The ABCs of Gene Cloning

Dominic W.S. Wong

The ABCs of Gene Cloning

Second Edition

Springer

Library of Congress Control Number: 2005931435

ISBN-10: 0-387-28663-2 e-ISBN: 0-387-28679-9
ISBN-13: 978-0387-28663-1

Printed on acid-free paper.

Camera ready copy provided by the author.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 2 1

springeronline.com

CONTENTS

Preface to the second edition

In the nine years since the First Edition, my contention remains that an effective approach to understand the subject of gene cloning is by learning the "vocabulary" and the "language". This book emphasizes the nuts and bolts on just how to do that - reading and speaking the language of gene cloning. It shows the readers how to distinguish between a gene and a DNA, to read and write a gene sequence, to talk intelligently about cloning, to read science news and to enjoy seminars with some degree of comprehension.

On the whole, the second edition is not any more advanced than the first, with the intent of keeping the book concise and not burdening the readers with unwarranted details. Nevertheless, changes were needed and new materials were incorporated in the revision. Part I has a new chapter to provide a tutorial on reading both prokaryotic and eukaryotic gene sequences. Part II consists of several additions, updating on new techniques and cloning vectors. The topics in Part III have been rearranged in separate sections - Part III now focuses on applications of gene cloning in agriculture, and Part IV is devoted entirely to applications in medicine. Chapters on gene therapy, gene targeting, and DNA typing have been thoroughly revised. Additional coverage is included on animal cloning and human genome sequencing. The heavy activity in rewriting and expanding Part IV reflects the rapid progress in the technology and the increased impact of gene cloning.

I enjoyed writing and revising this book with deep satisfaction. It has been an inspiring experience to witness the remarkable development in the field of gene cloning and the tireless dedication of thousands of scientists in making genes tick.

Preface to the first edition

Gene cloning has become a fast growing field with a wide-ranging impact on every facet of our lives. The subject of gene cloning could be intimidating to the novice with little formal training in biology. This book is not intended to give an elementary treatment of recombinant DNA technology, as there are already a number of books in this category. The objective of writing this book is to provide a genuine introduction in gene cloning for interested readers with no prior knowledge in this area to learn the vocabulary and acquire some proficiency in reading and speaking the "language".

In the process of writing this book, the author was continuously confronted with how to present the language of a complex field in a simple and accessible manner. I have chosen to devote Part I of this book to outlining some basic concepts of biology in a straightforward and accessible manner. My intention is to highlight only the essentials that are most relevant to understanding gene cloning. For those who want to pursue a thorough review of genetics or molecular biology, there are many excellent references available. Part II of the book describes cloning techniques and approaches used in microbial, plant, as well as mammalian systems. I believe that a discussion beyond microbes is a prerequisite to a better comprehension of the language and the practical uses of gene cloning. Part III describes selected applications in agriculture and food science, and in medicine and related areas. I have taken the approach to first introduce the background information for each application, followed by an example of cloning strategies published in the literature. The inclusion of publications is an efficient way to demonstrate how gene cloning is conducted, and relate it to the concepts developed in Parts I and II. Moreover, it enables the readers to "see" the coherent theme underlining the principles and techniques of gene cloning. Consistent with its introductory nature, the text is

extensively illustrated and the contents are developed in a logical sequence. Each chapter is supplemented with a list of review questions as a study-aid.

I hope that this book will succeed in conveying not only the wonderful language of gene cloning, but also a sense of relevance of this science in our everyday lives. Finally, I acknowledge the contributions of my teachers and colleagues, especially Dr. Carl A. Batt and Dr. Robert E. Feeney, to my persisting interest of biological molecules and processes. Special thanks are due to Dr. Eleanor S. Reimer who has been very supportive in making this book a reality.

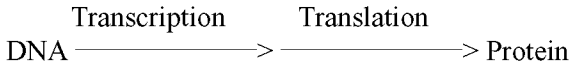
Part One

Fundamentals of Genetic Processes

CHAPTER 1

INTRODUCTORY CONCEPTS

The building blocks of all forms of life are cells. Simple organisms such as bacteria exist as single cells. Plants and animals are composed of many cell types, each organized into tissues and organs of specific functions. The determinants of genetic traits of living organisms are contained within the nucleus of each cell, in the form of a type of nucleic acids, called deoxyribonucleic acid (DNA). The genetic information in DNA is used for the synthesis of proteins unique to a cell. The ability of cells to express information coded by DNA in the form of protein molecules is achieved by a two-stage process of transcription and translation.



1.1 What is DNA and What is a Gene?

A DNA molecule contains numerous discrete pieces of information, each coding for the structure of a particular protein. Each piece of the information that specifies a protein corresponds to only a very small segment of the DNA molecule. Bacteriophage λ, a virus that infects bacteria, contains all its 60 genes in a single DNA molecule. In humans, there are ~31,000 genes organized in 46 chromosomes, complex structures of DNA molecules associated with proteins.

When, how, and where the synthesis of each protein occurs is precisely controlled. Biological systems are optimized for efficiency; proteins are made only when needed. This means that transcription and translation of a gene in the production of a protein are highly regulated by a number of control elements, many of which are themselves proteins. These regulatory proteins are in turn coded by a set of genes.

It is therefore more appropriate to define a gene as a functional unit. A gene is a combination of DNA segments that contain all the information necessary for its expression, leading to the formation of a protein. A gene defined in this context would include (1) the structural gene sequence that encodes the protein, and (2) sequences that are involved in the regulatory function of the process.

1.2 What is Gene Cloning?

Gene cloning is the process of introducing a foreign DNA (or gene) into a host (bacterial, plant, or animal) cell. In order to accomplish this, the gene is usually inserted into a vector (a small piece of DNA) to form a recombinant DNA molecule. The vector acts as a vehicle for introducing the gene into the host cell and for directing the proper replication (DNA -> DNA) and expression (DNA -> protein) of the gene (Fig. 1.1).

The process by which the gene-containing vector is introduced into a host cell is called "transformation". The host cell now harboring the foreign gene is a "transformed" cell.

The host cell carrying the gene-containing vector produces progeny all of which contain the inserted gene. These identical cells are called "clones".

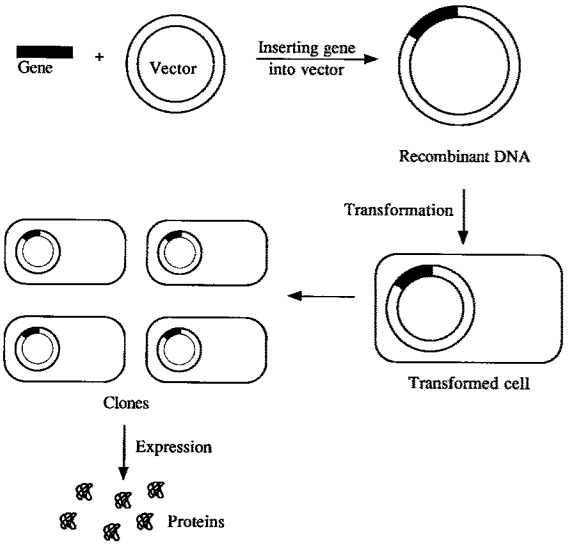


Fig. 1.1. General scheme of gene cloning.

In the transformed host cell and its clones, the inserted gene is transcribed and translated into proteins. The gene is therefore "expressed", with the gene product being a protein. The process is called "expression".

1.3 Cell Organization

Let us focus the attention for a moment on the organization and the general structural features of a cell, knowledge of which is required for commanding the language of gene cloning. Cells exist in one of two distinct types of arrangements (Fig. 1.2). In a simple cell type, there are no separate compartments for genetic materials and other internal structures.

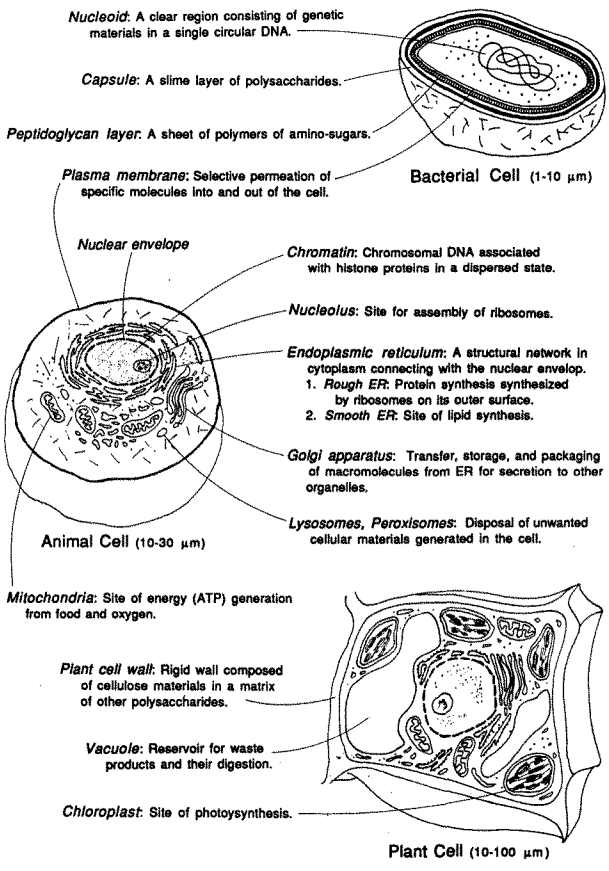


Fig. 1.2. Drawing of cells showing details of organelles.

Organisms with this type of cellular organization are referred to as prokaryotes. The genetic materials of prokaryotes, such as bacteria, are present in a single circular DNA in a clear region called nucleoid that can be observed microscopically. Some bacteria also contain small circular DNA molecules called plasmids. (Plasmids are the DNA used to construct vectors in gene cloning. See Section 9.1) The rest of the cell interior is called the cytoplasm and contains numerous minute spherical structures called ribosomes - the sites for protein synthesis. (Defined structures like ribosomes, are called organelles.) The rest (fluid portion) of the cytoplasm is the cytosol, a solution of chemical constituents that maintain various functions of the cell. All the intracellular materials are enclosed by a plasma membrane, a bilayer of phospholipids in which various proteins are embedded. In addition, some bacterial cells contain an outer layer of peptidoglycan (a polymer of amino-sugars) and a capsule (a slimy layer of polysaccharides).

In contrast, a vast majority of living species including animals, plants, and fungi, have cells that contain genetic materials in a membrane-bound nucleus, separated from other internal compartments which are also surrounded by membranes. Organisms with this type of cell organization are referred to as eukaryotes. The number and the complexity of organelles in eukaryotic cells far exceed those in bacteria (Fig 1.2). In animal cells all organelles and constituents are bound by a plasma membrane. In plants and fungi, there is an additional outer cell wall that is comprised primarily of cellulose. (In plant and fungal cells, the cell wall needs to be removed before a foreign DNA can be introduced into the cell in some cases as described in Sections 10.1 and 10.2.)

1.4 Heredity Factors and Traits

In an eukaryotic nucleus, DNA exists as complex with proteins to form a structure called chromatin (Fig. 1.3). During cell division, the fibrous-like chromatin condenses to form a precise number of well-defined structures called chromosomes, which can be seen clearly under a microscope.

Chromosomes are grouped in pairs by similarities in shape and length as well as genetic composition. The number of chromosome pairs varies among different species. For example, carrots have 9 pairs of chromosomes, humans have 23 pairs, and so on. The two similar chromosomes in a pair are described as homologous, containing genetic materials that control the same inherited traits. If a heredity factor (gene) that de-

termines a specific inherited trait is located in one chromosome, it is also found at the same location (locus) on the homologous chromosome. The two copies of a gene that are found in the same loci in a homologous chromosome pair are determinants of the same hereditary trait, but may exist in various forms (alleles). In simple terms, dominant and recessive alleles exist for each gene.

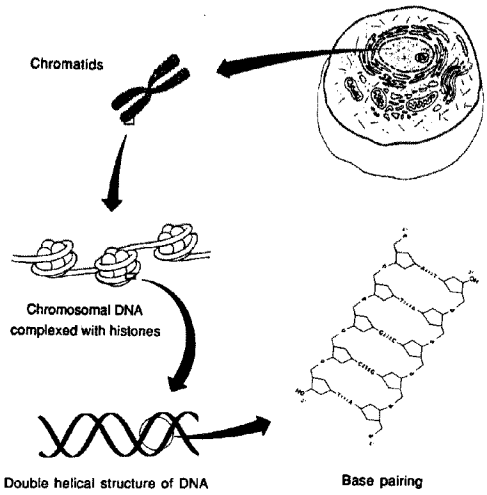


Fig. 1.3. Molecular structure of cellular chromosomes.

In a homologous chromosome pair, the two copies of a gene can exist in three types of combinations: 2 dominant alleles, 1 dominant and 1 recessive, or 2 recessives. Dominant alleles are designated by capital letters, and recessive alleles by the same letter but in lower case. For example, the shape of a pea seed is determined by the presence of the *R* gene. The dominant form of the gene is "*R*", and the recessive form of the gene is designated as "*r*". The homologous combination of the alleles can be one of the following: (1) *RR* (both dominant), (2) *Rr* (one dominant, one recessive) or (3) *rr* (both recessive). This genetic makeup of a heredity factor is called the genotype. A dominant allele is the form of a gene that is always expressed, while a recessive allele is suppressed in the presence of a dominant allele. Hence, in the case of the genotypes *RR* and *Rr*, the pea seeds acquire a round shape, and a genotype of *rr* will give a wrinkled seed. The observed appearance from the expression of a genotype is called its phenotype.

In our example, a pea plant with a genotype of *RR* or *Rr* has a phenotype of round shape seeds. When two alleles of a gene are the same

(such as *RR* or *rr*), they are called homozygous (dominant or recessive). If the two alleles are different (such as *Rr*), they are heterozygous. The genotypes and phenotypes of the offspring from breeding between, for example, two pea plants having genotypes of *Rr* (heterozygous) and *rr* (homozygous recessive), can be tracked by the use of a Punnett square (Fig. 1.4A). The offspring in the first generation will have genotypes of *Rr* and *rr* in a 1:1 ratio, and phenotypes of round seed and wrinkled seed, respectively.

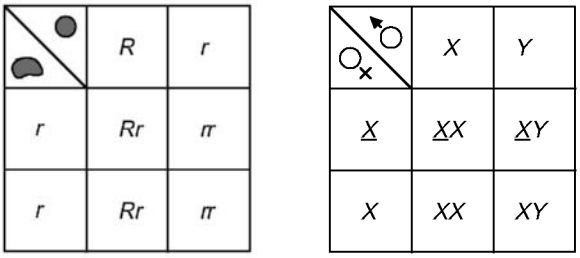


Fig. 1.4. Cross between (A) Rr and rr pea plants, and (B) carrier female and normal male.

The example of round/wrinkled shape of pea seeds is typical of one gene controlling a single trait. The situation is more complex in most cases, because many traits are determined by polygenes. Eye color, for example, is controlled by the presence of several genes. In some cases, a gene may exist in more than two allelic forms. Human ABO blood types are controlled by a gene with 3 alleles - I^A and I^B are codominant, and I^o is recessive. Additional variations are introduced by a phenomenon called crossing over (or recombination) in which a genetic segment of one chromosome is exchanged with the corresponding segment of the homologous chromosome during meiosis (a cell division process, see Sections 1.5 and 17.1).

A further complication arises from sex-linked traits. Humans have 23 pairs of chromosomes. Chromosome pairs 1 to 22 are homologous pairs, and the last pair contains sex chromosomes. Male has XY pair and female has XX chromosomes. The genes carried by the Y chromosome dictate the development of a male; the lack of the Y chromosome results in a female. A sex-linked gene is a gene located on a sex chromosome. Most known human sex-linked genes are located on the X chromosome, and thus are referred to as X-linked. An example of a sex-linked trait is color blindness, which is caused by a recessive allele on the X chromosome (Fig. 1.4B). If a carrier female is married to a normal male, the children

will have the following genotypes and phenotype- Sons: $\underline{X}Y$ (color blind) and XY (normal), and daughters: $\underline{X}X$ (normal, carrier) and XX (normal, non-carrier)

1.5 Mitosis and Meiosis

The presence of homologous chromosome pairs is the result of sexual reproduction. One member of each chromosome pair is inherited from each parent. In human and other higher organisms, autosomal cells (all cells except germ cells, sperms and eggs) contain a complete set of homologous chromosomes, one of each pair from one parent. These cells are called diploid cells ($2n$). Germ cells contain only one homolog of each chromosome pair, and are referred to as haploid (n).

A fundamental characteristic of cells is their ability to reproduce themselves by cell division - a process of duplication in which two new (daughter) cells arise from the division of an existing (parent) cell. Bacterial cells employ cell division as a means of asexual reproduction, producing daughter cells by binary fission. The chromosome in a parent cell is duplicated, and separated so that each of the two daughter cells acquires the same chromosome as the parent cell.

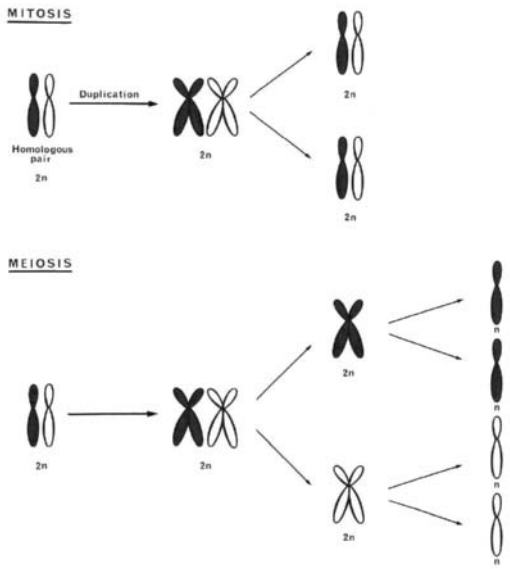


Fig. 1.5. Schematic comparison between mitosis and meiosis.

For eukaryotic cells, the process is not as straightforward. Two types of cell division, mitosis and meiosis, are identified. In mitosis, each chromosome is copied into duplicates (called chromatids) that are separate and partitioned into two daughter cells. Therefore, each of the two daughter cells receives an exact copy of the genetic information possessed by the parent cell (Fig. 1.5). Mitosis permits new cells to replace old cells, a process essential for growth and maintenance. In meiosis, the two chromatids of each chromosome stay attached, and the chromosome pairs are separated instead, resulting in each daughter cell carrying half of the number of chromosomes of the parent cell (Fig. 1.5). Note that at this stage, each chromosome in the daughter cells consists of 2 chromatids. In a second step of division, the chromatids split, resulting in 4 daughter cells each containing a haploid number of chromosomes, i.e. only one member of each homologous chromosome pair. Meiosis is the process by which germ cells are produced. After fertilization of an egg with a sperm, the embryo has complete pairs of homologous chromosomes.

1.6 Relating Genes to Inherited Traits

We will now examine how the preceding discussion of dominance and recessiveness, and genotypes and phenotypes, relate to how genes determine inherited traits. In simple terms, a gene can exist in a functional form, so that it is expressed through transcription and translation to yield a gene product (a specific protein) that exhibits its normal function. However, a gene can also be non-functional due to a mutation, for example, resulting in either the absence of a gene product, or a gene product that does not function properly. Therefore, a homozygous dominant genotype, such as *AA*, means that both alleles in the chromosome pair are functional. A genotype of *Aa* will still have one functional copy of the gene that permits the synthesis of the functional protein. A homozygous recessive (*aa*) individual does not produce the gene product (or produce a nonfunctional gene product). Since a gene controls an inherited trait via the synthesis of the protein it specifies, the presence or absence of a gene product due to a functional or nonfunctional (mutated) gene directly affect a particular inherited characteristic. Genes with multiple alleles can be explained by the difference in the efficiencies of the functions of the gene products. Another explanation is that one copy of the gene produces a lower amount of the gene product than the corresponding normal (functional) gene.

A simple example can be drawn from the genetic disorder of obesity in mice. Obese (*ob*) is an autosomal recessive mutation on mouse

chromosome 6. The normal gene encodes the Ob protein which functions in a signal pathway for the body to adjust its energy metabolism and fat accumulation (see Section 17.4). Mice carrying 2 mutant copies (*ob/ob*) of the gene develop progressive obesity with increased efficiency in metabolism (i.e. increase weight gain per calorie intake). Mice with *ob/ob* genotype apparently do not produce the gene product (Ob protein), because both copies of the *ob* gene are nonfunctional.

1.7 Why Gene Cloning?

The general objective of gene cloning is to manipulate protein synthesis. There are several reasons why we want to do this.

(1) To produce a protein in large quantity. Large-scale production of therapeutic proteins has been a primary locus of biotechnology. Many proteins of potential therapeutic values are often found in minute amounts in biological systems. It is not economically feasible to purify these proteins from their natural sources. To circumvent this, the gene of a desirable protein is inserted into a suitable host system that can efficiently produce the protein in large quantities. Examples of pharmaceuticals of this type include human insulin, human growth hormone, interferon, hepatitis B vaccine, tissue plasminogen activator, interleukin-2, and erythropoietin. Another area of great interest is the development of "transpharmers". The gene of a pharmaceutical protein is cloned into livestock animals, and the resulting transgenic animals can be raised for milking the protein.

(2) To manipulate biological pathways. One of the common objectives in gene cloning is to improve crop plants and farm animals. This often involves alteration of biological pathways either by (A) blocking the production of an enzyme, or (B) implementing the production of an exogenous (foreign) enzyme through the manipulation of genes. Many applications of gene cloning in agriculture and food belong to the first category. A well-known example is the inhibition of the breakdown of structural polymers in tomato plant cell wall, by blocking the expression of the gene for the enzyme involved in the breakdown process (using antisense technique). The engineered tomatoes, with decreased softening, can be left to ripe on the vine, allowing full development of color and flavor. Another example is the control of ripening by blocking the expression of the enzyme that catalyzes the key step in the formation of the ripening hormone, ethylene.

On the other hand, new functions can be introduced into plants and animals by introducing a foreign gene for the production of new proteins