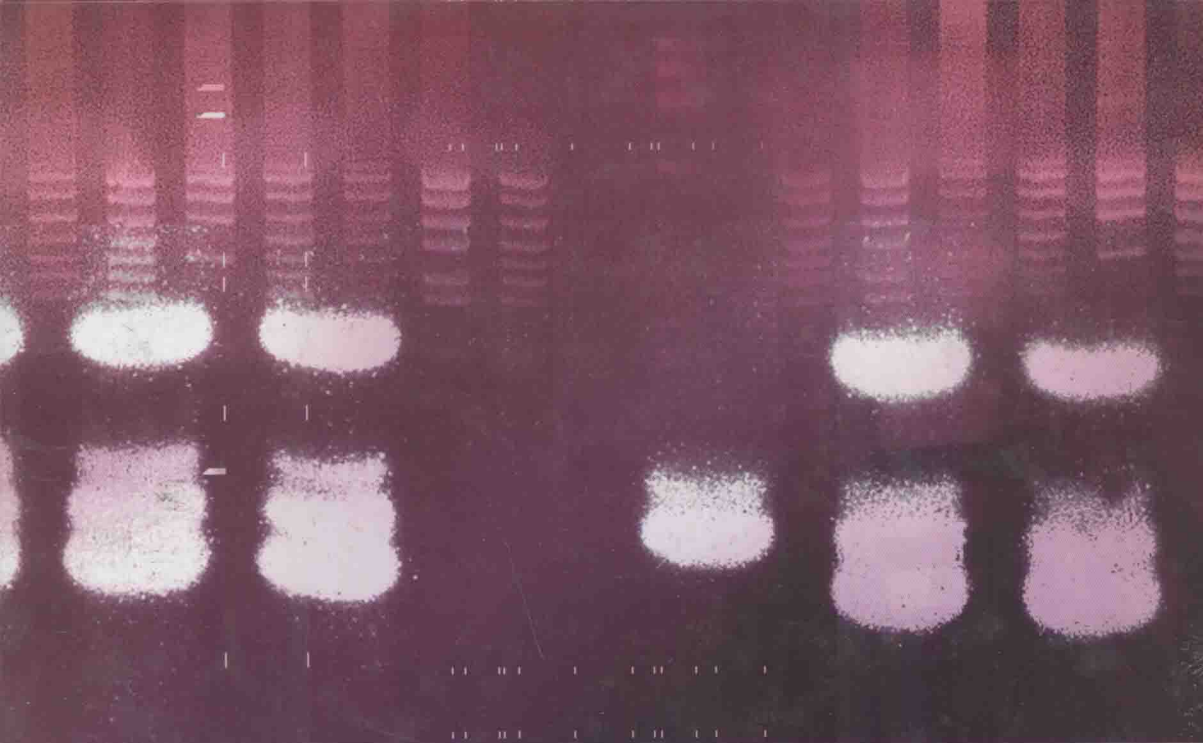


Edited by Jean-Louis Serre

Diagnostic Techniques in Genetics

WILEY

Diagnostic Techniques in Genetics

Edited by
Jean-Louis Serre
Université de Versailles

Translated by
Isabelle and Simon Heath

John Wiley & Sons, Ltd

First published in French as *Analyse de Génomes, Les Diagnostics Génétiques* © 2002 Dunod, Paris

Translated into English by Isabelle & Simon Heath.

This work has been published with the help of the French Ministère de la Culture-Centre national du livre

The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, Canada L5R 413

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Diagnostics génétiques. English.
Diagnostic techniques in genetics / edited by Jean-Louis Serre ;
translated by Isabelle and Simon Heath.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-470-87024-2 (cloth : alk. paper)
ISBN-10: 0-470-87024-9 (cloth : alk. paper)
ISBN-13: 978-0-470-87025-9 (pbk. : alk. paper)
ISBN-10: 0-470-87025-7 (pbk. : alk. paper)
1. Molecular genetics. 2. Molecular diagnosis. 3. Genetic screening.
I. Serre, Jean-Louis. II. Title.
[DNLM: 1. Genetic Techniques. 2. Molecular Diagnostic Techniques–methods.
3. Genetic Screening–methods. QZ 52 D536 2006a]
QH442.D5313 2006
616′.042–dc22 2006010655

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-87024-2 (HB) ISBN-13 978-0-470-87025-9 (PB)
ISBN-10 0-470-87024-9 (HB) ISBN-10 0-470-87025-7 (PB)

Typeset in 10.5/13pt Minion by TechBooks, New Delhi, India
Printed and bound in Great Britain by Antony Rowe Ltd., Chippenham, Wilts
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Preface

During the second half of the 20th century, in less than 50 years, there have been two revolutions in biology – one in protein biochemistry followed by one in molecular genetics – with major repercussions both for fundamental research and medical applications.

The development of protein biochemistry (at the end of the 1950s) has led to the purification of many proteins followed by the study – using crystallography – of their three-dimensional structure and, in parallel with this, their sequencing (identification of the linking of amino acids).

The study of variants for a peptide chain then allowed – by identification of amino acid substitutions, notably those caused by missense mutations in the coding sequence of a gene – the exhaustive definition of the causal relationships between each mutation, its molecular effect and its cellular, tissular and clinical effects in many pathologies. Of these, the most intensely studied relationship has been for the protein haemoglobin in the haemoglobinopathies.

Following this first revolution there were many medical applications, but these were unfortunately limited by the fact that the study of proteins can only be the study of gene products, which requires that these products are present and accessible for analysis.

The best example of this is the prenatal diagnosis of β-thalassaemia, a lethal haemoglobinopathy resulting from the absence of haemoglobin β-chains in children having loss of function mutations in both of their copies of the β-gene. The diagnosis was already challenging because it required taking a fetal blood sample *in utero* from a vein in the cord without contaminating it with maternal blood and without provoking a miscarriage and, above all, any success was tempered by the fact that because transcription of the β-gene only starts replacing that of the γ-genes in the fourth month of pregnancy, biochemical tests of the presence or absence of adult haemoglobin ($\alpha 2\beta 2$) in the essentially entirely fetal haemoglobin ($\alpha 2\lambda 2$) are pushed back to the fifth or sixth month, leading necessarily to late pregnancy terminations with obvious psychological, clinical and ethical consequences.

The consequences of the second revolution – in molecular genetics – have been more important in the domain of fundamental research than in its applications, simply because it gives direct access to the genes which direct biological phenomena

whereas before it was only possible, at best, to access their products or just the phenotypic perturbations of the phenomena resulting from mutations.

Molecular genetics has allowed the cloning of genes, their sequencing and, above all, the ability to isolate them *in vitro*, in vectors, to modify them in a targeted manner, and to reintroduce them into cells or organisms to study, in a controlled manner, the way in which they control one or other aspect of a phenomenon.

Genetics, in becoming molecular, has very rapidly become an integral part of most biological fields, not only due to its power to dissect a biological phenomenon but also because of its increasing ability to unify disciplines which had previously appeared different or unrelated.

In all disciplines such as physiology, embryology and oncology, development and aging, cell biology, population biology, ecology and evolution, the phenomena in question are associated with the expression of specific genes, which can now be localized and identified, cloned and modified at leisure by directed mutagenesis, and so the effects can be studied in a chosen context, *in vitro*, *ex* or *in vivo*. Molecular genetics is therefore a remarkable tool because potentially it allows the opening of all the 'black boxes' that in these disciplines represent pertinent, but essentially descriptive, approaches without taking account of the global view of the phenomena or the studied structures.

For human genetic pathologies, the molecular and cellular description of the physiopathology of a disease is the condition sine qua non for therapeutic research and, in future, for the possible development of pre- or post-natal genetic diagnosis or even genetic screening of heterozygotic carriers for recessive diseases.

Therefore, the question of the diagnosis of β-thalassaemia can nowadays be resolved without difficulty since direct study of the β-genes from fetal DNA shows whether the two copies are mutated or not in the first weeks of a pregnancy, well before protein biochemistry could indicate the presence or absence of haemoglobin A.

However, genetic diagnoses – mainly pre-natal for severe pathologies – are only one of the principal clinical applications of molecular biology. The molecular biology of DNA has also revolutionized many approaches such as the diagnosis of infectious diseases where the pathogenic agent, bacterium or virus, will be identified by the presence of sequences specific to its genome, or the traceability studies that can now be performed by searching for transgenic elements identifying genetically modified organisms (GMOs).

Even the screening for anomalies in chromosome number (monosomies, trisomies) that is normally performed by karyotyping, which is a classic and highly reliable technique, can sometimes be performed using quantitative techniques for estimating molecular dosage much faster (24 to 48 hours rather than 10 to 15 days) and much more cheaply, with nevertheless some limitations which will be described in this book.

In addition, some karyotypic anomalies, not accessible to conventional cytogenetics (microdeletions, uniparental disomies) can be detected or identified by

techniques such as fluorescent *in situ* hybridization (FISH) combining karyotyping and molecular biology.

Oncology itself relies heavily on molecular biology which has allowed, at the fundamental level, the identification of a very large number of genes implicated in tumour phenomena (oncogenes and antioncogenes) and, therefore, identification of mutations in these genes, either in the tumours of affected patients (of both diagnostic and prognostic value), or in healthy patients at risk from a hereditary form of cancer (screening and prevention).

Finally, there is a field where the applications of molecular biology are so revolutionary that they have contributed as much to the popularization of molecular biology as to their medical application; this field is that of genetic fingerprinting applied to forensic medicine (paternity testing or criminal identification).

This book therefore proposes, after a simple yet exhaustive description of the principal techniques of molecular biology, to review the applications in the different disciplines written by a recognized expert in that discipline, without neglecting ethical questions or those that are raised by society in regards to current and future applications.

Jean-Louis Serre
Université de Versailles

List of Contributors

Catherine Boileau, CHU Paris, Île-de-France Ouest

Nevine Boutros, Hôpital Saint-Vincent-de-Paul, Paris

Anne Casetta, Hôtel Dieu, Paris

Denis Cointe, Hôpital Antoine-Béclère, Clamart

Véronique David, CHU Pontchaillou, Rennes

Annick Diolez, Institut national de la recherché agronomique, Versailles

Emmanuelle Girodon, CHU Henri-Mondor, Créteil

Liliane Keros, CHU Paris-Sud, Antoine- Béclère, Clamart

Éric Le Guern, CHU Pitié-Salpétrière, Paris

Jean-Paul Moisan, CHU de Nantes

Étienne Mornet, Université de Versailles, Saint-Quentin-en-Yvelines

Olivier Pascal, CHU de Nantes

Véronique Pingaud, CHU Henri-Mondor, Créteil

Serge Pissard, CHU Henri-Mondor, Créteil

Jean-Louis Serre, Université de Versailles, Saint-Quentin-en-Yvelines

Brigitte Simon-Bouy, Université de Versailles, Saint-Quentin-en-Yvelines

Dominique Stoppa-Lyonnet, Institut Curie, Paris

Contents

1 Techniques and tools in molecular biology used in genetic diagnoses

Jean-Louis Serre, *Université de Versailles*

This chapter is not meant to be an exhaustive survey. Its purpose is to make the reader familiar with molecular biology techniques and tools used for genetic diagnosis. Some specific techniques will be described in the chapter describing the diagnoses for which they are used.

1.1 Nucleic acids

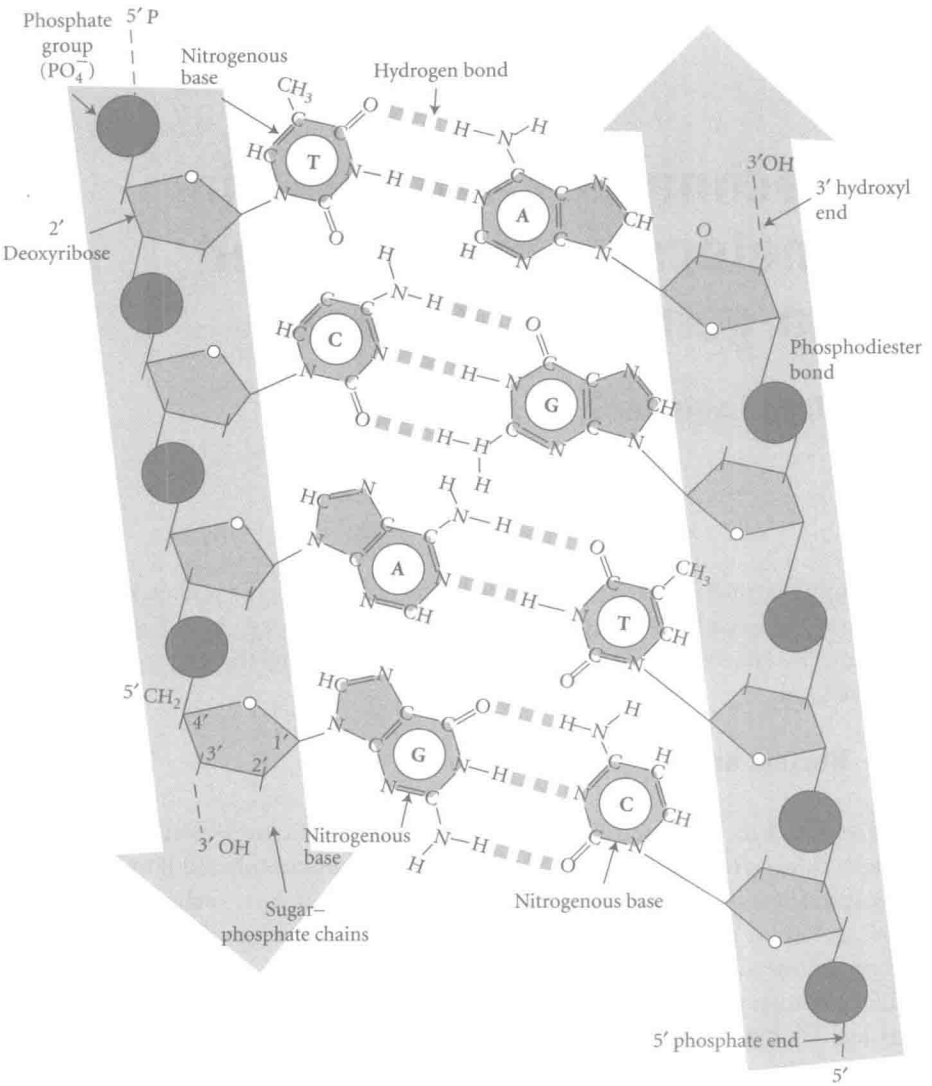
DNA and RNA molecules are linear molecules made up of a succession of nucleotides. A nucleotide consists of a sugar (pentose with five carbons numbered 1′ to 5′) carrying a phosphate on its 5′ carbon and a nitrogenous base on its 1′ carbon (Figure 1.1). RNA contains a ribose while DNA contains a deoxyribose (no hydroxyl on the 2′ carbon, Figure 1.1).

Each deoxyribose of the DNA carries one of the four following bases: adenine (A), guanine (G), cytosine (C) and thyamine (T); for RNA, each ribose carries one of the following bases: adenine (A), guanine (G), cytosine (C) and uracil (U, acting in place of the thyamine). Another nitrogenous base is sometimes present but at a very low frequency.

Nucleotides are linked by a phosphodiester (sugar-phosphate) bond (Figure 1.1) between the 5′ phosphate of one nucleotide and the 3′ hydroxyl of the previous nucleotide. DNA and RNA are synthesized from the 5′ to the 3′ end in order to keep the 5′P and the 3′OH extremities in the same orientation (Figure 1.1).

In most organisms (plants, animals and some bacteria), the DNA molecules carrying the genetic information consist of two complementary and anti-parallel strands

Diagnostic techniques in genetics Edited by Jean-Louis Serre



A DNA molecule consists of two strands held together by hydrogen bonds. Two hydrogen bonds in A/T base pairs and three in C/G pairs assuring the cohesion of the double helix also called a DNA duplex. Each strand consists of a nucleotide chain. The two strands are perfectly complementary. The two strands run antiparallel to each other.

Figure 1.1 Unwound structure of DNA (after J.-L. Rossignol (2000). *Génétique*. Dunod, Paris)

(Figure 1.1) linked by weak hydrogen bonds while RNA molecules are only single stranded. Bacteria and viruses are more diverse and have DNA or RNA genomes, single stranded or double stranded.

The two strands of a DNA molecule are complementary, meaning A is always associated with T and G is always associated with C, and vice versa. The complementary

Box 1.1 DNA molecules consist of two complementary strands

The two consecutive strands of the DNA always put a thymine in front of an adenine or a cytosine with a guanine, in such a way that knowing the sequence of one strand allows the deduction of the other one. This complementarity is a property used in natural mechanisms such as replication, transcription and translation but also in all *in vitro* technologies based on molecular hybridization, the capability of two single-stranded complementary sequences to associate and form a double-strand sequence.

nature of the two strands underlies the mechanism of replication: once the strands of a DNA molecule are separated, two new complementary strands are synthesized using each of the existing strands as a template.

The complementary nature of the two strands also underlies all molecular biology techniques using 'hybridization', meaning the ability of two complementary DNA strands to 'renature', or to 'hybridize' – the double-strand structure being thermodynamically more stable (lower free-energy level than the single-strand structure).

The two complementary strands of a DNA molecule are assembled in an antiparallel fashion, in an inverse orientation so that the 5′ and 3′ extremities are in opposite directions (Figure 1.1). This inversion of the 5′/3′ polarity of the strands is the basis for the helical structure of the DNA and for the different conformations of the binding sites for regulatory factors. It also provides information for the transcription machinery to distinguish between the coding and the non-coding strand at the level of the promoter. In addition, the translation machinery is affected by the 5′/3′ orientation of the messenger RNA, and knows how to identify useful sequences like the translation initiation codon. Therefore, when doing molecular biology, it is important to keep in mind that DNA fragments (genomic or not) which go through renaturation (hybridization) not only have complementary sequences but also opposite 5′/3′ polarity.

It is useful to remember that DNA synthesis, as well as RNA synthesis, consists of primer extension (RNA primer synthesized by a primase during replication) from its 3′OH extremity. Polymerization occurs in the 5′ to 3′ direction, using a DNA (or an

Box 1.2 The two strands of DNA molecules are also antiparallel

The 5′ phosphate extremity of a strand is always facing the 3′OH extremity of the other in a way that the sugar–phosphate chains are orientated in a reverse or antiparallel orientation. This reverse orientation of the two DNA strands constitutes a property used in the orientation of natural mechanisms of replication, transcription and translation, but also in most of the technologies based on molecular hybridization such as PCR, sequencing and cloning.

RNA) strand as a model, and to which the primer is hybridized in the opposite direction (3′ to 5′). The duplex formed is complementary and has the opposite polarity.

Apart from the messenger RNAs that are transiently present during gene expression, other RNAs belong to the cellular machinery and constitute the 'machine tools' involved in gene expression and its regulation (ribosomal and transfer RNA for translation, nuclear RNA for RNA maturation and transport). RNA molecules are single stranded, although they often fold to form partial duplex domains.

1.2 The different types of genetic material studied

For diagnosis, molecular biologists mostly study and manipulate DNA, but RNA can also be studied, for example in virology where the genomes of many viruses are made of RNA.

1.2.1 DNA origins and types

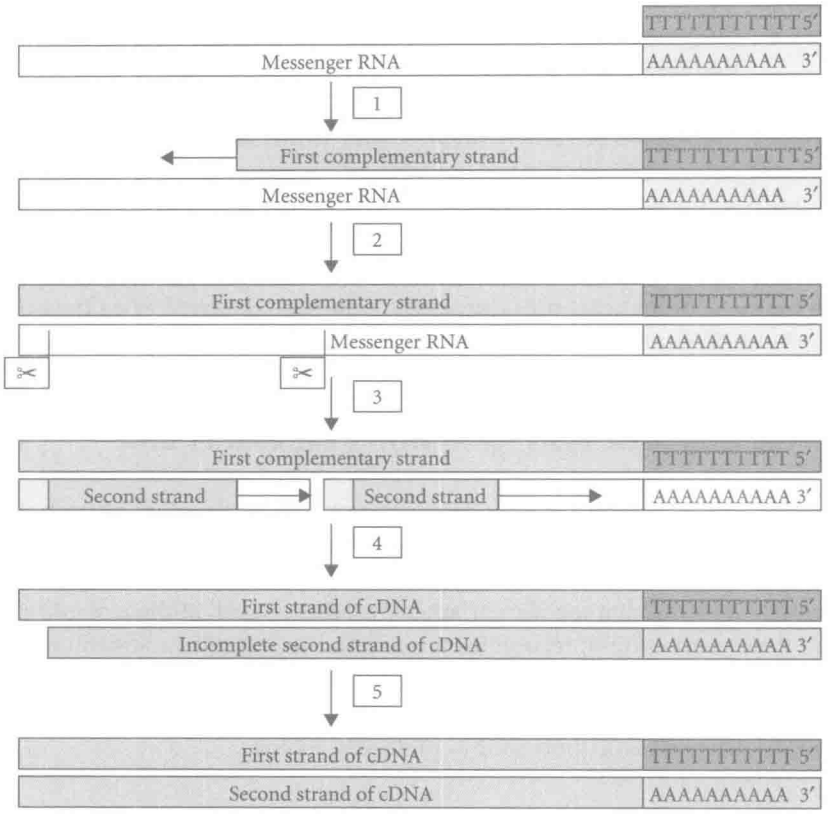
Individuals are formed of clones of cells, each cell containing the constitutional genome, the totality of the genetic information present at conception in the cell (zygote) arising from the fusion of the gametes. The DNA from the constitutional genome, also called 'constitutional DNA', can be extracted from all healthy tissue biopsies. In most cases in man, it is extracted from lymphocytes after a blood test.

It is also useful to specify the origin of the DNA when it comes from a particular tissue sample, for example 'somatic' DNA as opposed to 'germinal' DNA or 'fetal' DNA (constitutional DNA extracted from fetal cells taken from an embryo *in vitro* or *in vivo*, amniotic cells taken from amniocentesis after 14 weeks of amenorrhoea, or trophoblastic cells obtained through biopsy after 11–12 weeks of amenorrhoea). Likewise, as DNA from tumour cells has local mutations compared to constitutional DNA (the tumourigenesis resulting from multiple mutations affecting a certain number of genes), DNA extracted from tumour tissue will be called 'tumour' as opposed to 'constitutional' DNA extracted from healthy somatic tissue.

1.2.2 RNA and cDNA

These are often the molecules of interest for molecular biologists within the framework of research or diagnostics, messenger RNA or viral RNA (see RT–PCR technique).

Isolation of messenger RNA is well understood in genes that code for peptide chains. These mRNAs, for the most part, carry a long polyA tail, by which they can be trapped within an affinity column containing a polyT carrying resin, binding in this column to the free oligodT (the synthesized form of a oligodeoxynucleotide from a thymidine mono-strand).



1. The reverse transcriptase synthesizes a strand of DNA complementary to an mRNA template strand, by lengthening the oligo-dT primer.
2. The messenger RNA strand is cut randomly and specifically by the action of RNAase H.
3. A DNA-polymerase I (*E. coli* or T4) is utilized to synthesize the second DNA strand, using the first as a template, by growth of the RNA primer and digestion of the RNA primers further down the strand.
4. The cDNA obtained are often not complete at the 5′ region of the mRNA.
5. Certain methods allow generation of a complete cDNA.

Figure 1.2 Synthesis of cDNA

Messenger RNA can be studied directly after purification (see Northern technique), but can also be converted into complementary DNA (cDNA), that is to say a double-stranded molecule of DNA made *in vitro* and without a 'natural' corresponding molecule. This is done using a reverse transcriptase that, in the presence of free triphosphate nucleotides, will start from an oligo-dT primer hybridized to the polyA tail and synthesize a primary strand of DNA that is complementary to that of the mRNA. After this, a DNA polymerase will synthesize a second strand of DNA using the first as a template (Figure 1.2). There are several methods for replacing the mRNA by the second DNA strand in order to obtain cDNA. The majority produce an incomplete cDNA where the 5′ part of the mRNA (which is generally non-coding) is missing, but methods have been developed to have a complete cDNA, an integral copy of the original mRNA sequence, in the form of double-stranded DNA.

Box 1.3 Single-strand messenger RNA may be converted into double-stranded cDNA

The cDNA copies of the polyA mRNAs vary from one tissue to another and form a representative collection of the genes expressed, and therefore transcribed into mRNA, in each of the tissues. It is possible using the RT–PCR technique (see below) to obtain in the form of cDNA, and then to amplify, only the mRNAs from this collection. In addition, cDNA contains the continuous coding sequence, with the non-coding sequences (introns) having been removed when the isolated mRNAs underwent nuclear maturation.

1.3 The enzymatic tools for *in vitro* treatment of DNA

Molecular biology started from the identification and purification of bacterial and viral enzymes allowing the *in vitro* manipulation of DNA and RNA nucleic acids, such as their cutting, joining, cloning, measurement of their size, establishing their sequence and identifying specific variations (mutations). All of these manipulations lead to the construction of recombinant genomes (genetically modified organisms) that have been used as tools in research for nearly 30 years, and more recently have had industrial applications in agronomy or pharmaceuticals. Table 1.1 shows a list of some of the enzymatic tools used in molecular biology.

1.4 DNA fragmentation and study of the fragments

1.4.1 DNA fragmentation

DNA fragmentation is currently one of the operations executed on DNA *in vitro*. Physical methods (pressure and ultrasound) perform well but work randomly, and molecular biology really started on the day that bacterial endonucleases were discovered. These enzymes, called restriction enzymes (because the protection they give against viral genome infections is 'restricted' to certain stocks), are endonucleases capable of cutting the DNA double helix, at a specific recognition and binding site (type II endonucleases), so that several identical DNA molecules are cut in the same manner, giving the same collection of fragments called 'restriction fragments'.

Certain restriction enzymes function in a way that cuts uniformly leaving fragments with blunt ends; others perform an overhanging cut creating fragments with single-stranded extremities. These extremities are often present in the form of complementary sequences because the restriction enzyme used in molecular biology recognizes a site in the DNA that constitutes a palindromic sequence. These single-strand complementary sequences are also called cohesive sequences because they allow the reassociation of the fragments through molecular hybridization of