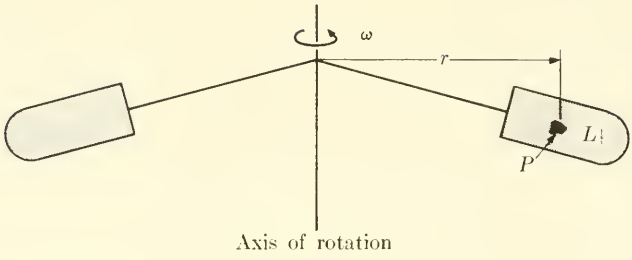


FIG. 38. Tubes spinning about an axis of rotation with an angular velocity ω and with the particle of interest at a distance r from the axis. The particle is at P and is suspended in a liquid L.

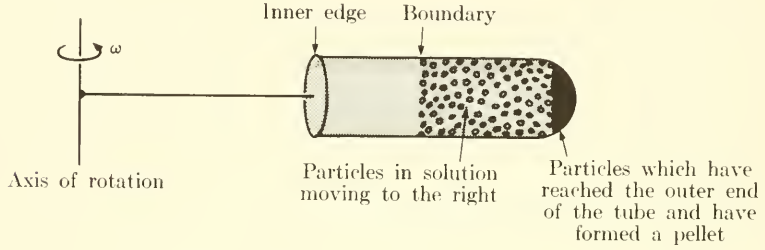


FIG. 39. A representation of the situation some time after a tube containing particles denser than the suspending liquid has been started spinning in a centrifuge. The particles which were at the inner edge at time zero move together to the right and thus form a boundary whose movement can be photographed for determination of the velocity of the movement.

(b) The rate at which a particle will move when being spun at a given frequency ω at a given radius r will depend on its density, the density and viscosity of the liquid, and on the size of the particle, as will be demonstrated below.

The individual particle cannot be studied, but the particles at the *inner* edge of the tube form a boundary which moves radially outward as a result of the centrifugation, as indicated in Fig. 39. This boundary can be made visible by optical methods, so that the rate of displacement of the particles can be deduced by photographing the transparent tube at various times after starting the centrifuge. Since the density and viscosity of the liquid can be measured, it is possible to compute the ratio of the mass of the particle (its molecular weight) to its size. If the particle is spherical, the molecular weight can be deduced directly. If it is not spherical, a somewhat more complicated procedure is required which gives us the size of the particle by measuring the diffusion of particles through a liquid. Then, by a combination of these measurements, the molecular weight can be deduced. The mathematics of this method will

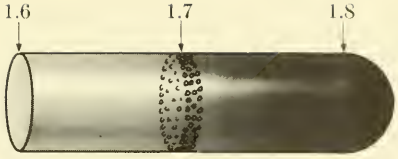


FIG. 40. The situation in a CsCl density gradient in which the solution density varies from 1.6 at the inner edge of the tube to 1.8 at the outer edge of the tube. The point of density 1.7 is indicated, and particles of that density will be found in a band at that point. If particles were to the left of the band, they would be more dense than the solution and thus would be moved to the right by the centrifuging force. If to the right of the band, they would be less dense than the solution and thus would be floated inward to the left. Thus, after many hours of spinning the mixture of particles and CsCl solution, the situation sketched will be obtained.

be presented below. This is called the method of the *sedimentation velocity.*

(c) If the centrifuge is run for a sufficiently long time (days), all the particles will tend to be at the bottom of the tube. But, due to the Brownian (heat) motion, some of the particles will actually go backward toward the axis of rotation. When these two processes (centrifugal sedimentation and backward diffusion) have come to equilibrium, there will be a gradient of particle concentration from essentially zero at the inner edge of the tube to a maximum at the outer edge. This gradient can be made visible by the optical methods referred to above. The bigger the particles, the more they will tend to be concentrated near the outer edge of the tube. From a measurement of the actual concentration gradient, then, the molecular weight of the particles may be directly deduced. This is called the method of *sedimentation equilibrium.*

(d) If a low molecular weight salt were in solution, the sedimentation equilibrium method would result in a very gradual gradient. Indeed, it should be obvious that a salt could be chosen with so low a molecular weight that it is hardly sedimented at all. Table salt (NaCl), for example, would hardly be expected to be recovered from a solution by this method. But it will in fact give a measurable gradient. In practice, for example, using a certain concentration of CsCl instead of NaCl, we can set up a salt gradient so that at the inner edge the density is 1.6 gm/ml and rises essentially uniformly to about 1.8 gm/ml at the outer edge. Other gradients may be obtained with other CsCl concentrations.

Now, if particles of an intermediate density, say 1.7 gm/ml, are placed in the solution before centrifuging, they will all end up in a band at their own characteristic density, as indicated in Fig. 40. The reasoning is that when the salt equilibrium has been established, those particles finding

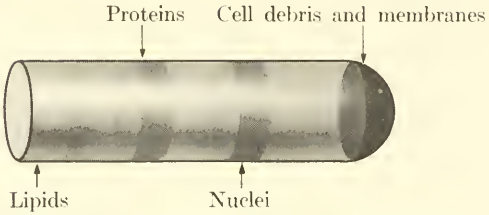


FIG. 41. A representation of the results of a centrifugation of a mixture of crushed cells layered onto the inner edge of a sucrose gradient which increases to the right. The fastest moving substances are cell membranes and internal membranes, which are represented as just having reached the outer edge of the tube. The next most rapidly moving parts are the nuclei, followed by the proteins, with the lipid material still floating at the starting place.

themselves to the left of the place where the density is 1.7 gm/ml will be forced centrifugally outward until they reach that place. Particles finding themselves to the right of that place will be less dense than the solution, and will therefore be floated inward until they reach the same place. Again, the density gradient and the particle band can be photographed by suitable optical methods, so that the particle density can be measured.

There is a further result from this *density gradient sedimentation* method. Because of the thermal (Brownian) motion of the particles, they will not be precisely in a thin band but will be in a band of some finite thickness. Indeed, the smaller these particles, the faster and further they will go at any given temperature. Thus it should be plausible that the width of the band is related inversely to the size of the particles: the bigger the particles, the smaller the band. This proposition can be demonstrated mathematically to be true, and therefore the molecular weight of the particles can also be obtained from the same photograph. This method then gives both molecular weight and density from a *single* photograph!

(e) Density gradient sedimentation can be used in another very fruitful way. The density gradient is first established (more usually with sugar than with salt, for technical reasons) and *afterwards* the solution to be studied, containing a mixture of particles, is placed on top of (at the inner edge of) the centrifuge tube. When the machine is then run for a while, the various particles in the solution will have reached different points in their travel towards the other end of the tube. If the centrifuge is stopped, the different kinds of particles will be found to have been separated into bands, as sketched in Fig. 41. In the sketch, we have assumed that some cells have been broken and that the method has separated the solution into the components indicated. By cutting the tube carefully, the various components can be separated from each other.

The purpose of the sugar (or salt) is to stabilize the contents so that minor shaking or convection will not remix the separated components. The viscosity of the sugar solution contributes greatly to the stabilization. Ordinary buffer solutions are not stable enough. This *density gradient separation* method gives no absolute quantitative data, but does give quantitative separation of components.

AN EXAMPLE OF THE USE OF CENTRIFUGATION IN BIOLOGY

One of the fundamental problems of genetics is whether the nuclear material, especially the nucleic acid, stays in one piece during replication. In principle three different fates are possible for the nucleic acid. First, it could always remain intact. This would imply that the new nucleic acid which goes to one of the daughter cells after mitosis is built by copying the parental nucleic acid without disrupting it at all. Since such a mechanism entirely conserves the parental molecules, it is called the *conservative* method of replication.

Second, the Watson-Crick model of DNA affords the possibility that the nucleic acid comes apart into just two pieces, which are separately replicated by pairing with the proper components. Thus each daughter cell would receive exactly half of the parental nucleic acid, in one single piece. If this model is correct, then in a second generation, resulting in four daughter cells, two would contain one each of the parental half-strands; the other two cells would contain none of the parental nucleic acid atoms. This scheme conserves half of the parental molecules of nucleic acid and is therefore called *semiconservative.*

Third, the nucleic acid might come apart into a number of pieces which are incorporated at random into various maturing DNA strands by some unknown mechanism. In this scheme the parental nucleic acid molecules are dispersed among the daughter cells and the method of replication is therefore called *dispersive.*

A combination of isotope utilization and density gradient centrifugation has permitted an unequivocal decision about the mechanism for bacteria. The bacteria in question were grown in a nutrient medium containing much heavy nitrogen (N^{15}) instead of normal nitrogen (N^{14}) and heavy carbon (C^{13}) rather than normal carbon (C^{12}). Thus the nucleic acids of such cells would be more dense than the nucleic acids of cells grown in a normal medium.

In this experiment, cells were grown for many generations in the dense medium and then transferred to a normal medium. Samples of cells were taken immediately upon transfer to the normal medium and at various times thereafter; the DNA of the sample cells was extracted and run in a CsCl density gradient. The results are shown in Fig. 42.

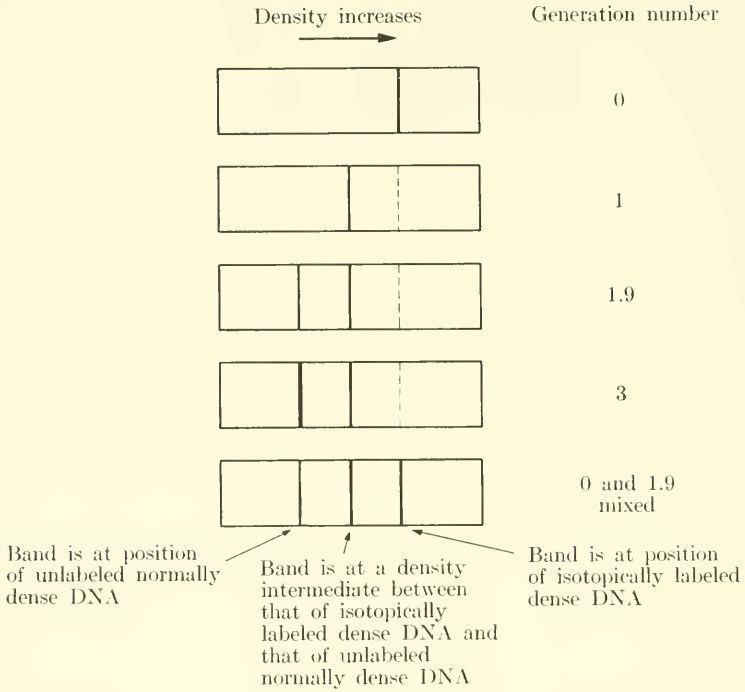


FIG. 42. Schematic drawings of photographs of DNA extracted from isotopically labeled dense organisms at various times after organisms were placed in unlabeled normally dense medium. The DNA was placed in CsCl and banded as explained in Fig. 40. Since the DNA absorbs highly specifically and strongly at 260 mμ, the band positions can be photographed by absorption of this light, as explained in Fig. 35.

In this sketch, each section represents a photograph of the centrifuge tube taken so as to show materials which absorb ultraviolet light, since nucleic acids absorb that part of the spectrum much more strongly than other biological materials. The single band at generation zero is due to the dense nucleic acid. After one generation, there are no molecules of the parental density, demonstrating unequivocally that a completely conservative mechanism of replication does not exist. The density of the nucleic acid molecules found after one generation is actually midway between that of the dense parental molecules and that of molecules of cells grown in a normal medium, as can be seen by looking at generations 1.9 or 3.0, in which most of the molecules are from cells which were formed after the transfer to a normal growth medium. It should be noted that at all times there are molecules of this intermediate density. Thus we see evidence that molecules which have obtained half their nucleic acid from the original parents and half from new normal density mate-

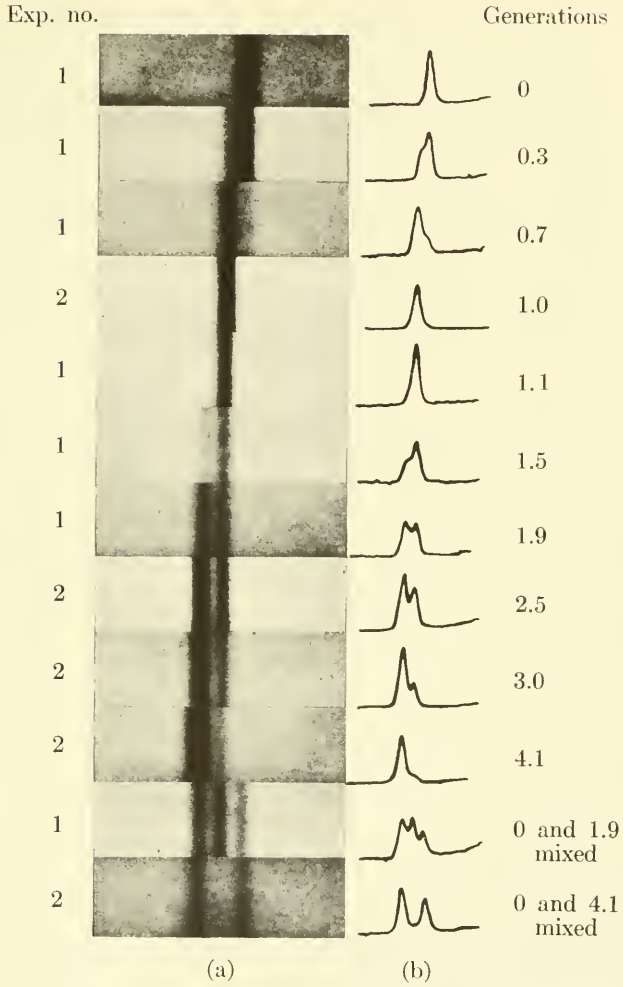


FIG. 43. The actual figure on which Fig. 42 is based. (From Meselson and Stahl, *Proc. Natl. Acad. Sci., U.S.*, **44**, 671, 1948; courtesy the authors and National Academy of Sciences.)

rials always persist. After the parental molecules have separated into two pieces, they do not separate further; there are no molecules of any density intermediate between these "half-dense" molecules and normally light molecules. Accordingly, we conclude that a semiconservative mechanism of replication is operating, and the Watson-Crick model has yet another piece of supporting data. For your study, the actual data published by the researchers is presented in Fig. 43.

It should be remembered that this evidence concerns only bacterial DNA. It is possible that bacteria which have no nucleus of the conventional type (there is no nuclear membrane and no evidence for mitosis) are a special case. Students should be cautious in extrapolating this result to all organisms.

THE MATHEMATICS OF THE SEDIMENTATION VELOCITY METHOD

Quantitative data are most frequently obtained by the use of the sedimentation velocity method. The essential features of the calculation may be derived from an analysis of the forces involved.

If a centrifugal force acts on a particle in solution, the particle will move if its density differs from that of the solvent. The particle will speed up but, in accelerating, the frictional resistance will be increased because (for low speeds) the resisting force is proportional to the particle speed. Thus the particle will accelerate until the radially outward centrifugal force is just balanced by the radially inward frictional force. From that point on the particle will move with constant speed outward, since there is no net force acting on it. This force equilibrium point may be written as

$$F_c = F_r,$$

where F_c is the net centrifugal force and F_r is the resistance force. As already mentioned, F_r is proportional to the velocity v:

$$F_r = fv,$$

where f is the proportionality constant which is known as the friction coefficient.

The centrifugal force is given by $F_c = mv^2/r = m\omega^2 r$, where m is the effective mass of a particle spinning at a rate of ω radians per second at a distance r from the axis of rotation. The effective mass is the difference between the mass of the particle and the mass of an equal volume of solvent liquid (if the particle had the same density as the solvent liquid, there would be no resultant motion):

$$m = m_p - m_L = m_p\left(1 - \frac{m_L}{m_p}\right) = m_p\left(1 - \frac{d_L}{d_p}\right),$$

where d is the density, and the ratio of the masses is replaced by the equivalent ratio of the densities. The force equilibrium condition may now be written as

$$m_p(1 - d_L/d_p)\omega^2 r = fv$$

or, alternatively, as

$$\frac{v}{\omega^2 r} = \frac{m_p(1 - d_L/d_p)}{f}.$$

The left-hand side of this expression is defined as the sedimentation coefficient s. It is in terms of s that many biological materials are described, since it is equivalent to stating the mass or molecular weight of the particle. Since s is a ratio between the velocity attained by a particle and the acceleration $\omega^2 r$ given it by the particular machine used, it will be the same for all centrifugation machines. Also, the right-hand term must be independent of the particular machine used, and it can be seen that it depends only on the properties of the particle and the solvent liquid.

To proceed further, we must, in effect, measure f. The usual procedure is to utilize a relation first derived by Einstein:

$$f = \frac{kT}{D},$$

where T is the absolute temperature, k is the molecular gas constant, and D is the so-called diffusion coefficient, which can be determined from the distance particles diffuse (by Brownian motion) in a measured time interval. Thus, by using Einstein's expression for f, we obtain an expression for s which depends on measurable quantities:

$$s = \frac{m_p(1 - d_L/d_p)D}{kT} = \frac{Nm_p(1 - d_L/d_p)D}{NkT},$$

where we have multiplied numerator and denominator by N, Avogadro's number. Here Nm_p is the molecular weight M of the particle and Nk is the more usual form of the (molar) gas constant R, as used in the gas law, $pV = RT$. Thus the final expression for the sedimentation coefficient is

$$s = \frac{M(1 - d_L/d_p)D}{RT}.$$

In practice, s, T, and d_L are obtained in one experiment and D in a separate experiment. The particle density d_p may be obtained by centrifuging the particles in liquids of various densities to determine the density at which the particles do not move; this liquid density must then equal the particle density. Thus, finally, we can obtain the molecular weight, which is the object of all this work and analysis.

For spherical particles, we have indicated that the diffusion coefficient, D, is inversely proportional to the cube root of the molecular weight.

Thus the sedimentation coefficient containing the product MD is proportional to the two-thirds power of the molecular weight. Accordingly, the sedimentation velocity method is much more sensitive to changes in molecular weight than is the diffusion method. Some typical sedimentation coefficients are given in the table below.

	Molecular weight	Sedimentation coefficient
Botulinus toxin	950,000	17
Southern bean mosaic virus	6,300,000	115
Bacteriophage T2	200,000,000	900

The unit in which the sedimentation coefficients is given is the *svedberg*, which is 10^{-13}/sec. This is the standard unit found in the literature.

4. *X-ray diffraction*

When x-rays impinge upon matter, most of the rays pass through unaltered and undeflected, if the matter is not very thick. Some of the rays, however, are scattered, and the pattern of scattering depends on the pattern of the atoms and molecules making up the piece of matter. Given the structure of the piece of matter, the scattering of the radiation can be computed, but given the pattern of the scattered radiation, it is much more difficult, and in some cases impossible, to deduce the structure of the scattering material. We shall first discuss the use of x-ray diffraction methods for determining the structures of crystalline materials. The usual experimental procedure yields a photograph of a pattern of lines or spots of varying spacing and intensity. The spacings will be shown to be related to the distances between the molecules (or atoms) in the crystal. Analysis of the intensities of the lines or spots can be made to yield the detailed structure of the (polyatomic) molecules themselves. The mathematical techniques needed to achieve these results are far beyond the scope of these writings, but an attempt will be made to suggest the essential factors involved in the methods.

We start with a standard diagram (Fig. 44) showing a beam of monochromatic x-rays hitting and bouncing off a crystal which is represented as an array of atoms, indicated by circles. The wavy lines are the paths of two of the rays making up the x-ray beam. The parallel lines are to be thought of as planes perpendicular to the paper and passing through the atoms as shown. Thus we are really talking about a three-dimensional crystal of which we see a plane section. The angle θ is that between the direction of the incident beam and the plane of the atoms.

The two rays are shown bouncing off atoms in such a way that the path of wave II becomes the same as that of I. This second ray travels

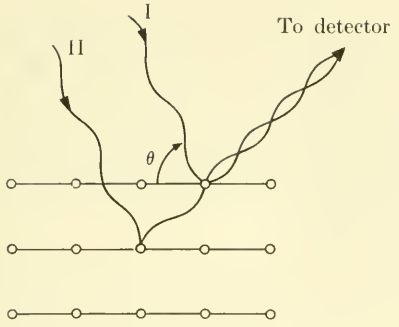


FIGURE 44

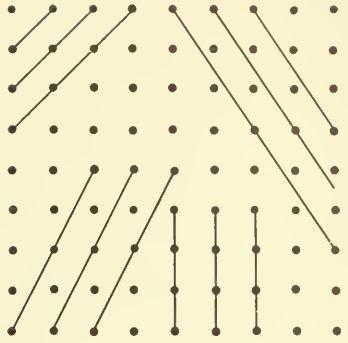


FIGURE 45

farther than the first ray by the amount indicated by the heavy line. A detector (such as a photographic film) into which these rays go will yield one of three possible results:

(a) If the extra distance is just equal to a whole wave, then the two rays arrive at the detector in the same phase of their motion (that is, crests are with crests, and troughs with troughs). They add up to give an intensity greater than that from a single ray.

(b) If the extra distance is just equal to half a wave, then the two rays arrive at the detector exactly out of phase (the crest of ray I would be superposed on the trough from ray II) and the rays cancel, giving zero intensity.

(c) If the extra distance is neither of these two alternatives there will also be a cancellation, because in actuality the detector receives many rays, not just two, and except for the two cases presented the rays will be at all phases of the motion with respect to each other and there will be an over-all cancellation.

For any spacing between planes, there will be precisely one angle θ which causes the rays to be exactly one wave apart, so that they reinforce each other; at any other angle general cancellation will occur.

Bragg was the first to point out that arrays of atoms in a crystal could be thought of as forming planes in the way shown. This is clearly an approximation, but the exact treatment of the problem by advanced methods yields precisely the same criterion for reinforcement as Bragg obtained from his simple model.

The relation for the condition of reinforcement is

$$\lambda = 2d \sin \theta,$$

where λ is the wavelength of the x-rays used and d is the spacing between planes. As will be indicated below, λ and θ are usually measured, which permits the computation of d.

There are in fact many possible sets of parallel planes that can be formed by arrays of atoms, as indicated in Fig. 45. Each of these sets of planes will produce reinforcement of incident x-ray beams scattered off them, provided only that the planes are at the proper angle to the incident beam. It should be readily appreciated that some planes have many more atoms than others, and since the over-all intensity obtained depends on the density of atoms in the planes, the more "natural" planes, in practice, yield the most intense reinforcements.

In general, the orientation of a single crystal will not be such as to produce reinforcement, but there are several means of obtaining the proper orientation. We shall first discuss the so-called *powder method.* As the name of the method implies, a crystalline material is first smashed into small pieces, and the resultant crystal powder is placed in the path of the x-ray beam. Since there are now very many small crystals, in essentially all possible orientations with respect to the beam, many will chance to be in the precise orientation needed to give ray reinforcement. These correctly oriented crystals will not all be parallel to each other, since a properly oriented crystal can rotate about the beam and still maintain its proper angle. As shown in Fig. 46(a), the result will be a reinforcement sufficient to produce circles of blackening of a photographic film.

If a strip of film (indicated by the dashed lines in the sketch) is used instead of a very large piece of film, the positions of the reinforcements will be given by the (somewhat curving) lines at various distances from the intense blackening produced by the undeflected x-ray beam. Such a strip is shown in part (b) of the sketch; it is the experimental result frequently presented in research papers.

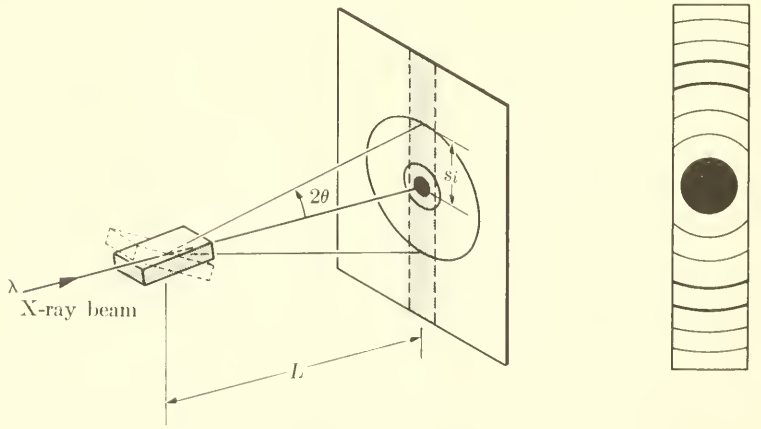


FIGURE 46

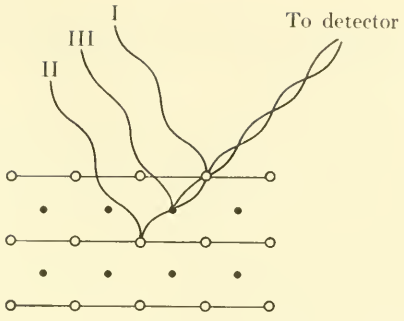


FIGURE 47

Consider the three measurable parameters indicated in the sketch:
L, the distance from the crystal to the film;
s_i, the distance from the undeflected beam to the individual lines;
λ, the wavelength of the x-rays.
Since in practice the angles are always small,

$$\sin 2\theta \approx 2 \sin \theta \approx \frac{s_i}{L}$$

and

$$\lambda = 2d \sin \theta \approx 2d \frac{s_i}{2L} \approx \frac{ds_i}{L}.$$

Thus, as indicated earlier in this section, the spacing d can readily be computed from the measured quantities. For the simple monatomic crystal we are discussing, it is only moderately difficult to relate the calculated spacings to the distances between atoms, but it is beyond the scope of this presentation.

It remains to indicate how the intensities can yield the structures of molecules placed at the crystal positions. To this end we return to Fig. 44 and, as in Fig. 47, insert atoms midway between the planes. It is true that there will now be some new powder lines, but we shall neglect them for the present and consider only what happens to the already existing powder lines. In the case presented in the sketch, the extra distance traveled by ray III is just one-half that for ray II. Since the latter distance is exactly one whole wave more than for ray I in the case of ray reinforcement, ray III travels exactly *one half wave* more than ray I. Thus it cancels ray I because its troughs are superposed on the crests of ray I, and vice versa. In this instance, then, the line will be missing. Now, if the extra atoms are not placed so that their contribution is precisely one half wave different, then there will not be a total cancellation of ray I. In general, then, the insertion of the extra atoms will diminish

some of the line intensities. It should not be difficult to construct the argument leading to the conclusion that other lines could actually be enhanced.

Now imagine that these extra atoms are moved so that they touch the original crystal atoms; the result will still be increases and decreases in the intensities of some of the lines. When two atoms are close together at a crystal position, the situation is entirely equivalent to having a diatomic molecule at that position. Accordingly, we have deduced that the effect of making a crystal out of diatomic molecules is to produce a pattern of lines that is the same as that of a single atom crystal, but with a different set of intensities. Continuing in this vein, we argue that if we place polyatomic molecules at the crystal positions, once again the basic pattern of lines is obtained, but the intensities will differ in a very complicated fashion. From the lines and their intensities, it should be possible to make deductions about the structures of the polyatomic molecules.

To say this more directly, if we make crystals of enzymes (or any other polyatomic molecules), we obtain patterns of lines and intensities which apparently can be used to deduce the three-dimensional structure of the molecules. We say *apparently* because in practice the difficulty increases enormously as the number of atoms in the molecules increases. For biologically important molecules, the unraveling of the structures is presently a matter of years, even utilizing the fastest of modern computers for the mathematical work and using some important new devices developed for such studies.

It is not always possible, even theoretically, to deduce the structures of molecules from the intensities. The theoretical difficulty stems from the so-called *phase problem*, which we shall now attempt to picture for you.

Consider part (a) of Fig. 48. We have indicated that the scattering from the two black atoms is such that crests are leaving them at the particular instant at which a trough is leaving the uncolored atom, which we will use as the reference atom. In part (b) the black atoms have been moved so that they are emitting troughs, but you will notice that at their old positions, indicated by the dotted circles, crests are present. Therefore the blackening on the film will be the same in both instances. Thus we have indicated that a given blackening can be produced by two (and of course there are many other possibilities) different configurations of the atoms within the molecule. If we knew the relative phases of the scattering (whether crests or troughs are being emitted when the reference atom is emitting a trough) we could deduce the structure directly. But since we don't know the phases, we can't know the positions of the atoms.

In parts (c) and (d) of the figure we show a way of getting around this difficulty. The molecule will be part of many different planes and,

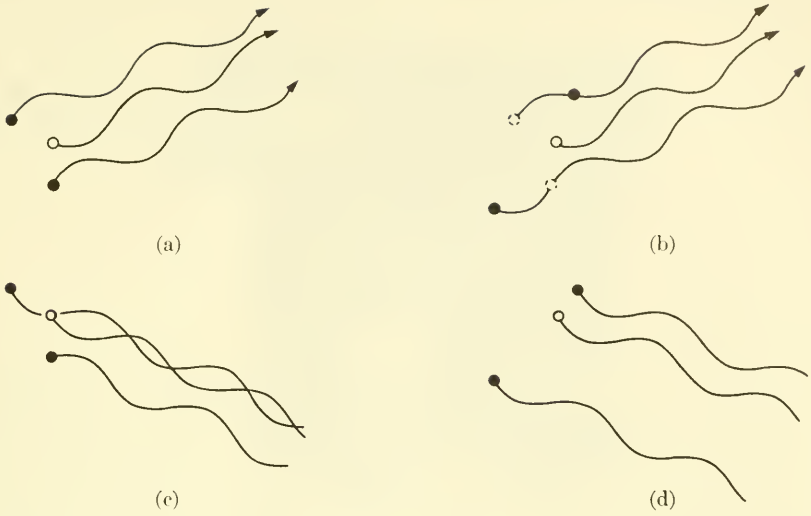


FIGURE 48

depending on where they are, the contributions of the two atoms differ widely. Thus, by combining the results of several different reinforcement situations, it is possible to deduce the correct structure of molecules containing a few atoms. However, if the number of atoms in the molecule is very large, as it is for almost all biologically important molecules whose structure is unknown, it is not feasible to deduce the structure from the observed intensities alone.

We now turn briefly to a method other than the powder method for obtaining properly oriented crystals. If a single crystal is placed in the path of the x-ray beam, it will not in general be at the proper angle to give reflections. But if the crystal is rotated slowly, it will at some instant be in the proper position to give reinforcement for a particular set of planes, and the x-rays will form a spot on the photographic film. When rotated to a somewhat different position, still another set of planes will be at the proper angle for reinforcement, and a spot will be formed at a different place on the film. Thus, by rotating the crystal, a set of spots of various intensities is obtained. The analysis is similar to that for the powder method, in that the distances between spots are related to the crystal plane spacings and the intensities of the spots are related to the structures of the (polyatomic) molecules at the crystal positions. The materials of biological interest usually do not lend themselves to the powder method, and other methods, especially rotating crystal methods, are usually employed.

As pointed out in the first paragraph of this section, it must be realized that, given the structure, the x-ray diffraction pattern is com-

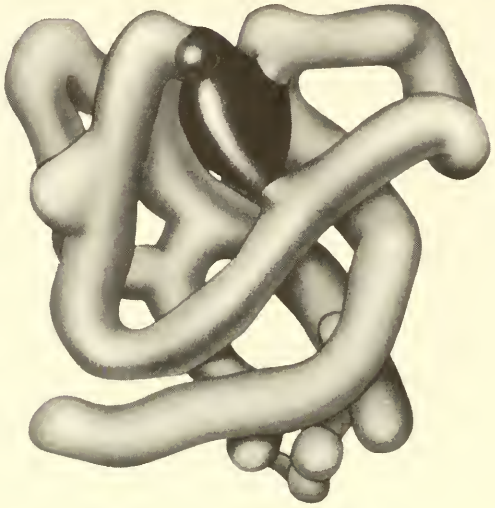


FIG. 49. Drawing of a model of the myoglobin molecule. The model shows the shape of the chain seen at 6 A resolution. The dark disc-shaped region is the haem group.

putable in a straightforward way. Thus if the structure of the molecule being studied is either known or suspected, the pattern can be computed and compared with the experimental pattern. A skilled worker can use the comparison to make reasonable guesses about what modifications may be needed to make the theoretical pattern agree better with the observed pattern. Thus, by a series of trial and error modifications, it is frequently possible to obtain a structure which fits the data very well. The phase problem still exists, but known bond lengths and angles and auxiliary information from infrared spectroscopy and other chemical means can frequently decide uniquely among alternatives.

Through various clever stratagems and much hard work, the structures of several complex molecules of biological interest have been obtained in recent years. One of these is sketched in Fig. 49. The number of molecules whose structures have been determined is too small for any generalizations about three-dimensional structures to have emerged. But despite the complexities and the time needed for such x-ray diffraction studies of molecular architecture, there is rapidly increasing interest in these methods because at present the information can be obtained in no other way.

CHAPTER 8 *Isotope Methods*

INTRODUCTION

There are two kinds of isotopes: radioactive and nonradioactive. Both kinds have important roles in biological studies. They may both be used as tracers in a given substance during its uptake and metabolism by any organism, and also in test-tube syntheses. Radioactive isotopes have two further uses. First, they may be used as indicators of localization within cells or within organisms *without* having to take the organism apart. Second, their disintegration may be used to effect an alteration in the functioning or the structure of biological systems. We will give examples of all three uses later in this chapter.

Some isotopes have a special use because they occur chiefly in one or a few types of biologically important compounds. The following table gives some examples, along with the lifetimes of the radioactive isotopes.

Isotope	*Primary occurrence in*	*Lifetime (half-life)*
Carbon-14	everything	5700 years
Phosphorus-32	nucleic acids	14.3 days
Tritium-3	everything	12.5 years
Iodine-131	thyroid compounds	8.1 days
Sulfur-35	proteins	87.1 days
Nitrogen-15	most compounds	stable
Oxygen-18	most compounds	stable
Deuterium-2	everything	stable

Even though a number of these isotopes are seen to occur in many biological compounds, they may nevertheless be used in a specific way by synthesizing special compounds containing them. For instance, if thymidine is synthesized with tritium, it can be used to study nucleic acids, because thymidine is known to be incorporated almost entirely by deoxyribonucleic acid. The same isotope could be used to study RNA by using uridine which had been synthesized in the presence of tritium.

The basic relation for the survival of undecayed radioactive atoms may be obtained by methods readily available to us. If there has been

an average of d disintegrations, then the zero class of the Poisson distribution lets us write the fraction of atoms in which there has been no decay as

$$\frac{N}{N_0} = e^{-d}.$$

Since the number of disintegrations increases steadily with time, it may be written as

$$d = \lambda t.$$

Thus the final expression for the survivors of radioactive decay is

$$\frac{N}{N_0} = e^{-\lambda t}.$$

In a time equal to $1/\lambda$, the surviving fraction is e^{-1}, or 37%. For purely historical reasons, $1/\lambda$ is not used directly as the index of the average lifetime of these atoms, but rather the index is taken as the time to reach 50% survival. From the Poisson formula, this value may be shown to be related to λ by the expression

$$\tau = \frac{0.693}{\lambda},$$

where τ is the time for half of the atoms to disintegrate; it is τ, the *half-life,* which is listed in the table above.

The units in which radioactivity are measured are, basically, the number of disintegrations per second, chosen because it is proportional to the number of radioactive atoms present. Physicists chose for the unit of activity that of one gram of radium: 3.7×10^{10} disintegrations per second; this unit is called a *curie.* For biological purposes, this unit is between 1000 and 1,000,000 times too high to be practical, and you will find reference to *millicuries* and *microcuries.*

In any given experiment, the total number of disintegrations per second need not be the most important aspect, however, for one can have the same number with different total amounts of the chemical species. For example, the fraction of all carbon atoms which is radioactive could vary from 100% downward (if pure isotope were being used). The fraction of atoms which is radioactive is ordinarily expressed in terms of the *specific* activity of the substance, i.e., disintegrations per second per gram, although in biological studies it is more usual to express it in terms of disintegrations per minute per microgram of substance.

There are difficulties involved in the detection of disintegrations, and the number registered by the counting apparatus is generally consider-

ably less than the total number that actually occur. As a result, activity has to be corrected for the efficiency of the counter (by using a standard source of radioactivity), or else the results are given as counts per minute per microgram, which is entirely valid so long as only relative radioactivity levels are necessary for the conclusions of the study.

1. *Tracer experiments*

(a) One of the most significant metabolic studies using the tracer aspects of isotopes has been the determination of the pathway of uptake of CO_2 by plants. In these experiments, $C^{14}O_2$ was bubbled into chambers containing photosynthesizing algae, and the algal metabolism was stopped after various times by taking aliquots and placing them in metabolic poisons. The algae were then fractionated by standard biochemical techniques, and the amount of radioactive carbon in the various compounds was determined. In this fashion it was found that the carbon first appears in the 3-carbon molecule of phosphoglyceraldehyde. The subsequent metabolic journey could be deduced by noting which compounds contained the tracer in the next aliquot, etc. Finally, a substantial proportion of the metabolic pathway was inferred. The pathway, of course, then had to be checked by establishing the existence of the requisite enzymes and of energy available in the proper amounts and in the proper forms (i.e., in the proper energy-rich compounds) to effect the chemical conversions indicated by the tracer study.

(b) An equally significant genetic study was carried out using sulfur and phosphorus labeling of bacteriophages. At the time of this study (1952) it was not known with compelling certainty that the nucleic acids carried the genetic information for the entire organism. Hershey and Chase grew two batches of phages, one in S^{35}, the other in P^{32}. They infected unlabeled bacteria with the S^{35} phages, and at various times after infection, aliquots were taken to be placed in a blender for study. Operation of the blender was shown to strip phages from the bacterial surface. In the experiment, about 95% of the sulfur-labeled protein could be stripped at all times during the latent period. If the blenderizing was done just after adsorption of the phages (within 2 or 3 minutes) the number of cells yielding progeny bacteriophages was very small. By about 3 minutes after adsorption was initiated, the stripping had no effect on the infective process; all cells yielded progeny phages. This experiment showed that the protein is almost irrelevant (to within about 5% of the total protein), since normal phage synthesis could ensue without the presence of about 95% of the protein.

When this experiment was repeated with the P^{32} phages, the amount of strippable P^{32} decreased steadily from 100% when the phages were

first added to practically 0% after about 3 minutes of adsorption; the phosphorus after that time was always found firmly attached to, or within, the bacteria. Here again, after 3 minutes there was 100% infection of the cells. But during the first 3 minutes, the percentage of cells yielding progeny phages was strictly equal to the percentage of the P^{32} which could no longer be stripped. Thus, since the P^{32} is contained entirely in the DNA of the phages, it was deduced that the nucleic acid was all injected into the cells within the first 3 minutes, and after that time the S^{35}-labeled protein was all that remained on the bacterial surface. Thus the injected nucleic acid was able to replicate itself and the normal protein coats of the phage. The genetic information and the control of metabolic functioning were thus shown both to reside entirely in the DNA (except for the possible 5% of the protein which was not strippable).

(c) The lifetime of human red blood cells has been determined through the use of the nonradioactive isotope N^{15}. The amino acid glycine was synthesized with the isotope, and fed to human beings. This amino acid is incorporated into the protein hemoglobin, so that within a few days, there was a significant amount of N^{15} in the red blood cells of the subjects. The amount of N^{15} in the red blood cells began to show a significant decrease only after about 100 days, and reached half its peak value after about 140 days. More refined calculations, taking into account the time required to build the N^{15}-containing glycine into the hemoglobin molecule, yield a result of 127 days for the average life span of human red blood cells. The fact that the content of N^{15} drops, as expected from the complete loss of label, leads to the conclusion that very little, if any, of the hemoglobin molecule is re-utilized for the synthesis of new erythrocytes.

2. *Localization experiments*

In these experiments, the information sought is generally of a strictly qualitative nature, although some quantitative information is not infrequently obtained in addition. The most obvious experiments have to do with cytological studies of localization within cells, for if the cell is large enough, the information can be obtained by radioautography.

Radioautography involves feeding the isotope to the organism and, at suitable later times, stopping the metabolism by appropriate agents. The intact cells are used, if possible, or else the cells are sectioned by the now-standard techniques for preparation of specimens for light and electron microscopy. The material is then placed in contact with a photographic emulsion sensitive to the radiation emitted by the isotope in use. The material sits in contact with the film for a period found to be

sufficient for the emitted rays to have darkened the film adequately. The film is then developed, and the material is studied with the microscope. If developed grains occur over a particular part of the cell, it follows that the isotope is located in that part. It is necessary to focus alternately on the material and on the film to make the identification.

This technique was used by Taylor in a study of the fate of the chromosome materials during cell division. Taylor fed onion root tips with tritium-labeled thymidine for a while, and then stopped further detectable labeling by adding a large excess of unlabeled thymidine. The cells were allowed to go through controlled numbers of divisions, and after each division, the roots were sectioned and radioautographs obtained. It was found that grains appeared over both chromatids at the beginning of the experiment. After all rounds of cell division, chromosomes usually contained either one labeled and one unlabeled chromatid, or else completely unlabeled chromatids. There was some chromosome breakage, and in such instances the part of one chromatid which was unlabeled was matched precisely with an equal length of labeled sister chromatid. This experiment shows directly the doubled nature of the chromosome and that the chromatids are retained intact during mitosis except for some occasional breakage and reunion.

By similar methods it has been demonstrated that strontium is deposited almost entirely in bone, while other atoms, such as phosphorus, are much more generally deposited. The very high localization of iodine in the thyroid has been amply confirmed by these techniques. Localizations of this kind are extremely useful clinically. For example, the administration of radioactive strontium could be of use in treatment of bone diseases, and treatment of thyroid disorders by giving radioactive iodine is so standard that thyroidectomy is now very infrequent.

Radioautography is limited for several reasons, the most important of which is the limitation on the resolution due to the finite size of the grains in the photographic emulsion and to the fact that the radiations are emitted in all directions, so that to each source of radiation there is a much larger zone of blackening of the film.

3. *Studies utilizing the disintegration itself*

We have already mentioned the use of localization of various atoms in the clinical treatment of disease. There has been important use of disintegration in fundamental biological studies. One of the most significant series of studies has been on the effect of P^{32} decay on the course of virus infection. It was first shown that P^{32}-labeled phages, kept in a refrigerator, exhibited a progressive decrease of viable phages. By placing unlabeled phages in a solution of P^{32} which would give them about

the same amount of irradiation, it was shown that this decrease was due to the disintegration and not to the bombardment of the other phages by the emitted beta particles. This kind of experiment is called a reconstruction experiment, and is frequently used to test hypotheses. In the picturesque jargon appropriate to the experiment, it was shown that very little "murder" occurred in the reconstruction experiment, so that, by inference, the deaths of the P^{32}-labeled phages had to be due to "suicide" coming from the disintegration of P^{32} which had been incorporated into the phages.

These P^{32} phages were used to infect unlabeled cells in an unlabeled medium, and at various times many aliquots were taken. The aliquots were frozen rapidly, using liquid nitrogen as coolant, and stored at liquid nitrogen temperature (about $-200°C$). Each day an aliquot of each infection time was warmed, and the number of cells yielding phage progeny was determined. In this way, for instance, it was found that after one minute of infection, the ability to yield phage progeny decreased at the same rate as the viability of the free phage particle. That is, the injected nucleic acid had apparently not been altered at all, for it suicided at the same rate as it would have suicided outside the cell. After about one-third of the latent period, the cell was completely insensitive to the suicide, in that there was no decrease at all in the number of cells yielding viable phage progeny. The obvious interpretation is that by this time in the latent period the nucleic acid has replicated itself out of nonradioactive materials, as is necessarily the case since there are only nonradioactive materials present.

A complementary experiment was made using nonradioactive phages to infect radioactive cells in a radioactive medium. In this case, the cells were always stable to disintegrations, since the infecting phage contained no radioactive atoms. The stability of the phage-yielding is additional evidence that "murder" makes very little contribution to the total effect, else the nonradioactive phage DNA would have been killed.

It is of great interest that the designers of this experiment decided to do a complete experiment, however foolish it may have seemed before the results were known. They infected radioactive cells with radioactive phages in a radioactive medium. Since all nucleic acids are now radioactive, it is evident that there can never be any stabilization of the cell's ability to yield phages. Nevertheless, when the experiment was made, a stabilization very similar to that in the first of the experiments was found: when about one-third of the latent period had passed, there was no decrease in the number of cells yielding viable phage progeny! This experiment remains to be satisfactorily interpreted.

The suicide method has been used also in studies of bacterial growth and function. In addition, suicide has been found to occur when atoms

other than phosphorus were used as the source of the disintegration. The mechanism of suicide is only incompletely known. Using P^{32}, when single-stranded RNA and DNA viruses are labeled, it is understandable that death will ensue because the nucleic acids are held together by a phosphorus backbone which is broken by the recoil of the S^{32} to which the P^{32} is transformed by the disintegration. For double-stranded DNA, the efficiency of the suicide is much less (about 0.1 efficient) and is probably due to the fraction of times there occurs a scission of the opposite strand originating from the recoiling S^{32} atom.

CHAPTER 9 *Radiobiology*

INTRODUCTION

Radiation is used in biology in several ways. First, light of various visible wavelengths is used as a source of energy for biological functions. Second, as detailed in the section on light absorption effects, light may be used to analyze such diverse subjects as the pigments responsible for the reception of biologically important light and for analysis of the concentration and shape of molecules. The two kinds of light which have not yet been discussed here are the ionizing and the infrared radiations. The latter, since they have thus far given few if any biologically important results, will not be discussed. The former include short wavelength ultraviolet radiation (which has been briefly treated) and the radiation composed of the physical particles: electrons, neutrons, protons, alphas, and photons in the x-ray and gamma-ray energies. These particles have the distinguishing property of being so energetic that they are easily able to ionize one or more of the atoms in the organisms, thereby creating a disturbance great enough to break one or more chemical bonds. A drastic effect might then be expected to ensue, in terms of the alteration of the biological integrity of the irradiated material. Indeed, most studies with these ionizing particles have concerned themselves with the killing of the target organisms. We will see, however, that the particles can do many other things.

The fundamental mechanism of action of these ionizing radiations rests on the ejection of an electron, leaving an ionized atom or molecule behind. Since the outermost electron is most weakly bound to its atom or molecule, it is this electron which is usually ejected. And, since it is this same outer electron which is involved in chemical bonds, the bond is readily broken by the ionization.

The ejection of the electrons takes place by two mechanisms, depending on the type of particle involved. Electrically charged particles (electrons, protons, and alphas), when flying past an atom or molecule, exert an electric force on the electrons, and thereby effect their ejection. The uncharged particles (x-rays, gamma-rays, and neutrons) then must have a somewhat less direct mode of action. For those who have studied physics, it is only necessary to say that the photoelectric effect and the

Compton effect play the major role for the photons. For those who have not studied physics, it may suffice to say that the photons, acting like billiard balls, strike the electrons and knock them out of their orbits. Even if an inner electron is ejected, the result will normally be an ionization and breaking of bonds because the electrons outside the orbit of the ejected electron will now cascade inward, and finally the outermost (bonding) electron will fall into an inner orbit and the bond will be broken. Neutrons act primarily also in a billiard-ball fashion, hitting protons (hydrogen nuclei) which have an almost identical mass, and knocking them out of their places. These protons now serve as agents of charged-particle irradiation of other atoms and produce the effects observed.

In all these instances, the ejected electrons (and the occasionally occurring photons) will also produce effects, termed secondary effects, which must be taken into account in any complete reckoning of the results of irradiation.

Up to this point, we have treated only direct action of radiation on the biological materials of interest. Radiation experiments are usually done on organisms suspended in water. Also, even single cells are primarily water in makeup, and complicated organisms contain internal fluids within vessels. Therefore, if radiations can produce effects on water, there is another source of biological effects if the products of water irradiation can react with the biological materials.

There has been extensive study of the effects of radiations on water. Water has been split into two kinds of products. The ions H^+ and OH^- are the well-known kind. The kind less known to students in their introductory studies of science is known as the free radical. An atom of hydrogen is not normally found alone in nature, because it is too reactive a substance to endure as such very long. Indeed, the stable elementary form of hydrogen is the molecule with two atoms of hydrogen. The hydrogen atom or radical is not normally free, so that when it is found uncombined, it is called a free radical. This particular radical, H, differs from H^+ in that it has an electron in orbit about it, and is therefore electrically neutral. A more general definition of free radicals is based on the electrically neutral form of a normally electrically charged atom or group of atoms. Thus any neutralized ion would be termed as free radical. And, if this is not its normal state, it is very reactive chemically.

Water can be split into its component free radicals: H and OH. The latter is a powerful oxidizing agent, pulling electrons strongly to itself in order to make itself into the stable OH^- ion. Thereby it will break chemical bonds and produce consequent biological effects. Experiments have shown that these free radicals are indeed produced by ionizing radiation, so that indirect effects of radiation may be expected, due to

the migration of these radicals to the biological materials. If the purpose of the irradiation is to produce a biological effect, it makes no difference that there are both direct and indirect actions. If, however, we wish to study some property of the biological substance itself, it is useful to restrict the situation to one of direct actions only.

This can be done in several ways, the most important of which is to add biologically inert materials to the solution of organisms in such concentration that the diffusing radicals and ions are much more likely to encounter the added materials than the organisms in question. For example, if we add glycerol to a solution of viruses, a not atypical situation would find 0.01 M glycerol and 10^{13} viruses per liter. The glycerol would then be present in a concentration of 10^{21} molecules per liter, or 10^{8} more concentrated than the viruses. So, neglecting the size difference between virus and glycerol, any diffusing radical or ion will encounter (and give up its energy to) 10^{8} glycerol molecules before a single virus will be affected, on the average. Thus the direct hits on the viruses will far outnumber the indirect hits, and so we may neglect these indirect effects.

In practice, broth, gelatin, and many small sulfur-containing molecules have been found to be highly efficient protective agents. The protective action of these agents is still incompletely understood, and forms an important part of current research in radiobiology.

THE TARGET THEORY

In this section, we assume that only direct effects of radiations occur. If ionizing radiation impinges on a solution of organisms, the probability of producing an effect will depend, first, on the probability that the incident particle goes through the organism. Clearly, the bigger the organism, the greater the probability that the line of flight of the incident particle will pass through it. Indeed, the probability is proportional to the projected cross section of the organism. Second, given that the line of flight is through the organism, there will be an effect only if an ionization is produced within it. If the incident particle produces an average of one ionization every micron of its path, the probability of producing an ionization within the organism will depend strongly on its thickness. If, for instance, the organism is 0.1 micron thick, then only one organism in ten will be ionized. If, on the other hand, the organism is 10 microns thick, then each organism will be ionized 10 times, on the average. Thus we are led to an important parameter of radiations: the number of ionizations per unit of path length. Currently, this aspect of ionizing particles is described in terms of the LET (linear energy transfer), which is the amount of energy transferred per unit of path length. This

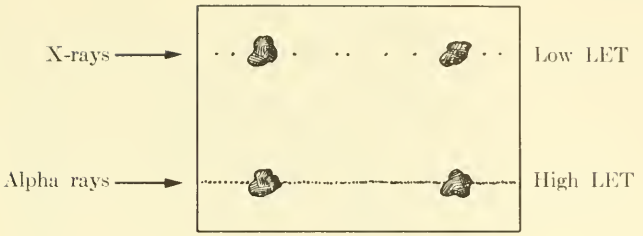


FIG. 50. A diagrammatic presentation of the difference between radiations. The densely ionizing alpha rays produce a high linear energy transfer (LET). The sparsely ionizing x-rays produce a low linear energy transfer. Thus every alpha ray produces many ionizations within a particle, while a particle would have to be very large for x-rays to produce at least one ionization in each particle. More often, the x-ray produces no more than a single ionization (if that) in a particle of a size found in typical biological experiments.

LET concept differs from the number of ionizations per unit path length by the factor of the energy transferred per ionization.

For low LET (sparsely ionizing radiations, such as very high-energy charged particles, x-rays, and gamma-rays) the probability of an ionization's occurring within the organism is directly proportional to its thickness, since the thicker the organism the more likely it is that an ionization will occur before the incident particle has passed entirely through the organism. (See Fig. 50.) Thus we can formulate the probability of producing an effect by irradiation as

P = probability that the line of flight goes through the organism
$\times$ the probability that an ionization will occur within the organism.

The first probability, as we have already argued, is proportional to the cross-sectional area A of the organism, and the second probability is proportional to its thickness t. The total probability is proportional to their product, At, which equals the volume of the organism. Thus low LET particles produce an effect which is proportional to the volume of the organism.

If we utilize radiation at the other extreme of high LET (densely ionizing particles, such as low-energy charged particles), then, as in Fig. 50, every incident particle whose line of flight goes through the organism will produce an ionization and therefore the biological effect being observed. This amounts to setting the second probability factor equal to unity. Then the total probability of producing an effect is proportional to the area A of the organism.

Consequently, if we do two separate irradiations with low LET particles and with high LET particles, we obtain results whose ratio is

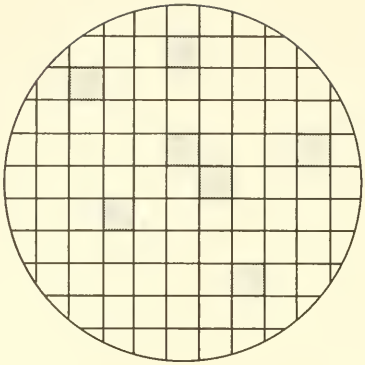


FIG. 51. Each shaded block represents a virus particle. The other areas are volumes equal to the virus particle volume but within which there are no virus particles.

proportional to the thickness of the organism. In this way we can obtain estimates of the size and shape of the organism.

If we knew enough about the physics of the incident particles, we could also deduce the actual volume of the organism. We now look at what is involved in this deduction. Consider a solution containing some viruses and imagine that the solution is broken up into subvolumes each equal to the volume of the virus particle. As indicated in Fig. 51, most of these hypothetical volumes will contain no viruses, for the virus concentration will never be more than, say, 10^{12} per ml, which is approximately 10^{-8} molar. Suppose next that we shine enough sparsely ionizing (low LET) radiation so that we have created an average of one ionization in each of these small compartments. This situation is entirely similar to the one we have previously dealt with in our discussion of biometry, when we were expounding the Poisson distribution. When there is an average of one ionization per compartment, there will be $e^{-1} = 0.37$ of the compartments which have no ionizations; the others all have one or more ionizations per compartment.

If the viruses are randomly placed in the solution, as of course they actually are, then 37% of the viruses, too, will escape ionizations. Thus the dose which leaves 37% of the viruses in a viable condition will be that dose which gives an average of one ionization per compartment the size of the virus. The way in which the 37% dose is actually found is by giving a series of doses, measuring the surviving fraction of viruses at each dose, and then interpolating to the 37% survival values, as in Fig. 52.

If we know the physics of the radiation, we can actually calculate the radiosensitive volume of the virus. Physicists have measured the ionizing

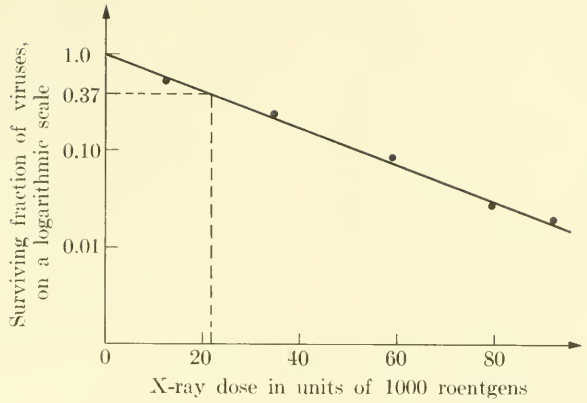


FIG. 52. The x-ray inactivation of bacteriophage alfa. The dose for producing 37% survival is found by the dashed lines to be about 23,000 roentgens.

properties of all the radiations we have mentioned earlier, and the number of ionizations produced per cc of substance is a number which can be found by looking in the proper books on radiation biology. It depends on the nature of the substance, for it depends on how many electrons are contained per unit volume. For typical biological material (wet tissue) there will be something like 10% hydrogen, 12% carbon, 4% nitrogen, 73% oxygen, 0.1% sodium, 0.2% phosphorus, 0.2% sulfur, 0.35% potassium, 0.1% chlorine, 0.04% magnesium, and 0.01% calcium. Knowing this, it is possible to compute the number of ionizations produced per cc by each kind of ionizing particle. The unit of dose was chosen by the physicists, who called the amount of radiation that will produce 1.8×10^{12} ionization per cc in air one *roentgen*. Note that the unit is expressed as ionizations *per cc*. For these sparsely ionizing radiations, such a unit is reasonable, since we have indicated that the probability of producing an ionization is proportional to the *volume* traversed. For substances such as water or normally occurring organisms, the density of the substances is about 1000 times greater than that of air; therefore this density factor must be taken into account in computing the number of roentgens delivered to the substance.

The upshot of the physicists' calculations is that the number of ionizations per cc of substance can be computed from the nature of the radiation and of the substance. Thus, suppose that the dose given to produce 37% survival of our viruses corresponds to 3×10^{18} ionizations per cc. If these are distributed so as to average one ionization per volume equal to the virus volume, the virus volume must be $1/(3 \times 10^{18})$ cc. Thus, from the biological measurement of the 37% survival dose (D_{37} in

roentgens) and a knowledge of the ionizations produced by the radiations utilized, we can deduce the volume of the virus.

There are important corrections to be made before the number of ionizations can be used as in the illustrative calculation above. The most important is the correction for the clustering of ionizations. It turns out that the (secondary) ejected electrons produce enough ionizations near their own origin that the ionizations are clustered in threes, on the average. This effect depends, however, on the size of the organism. For large organisms, we would simply divide the total number of ionizations by three to find the effective number. For very tiny organisms, it is possible that the clustering is on a scale larger than the virus diameter, so that we would actually count all the ionizations. Since this and other corrections require specialized knowledge of the physics of radiations, we will not deal with them further.

It is extremely important, however, to point out that we have tacitly assumed that a single ionization anywhere in the virus (or whatever biological organism is being used) will produce the effect being measured. This is simply not generally true, any more than it would be true that a single bullet passing through any part of your body would kill you. There are parts of any organism which may be damaged without affecting some of the properties of the organism. For viruses, for example, it turns out that the volume deduced from irradiation experiments is definitely less than the volume of the virus; it turns out to be equal to the volume of the nucleic acid of the virus. This is to be expected, since radioactive tracer experiments have shown that the nucleic acids of viruses can, by themselves, produce new viruses, complete with protein coats. Accordingly, we should refine our thinking and our terminology by talking about the *radiosensitive volume* — the volume within which an ionization will produce the effect being studied.

This definition includes a second refinement in our thinking. It is not at all necessary to study only viability. There have been studies of the radiosensitive volume associated with many other properties, such as the ability of the virus to kill a cell, the ability to adsorb to a cell, the ability of viruses to agglutinate red blood cells, the ability of cells to make enzymatic adaptations, etc.

We have discussed how *sparsely* ionizing radiations are used. If we use *densely* ionizing radiations, we need a physical device to count the number of incident particles per cm^2. Then, as before, we give a series of doses of radiation, interpolate to the 37% survival point, and find the number of particles per cm^2 which give this survival. In a way entirely similar to that used for volume calculations, we here calculate the area of the target organism or, more accurately, the radiosensitive area. As we mentioned above, use of the two kinds of radiation in parallel

experiments gives us the actual volume, area, and thickness of the target. From a knowledge of the structure of the organism, it may be possible to use this knowledge of shape to identify the part of the organism involved in the effect being studied.

Still another way in which these radiations have been employed involves the use of electrons so low in energy that they can penetrate only small thicknesses of materials. By increasing the energy of these electrons, an energy can be found at which the electrons begin to produce a particular effect. From the known penetration thickness of such electrons, we then deduce how far under the surface of the organism lies the portion associated with the effect being studied.

MULTITARGET ANALYSIS

For those who can follow a little algebra, it is possible to show how the number of targets per organism can be discovered. Such a situation could obtain, for example, in polynucleate cells. If it is necessary to inactivate all the nuclei to kill the organism, several targets must be hit *in each cell*. Below are given the few mathematical steps needed to make this analysis for a cell with n nuclei:

Let p be the probability that a nucleus survives a given dose D of radiation.
$(1 - p)$ is the probability that it does *not* survive this dose D.
$(1 - p)^n$ is the probability that *all* n nuclei do *not* survive this dose D.
$1 - (1 - p)^n$ is then the probability that *at least one* of the n nuclei survives the dose D.

This last expression then represents the fraction of cells which survive the dose D because *each* of *such cells* has *at least one* intact nucleus. We write this as

$$\frac{N}{N_0} = 1 - (1 - p)^n,$$

where N is the number of cells surviving a dose D if there were N_0 cells at zero dose.

If there were only one nucleus ($n = 1$), this would reduce to p, as expected. When very large doses have been given, the survival probability p becomes very small. In this case, we can expand the term in parentheses by the binomial theorem to give

$$(1 - p)^n = 1 - np + n(n - 1)p^2/2 - \cdots$$

If p is very small, then the right-hand side of this expression is ad-

equately approximated by the first two terms: $1 - np$. Thus the surviving fraction of cells becomes

$$\frac{N}{N_0} = 1 - (1 - np) = np.$$

That is, the surviving fraction is n times what it would have been for a mononucleated cell. This is entirely understandable, because n times as much radiation is required to hit n times as many targets.

Since p is the survival probability corresponding to a dose D, if we express D in terms of the average number of ionizations produced in the target, we can at once write the expression for p, for p is the fraction of cells having *no hits*, if the average is D hits. This is just the zero class of the Poisson distribution again, so

$$p = e^{-D}.$$

Thus for mononucleate cells the surviving fraction is

$$\frac{N}{N_0} = p = e^{-D},$$

or

$$\log \frac{N}{N_0} = -D.$$

For multinucleate cells, we had

$$\frac{N}{N_0} = np = ne^{-D},$$

so that

$$\log \frac{N}{N_0} = \log n - D.$$

Both of these expressions have a slope of $-D$. As shown in Fig. 53, where $\log N/N_0$ is plotted against D for several values of n, the asymptotic slopes are parallel and are n times higher than the value for the mononucleate cell. Thus, if

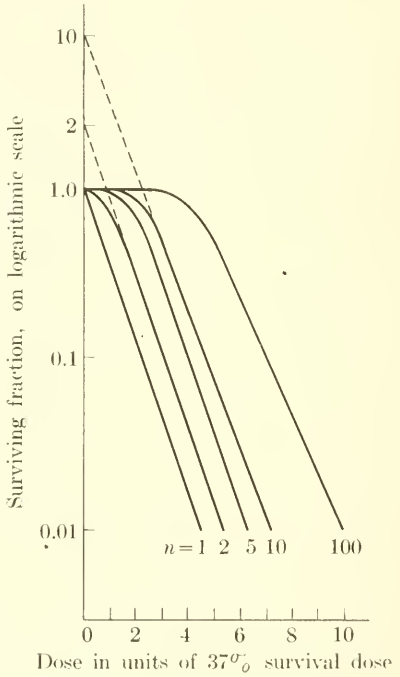


FIG. 53. Theoretical survival curves where there are n equally radiosensitive nuclei per cell. The curves are plots of the expression in the text: $1 - (1 - e^{-D})^n$, where n is the number of nuclei and D is the radiation dose in units of the 37% survival dose. The dashed-line extrapolations of the straight portions of the curves give the number of nuclei as the intercepts on the zero dose axis.

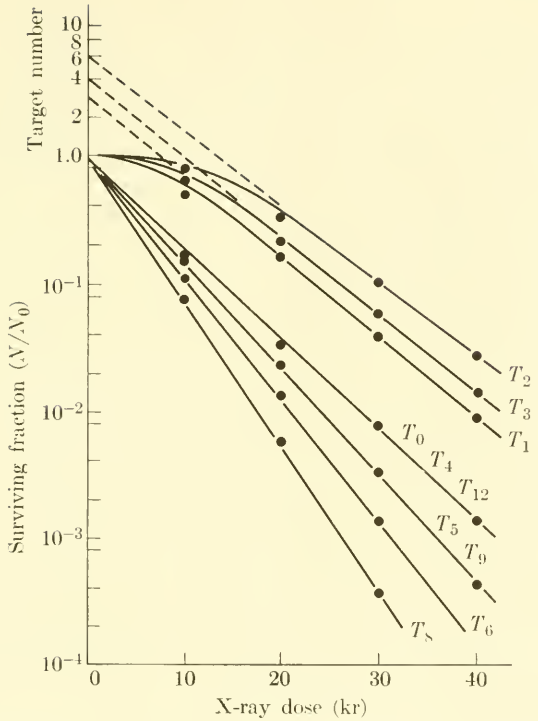


FIG. 54. Bacterial survival after various x-ray doses. The age of the culture in hours is listed with the survival curve for bacteria of that age. (From Stapleton, *J. Bacteriology*, **20**, 357, 1955; courtesy the author and Williams and Wilkins Co., Baltimore, Md.)

we obtain a multitarget curve as a result of an experiment, we simply draw a line of the same slope so as to pass through the 100% survival point, and we can then calculate the value of n. In addition, from the mononucleate survival curve, we can find the 37% survival dose, and then obtain the radiosensitive volume as before.

The value of n can also be found by extrapolating the straight-line portion of the survival curve back to the zero dose line; the intercept is the nuclear number itself. That this is true can be seen from the last equation: $\log (N/N_0) = \log n - D$.

If we set D equal to zero, we obtain $\log (N/N_0) = \log n$. This way of finding n is illustrated in Fig. 53 by extrapolating the $n = 2$ and $n = 10$ curves back to the zero dose line.

The student should be warned that there are serious possible complications which can arise in a radiobiological experiment. For example, there are cases of irradiation of polyploid yeasts in which the target

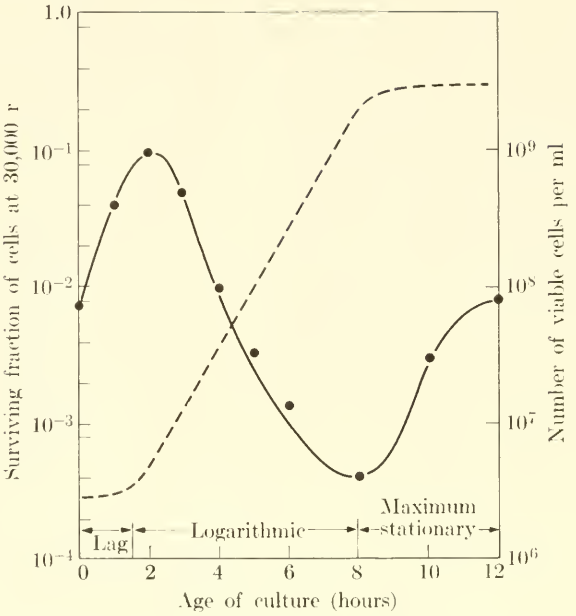


FIG. 55. A replotting of some of the data of Fig. 55, showing the cycling of radio sensitivity. The cell concentration is shown as a dashed line. (From Stapleton, *J. Bacteriology,* **20**, 357, 1955; courtesy the author and Williams and Wilkins Co., Baltimore, Md.)

numbers simply do not follow the ploidy. Another example is the variation in inactivation curves of bacteria in various stages of their growth cycle. The experimental results are given in Fig. 54, taken from Stapleton's work. This information can be summarized as in Fig. 55, also taken from Stapleton, which shows clearly the cyclic variation in radiosensitivity. This variation is not altogether out of the purview of target theory, for rapidly dividing cells could well have a lesser number of nuclei than lagging or stationary phase cells in which nuclear division may have occurred without corresponding cell division. Furthermore, the metabolic activity of the nucleic acids of the nuclear apparatus could also result in different sensitivities at different times in the growth cycle. However, there still remain the problems of a change in target number from unity to six, which seems outside any reasonable increase in nuclear number.

Opponents of the use of target theory maintain that such failures show the terrible weakness of the theory, and some even go so far as to say that target theory should therefore really not be used at all. Our point of view is that since there are many instances in which target theory gives well-documented success, there is nothing basically wrong with the theory, but that where the theory fails it must be due to the

inapplicability of one or more of its assumptions. From an analysis of the failures, one should then learn even more about the nature of the biological entities and about radiobiology.

The target theory may be applied, in part, to studies utilizing ultraviolet radiation, for these are also energetic enough to ionize atoms. The restriction stems from the fact that although the probability of an ultraviolet photon's being absorbed by matter is proportional to the electron density (and thus to the volume of the matter), there is no theoretical connection between the radiosensitive volume and the volume of any substantial portion of the organism involved. The photon's absorbing element is really the electron taking part in a bond, and its volume is not greatly dissimilar to that of the bond itself. Therefore, although a volume can be computed, it doesn't teach us anything. Furthermore, if we break our substance into pieces, the absorption probabilities of the pieces do not total to the probability of the united substance. In sum, then, one cannot compute a volume from ultraviolet inactivation data. However, since the action appears to be due to a single ionization process, the kind of target analysis we have presented may still be applied to deduce the *number* of targets. Also, from the asymptotic slopes of the inactivation curves, we can say whether the targets are the same or not.

BIOLOGICAL EFFECTS OF RADIATION

Up to this point, we have been analyzing the physical action of radiations on biological systems. There remains the subject of the biological effects. We shall discuss the following examples of biological effects: mutation, ultraviolet-light inactivation and reactivation.

1. *Mutation*

Since the initial demonstration by Muller that x-rays can induce mutations, an extensive body of literature has developed on this topic. Aside from the straightforward use of radiations to induce mutations for the sake of obtaining mutant organisms, the kinds of mutations induced permit deductions as to the nature of mutation itself. X-ray mutagenesis has been found to include chromosome breaks, translocations, and other cytological effects which offer strong evidence that the x-ray simply breaks the chromosome. Ultraviolet light mutagenesis has recently been given an enormous stimulus by the finding that physiological doses effect the formation of dimers of one of the four bases making up nucleic acids. In such a situation, as the nucleic acid is copied during its replication, the copying mechanism is very likely to err when it reaches the dimer, thereby producing a mutation. Such a mechanism is restricted to the locations having two identical bases in series along the nucleic acid chain. It is not known whether x-rays can also act this

way, but since a respectable fraction of the x-ray energy is dissipated in amounts similar to those of ultraviolet light photons, it is not improbable that it sometimes does.

Since nucleic acids have been shown to include double-stranded structures as the major component of the genetic apparatus, it is conceivable that one strand could be affected without at all altering the other strand. Then, when copying of both strands ensues, if the copying mechanism reaches an ionized base it could make an error by inserting one of the other three bases, thereby producing a mutation.

2. *Ultraviolet-light inactivation and reactivation*

The formation of standard chemical bonds across the two strands of the DNA would surely provide adequate impedance to DNA replication, thereby inducing death of the organism at the time of its attempt to replicate the DNA. This picture is supported by the finding that UV-inactivated DNA does not separate into two strands on prolonged heating, indicating a "sewing together" of the strands. In addition, dimerizations within a strand, as discussed above, can lead to mutation, and some of these may be lethal mutations in that they affect an indispensable function of the DNA.

Two types of reactivation of UV inactivation have been discovered. The first type is called *photoreactivation.* If UV-inactivated organisms are subsequently exposed to visible light, a substantial fraction of the organisms is found to be viable again. In microbial systems, this has been shown to be due to the light supplying the energy needed for a photoreactivating enzyme to work. This enzyme has been greatly purified, and it is found to work in the test tube. If DNA is treated with UV, as mentioned above, it no longer separates on being heated, but if it is then treated with the photoreactivating enzyme (in the light), its separability is restored.

The second type of reactivation is potentially more instructive, although its promise has not yet been realized. Bacteriophages inactivated by UV have been shown to be able to cooperate to produce live progeny. Specifically, cells infected with two or more UV-inactivated phages are found to yield viable phage progeny, even though each phage, by itself, could produce no progeny. Presumably, the phage DNA which is injected into the cells can undergo a reaction which allows the DNA copying mechanism to produce one viable copy by copying the good parts of the two UV-damaged phages. Such an interaction should provide us with great insight into the functioning and replication of DNA, but thus far there has been no really successful theory of reactivation in multiple infection or, as it is normally called, *multiplicity reactivation.*

References

GENERAL REFERENCES

Biophysical Science, by Eugene Ackerman. Prentice Hall, Inc., Englewood Cliffs, N. J., 1962. This is the best general textbook to appear to date. There are some limitations which are undoubtedly due to the fact that a single author has attempted to cover the entire broad scope of biophysics.

Molecular Biophysics, by Richard Setlow and Ernest Pollard. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1962. Some parts of this text are very clever and illustrate the point of view of biophysicists very much better than does Ackerman's text. However, some important concepts are not defined, and the text is less up to date and less inclusive than Ackerman's.

Medical Physics, edited by O. Glasser. Year Book Publishers, Inc., Chicago. This collection of short articles is intended for those whose background is not greatly different from that of beginning science students. It includes many aspects of science which are within its intended scope of *medical* physics as distinguished from *bio*physics.

SELECTED REFERENCES

Chapter 1

Errors of Observation and Their Treatment, by J. Topping. Reinhold Publishing Corporation, New York, 1958. An excellent short treatment of many aspects of statistical analysis. Students should own this little book, which will serve them well throughout their science careers.

Facts from Figures, by M. J. Moroney. Penguin Books Inc., Baltimore, 1958. A wordy (but interestingly written) exposition of the ideas of elementary statistics.

Chapter 2

So far as I know the only book that treats the material discussed in this chapter on the desired level is the text by Setlow and Pollard. However, the mathematical and physical discussions of the topics is beyond the background of most beginning students. A much more suitable treatment will be contained in a book being readied for publication by Addison-Wesley: *Biophysical Principles of Structure and Function*, by Fred M. Snell, Sidney Shulman, Richard P. Spencer, and Carl Moos.

Chapter 3

There are good discussions of this topic in both Ackerman, and Setlow and Pollard.

Vision and the Eye, by M. H. Pirenne. Chapman and Hall, London, 1948. An excellent book for an introduction to a substantial fraction of the work in this area. It is now a little out of date, but is recommended both for its content and for the interesting presentation of an interesting subject.

Chapter 4

There is a good presentation of this topic in Ackerman.

Theory of Hearing, by Ernest G. Wever. John Wiley and Sons, Inc., New York, 1949. This interestingly written book covers the historical aspects of the subject and then presents the details of current research. Although somewhat out of date, its only real defect is the inadequate presentation of the work of Von Bekesy. The level is fine for beginning students.

"The Ear," by G. von Bekesy. *Scientific American*, August, 1957.

Chapter 5

Both the Setlow and Pollard, and the Ackerman texts have adequate presentations of these topics. The one in Setlow and Pollard is better written.

Chapter 6

Ackerman and Setlow and Pollard have sections on this subject. Ackerman's is longer, more detailed, and better written.

Essentials of Biological and Medical Physics, by Ralph W. Stacy, David T. Williams, Ralph E. Worden, and Rex O. Morris. McGraw-Hill Book Company, Inc., New York, 1955. This is a general textbook. I have not listed it among the general references because it contains little more than an outline of most of the topics I have chosen to present in this monograph. The excellent chapter on muscle is almost entirely understandable by beginning students.

Chapter 7

There are good treatments of most of these topics in Ackerman, Setlow and Pollard, and Glasser.

Chapter 8

Both Ackerman, and Setlow and Pollard contain reasonably inclusive, well-written expositions of this subject.

Isotopic Tracers in Biology, by Martin Kamen. Academic Press Inc., New York, 1957. This book, written by one of the pioneers in the field, contains a wealth of information on all aspects of the subject. It is especially good for beginning students, who will gain an appreciation of the scope of isotope methods.

Chapter 9

There are good treatments of this topic in both Ackerman and in Setlow and Pollard. The presentation in the text by Setlow and Pollard is excellent for students seeking an inclusive treatment, since it contains an exposition of the methods used and results obtained in the laboratory directed for a decade by Prof. Pollard.

Actions of Radiations on Living Cells, 2nd edition, by Douglas Lea. Cambridge University Press, New York, 1955. A classic in the field which almost created the field when it appeared in 1947. A revision due to the late Douglas Lea's associates altered very little of the original, as would be expected for a classic. Many parts are understandable by beginning students, even though the mathematics and physics may be beyond their grasp.

Index

THE AUTHOR

Hern
Univer
Mich
Prof
and
ph

iophysics at Brandeis
the University of
ity of Pittsburgh.
nterest are virus genetics
ember of the Bio-
tion for the Ad-
of University

is intended
ology. Its purpo
ics into the elementa
contention that the trend in biol
chemical biology, and today's students will be bette
if they are taught to think in physio-chemical terms
with their first course in biology.

Among the topics covered are statistics (from the point of view of the problems faced in evaluating measurements); physical forces and chemical bonds; physics of vision, hearing, and muscles; and biophysical methods.

ADDISON-WESLEY PUBLISHING COMPANY
READING, MASSACHUSETTS · PALO ALTO · LONDON · DON MILLS, ONTARIO

1881

PRINTED IN U.S.A.