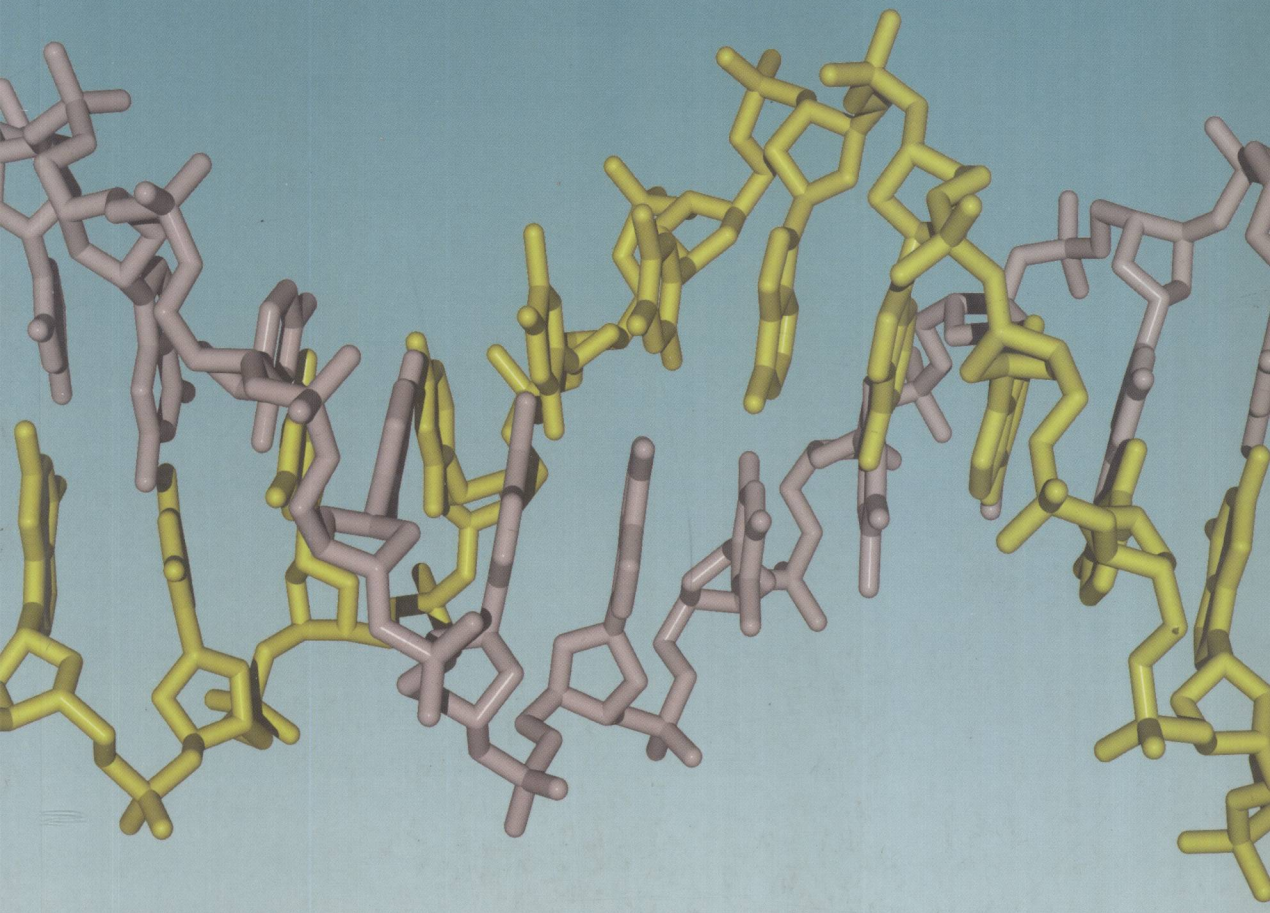
Gene Cloning

Principles and Applications



Julia Lodge, Pete Lund & Steve Minchin

Gene Cloning

Julia Lodge, Pete Lund & Steve Minchin

School of Biosciences
University of Birmingham
Edgbaston
Birmingham
UK

Taylor & Francis
Taylor & Francis Group

Published by:

Taylor & Francis Group

In US: 270 Madison Avenue
New York, N Y 10016
In UK: 2 Park Square, Milton Park
Abingdon, OX14 4RN

ISBN: 0-7487-6534-4

A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Lodge, Julia.
Gene cloning : principles and applications / Julia Lodge, Pete Lund & Steve Minchin.
p. ; cm.
Includes bibliographical references and index.
ISBN 0-7487-6534-4 (alk. paper)
1. Molecular cloning. I. Lund, Peter A. II. Minchin, Steve. III. Title.
[DNLM: 1. Cloning, Molecular. 2. Gene Library. 3. Genomics --methods. QU 450 L822g 2007]
QH442.2.L63 2007
660.5'5--dc22

Editor: Elizabeth Owen
Editorial Assistant: Kirsty Lyons
Production Editor: Georgina Lucas
Typeset by: Phoenix Photosetting, Chatham, Kent, UK
Printed by: MPG BOOKS Limited, Bodmin, Cornwall, UK

Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

T&F informa

Taylor & Francis Group, an informa business

Visit our web site at http://www.garlandscience.com

Gene Cloning

Contents

1 Introduction

1.1 The Beginning of Gene Cloning

In November 1973, a five-page paper was published in the prestigious journal *Proceedings of the National Academy of Sciences USA* by Stanley Cohen, Annie Chang, Herb Boyer and Robert Helling from Stanford University in California. The title of the paper was "Construction of biologically functional plasmids *in vitro*", and it described for the first time the production of an organism into which a DNA molecule had been introduced which consisted of DNA sequences from two different sources, joined together in the test tube. Although this work itself built on an earlier body of research, it may justifiably be seen as the paper which marked the birth of a scientific revolution which has continued to this day.

Great changes in science come about in different ways. Sometimes, they are the result of new concepts that transform our way of looking at things, or give us new insights into areas of knowledge which had previously been obscure. Such a revolution in biology had already occurred in the two decades before Cohen's paper, with the realization that the fundamental stuff of inheritance is DNA, with the discovery of DNA's remarkable structure, and with the unscrambling of the genetic code. Other dramatic changes in science have been more technical than conceptual, and are no less important for that. Cohen's paper describes the first methods for manipulating DNA in ways that began to give the experimenters a measure of control over these molecules, hence enabling manipulation of the genetic properties of the organisms that contain them. Humans have, of course, been selectively breeding organisms for particular traits for millennia: domestication of wild plants for crops was a hugely successful early experiment in genetic engineering. But with the advent of what are commonly called recombinant DNA techniques, the degree to which we can produce predetermined genetic changes with high precision has grown to the point where now it is commonplace to make bacteria or plants that produce human proteins, to tinker with the basic structures of enzymes to

alter their activity or stability, or to pull a single gene from the tens of thousands present in a human chromosome and identify within in it a single changed base that may give rise to a crippling genetic disease.

One of the hallmarks of the maturity of a technology is the extent to which it works its way through the system. It begins as the preserve of one or a few specialized laboratories; then, it becomes more widely available and used, and commercial applications begin to appear; ultimately, it makes its way onto undergraduate and even school curricula. Experiments that were once the material of Nobel prizes become the topic of routine practicals. For many years we have run, here in our own school, a practical for second-year undergraduates which is, in essence, not that dissimilar to the breakthrough experiment described by Cohen and his colleagues in 1973. Many hundreds of our undergraduates have become gene cloners by the end of their second year at university, and this is also true in thousands of institutions all over the world, including probably the one where you are studying. In addition many aspects of the biology that are taught are only known and understood today because of the incredible power that recombinant DNA methods give us to answer fundamental questions about the nature of life and the functions of cells and organisms. Many of our graduates go on to use these methods extensively in their own careers as research scientists in academia and industry.

If the basic technique described by Cohen *et al.* was all there was to recombinant DNA methods, our life as academics teaching molecular biology would be very simple. But, in fact, Cohen and his colleagues were in many ways the Wright brothers of the field, and in little more than three decades since they published their paper, we have moved from fragile biplanes to jumbo jets. Today's research laboratories have access to hundreds of different approaches to biological investigation which use aspects of recombinant DNA techniques, and whole industries are founded upon their exploitation. Methods have been introduced and refined at a dizzying pace that shows no obvious sign of relenting. Some – such as the ability to exponentially amplify vanishingly small amounts of DNA in a test tube, to manipulate the germ line of complex multicellular organisms, or to determine the complete sequences of the genomes of many organisms – have been revolutions in their own right. Others represent incremental improvements to basic techniques which have nonetheless transformed complex methods into processes that can be done using off-the-shelf kits, or (increasingly) performed by robots. This presents something of a problem for us as teachers, or rather two problems. The first is that the pace of change is such that it is difficult to know what to include and what to omit from undergraduate courses on the subject, and hard to find text books at the right level that are up to date. The second is that the essentially technical nature of the recombinant DNA revolution means it is important to present the subject in such a way that it is not just a dry list of methods, but

which also conveys a sense of the excitement and insight that these approaches have brought to so many different areas, not only in the research laboratory but also in everyday applications. Hence the book that you are now reading. In it we have tried to present a selection of what we regard as the key concepts and methods that underlie gene cloning, at a level which should be easily understandable to a typical undergraduate in a bioscience or medical subject, and to illustrate these as much as possible with examples drawn from the laboratories of universities and companies around the world. Our aim throughout has been to be as comprehensive as possible both with basic methods and with their more advanced applications, subject to space constraints. Inevitably, we have had to be selective in the material that we have covered, and even during the course of writing the book we have had to go back and revise or add to early material as new methods have been published. But we believe that the major aspects of the subject are all here, presented in a form that you will find easy to understand, and which will interest and enthuse those of you that read and use it.

1.2 How To Use This Book

The layout of the book is quite traditional, with the different chapters dealing with methods and concepts of increasing complexity through the book. Although we have made the individual chapters self-contained, and used extensive cross-referencing between chapters, we expect most people will start at the beginning and work their way through the book as needed according to the course they are studying. Each chapter starts with a list of "learning outcomes" – that is, a list of the things you should be able to do once you have read and understood the material in the chapters. These should help you to assess whether you have understood what the chapter is all about. By way of an introduction to the book we present some information about the way genomes are organized in both prokaryotic and eukaryotic organisms. There follows a group of chapters which present basic details about the enzymes and reactions used in simple gene manipulations, and then go on to talk about how genes are actually cloned and identified. We have gone into the details of how clones of particular genes are found, since our experience has been that this is an area that students often find difficult to understand. The advent of high-throughput genome sequencing and the consequent availability of huge amounts of gene sequence data online means that approaches to gene cloning have changed a lot in recent years, but we feel it is still important for you to understand the "traditional" (i.e. more than 10 years old!) methods, even though the use of gene libraries is becoming less common.

It would be ridiculous in a book of this nature, however, not to give a good deal of weight to the topic of genomics (i.e. all aspects of studying organisms at the whole genome level), since this constitutes one of the more recent revolutions in the methodology of the biosciences. Two chapters

describe how DNA is sequenced and how the large amounts of sequence data deposited in international databases can be mined and analyzed – although we have not gone into this latter area in too much technical detail, since this is a whole new discipline in its own right and requires skills in mathematics and computer programming which are beyond the remit of this book.

We then turn to more applied aspects of recombinant DNA methods, including how cloned genes can be used as the source of large amounts of proteins, and how genes can be manipulated and introduced into higher organisms to produce so-called transgenic organisms, the uses of some of which are described as case studies. We discuss also some of the powerful research uses of these methods, such as deepening our understanding of how genes are regulated in cells, and enabling us to functionally dissect proteins. Finally, we conclude with a chapter which discusses some further applications, mainly in a medical context, to add to those used as illustrations in earlier chapters.

Although most of the text will be self-explanatory, assuming you have a degree of basic knowledge (the things we expect you to already know are listed in the next section), there are some places where particular general concepts seemed to us to be sufficiently important that we have put them in boxes, separate from the rest of the text.

One thing that you will notice in the book is the use of large numbers of examples, based on genuine experiments and published results, to illustrate the points that we are making. One of the features of molecular biology is that the methods can be applied to all living organisms, and you will find that in some cases, our examples will be based on bacterial systems (prokaryotes), and on others they will refer to eukaryotes, ranging from single-celled organisms such as yeast all the way to humans. We (the authors) have research and teaching experience both with prokaryotes and eukaryotes, and it has been our experience that it is often best to discuss the simpler prokaryotic systems first, to introduce basic concepts, before going on to talk about the more complex eukaryotes. At the end of each chapter, we have included references for the papers which are referred to in the case studies discussed in that chapter. In most cases, these are available online; if not, they should be in your institution's library. Reading these papers should add to your understanding of the methods and their applications discussed in this book. Some of the papers are quite straightforward, while others are complex and may be tricky to follow in places. However, learning to read the scientific literature is an essential part of any undergraduate degree, and we encourage you to read as many of these papers as you are able. A key feature of the book is the questions that are included in the text of each chapter. Some of these are simply designed to make sure that you have taken in what you have just read, by (for example) setting simple problems based on the previous sections. Others do require a bit of extra

thought, or the bringing together of several different topics. As it is only by trying to answer questions on it that you can really tell how good your understanding of a subject is, we encourage you to persevere with these questions, even if they appear difficult at first, before turning to our answers at the end of each chapter.

1.3 What You Need To Know Before You Read This Book

In writing this book, we have tried to pitch it roughly at a level that would be understandable by undergraduates in the UK in their second year, although some of the more advanced material would perhaps be left until the following final year, and these are the levels at which we have experience of teaching these topics. It is important to be clear, therefore, that this is not a textbook about fundamental concepts in genetics, cell biology, or biochemistry, and it is assumed that you will already know these before you start. We take it that anyone studying this book will be familiar with:

- The structure of DNA
- The nature of the genetic code
- The way in which information flows from DNA via RNA to proteins, and the basic nature of the mechanisms (transcription and translation) by which this happens
- The nature of proteins, including the way in which their structure determines their function, and their different roles in cells
- Basic cell biology of both prokaryotes and eukaryotes

If these are not areas that you are familiar with, then much of the material in this book will be hard to follow, and it would be better to study a more basic text first before trying to use the current book.

1.4 A Request From the Authors

We have tried very hard to make this book precise, informative, interesting and correct. Some of the material has been tested extensively on undergraduates here in Birmingham or has grown from material that we have been teaching for many years. Other material is relatively new, and has involved us in a great deal of research of our own, reading original papers and talking with people using methods with which we ourselves were not directly familiar. It is inevitable, however, that the book will contain flaws, and we genuinely do want to hear about these so that in the event that future editions are needed, we can incorporate any suggestions which are made by you for the benefit of other readers. If you have comments or corrections to make, do please send them by e-mail to gene_cloning@bham.ac.uk. We look forward to reading your comments, and we hope you find the book a valuable aid to studying the fascinating and important topic of gene cloning.

Further Reading

Construction of biologically functional bacterial plasmids *in vitro*. (1973) Cohen SN, Chang AC, Boyer HW and Helling RB. Proc Natl Acad Sci USA, Volume 70 Pages 3240–3244.

The first paper to describe the production of an organism that contained DNA sequences from two different sources.

2 Genome Organization

Learning outcomes:

By the end of this chapter you will have an understanding of:

- *the genomic organization of prokaryotes and eukaryotes, and in particular the human genome*
- *the different types of sequence within the eukaryotic genome: coding, non-coding, non-repetitive and repetitive*
- *the physical characteristics of chromosomes*
- *why an appreciation of genome organization is important in the context of gene cloning*

2.1 Introduction

The genome of an organism can be defined as "the total DNA content of the cell", and as such it contains all the genetic information required to direct the growth and development of the organism. For all multicellular organisms this growth and development starts from a single cell, the fertilized egg. In the case of humans the egg develops into an adult comprising approximately 10^{12} cells made up from over 200 different cell types.

As you will be aware the gene is the basic unit of biological information. Most genes code for a protein product: the gene is transcribed to RNA and this RNA messenger is then translated to the protein product. In addition to genes which encode proteins, there are many genes which encode stable RNAs such as ribosomal RNA and transfer RNA. The number of genes contained within the genome of an organism ranges from around 500 for the bacterium *Mycoplasma genitalium* to over 50,000, predicted to be present in most plants.

In bacteria the genetic information is normally carried on one circular DNA molecule referred to as the bacterial chromosome, which may be supplemented with several small self-replicating DNA molecules, also known as plasmids. Eukaryotic cells contain several linear chromosomes within the nucleus. Human cells, for example, contain 23 pairs of chromosomes. In addition to the DNA present in the nucleus, mitochondria and chloroplasts

contain DNA that encodes a fraction of the functions of these organelles. For multicellular organisms the genome content is identical for all cells with only a few exceptions (such as red blood cells, which contain no nuclei and hence no nuclear DNA).

Although the genome and gene content of an organism is necessary for the development and survival of that organism, it is not sufficient. There are important proteins in the fertilized egg whose function is to control how theses genes are used. Because of this no free-living organism could be created from its DNA alone. All cells present on Earth today have arisen from pre-existing cells. Genetic engineering cannot lead to the generation of novel organisms from basic components, genetic engineering can only modify the genetic make-up of pre-existing cells by adding or removing functions from the organism's genome.

In order to understand how to manipulate DNA it is important to understand the way it is organized in different organisms. In this chapter we will discuss the main features of the genome of higher eukaryotic organisms using the human genome as our primary example. We will also discuss bacterial and viral genomes so as to understand how they differ from those of eukaryotic organisms.

2.2 The C-value Paradox

The C-value is a measure of genome size, typically expressed in base pairs of DNA per haploid genome. The use of the term haploid genome refers to a single copy of all the genetic information present in the nucleus. Diploid nuclei of organisms produced sexually will of course contain two complete,

Table 2.1 Characteristics of the genomes of example organisms

Organism	Genome Size (bp)	Chromosome Number (*n*)	Predicted Number of Genes
Mycoplasma genitalium	580,000	1	500
Escherichia coli K12	4,639,000	1	4,500
Saccharomyces cerevisiae (yeast)	12,069,000	16	6,000
Caenorhabditis elegans (worm)	97,000,000	6	20,000
Drosophila melanogaster (fly)	137,000,000	6	15,000
Oryza sativa (rice)	420,000,000	12	40,000
Arabidopsis thaliana (weed)	115,000,000	5	28,000
Fugu rubripes (pufferfish)	390,000,000	22	25,000
Mouse	2,500,000,000	20	25,000
Humans	3,300,000,000	23	25,000