The Chemical Foundations of Molecular Biology

ROBERT F. STEINER

The Chemical Foundations of Molecular Biology

ROBERT F. STEINER, Ph. D

Head, Biological Macromolecules Branch
Physical Biochemistry Division
Naval Medical Research Institute
Bethesda, Maryland

D. VAN NOSTRAND COMPANY, INC.
Princeton, New Jersey
Toronto New York London

Preface

The author of a book on molecular biology is confronted at the onset by a problem in terminology. The term "molecular biology" is of relatively recent origin and appears to be used in two quite different senses. Many authors have chosen to restrict its area of application to the chemistry of genetic processes and the genetic control of metabolism. By this definition the proper domain of molecular biology would be confined largely to the replication of nucleic acids and the directed synthesis of proteins.

I regard this usage as unfortunate and prefer to give the term its most obvious and literal meaning, namely, those aspects of biology which can be described at the molecular level. This definition is broad enough to include the topics of replication and protein synthesis, which retain their central importance.

Various practical considerations, including the desirability of avoiding a book of excessive length and price, have made it necessary to limit the coverage by excluding many subjects which could, by the above definition, be regarded as falling within the discipline of molecular biology. The additional criteria for inclusion have been, first, the *timeliness* of the subject, or the intensity of current scientific interest; and second, the extent to which it is already covered by existing texts.

The book has been written at the advanced undergraduate level. It is hoped that this will not preclude its being of use to less advanced students or to fully qualified scientists from other disciplines. The reader should be acquainted with elementary general chemistry and biology, as well as with organic chemistry, and preferably should have some knowledge of physical chemistry. For parts of Chapter 4 a knowledge of basic calculus is essential.

The author has no illusions that the choice of material will please everyone or satisfy all needs of the aspiring molecular biologist. The intended function of the book is to serve in a manner somewhat analogous to the first stage of a multistage rocket, in that it leaves the later phases of the journey to be completed in more specialized craft.

R. F. S.

Contents

1 INTRODUCTION 1

1-1 The Meaning of Molecular Biology 1
1-2 The Major Biopolymers 2
1-3 Cellular Organization 3
1-4 A Rudimentary Account of Genetic Principles 5
1-5 Substructure of the Gene 21

2 THE AMINO ACIDS 26

2-1 Preliminary Remarks 26
2-2 Optical Activity of Amino Acids 27
2-3 The Neutral Amino Acids 33
2-4 Acidic and Basic Amino Acids 39
2-5 Amino Acid Analogs 41

3 THE CHEMICAL STRUCTURE OF PROTEINS 43

3-1 Preliminary Remarks 43
3-2 The Manner of Linkage of Amino Acids in Proteins 44
3-3 The Sequential Arrangement of Amino Acids in Proteins 47
3-4 Techniques of Sequence Determination 51
3-5 The Primary Structures of Individual Proteins 62
3-6 Variations in Primary Structure 69

4 THE SIZE, SHAPE, AND ELECTRIC CHARGE OF PROTEIN MOLECULES 76

4-1 Introduction 76
4-2 Electron Microscopy 78
4-3 Methods Yielding Average Values of Molecular Weight Directly: Osmotic Pressure 81
4-4 Methods Yielding Average Values of Molecular Weight Directly: Light Scattering 83
4-5 General Aspects of the Hydrodynamic Methods 90
4-6 Methods Yielding a Measure of the Size and Shape, Subject to the Assumption of a Geometrical Model: Viscosity 99

4-7 Methods Yielding a Measure of the Size and Shape, Subject to the Assumption of a Geometrical Model: Fluorescence Polarization 102
4-8 Methods Yielding a Measure of the Size and Shape, Subject to the Assumption of a Geometrical Model: Diffusion 105
4-9 Methods Yielding Both the Molecular Weight and a Measure of the Homogeneity—Ultracentrifugation 109
4-10 Methods Yielding a Measure of the Homogeneity: Electrophoresis 120

5 THE SPATIAL ORGANIZATION OF PROTEINS 131
5-1 General Remarks 131
5-2 Forces Involved in the Stabilization of Protein Structure 133
5-3 Polypeptide Conformations 138
5-4 The Determination of the Conformation of Polypeptides in the Solid State 143
5-5 The Estimation of Helical Content for Proteins and Polypeptides in Solution 147
5-6 The Helix-Coil Transition for Polypeptides 152
5-7 The Tertiary Structure of Proteins 157

6 STRUCTURE AND FUNCTION OF CERTAIN IMPORTANT PROTEINS 162
6-1 General Remarks 162
6-2 Structural Proteins: Collagen 163
6-3 Contractile Proteins: Myosin and Actin 171
6-4 Antibodies 178
6-5 Fibrinogen 188
6-6 Plasma Proteins: Serum Albumin 194
6-7 Hemoglobin 197
6-8 Myoglobin 204

7 CATALYTIC PROTEINS: THE ENZYMES 208
7-1 Introduction 208
7-2 Enzyme Kinetics 215
7-3 Structure and Activity of Trypsin 229
7-4 Structure and Activity of the Chymotrypsins 232
7-5 Acetylcholinesterase 238
7-6 Pancreatic Ribonuclease 242

8 THE NUCLEOTIDES 247
8-1 General Remarks 247
8-2 The Nucleotide Bases 250

8-3 The Nucleosides 256
8-4 The Nucleotides 259
8-5 Properties of the Nucleotides 261

9 THE DEOXYRIBONUCLEIC ACIDS 266
9-1 General Remarks 266
9-2 The Primary Structure of DNA 267
9-3 The Secondary Structure of DNA 272
9-4 Solution Properties of DNA 277
9-5 The Atypical DNA of $\phi \times 174$ Virus 279
9-6 The Denaturation of DNA 280
9-7 The Renaturation of Denatured DNA and the Formation of Molecular Hybrids 286
9-8 Biosynthetic DNA 292
9-9 The Biological Replication of DNA 295
9-10 Transforming Principle 299

10 THE RIBONUCLEIC ACIDS AND THE BIOSYNTHESIS OF PROTEINS 307
10-1 General Remarks 307
10-2 The Primary Structure of RNA 310
10-3 The Secondary Structure of RNA 314
10-4 Biosynthetic Polyribonucleotides 322
10-5 The Guided Synthesis of RNA 326
10-6 Protein Biosynthesis 329
10-7 Enzyme Induction 338
10-8 The Linear Assembly of the Polypeptide Chains of Hemoglobin 341

11 THE NUCLEIC ACIDS OF VIRUSES AS CARRIERS OF BIOLOGICAL INFORMATION 347
11-1 General Remarks 347
11-2 The Infectious RNA of Tobacco Mosaic Virus 352
11-3 Infectious RNA's of Other Viruses 360
11-4 The T-Even Bacterial Viruses 361
11-5 The Temperate Bacteriophages 379

12 THE CARBOHYDRATES AND THEIR BIOSYNTHESIS 383
12-1 The Monosaccharides 383
12-2 The Polysaccharides 389
12-3 The Biosynthesis of Carbohydrates 396
12-4 The Biosynthesis of the Polysaccharides 408

13 ENERGY TRANSFORMATIONS BY BIOLOGICAL SYSTEMS 412
13-1 Energy Storage and Utilization 412
13-2 The Coenzymes of Oxidative Metabolism 418
13-3 The Initial Phase of the Oxidative Metabolism of Carbohydrates 426
13-4 The Mitochondria 432
13-5 The TCA Cycle 433
13-6 Oxidative Phosphorylation 438

APPENDIX A The Primary Structure of the B-Chain of Insulin 443
APPENDIX B Basic Thermodynamic Concepts 447
APPENDIX C Synthesis of Polypeptides 453
APPENDIX D Biological Oxidation and Reduction 457

INDEX 463

1

Introduction

1-1 THE MEANING OF MOLECULAR BIOLOGY

General remarks. It is desirable to begin this introductory chapter with a definition of terms. The term "molecular biology" would have been meaningless 50 years ago and would have represented little more than a tantalizing aspiration as recently as 25 years ago. Biological systems can be described from many viewpoints which do not begin to approach the molecular level. Indeed, classical biology developed in a manner which paralleled, rather than stemmed from, the contemporaneous evolution of the physical sciences. For example, the science of genetics attained a high degree of success in rendering coherent an enormous body of superficially chaotic data long before more than the dimmest notion was held of the chemical nature of genetic determinants.

Molecular biology represents an effort to account for, in detail, biological events in terms of the established principles of physics and chemistry. Since molecules are the most complex physical entities which can be described in a sharply definitive manner, such an endeavor must necessarily be concerned with the properties of molecules which occur in biological systems. Its hopes for success depend on the assumption, originally largely a matter of faith, that biological systems differ only in complexity from the simple systems which furnished the earlier triumphs of chemistry and that they obey the same laws.

In the broadest sense, molecular biology may be regarded as the study of those aspects of biological systems, or components thereof, which can be described at the molecular level. Since the area for which this is the case tends to enlarge progressively, it can be foreseen that all of biology will ultimately fall into this category. The term is also often used in a more restrictive sense to denote the molecular aspects of genetic mechanisms and the control of metabolic processes by genetic material.

This book will adhere to the more general definition. For didactic pur-

poses, the discussion will be largely centered about the complicated giant molecules, or *biopolymers*, which occur only in living systems.

The material to be presented may be regarded as an effort to answer the following questions:

(1) What is the nature of the principal biopolymers? A complete answer to this question in any particular case would include an account of its gross size and shape, its purely chemical structure, and its geometrical organization.

(2) What is the biological function of each biopolymer and how is function related to structure? A primary objective of molecular biology is the correlation of definite biological processes with chemical events at the molecular level.

(3) How are the biopolymers positioned and oriented in the cells of living systems? This question is of course related to (2).

(4) How are the biopolymers formed from their constituents? In particular, how are the formidable energetic and entropic obstacles to their synthesis overcome by living systems?

The overwhelming volume of the available information naturally makes necessary some degree of selectivity in the choice of topics to be covered. It is inevitable that the selection should be somewhat biased, both by the particular interests of the author and by the fortuitous availability of much more information in some areas than in others.

The balance of this introductory chapter will be devoted primarily to background information which, while not in the domain of molecular biology proper, provides a needed frame of reference for much of the discussion to follow. The presentation here will be elementary, condensed, and oversimplified. It is hoped that the reader will supplement it with the more complete presentations to be found in the references at the end of this chapter.

1-2 THE MAJOR BIOPOLYMERS

Types of biopolymers. *Polysaccharides* are perhaps the simplest of the biopolymers in structure and function. They are polymers of simple sugars, or *monosaccharides*, or derivatives thereof. The biological function of the polysaccharides is relatively restricted and passive. *Starch* and *glycogen* serve as nutritional reservoirs. *Cellulose* has a function which is primarily structural. Other polysaccharides have functions which are relatively limited and specialized.

The *nucleic acids* are linear polymers of *nucleotides*, which consist of a sugar, a nitrogenous base, and phosphoric acid. They occur in two

forms called *ribonucleic* and *deoxyribonucleic* acid and universally referred to as RNA and DNA, respectively. These occupy a central position in molecular biology as the directive agents for control of genetic processes and protein synthesis. A major fraction of this book will be centered about their structure and properties.

The *proteins* account for the largest portion of living systems and have the greatest diversity of structure and function. Their basic chemical structure corresponds to linear polymers of *amino acids,* which may be cross-linked. Various schemes for the classification of proteins have been proposed, but their diversity is such that no system is completely satisfactory. With regard to *function,* the proteins may be roughly grouped as follows:

(1) Structural: Proteins of this class are generally rather inert with respect to biochemical processes. They serve to maintain the form and position of organs, as components of container walls for biological fluids, as means of attachment of tissues to the skeleton, and so on.

(2) Contractile: These, by virtue of their property of contractibility, supply to the living organism the capacity of motion or of external work. The muscle proteins are of course the most familiar examples. Their contractile properties appear to reflect their ability to undergo an internal configurational change which results in a change in extension.

(3) Catalytic: The properties of living cells depend upon the continual occurrence of an extraordinary variety of chemical processes. The regulation of the rates of these reactions is the province of a class of proteins with catalytic properties. These are called *enzymes.*

(4) Transport: The transfer of an essential biological factor, from a point in the organism where it becomes available from an external source to a point where it is biochemically utilized, may be accomplished by reversible combination with a carrier protein. In this manner oxygen is transported by hemoglobin.

1-3 CELLULAR ORGANIZATION

Structural elements. The reader should be warned that the description of cellular elements to be presented here will be held to an extreme minimum and should, if possible, be supplemented by the fuller descriptions to be found in the references cited at the end of this chapter (1-5)

The cell is the basic unit of which all living organisms are constructed. The recognition of its existence and of its fundamental importance dates from the nineteenth century and the classical investigations of Schwann and Virchow.

There are limits to the extent with which one can generalize about living cells. Their size and morphology naturally vary considerably according to their function. Nevertheless, there are many features which recur for the cells of almost all higher species. Examination of cells with a microscope, using appropriate stains, reveals the presence of many differentiated bodies, which respond in different ways to stains as a consequence of their varying chemical composition. The cells of living organisms are endowed with their characteristic functions by the presence of hierarchies of biological polymers. The distribution of these within the cell is governed by the localization of particular functions in specific regions.

A basic division of the cell exists between the *nucleus* and the surrounding *cytoplasm.* Both are surrounded and confined by membranes. The cell membrane, which encloses the cytoplasm (Fig. 1-1a), possesses a complex structure and is capable of considerable specificity in regulating the passage of ions and molecules.

Within the cytoplasm, examination with the electron microscope has revealed the presence of numerous preformed bodies. These may be isolated from the nucleus and from each other by differential centrifugation of disintegrated cells. Among the most prominent of these granules are the *chloroplasts* of green plant cells and the *mitochondria,* which occur in both animal and plant cells. The former (Chapter 12) utilize the radiant energy of sunlight to form compounds essential for the synthesis of carbohydrates from water and atmospheric carbon dioxide. The mitochondria fill the role of a "powerhouse" for the cell and supply in usable form the energy required for such processes as the synthesis of proteins and nucleic acids, the transport of essential materials, and the performance of mechanical work. The source of this energy is the oxidative metabolism of foodstuffs, which is mediated by an array of enzymes localized in the mitochondria.

Another organized structural element of the cytoplasm is the *lysosome.* This contains the digestive enzymes which break down biopolymers into their smaller constituents, which can be oxidized by the mitochondrial enzyme system. Rupture of the lysosomal membrane and release of its contents result in a rapid lysis of the cell.

The *centrosomes* (or *centrioles*) become readily visible by light microscopy only when the cell is beginning to divide. They are replicated during mitosis and appear to have a directive function in this process (section 1-4).

The cytoplasm, in addition to the particulate bodies mentioned above, appears to possess a system of internal membranes which are not visible in the ordinary light microscope. These vary greatly in complexity from

cell to cell. This membrane system has been called the *endoplasmic reticulum.* A large number of small granules, called *ribosomes,* occur on its surface. These are roughly spherical in shape and contain a high proportion of one form of nucleic acid, ribonucleic acid (RNA). The ribosomes are believed to be the primary sites of protein synthesis within the cell. The membrane may be synthesized by certain mysterious structures called *Golgi bodies.*

A considerable fraction of the contents (other than water) of the cytoplasm fails to sediment in a centrifugal field strong enough to deposit the ribosomes, mitochondria, and other particulate forms. This material is called the *cell sap,* or *soluble fraction,* as opposed to the particulate entities described above. It consists mainly of proteins, fats, RNA, and small molecules which are not integrated into a preformed structural element.

The cytoplasm is separated from the nucleus by the *nuclear membrane.* The nucleus contains the vitally important filaments of *chromatin,* to which the deoxyribonucleic acid (DNA) of the cell is confined. These are rather indistinct in the interval between cell divisions, but become much more compact and visible during division. During this process they appear as distinct bodies called *chromosomes.*

Another prominent preformed granule is the *nucleolus.* This undergoes cyclic changes in appearance, disappearing during cell division and reappearing at the end of division. It is rich in RNA and may be the site of protein and RNA synthesis.

1-4 A RUDIMENTARY ACCOUNT OF GENETIC PRINCIPLES

General remarks. A complete account of the discipline of genetics would require many volumes, and no attempt will be made to do justice to the subject in this book, except in the as yet relatively restricted areas where a discussion on a molecular level is feasible. However, since certain very basic principles provide a necessary background for much of the discussion to follow and since many readers may be without formal training in genetics, a brief account of these will be given here.

Mendelian genetics. The original experiments of Mendel are still the best point of departure for a discussion of classical genetics (6-9). Mendel crossed strains of garden peas (*Pisum sativum*), which differed with respect to certain easily recognizable external characteristics, such as the color of pigmentation of the flowers and color and surface form of the seeds. After protracted trial and error, 22 varieties were selected for

further experimentation. All of these were "pure" strains which bred true; that is, each generation of progeny derived from a single strain contained only plants which uniformly resembled their parents. These supplied seven distinct differences of character (or *phenotype*), which was adequate for his purpose (6-9).

It was found that, if two different strains were crossed, that in each case *all* the seedlings of the first filial (F_1) generation uniformly resembled one of the parents with respect to each of the specific characteristics.

For example, a cross of a strain of peas with purple flowers with a strain producing white flowers yielded an F_1 generation whose members produced, not light purple or mottled flowers, but flowers of the same purple color as one of the parental strains.

This held for each of the seven phenotypes studied by Mendel. The type which was manifest in the F_1 generation following the cross, as purple pigmentation in the above instance, was called *dominant* and its latent alternative, which was in this case white flower color, was termed *recessive*. No intermediates appeared in the F_1 generation, whose members uniformly resembled one or other parent.

If a second (F_2) generation was raised from the F_1 generation by self pollination, the recessive types reappeared. Intermediates were again absent. The ratio of dominants to recessives was the same in the case of each of the seven characteristics tested for and was equal to 3. In the above-mentioned case of flower pigmentation, three-fourths of the F_2 progeny were purple and one-fourth white.

The results of Mendel suggested that a *segregation* occurred of the factors responsible for each of the characteristics so that they existed as discrete genetic units which were transmitted in intact form. Each genetic determinant existed in two stable *allelic* forms which retain their identity in the F_1 gencration, to separate and form new combinations in the F_2 generation.

These genetic determinants, or units of heredity, were subsequently termed *genes*. Each individual of the F_1 generation is the product of a fusion of the germ cells of the parents and receives half of its determinants from each parent. If one characteristic is singled out, as flower pigmentation, then each individual of every generation possesses two corresponding genes which control pigmentation.

An organism possessing a double quantity of a single allelic form of a gene is said to be a *homozygote*. This was the case for the original pure strains of Mendel. One which contains two different allelic forms, like the members of the F_1 generation, is called a *heterozygote*.

Thus if the dominant allelic form (for example, purple) is designated

as *A* and the alternative recessive form (for example, white) as *a*, then the original strains are *AA* and *aa*. The members of the F_1 generation are all heterozygotes of the *Aa* type.

The F_2 generation contains *AA*, *Aa*, and *aa* forms. The ratio of individuals with dominant characteristics to those with recessive characteristics is exactly what would be expected for a separation and *random* recombination of allelic forms. Thus a random reshuffling would produce individuals of types *AA*, *Aa*, and *aa* in a ratio of 1:2:1. Since *Aa* forms would have dominant characteristics, this distribution would predict a 3:1 ratio of individuals having dominant characteristics (*AA* or *Aa*) to those displaying recessive traits (*aa*).

This pattern was followed for successive generations. For example, if the purple-flowered plants of the F_2 generation, which contained *AA* and *Aa* types in the ratio 1:2, were allowed to produce an F_3 generation, it was found that one-third of the F_2 plants (the *AA* forms) produced only purple progeny, while the remaining two-thirds (the *Aa* form) produced both purple (*Aa* or *AA*) and white (*aa*) progeny in a 3:1 ratio.

Mendel also made crosses between strains differing in two of his seven selected hereditary traits. The distribution of progeny was in each case that expected for an *independent* assortment of genetic traits. For example, two strains were crossed which differed in the *color* (yellow versus green) and *form* (smooth versus wrinkled) of their seeds. The dominant allelic forms were yellow (*A*) and smooth (*B*), while the green (*a*) and wrinkled (*b*) types were recessive (Fig. 1-1).

If a pure strain with yellow-smooth seeds (*AABB*) was crossed with a pure strain with green-wrinkled seeds (*aabb*), all members of the F_1 generation had yellow-smooth seeds, since all of the heterozygotes were of the *AaBb* type and contained both dominant genes.

The F_2 generation showed all four combinations of phenotype in the ratio expected for *random* and *independent* reassortment of genes (Fig. 1-1b). That is, the probability that a given individual would have a smooth or wrinkled seed was entirely independent of whether the seed was yellow or green.

The relative frequencies of occurrence of the four possible combinations were as follows:

Properties	Gene Composition	Relative Frequency
Yellow-smooth	*AaBB*; *AABB*; *AaBb*; *AABb*	9
Green-smooth	*aaBB*; *aaBb*	3
Yellow-wrinkled	*AAbb*; *Aabb*	3
Green-wrinkled	*aabb*	1

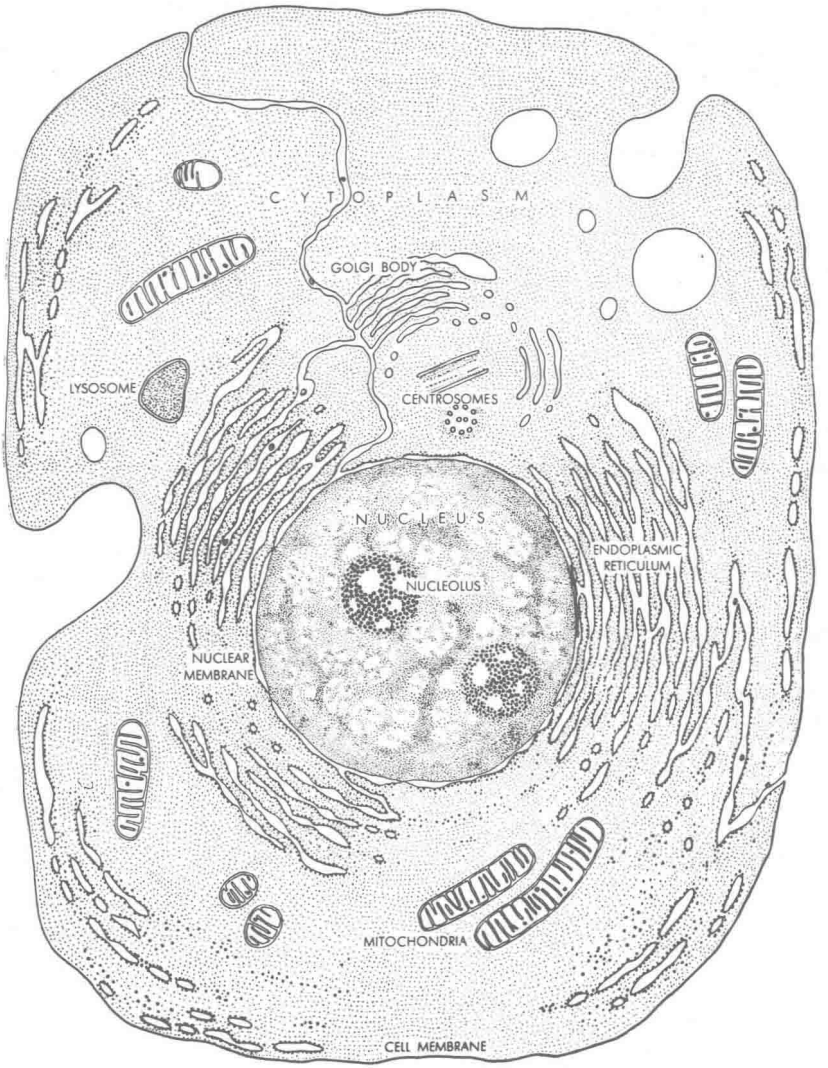

Fig. 1-1a. Schematic representation of a generalized living cell, showing the differentiated structures of the nucleus and cytoplasm. (This illustration is redrawn from J. Brachet, *Sci. Amer.*, 205, 50 [1961].)

The over-all picture of the gene which emerged from these experiments was that of a *stable* and *indivisible* unit which could exist in two or more sharply differentiated forms, to which no intermediate forms existed. Genes controlling different properties were thought to be transmitted entirely independently.

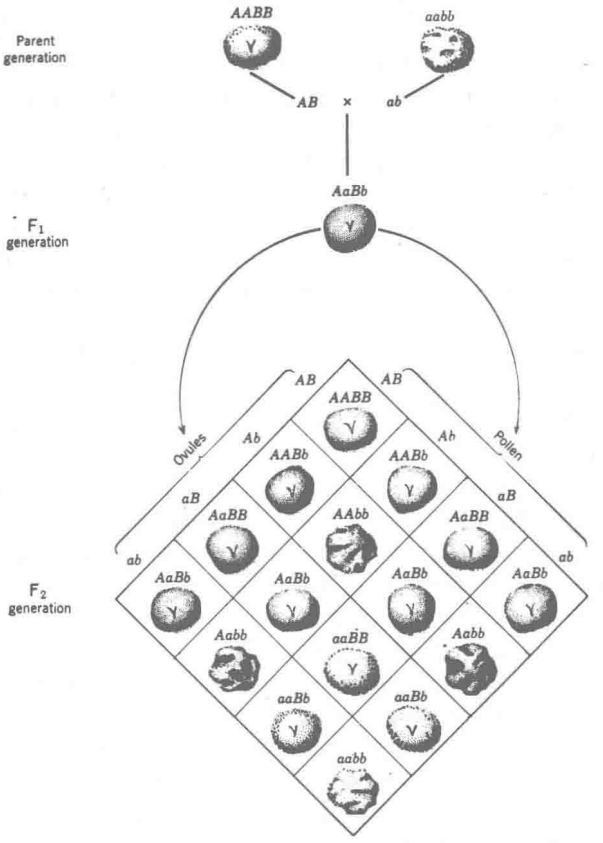

Fig. 1-1b. **Illustration of the independent assortment of genes (7). *A* and *a* represent the genes for yellow and green colors, respectively, while *B* and *b* represent those for smooth and wrinkled seed surfaces.**

Part of the success of this simple model in explaining Mendel's results was due to his fortunate choice of inheritable traits. In actuality there are many instances of sets of genes which do not migrate independently and which instead behave as if they were linked together in a common genetic bundle. However, a discussion of linked genes should be prefaced by an account of cell division and chromosomal movements.

Chromosomes and cell division. The existence of visible elongated structures which appear in the nucleus during cell division has been recognized since the last century. In the intervals between cell division, the chromosomes become so indistinct that their persistence was long questioned. However, it is now believed that the chromosomes retain their identity

throughout the cycle of the cell, although their form and appearance change.

The number of chromosomes is a well-defined constant for the cells of each species of higher organism. In general, the ordinary *somatic* cells of higher organisms, which are not directly concerned with reproduction of the organism as a whole, possess a double set of chromosomes and are said to be *diploid*. One-half of the double set of chromosomes is derived from each parent. For example, the somatic cells of the human species contain 46 chromosomes consisting of two homologous sets of parental origin, with 23 chromosomes each. The latter are termed *haploid* sets. Each somatic cell contains two haploid sets.

The germ cells, or *gametes*, contain only a *single* haploid set. During the fertilization process a fusion of gametes from each parent occurs to form a diploid, consisting of two homologous haploid sets of paternal and maternal origin.

Replication of the somatic cells occurs by a process called *mitosis* (Fig. 1-2). The central event of mitosis is the exact duplication of each chromosome, followed by a division into two new cells, each of which receives a normal complement of chromosomes.

In terms of Fig. 1-2, the replication of chromosomes occurs at some time between stages 11 (telophase) and (b) to (d) (prophase). By the time the chromosomes become clearly visible they are already differentiated into two distinct strands called *chromatids*. As mitosis continues these become progressively shorter and thicker (stages [e] to [h]). The clear circles of Fig. 1-2 represent the *centromeres* which appear to have an essential function in the mitotic process. The two chromatids, into which each chromosome is split during the earlier phases of mitosis, continue to share a single centromere until the actual separation of chromatids occurs.

During the later stages of mitosis the chromatids separate, each with its own centromere, to produce two diploid sets of chromosomes (stages [i] and [j]). This is followed by the segregation of the two sets and the final division into two cells, each of which has a complete diploid set of chromosomes (stages [j] to [11]).

The gametes (ova and spermatozoa) of the higher organisms are formed by the diploid cells of the reproductive tract. They arise by a special kind of cell division called *meiosis*, in which the normal diploid set of chromosomes is reduced to a single haploid set.

Essentially, meiosis consists of two nuclear divisions occurring in rapid succession, *while the chromosomes divide only once* (Fig. 1-3). The result of this process is the formation of four cells each of which has a haploid number of chromosomes (Fig. 1-3).