现代生物技术前沿

From Genes to Genomes:
Concepts and Applications of
DNA Technology

〔美〕J. W. 戴尔 M. 冯尚茨 著

从基因到基因组
——DNA技术概念和应用

（影印版）

科学出版社
www.sciencep.com

现代生物技术前沿

From Genes to Genomes:
Concepts and Applications of DNA Technology

〔美〕J. W. 戴尔 M. 冯尚茨 著

从基因到基因组
——DNA技术概念和应用

（影印版）

科学出版社
北京

图字 01-2003-7121

内 容 简 介

基因技术、遗传工程、重组 DNA 和基因克隆等技术的快速发展，把分子生物学推到了生命科学的前沿。本书对分子生物学核心技术的原理及其使用进行了简明扼要的阐释。以对分子生物学的基本概念的简短介绍开篇，然后介绍核心的分子生物学方法及其整合运用，从单个基因的克隆与研究到整个基因组的测序，以及基因组信息的分析。最后，本书还介绍了这些技术在生物技术、医药和农业以及整个生命科学研究中的应用。

本书内容涵盖了分子生物学基础、基因克隆、核酸的纯化与分离、DNA 的剪切与连接、载体、基因组文库和 cDNA 文库、聚合酶链反应、DNA 测序、序列数据分析、基因变异分析、基因表达分析、基因功能分析、医药应用、转基因等等方面。

本书适于生物化学、分子生物学、遗传学、细胞生物学、生物技术、生物工程等相关学科及研究领域的本科生、研究生及教学科研人员参考使用。

书名原文：

From Genes to Genomes: Concepts and Applications of DNA Technology

图书在版编目(CIP)数据

从基因到基因组——DNA 技术概念和应用/(英)戴尔(Dale, J. W.)，冯尚茨(von Schantz, M.)著. —影印本. —北京：科学出版社，2004.1
(现代生物技术前沿)
ISBN 7-03-012477-4

Ⅰ. 从… Ⅱ. ①戴… ②冯… Ⅲ. 基因-理论-英文 Ⅳ. Q343.1

中国版本图书馆 CIP 数据核字(2003)第 124197 号

责任编辑：莫结胜
责任印制：刘士平/封面设计：王 浩

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
丽源印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2004 年 1 月第 一 版 开本：B5 (720×1000)
2004 年 1 月第一次印刷 印张：23 1/4
印数：1—4 000 字数：454 000

定价：46.00 元

(如有印装质量问题，我社负责调换〈新欣〉)

Preface

Over the last 30 years, a revolution has taken place that has put molecular biology at the heart of all the biological sciences, and has had extensive implications in many fields, including the political arena. A major impetus behind this revolution was the development of techniques that allowed the isolation of specific DNA fragments and their replication in bacterial cells (*gene cloning*). These techniques also included the ability to engineer bacteria (and subsequently other organisms including plants and animals) to have novel properties, and the production of pharmaceutical products. This has been referred to as *genetic engineering*, *genetic manipulation*, and *genetic modification* – all meaning essentially the same thing. However, many of the applications extend further than that, and do not involve cloning of genes or genetic modification of organisms, although they draw on the knowledge derived in those ways. This includes techniques such as *nucleic acid hybridization* and the *polymerase chain reaction* (*PCR*), which can be applied in a wide variety of ways ranging from the analysis of differentiation of tissues to forensic applications of DNA fingerprinting and the diagnosis of human genetic disorders. In an attempt to cover this range of techniques and applications, we have used the term *DNA technology* in the subtitle.

The main title of the book, *From Genes to Genomes*, is derived from the progress of this revolution. It signifies the move from the early focus on the isolation and identification of specific genes to the exciting advances that have been made possible by the sequencing of complete genomes. This has in turn spawned a whole new range of technologies (*post-genomics*) that are designed for genome-wide analysis of gene structure and expression, including computer-based analyses of such large data sets (*bioinformatics*).

The purpose of this book is to provide an introduction to the concepts and applications of this rapidly-moving and fascinating field. In writing this book, we had in mind its usefulness for undergraduate students in the biological and biomedical sciences (who we assume will have a basic grounding in molecular biology). However, it will also be relevant for many others, ranging from research workers who want to update their knowledge of related areas to

anyone who would like to understand rather more of the background to current controversies about the applications of some of these techniques.

We wish to acknowledge the valuable feedback received from anonymous reviewers who read the manuscript, and students at the School of Biomedical and Life Sciences at the University of Surrey who used it as a textbook. Needless to say, any remaining errors and omissions are our own. Last but not least, we wish to thank our editor, Nicky McGirr, for her help and encouragement.

Jeremy W Dale
Malcolm von Schantz

Contents

1 Introduction

This book is about the study and manipulation of nucleic acids, and how this can be used to answer biological questions. Although we hear a lot about the commercial applications, in particular (at the moment) the genetic modification of plants, the real revolution lies in the incredible advances in our understanding of how cells work. Until about 30 years ago, genetics was a patient and laborious process of selecting variants (whether of viruses, bacteria, plants or animals), and designing breeding experiments that would provide data on how the genes concerned were inherited. The study of human genetics proceeded even more slowly, because of course you could only study the consequences of what happened naturally. Then, in the 1970s, techniques were discovered that enabled us to cut DNA precisely into specific fragments, and join them together again in different combinations. For the first time it was possible to isolate and study specific genes. Since this applied equally to human genes, the impact on human genetics was particularly marked. In parallel with this, *hybridization* techniques were developed that enabled the identification of specific DNA sequences, and (somewhat later) methods were introduced for determining the sequence of these bits of DNA. Combining those advances with automated techniques and the concurrent advance in computer power has led to the determination of the full sequence of the human genome.

This revolution does not end with understanding how genes work and how the information is inherited. Genetics, and especially modern molecular genetics, underpins all the biological sciences. By studying, and manipulating, specific genes, we develop our understanding of the way in which the products of those genes interact to give rise to the properties of the organism itself. This could range from, for example, the mechanism of motility in bacteria to the causes of human genetic diseases and the processes that cause a cell to grow uncontrollably giving rise to a tumour. In many cases, we can identify precisely the cause of a specific property. We can say that a change in one single base in the genome of a bacterium will make it resistant to a certain antibiotic, or that a change in one base in human DNA could cause debilitating disease. This only scratches the surface of the power of these techniques, and indeed this book can only provide an introduction to them. Nevertheless, we hope that by the time

you have studied it, you will have some appreciation of what can be (and indeed has been) achieved.

Genetic manipulation is traditionally divided into *in vitro* and *in vivo* work. Traditionally, investigators will first work *in vitro*, using enzymes derived from various organisms to create a *recombinant DNA molecule* in which the DNA they want to study is joined to a *vector*. This recombinant vector molecule is then processed *in vivo* inside a *host* organism, more often than not a strain of the *Escherichia coli* (*E. coli*) bacterium. A *clone* of the host carrying the foreign DNA is grown, producing a great many identical copies of the DNA, and sometimes its products as well. Today, in many cases the *in vivo* stage is bypassed altogether by the use of PCR (polymerase chain reaction), a method which allows us to produce many copies of our DNA *in vitro* without the help of a host organism.

In the early days, *E. coli* strains carrying recombinant DNA molecules were treated with extreme caution. *E. coli* is a bacterium which lives in its billions within our digestive system, and those of other mammals, and which will survive quite easily in our environment, for instance in our food and on our beaches. So there was a lot of concern that the introduction of foreign DNA into *E. coli* would generate bacteria with dangerous properties. Fortunately, this is one fear that has been shown to be unfounded. Some natural *E. coli* strains *are* pathogenic – in particular the O157:H7 strain which can cause severe disease or death. By contrast, the strains used for genetic manipulation are harmless disabled laboratory strains that will not even survive in the gut. Working with genetically modified *E. coli* can therefore be done very safely (although work with *any* bacterium has to follow some basic safety rules). However, the most commonly used type of vector, plasmids, are shared readily between bacteria; the transmission of plasmids between bacteria is behind much of the natural spread of antibiotic resistance. What if our recombinant plasmids were transmitted to other bacterial strains that *do* survive on their own? This, too, has turned out not to be a worry in the majority of cases. The plasmids themselves have been manipulated so that they cannot be readily transferred to other bacteria. Furthermore, carrying a gene such as that coding for, say, dogfish insulin, or an artificial chromosome carrying 100 000 bases of human genomic DNA is a great burden to an *E. coli* cell, and carries no reward whatsoever. In fact, in order to make them accept it, we have to create conditions that will kill all bacterial cells *not* carrying the foreign gene. If you fail to do so when you start your culture in the evening, you can be sure that your bacteria will have dropped the foreign gene the next morning. Evolution in progress!

Whilst nobody today worries about genetically modified *E. coli*, and indeed diabetics have been injecting genetically modified insulin produced by *E. coli* for decades, the issue of genetic engineering is back on the public agenda, this time pertaining to higher organisms. It is important to distinguish the *genetic*

modification of plants and animals from *cloning* plants and animals. The latter simply involves the production of genetically identical individuals; it does not involve any genetic modification whatsoever. (The two technologies can be used in tandem, but that is another matter.) So, we will ignore the cloning of higher organisms here. Although it is conceptually very similar to producing a clone of a genetically modified *E. coli*, it is really a matter of reproductive cell biology, and frankly relatively uninteresting from the molecular point of view. By contrast, the genetic modification of higher organisms is both conceptually similar to the genetic modification of bacteria, and also very pertinent as it is a potential and, in principle, fairly easy application following the isolation and analysis of a gene.

At the time of writing, the ethical and environmental consequences of this application are still a matter of vivid debate and media attention, and it would be very surprising if this is not still continuing by the time you read this. Just as in the laboratory, the genetic modification as such is not necessarily the biggest risk here. Thus, if a food crop carries a gene that makes it tolerant of herbicides (weedkillers), it would seem reasonable to worry more about increased levels of herbicides in our food than about the genetic modification itself. Equally, the worry about such an organism escaping into the wild may turn out to be exaggerated. Just as, without an evolutionary pressure to keep the genetic modification, our *E. coli* in the example above died out overnight, it appears quite unlikely that a plant that wastes valuable resources on producing a protein that protects it against herbicides will survive long in the wild in the absence of herbicide use.

Nonetheless, this issue is by no means as clear-cut as that of genetically modified bacteria. We cannot test these organisms in a contained laboratory. They take months or a year to produce each generation, not 20 minutes as *E. coli* does. And even if they should be harmless in themselves, there are other issues as well, such as the one exemplified above. Thus, this is an important and complicated issue, and to understand it fully you need to know about evolution, ecology, food chemistry, nutrition, and molecular biology. We hope that reading this book will be of some help for the last of these. We also hope that it will convey some of the wonder, excitement, and intellectual stimulation that this science brings to its practitioners. What better way to reverse the boredom of a long journey than to indulge in the immense satisfaction of constructing a clever new screening algorithm? Who needs jigsaw and crossword puzzles when you can figure out a clever way of joining two DNA fragments together? And how can you ever lose the fascination you feel about the fact that the drop of enzyme that you're adding to your test tube is about to manipulate the DNA molecules in it with surgical precision?

2 Basic Molecular Biology

In this book, we assume you already have a working knowledge of the basic concepts of molecular biology. This chapter serves as a reminder of the key aspects of molecular biology that are especially relevant to this book.

2.1 Nucleic Acid Structure

2.1.1 The DNA backbone

Manipulation of nucleic acids in the laboratory is based on their physical and chemical properties, which in turn are reflected in their biological function. Intrinsically, DNA is a very stable molecule. Scientists routinely send DNA samples in the post without worrying about refrigeration. Indeed, DNA of high enough quality to be cloned has been recovered from frozen mammoths and mummified Pharaohs thousands of years old. This stability is provided by the robust repetitive phosphate–sugar backbone in each DNA strand, in which the phosphate links the 5′ position of one sugar to the 3′ position of the next (Figure 2.1). The bonds between these phosphorus, oxygen, and carbon atoms are all *covalent bonds*. Controlled degradation of DNA requires enzymes (nucleases) that break these covalent bonds. These are divided into *endonucleases*, which attack internal sites in a DNA strand, and *exonucleases*, which nibble away at the ends. We can for the moment ignore other enzymes that attack for example the bonds linking the bases to the sugar residues. Some of these enzymes are non-specific, and lead to a generalized destruction of DNA. It was the discovery of *restriction endonucleases* (or *restriction enzymes*), which cut DNA strands at specific positions, that opened up the possibility of *recombinant DNA technology* ('*genetic engineering*'), coupled with *DNA ligases*, which can join two double-stranded DNA molecules together.

RNA molecules, which contain the sugar ribose (Figure 2.2), rather than the deoxyribose found in DNA, are less stable than DNA. This is partly due to their greater susceptibility to attack by nucleases (*ribonucleases*), but they are also more susceptible to chemical degradation, especially by alkaline conditions.

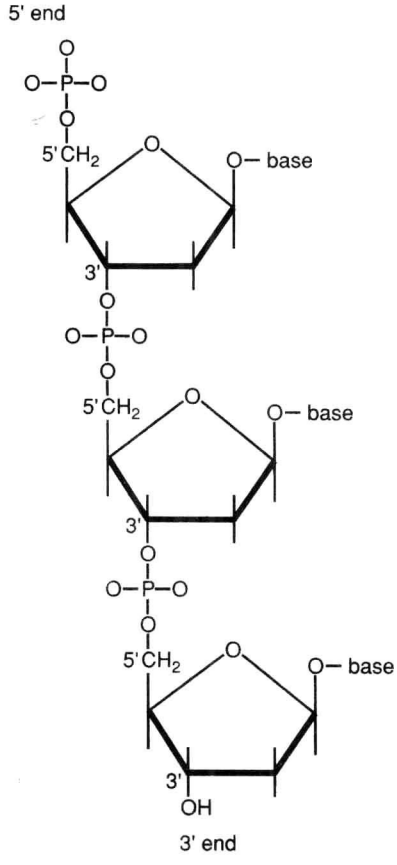

Figure 2.1 DNA backbone

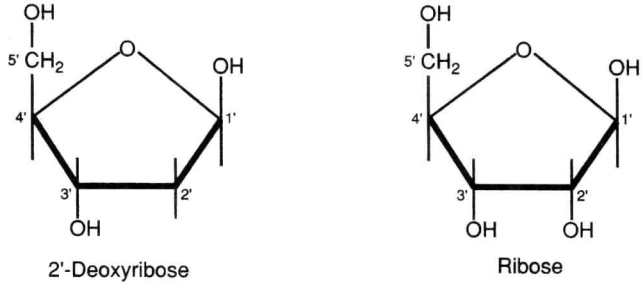

Figure 2.2 Nucleic acid sugars