Introduction to Molecular Biology and Genetic Engineering

Book · January 2011

CITATIONS
2

READS
23,049

1 author:

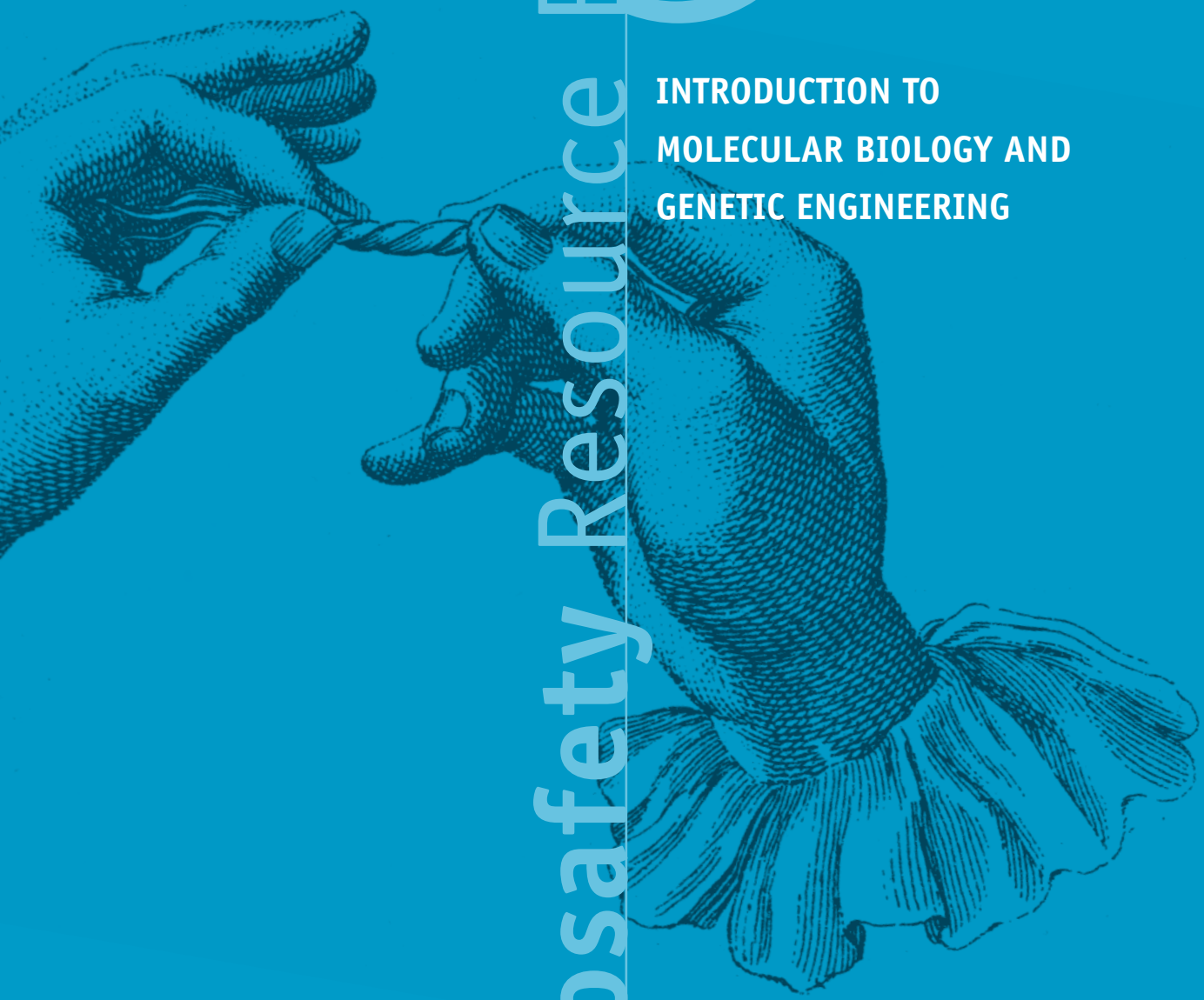
Zephaniah Dhlamini
National University of Science and Technology, Bulawayo
16 PUBLICATIONS **130** CITATIONS

SEE PROFILE

Biosafety Resource Book

MODULE a

INTRODUCTION TO MOLECULAR BIOLOGY AND GENETIC ENGINEERING



FIAT PANIS

MODULE

Biosafety Resource Book

INTRODUCTION TO MOLECULAR BIOLOGY AND GENETIC ENGINEERING

Oliver Brandenberg
Zephaniah Dhlamini
Alessandra Sensi
Kakoli Ghosh
Andrea Sonnino

Food and Agriculture
Organization of
the United Nations
Rome, 2011

FOREWORD

During the period 2002–10, FAO undertook an intense activity of biosafety capacity development, largely centred on enhancing the capacities of regulators and other technical staff involved in the implementation of biosafety frameworks, along with other components. The training programme was tailored to meet the needs of a very specific audience: biosafety regulators, policy-makers and members of national biosafety committees, with diverse educational backgrounds, not necessarily well versed in all the biosafety-related fields. The training courses therefore aimed to: i) offer background knowledge critical in the process of reviewing biosafety dossiers and biosafety-related decision-making; ii) provide acquaintance with concepts and methodologies relevant to risk analysis of GMO release and biosafety management.

The training programme consisted of the following modules:

MODULE A

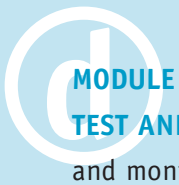
MOLECULAR BIOLOGY AND GENETIC ENGINEERING, which reviews the very basic scientific concepts and principles employed in producing GMOs, and provides a brief description of current and emerging uses of biotechnology in crops, livestock and fisheries.

MODULE B

ECOLOGICAL ASPECTS, which provides the necessary background information on ecology and evolution needed to analyse and understand the consequences of introducing GMOs into the environment.

MODULE C

RISK ANALYSIS, which provides basic information on biological risks, concepts, principles, and methodologies of risk assessment, risk management and risk communication. It focuses on crop biotechnology and environmental risk assessment of GM crops since these are of immediate interest to most countries.



MODULE D

TEST AND POST-RELEASE MONITORING OF GMOs, which addresses the use and monitoring of GMOs under containment, confinement and limited field trials, as well as the monitoring of commercially released GMOs. It also covers surveillance and emergency planning.

MODULE E

LEGAL ASPECTS, which provides an overview of the existing legal tools and frameworks on biotechnology and biosafety, and offers a thorough description of the international instruments that regulate biosafety and their interactions.

FOREWORD

This Biosafety Resource Book stems from experience gained in biosafety capacity development projects and is based on the materials developed by the lecturers who have taught in the training courses organized to date. The Resource Book has been prepared in response to an expressed need, with the purpose of being used as a training tool in future activities. The Resource Book also aims at providing biosafety regulators, policy-makers and members of national biosafety committees with reference materials that can be readily consulted beyond the training events, when the need arises. Special attention has been paid to avoid technical jargon and to keep the modules scientifically accurate as well as accessible to non-specialists.

FAO's biosafety capacity building activities are the result of a collaborative effort, involving numerous institutions, including national biosafety committees of many countries, ministries, universities and research institutes, NGOs and the private sector. The precious contribution of national project coordinators, national and international consultants, as well as FAO officers from headquarters and decentralized offices, is gratefully acknowledged. The enthusiasm and dedication of the participants in the training activities were crucial for their success.

Andrea Sonnino
Chief,
Research and Extension Branch
Office of Knowledge Exchange,
Research and Extension

LIST OF CONTRIBUTORS

This module has been prepared based on the lectures and contributions of:

Oliver Brandenberg
Research and Extension Branch
Food and Agriculture Organization of
the United Nations (FAO)
Rome, Italy
(present address:
Institute of Medical Virology,
University of Zürich,
Zürich, Switzerland)

Zephaniah Dhlamini
Department of Applied Biology & Biochemistry
National University of Science and Technology
Ascot, Bulawayo, Zimbabwe

Richard Edema
Department of Biotechnology
Makerere University
Kampala, Uganda

Hrvoje Fulgosi
Department of Molecular Biology
Rudjer Boskovic Institute
Zagreb, Croatia

Kakoli Ghosh
Plant Protection and Production Division
Food and Agriculture Organization of
the United Nations (FAO)
Rome, Italy

Antonietta Gutiérrez-Rosati
Universidad Nacional Agraria
La Molina, Lima, Perú

Andrew Kiggundu
National Agricultural Biotechnology Laboratory
Kawanda, Uganda

Jesse Machuka
Department of Biochemistry and Biotechnology
Kenyatta University
Nairobi, Kenya

Hosea Ken Matengo
Department of Molecular Biology and
Biotechnology
University of Dar-es-Salaam
Dar es Salaam, United Republic of Tanzania

Emmarold E. Mneney
Mikocheni Agricultural Research Institute (MARI)
Dar es Salaam, United Republic of Tanzania

Sam Kwame Offei
Department of Crop Science
University of Ghana
Legon, Ghana

Leonard O'Garro
Department of Biological and Chemical Sciences
Faculty of Pure and Applied Sciences
University of West Indies
Cave Hill, Barbados

Ajaj Parida
M. S. Swaminathan Research Foundation (MSSRF)
Chennai, India

Alessandra Sensi
Research and Extension Branch
Food and Agriculture Organization of
the United Nations (FAO)
Rome, Italy
(present address:
EuropeAid Co-operation Office, Unit A.3
Centralised Operations for Europe, the
Mediterranean and the Middle East
Brussels, Belgium)

Andrea Sonnino
Research and Extension Branch
Food and Agriculture Organization of
the United Nations (FAO)
Rome, Italy

CONTENTS

MODULE a

LIST OF ABBREVIATIONS

A	Adenine
AI	Artificial insemination
ALS	Acetolactate synthase
Amp	Ampicillin
ARS	Autonomously replicating sequence
ATP	Adenosine triphosphate
A. tumefaciens	*Agrobacterium tumefaciens*
BAC	Bacterial artificial chromosome
bp	base pair
BSE	Bovine Spongiform Encephalopathy
Bt	*Bacillus thuringiensis*
C	Cytosine
CaMV	Cauliflower mosaic virus
CAT	Chloramphenicol acetyltransferase
CBD	Convention on Biological Diversity
CTAB	Cetyl trimethylammonium bromide
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
E. coli	*Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
EPSPS	5-enolpyruvylshikimate-3-phosphate synthase
ES cell	Embryonic stem cell
ET	Embryo transfer
FAO	Food and Agriculture Organization of the United Nations
FDA	U.S. Food and Drug Administration
G	Guanine
GFP	Green fluorescent protein
GMO	Genetically modified organism
GUS	Beta-glucuronidase
HPLC	High performance liquid chromatography
ISO	International Organization for Standardization
ISTA	International Seed Testing Agency
IVF	*In vitro* fertilization
kb	kilobase
MAS	Marker-assisted selection
Mb	Megabase
MCS	Multiple cloning site
mRNA	Messenger RNA
OPU	Oocyte pick-up
ORF	Open reading frame
ori	Origin of replication
PAT	Phosphinothricin acetyltransferase
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PPT	Phosphinothricin
QTL	Quantitative trait loci
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT-PCR	Reverse transcriptase PCR
S. aureus	*Staphylococcus aureus*
SCNT	Somatic cell nuclear transfer
SMG	Selectable marker gene
snRNP	small nuclear ribonucleoproteins
ST	Somatotropin
T	Thymine
T-DNA	Transfer DNA
Ti	Tumour-inducing
TMV	Tobacco mosaic virus
tRNA	Transfer RNA
U	Uracil
UTR	Untranslated region
vir genes	virulence genes
YAC	Yeast artificial chromosome

For further explanation of terminology and other abbreviations, please refer to the FAO "Glossary of biotechnology for food and agriculture" (FAO, 2007), which is also available online at: http://www.fao.org/biotech/index_glossary.asp?lang=en

CHAPTER 1

INTRODUCTION TO BIOTECHNOLOGY: BASIC CONCEPTS AND DEFINITIONS

1.1 DEFINITION OF BIOTECHNOLOGY

The term biotechnology was coined in 1919 by Karl Ereky, a Hungarian engineer. At that time, the term included all the processes by which products are obtained from raw materials with the aid of living organisms. Ereky envisioned a biochemical age similar to the stone and iron ages.

Nowadays, according to the Convention on Biological Diversity (CBD), **Biotechnology** is defined as "any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use" (CBD, 1992). The living organisms or derivatives thereof most frequently used include micro-organisms, animals and plants (or their isolated cells) as well as enzymes. They can be utilized to process substances, usually other natural, renewable materials, or serve themselves as sources for valuable substances or goods. Several branches of industry rely on biotechnological tools for the production of food, beverages, pharmaceuticals and biomedicals. The CBD definition is applicable to both "traditional" or "old" and "new" or "modern" biotechnology.

BIOTECHNOLOGY
Any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use.

Long before the term biotechnology was coined for the process of using living organisms to produce improved commodities, people were utilizing living micro-organisms to obtain valuable products, for example through the process of fermentation.

A list of early biotechnological applications is given below in Table 1.1:

Table 1.1 | **Traditional applications of biotechnology**

Providing bread with leaven	Prehistoric period
Fermentation of juices to alcoholic beverages	Prehistoric period
Knowledge of vinegar formation from fermented juices	Prehistoric period
Manufacture of beer in Babylonia and Egypt	3rd century BC
Wine manufacturing in the Roman Empire	3rd century AD
Production of spirits of wine (ethanol)	1150
Vinegar manufacturing industry	14th century AD
Discovery of the fermentation properties of yeast	1818
Description of the lactic acid fermentation by Pasteur	1857
Detection of fermentation enzymes in yeast by Buchner	1897
Discovery of penicillin by Fleming	1928
Discovery of many other antibiotics	≈1945

Since the middle of the twentieth century biotechnology has rapidly progressed and expanded. In the mid-1940s, scale-up and commercial production of antibiotics such as penicillin occurred.

The techniques used for this development were:

- isolation of an organism producing the chemical of interest using screening/ selection procedures, and
- improvement of production yields via mutagenesis of the organism or optimization of media and fermentation conditions. This type of biotechnology is limited to

chemicals occurring in nature. It is also limited by its trial-and-error approach, and requires a lengthy procedure over years or even decades to improve yields (Rolinson, 1998).

About three decades ago, with the advance of molecular biology, biotechnology became more of a science than an art. Regions of deoxyribonucleic acid (**DNA)** (called genes) were found to contain information that directs the synthesis of specific proteins. Proteins can therefore be considered as the final product of a gene; they are the molecules that carry out almost all essential processes within a cell. Each protein has its own identity and function: many are so-called enzymes that catalyse (facilitate) chemical reactions, others are structural components of cells and organs (Morange and Cobb, 2000). Today it is possible to express a gene, regardless of its origin, in a simple bacterium such as *Escherichia coli (E. coli)*, so that the bacterium produces large quantities of the protein coded for by the gene. The same principle can be applied to many other micro-organisms, as well as to higher organisms such as plants and animals.

DNA
Acronym for Deoxyribonucleic Acid: The material in which our hereditary information is stored.

The techniques used for this purpose include:

- isolation of the gene coding for a protein of interest;
- cloning (i.e. transfer) of this gene into an appropriate production host;
- improving gene and protein expression by using stronger promoters, improving fermentation conditions etc. (Gellisen, 2005). Together, these techniques are known as **recombinant DNA technology** and will be discussed at some length throughout this resource book.

RECOMBINANT DNA TECHNOLOGY
The process of constructing and manipulating DNA sequences that do not occur naturally, by combining DNA fragments from different sources.

About two decades ago, protein engineering became possible as an offshoot of the recombinant DNA technology. Protein engineering differs from "classical" biotechnology in that it is concerned with producing new (engineered) proteins which have been modified or improved in some of their characteristics (Park and Cochran, 2009).

The techniques involved in protein engineering are essentially based on recombinant DNA technology and involve:

- various types of mutagenesis (to cause changes in specific locations or regions of a gene to produce a new gene product);
- expression of the altered gene to form a stable protein;
- characterization of the structure and function of the protein produced;
- selection of new gene locations or regions to modify for further improvement as a result of this characterization.

The commercial implications of the technical developments listed above are that a large number of proteins, existing only in tiny quantities in nature, can now be produced on an industrial scale. Furthermore, the yields of biochemical production can be increased much faster than what was originally possible with classical fermentation.

GENETICALLY MODIFIED ORGANISM
An organism whose genetic material has been modified, for example by introducing foreign DNA sequences obtained by recombinant DNA technology.

Importantly, the production of transgenic animals and plants that contain genetic elements from foreign sources and possess novel traits and characteristics is also based on the techniques outlined above. As all these approaches result in the creation of **genetically modified organisms** (GMOs) that can be potentially harmful to the environment and human health, the part of biotechnology that deals with GMOs is strictly regulated by biosafety laws and guidelines. The main thrust of this resource book is on the development and enforcement of such regulatory frameworks at domestic and international levels.

Biotechnology applications are developed by a collection of multidisciplinary research activities, commonly referred to as *enabling technologies*. Apart from fermentation and genetic engineering/recombinant DNA technology, other important enabling technologies are plant and animal cell culture technology and enzyme technology.

The basis of these enabling technologies are the scientific disciplines of molecular biology, genetics, microbiology, biochemistry, protein chemistry, chemical and

process engineering and computer science. An overview of important events in the development of modern molecular biology and recombinant DNA technology is provided in Table 1.2:

Table 1.2 | **An overview of recombinant DNA-based biotechnology**

Double helix structure of DNA is first described by Watson and Crick	1953
Cohen and Boyer, amongst others, develop genetic engineering	1973
The first human protein (somatostatin) is produced in a bacterium *(E. coli)*	1977
The first recombinant protein (human insulin) approved for the market	1982
Polymerase chain reaction (PCR) technique developed	1983
Launch of the Human Genome Project	1990
The first genome sequence of an organism *(Haemophilus influenzae)* is determined	1995
A first draft of the human genome sequence is completed	2000
Over 40 million gene sequences are deposited in GenBank, and genome sequences of hundreds of prokaryotes and dozens of eukaryotes are finished or in draft stage	2005

1.2 OVERVIEW OF APPLICATIONS OF BIOTECHNOLOGY

Since the advance of recombinant DNA technology, several techniques and applications have been developed that are benefiting humankind in the areas of agriculture, medicine, environment, industry and forensics. The following sections briefly describe some of these applications and their potential benefits to society.

1.2.1 Industry

Biotechnology can be used to develop alternative fuels; an example is the conversion of maize starch into ethanol by yeast, which is subsequently used to produce gasohol (a gasoline-ethanol mix). Bacteria are used to decompose sludge and landfill wastes (Soccol *et al.*, 2003). Through biotechnology, micro-organisms or

their enzymes can be adapted to convert biomass into feed stocks, or they can be used for manufacturing biodegradable plastics (bioplastics). Other organisms (micro-organisms, plants and mammals) are used as bioreactors for producing chemical compounds that are extracted from them and processed as drugs and other products. Plant and animal fibres are used for producing a variety of fabrics, threads and cordage. Biotechnology is applied to improve the quality and quantity of these products. Biopulping is a technique whereby a fungus is used to convert wood chips into pulp for papermaking (Gavrilescu and Chisti, 2005).

1.2.2 Health and medicine

In the area of health and medicine, biotechnology has numerous and important functions. Biotechnologies are used to develop diagnostic tools for identifying diseases.

Biotechnology is also used to produce more effective and efficient vaccines, therapeutic antibodies, antibiotics, and other pharmaceuticals. Biotechnology is a USD 70 billion a year industry that has produced several blockbuster drugs and vaccines, i.e. drugs with sales volumes exceeding USD 1 billion per year (Lawrence, 2007). Furthermore, there are more than 370 drug products and vaccines obtained through biotechnology currently in clinical trials, targeting more than 200 diseases including various cancers, Alzheimer's disease, heart disease, diabetes, multiple sclerosis, AIDS and arthritis (Sullivan *et al.*, 2008).

Through the biotechnology of gene therapy, scientists are making efforts at curing genetic diseases by attempting to replace defective genes with the correct version. A revolutionary strategy is being developed whereby staple foods such as potatoes, bananas, and others are used as delivery vehicles to facilitate the immunization of people in economically depressed regions of the world (Tacket, 2009).

1.2.3 Environment

Development and usage of alternative fuels that burn cleaner and improve air quality through reduced pollution of the environment is possible by biotechnological means. Micro-organisms are used to decompose wastes and clean up contaminated sites by the technology of bioremediation. The use of disease-resistant cultivars can make crop production less environmentally intrusive by reducing the use of agrochemicals (Chatterjee *et al.*, 2008).

1.2.4 Forensics

Since the DNA profile, i.e. the nucleotide sequence of the genome, is unique in every individual, it can be used as a powerful basis of identifying individuals in a population. DNA-based evidence is used in cases involving paternity disputes and family relationships. Furthermore, it is used in health care and judicial systems. In the judicial system, forensic experts use DNA profiling to identify suspects in criminal cases, especially when body fluids and other particles like hair and skin samples can be retrieved (Jobling and Gill, 2004).

1.2.5 Agriculture

Biotechnology can complement conventional breeding for crop and animal improvement. Instead of extensive re-arrangement of genes, as occurs in conventional breeding, biotechnology enables targeted gene transfer to occur. The genome of the recipient individual remains intact, except for the introduced gene (or genes), thus accelerating breeding programmes and the development of organisms with desirable characteristics. Furthermore, biotechnology enables gene transfer across natural breeding boundaries, overcoming mating barriers and creating a "universal gene pool" or "universal breeding population" accessible to all organisms. Likewise, it is possible to specifically introduce novel, desirable traits and characteristics into

existing species. This biotechnological application is used to improve the yield of crop and animal species and their product quality such as nutritional value and shelf life (Shewry *et al*, 2008). In addition to these benefits, this methodology reduces the need for agrochemicals by creating disease and pest-resistant species, thereby reducing environmental pollution from chemical runoff. Increased yields and higher food quality can contribute to reducing world hunger and malnutrition (FAO, 2004).

AGRICULTURAL BIOTECHNOLOGY
All biotechnological applications developed for a potential use in agriculture.

Several technologies in the field of **agricultural biotechnology** exist that do not rely on the creation of GMOs. Molecular techniques are being used to monitor breeding populations and to diagnose animals and plants infected with diseases. Micropropagation techniques are being widely used to generate clonal plant materials, allowing rapid large-scale clonal propagation of many plant species including trees. Biofertilizers and biopesticides can be applied in place of conventional fertilizer and pesticides to promote plant growth and health in an environmentally sustainable way (FAO, 2001).

To summarize, the field of biotechnology is very diverse, both in terms of methodologies and techniques applied and the potential applications and outcomes. Biotechnology has the potential to contribute to a worldwide sustainable development and the reduction of world hunger, including the branches of biotechnology concerned with agricultural research and development (FAO, 2004). Importantly, biotechnology is not only based on GMOs, but offers several important and well established techniques that are not dependent on or derived from genetic modifications. However, the focus of this publication is on GMOs and related products. Following an introduction to the molecular background and the scientific basis of GMO development and to the aims and prospects of this research, the main part of this book will introduce biosafety concepts related to the use of GMOs.

CHAPTER 2

STRUCTURE AND FUNCTION OF GENES

2.1 GENES AND HEREDITY

The study of genes and heredity is called **genetics**. Heredity phenomena have been of interest to humans since long before the underlying principles were scientifically investigated and understood. Ancient peoples were improving plant crops and domesticating animals by selecting desirable individuals for breeding. Genetics as a set of scientific principles and analytical procedures emerged in the 1860s when the Augustinian monk Gregor Mendel performed a set of experiments that revealed the existence of biological "factors" responsible for transmitting traits from generation to generation. These factors were later called genes, following the discovery of chromosomes and genetic linkage in the early twentieth century. Up to this point genetics looked at genes as abstract entities that somehow control hereditary traits. Through genetic analyses the inheritance of different genes was studied, but the physical and biochemical nature of the gene remained unknown. Further work revealed that chromosomes consist of DNA and protein, and subsequent studies allowed the conclusion that DNA is, in fact, the hereditary material (Morange and Cobb, 2000).

GENETICS
The science of genes and heredity.

DNA was thought to be a simple molecule, thus many scientists did not believe that it indeed carried and stored the information for an entire organism. How can such huge amounts of information be contained and passed on from one generation to the next? Clearly, the genetic material must have both the ability to encode specific information and the capacity to duplicate that information precisely during every cell division. What kind of molecular structure could allow such complex functions?

2.2 THE STRUCTURE OF DNA

Although the exact DNA structure was not known until 1953, its basic building blocks had been known for many years. It had been shown that DNA is composed of four basic molecules called *nucleotides*, which are identical except that each contains a different nitrogen-containing base. Each nucleotide is made up of a phosphate group, a sugar (of the deoxyribose type), and one of the four bases. The four bases are adenine (A), guanine (G) (the purines) and cytosine (C) and thymine (T) (the pyrimidines; see also Figure 2.1).

DNA STRUCTURE
In 1953, the structure of DNA was deciphered and found to be a double helix. This model offered possible explanations for the processes of DNA replication and gene expression.

In 1953 James Watson and Francis Crick were the first to succeed in putting the building blocks together and came up with a reasonable **DNA structure**. They used DNA X-ray diffraction patterns produced by Rosalind Franklin and Maurice Wilkins and data from Erwin Chargaff. The X-ray data showed the DNA molecule to be long, thin and helical (spiral-like) in shape.

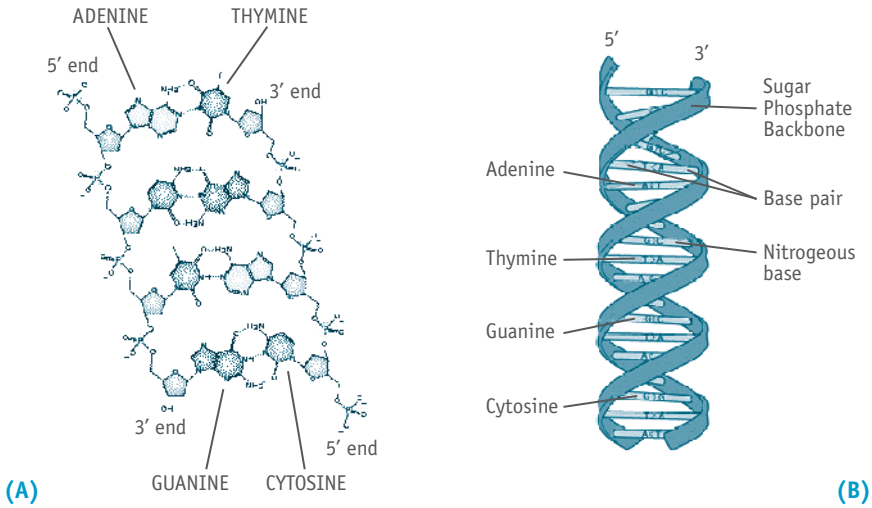
Chargaff had established certain empirical rules about the amounts of each component of DNA:

- The total amount of pyrimidine nucleotides (T + C) always equals the total number of purine nucleotides (A + G).
- The amount of T always equals the amount of A, and the amount of C always equals the amount of G. But the amount of A + T is not necessarily equal to the amount of G + C.

The structure that Watson and Crick derived from these clues is a double helix (Figure 2.1). Each helix is a chain of nucleotides held together by phosphodiester bonds, in which a phosphate group forms a bridge between -OH groups on two adjacent sugar residues. The two DNA chains (helices) are running in an antiparallel direction and are held together by hydrogen bonds between opposing bases, thus forming a double helix. Each base pair (bp) consists of one purine and one pyrimidine base, paired according to the following rule: G pairs with C, and A pairs with T (Watson *et al.*, 2008).

Figure 2.1 | **The structure of DNA**

In part (A), the four bases, the pairing of the bases and the connection of the bases through the sugarphosphate-backbone are depicted. Note that the two DNA strands are held together through base pairing and are running in opposite direction, labelled 3′ and 5′ end respectively (read: three prime and five prime). In part (B), a schematic drawing of the actual DNA double helix structure is depicted, containing the same elements in simplified form and labelling as in (A).



Elucidation of the structure of DNA caused a lot of excitement in the scientific community for two major reasons. First, the structure suggests an obvious way in which the molecule can be duplicated, or replicated, since each base can specify its complementary base by hydrogen bonding. Thus each strand can serve as a template for the synthesis of a complementary strand. Second, the structure suggests that the sequence of nucleotide pairs in DNA is dictating the sequence of amino acids in a protein encoded by a gene. In other words, some sort of genetic code may comprise information in DNA as a sequence of nucleotide pairs, which can be translated into the different language of amino acid sequence in protein.

2.3 THE FLOW OF GENETIC INFORMATION: THE CENTRAL DOGMA

THE CENTRAL DOGMA
Francis Crick proposition that DNA is transcribed to a messenger molecule (mRNA) which is subsequently translated to protein: DNA makes RNA makes protein.

In the early 1950s, Francis Crick suggested that there is a unidirectional flow of genetic information from DNA through ribonucleic acid (RNA) to protein, i.e. "DNA makes RNA makes protein". This is known as **the central dogma** of molecular biology, since it was proposed without much evidence for the individual steps. Now these steps are known in detail: DNA is transcribed to an RNA molecule (messenger RNA [mRNA]), that contains the same sequence information as the template DNA, and subsequently this RNA message is translated into a protein sequence according to the genetic code (Miller *et al.*, 2009).

2.4 THE GENETIC CODE

THE GENETIC CODE
The relation between the sequence of bases in DNA, in 64 possible nucleotide triplet combinations, and the sequence of amino acids in protein.

The basic building blocks of DNA are the four nucleotides; the basic building blocks of proteins are the amino acids, of which there are 22 that naturally occur in proteins (the so-called proteinogenic amino acids). **The genetic code** is the correspondence between the sequence of the four bases in nucleic acids and the sequence of the 22 amino acids in proteins. It has been shown that the code is a triplet code, where three nucleotides (one codon) encode one amino acid. Since

BOX 2.1

RNA

The ***ribonucleic acids (RNA)*** are an important class of molecules in the flow of genetic information. Some viruses even use RNA, instead of DNA, to carry their genetic information. All other organisms that use DNA as the genetic material must first transcribe their genetic information into RNA, in order to render the information accessible and functional.

RNA is similar in composition to DNA. It is a long linear molecule (polymer) that is made up of a limited number of monomers, the nucleotides. As in DNA, each nucleotide is composed of a sugar, a phosphate, and a base. The sugar, however, is ribose instead of deoxyribose as seen with DNA, hence the names ribo- and deoxyribonucleic acids. Unlike DNA, RNA molecules are usually single stranded and do not form double helices. RNA molecules are made up of the same bases as DNA, except that the DNA base thymine (T) is replaced by uracil (U) in RNA.

The cell contains different kinds of RNA, most importantly messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). These three RNA classes correspond to the three basic roles RNA plays in the cell.

First, RNA serves as the intermediary in the flow of information from DNA to protein. The DNA is transcribed (copied) into mRNA via an enzyme (RNA polymerase) and subsequently the mRNA is translated into protein. In the latter process, translation of mRNA to protein, tRNA and rRNA play important roles. tRNA molecules serve as adaptors that translate the information in the nucleic acid sequence of mRNA into the sequence of amino acids, the constituents that make up a protein. Finally, the rRNA molecules are the major functional components of the molecular machines, the so-called ribosomes, which carry out the translation process.

RNA
Acronym for ribonucleic acid, the second class of nucleic acids in the cell besides DNA. RNAs occupy several crucial functions in the flow of genetic information from DNA to protein.

there are only 22 amino acids to be specified and 64 different codons (4^3 =64), most amino acids are specified by more than one codon and the genetic code is said to be degenerate, or to have redundancy. The genetic code has colinearity, which means that the order of the bases in the DNA corresponds directly to the order of amino acids in the protein (Watson *et al.*, 2008).

Clearly, if the genetic code is to be read as we would read a sentence in a book, we need to know where to start and stop. The codon AUG serves as a start signal, encoding the amino acid methionine, which is therefore the first amino acid incorporated into all proteins. However, methionine is also found elsewhere, not only at the beginning. Therefore, the translational machinery has to find the correct methionine codon to start and not just any given AUG codon anywhere in the gene sequence. This process is facilitated by sequences surrounding the initiation AUG codon. These sequences are therefore highly important for the translation process. The end of the translated region is specified by one of three codons which encode "stop". These are UAA, UAG and UGA. If **mutations**, i.e. unintended changes in the DNA sequence, take place that create one of the stop codons instead of an amino acid encoding codon, the results may be severe as the resultant protein will be shorter than intended. Such proteins are referred to as being truncated, and are very likely non-functional. Other mutations alter one codon to another, resulting in the replacement of the original amino acid by a different one, which can have severe or negligible effects, depending on the importance of the amino acid, for the entire protein. The addition or deletion of a single nucleotide can also have a severe effect, since all following codons will be shifted by one nucleotide, resulting in a very different message – a so-called frameshift mutation. The region between the start-methionine and the first stop codon is referred to as the open reading frame (ORF).

MUTATION
Random changes in DNA sequence, induced by replication errors, mutagenic substances or physical factors such as UV light and radioactive irradiation.

Finally, the genetic code is virtually universal, i.e. it is the same in all organisms living on this planet. Genes taken from plants can be decoded by animal cells, while genes from prokaryotes can be decoded by eukaryotic systems, and vice versa.

Without such a universal nature of the code, genetic manipulation and genetic engineering would be much more difficult (Voet and Voet, 2004).

2.5 THE GENE

Historically, a **gene** is defined as a heritable unit of phenotypic variation. From a molecular standpoint, a gene is the linear DNA sequence required to produce a functional RNA molecule, or a single transcriptional unit (Pearson, 2006). Genes can be assigned to one of two broad functional categories: structural genes and regulatory genes. It is the function of the end product of a gene that distinguishes structural and regulatory genes.

GENE
Broadly defined as a sequence of DNA encoding a functional product. This includes the coding region itself as well as all assoctiated regulatory regions.

- *Structural genes* code for polypeptides or RNAs needed for the normal metabolic activities of the cell, e.g. enzymes, structural proteins, transporters, and receptors, among others.
- *Regulatory genes* code for proteins whose function is to control the expression of structural genes. With regard to molecular composition both classes of genes are similar.

A gene usually occupies a defined location on a chromosome, of which there are 46 in every human cell and which contain the entire human **genome** (see below). The exact chromosomal gene location is defined by specific sequences for the start and termination of its transcription. Each gene has a specific effect and function in the organism's morphology or physiology, can be mutated (i.e. changed), and can recombine with other genes. It is a store of information (in the form of nucleotide base sequence); consequently it does not initiate any action, but is acted upon, e.g. during the process of gene expression. The complete set of genes of an organism, its genetic constitution, is called the genotype. The human genome, for example, contains an approximate number of 25 000 protein-coding genes. The physical manifestation, or expression, of the

GENOME
The complete set of genetic information of an organism encoded in its DNA.

PHENOTYPE
All observable characteristics and traits of an organism. The phenotype is the result of the organism's genotype and environmental influences.

genotype is the **phenotype** (i.e. the organism's morphology and physiology). If a particular characteristic, such as brown eye colour, is part of an organism's phenotype, one can conclude that the individual carries the gene(s) for that characteristic. If, however, a particular characteristic is not expressed, one cannot implicitly conclude that the particular gene is absent because expression of that gene might be repressed. Different varieties of the same gene, resulting in different phenotypic characteristics, are called **alleles** (Griffiths *et al.*, 2007).

ALLELE
Different versions of the same gene. In a population, usually many versions of the same gene can be found.

Genes may be located on either strand of the double-stranded DNA. But, regardless of which strand contains a particular gene, all genes are read in a 5′ to 3′ direction, and the strand containing the particular gene is referred to as the sense or coding strand.

CHROMOSOME
A single DNA molecule, associated with specific proteins. The storage form of DNA within the cell.

As stated above, every cell of a human body, except germ line cells, contains 46 **chromosomes**. From each parent, we inherit 23 chomosomes, representing the complete genome. Thus, each body cell is *diploid*, i.e. contains two copies of the human genome and likewise two copies (alleles) of each gene.

The haploid set of the human genome (23 chromosomes), consists of approximately 3 200 megabases (Mb; 1 Mb = 10^6 bp) and contains an estimated number of 20 000 to 25 000 protein-coding genes (International Human Genome Sequencing Consortium, 2004). In fact, protein-coding DNA sequences only represent approximately 1.5 percent of the total genome; the remaining majority of DNA represent regulatory sequences, RNA encoding genes, or simply DNA sequences that have not yet been assigned to a certain function (sometimes inappropriately referred to as "junk DNA"). Interestingly, the estimated number of proteins is somewhat higher than the number of genes, due to alternative splicing (see 2.6 and 2.9.1) and other variations in gene expression.

In comparison, the genome of *E. coli*, a widely used model bacterium, consists of one chromosome of 4.6 Mb in size, encoding approximately 4 400 genes in

total. The genome of *Arabidopsis thaliana*, probably the most important model plant, consists of five chromosomes, of 157 Mb in size and encodes approximately 27 000 genes. Importantly, there is no straight connection between genome size, number of genes and organism complexity; some plants, vertebrates and even protozoans (single-cell organisms) have significantly larger genomes than the human genome (Patrushev and Minkevich, 2008).

2.6 THE ARRANGEMENT AND LAYOUT OF GENES

In **eukaryotic organisms** each cell contains more than one DNA molecule packaged into individual chromosomes; a diploid human cell, as stated, contains 46 chromosomes. Along the length of each DNA molecule/chromosome one can find thousands of genes, with more or less random spacing. In bacteria, one can frequently find clusters of genes that are related, in the sense that the proteins encoded by these genes are required in the same metabolic pathway. Therefore, as the cell needs all the gene products more or less simultaneously in order to keep that pathway running, it is appropriate for the cell to arrange these genes in clusters and employ a mechanism to express them together. These clusters of genes are known as operons; the most studied **operon** is the lactose operon in *E. coli*. This operon contains three genes which are adjacent on the DNA and are required for the utilization of lactose as a metabolic energy source in the cell. The operon also contains all the control sequences (repressor, promoter and operator, see Figure 2.2) needed to ensure efficient expression of the genes as an ensemble (Reznikoff, 1992). Operons do not occur in higher organisms but related genes are sometimes found in clusters as well, and comparable regulatory mechanisms are found.

EUKARYOTES
All organisms that possess a cellular structure, called nucleus, in which the DNA is contained within each cell. This includes all organisms except bacteria and archaebacteria, which do not posess a nucleus and are referred to as Prokaryotes.

OPERON
An arrangement of genes and certain regulatory regions to ensure expression of the genes as an ensemble in a controlled manner.

Many genes in eukaryotes have a distinctive structural feature: the nucleotide sequence contains one or more intervening segments of DNA that do not code for the amino acid sequence of the protein. These non-translated sequences interrupt the otherwise co-linear relationship between the nucleotide sequence of the gene

and the amino acid sequence of the protein it encodes. Such non-translated DNA segments in genes are called introns. The pieces that constitute mature mRNA, and therefore ultimatively for protein, are referred to as exons. During and after transcription the exons are spliced together from a larger precursor mRNA that contains, in addition to the exons, the interspersed introns. The number of exons that constitute a final mRNA molecule depends on the gene and the organism, but can range from as few as one to as many as fifty or more. The origin of **intron/exon structure** is a matter of scientific debate. To date it is not clear whether it predated the divergence of eukaryotes and prokaryotes with the subsequent loss of introns in prokaryotes, or if introns and the splicing mechanism evolved in eukaryotes after their evolutionary separation from prokaryotes (Mattick, 1994).

INTRONS AND EXONS
In eukaryotes, genes often consist of coding regions (exons) with interspersed non-coding regions (introns) which are removed during the process of gene expression.

In addition to introns and exons, the structural features of the eukaryotic gene include regulatory elements, a promoter region, a transcription start site and a transcription termination site (Figure 2.2). Specific proteins in the cell nucleus, the cellular compartment where DNA is stored, can bind to regulatory element sequences of a gene, thus controlling the expression of that gene. The promoter region is the sequence of the gene where the transcription machinery (the assembly of proteins required for transcription) binds to the DNA in order to start transcription to RNA. The start site indicates to the transcription machinery where to start and the termination site indicates were to stop transcription of the gene.

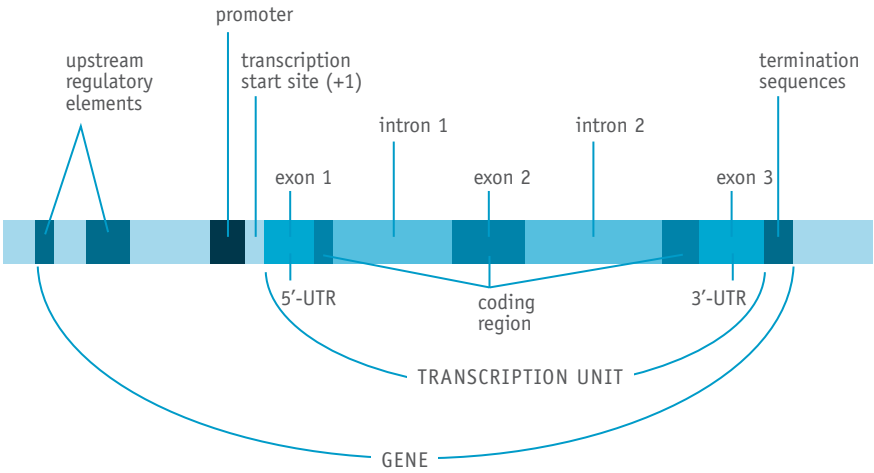
2.7 GENE EXPRESSION

GENE EXPRESSION
The process of converting the information stored in a gene to a functional product. Correct regulation of gene expression is crucial for the correct development and function of an organism.

Genes exert their function through a process called **gene expression**, a process by which heritable information from a gene, encoded on DNA, is transformed into a functional gene product, such as protein or RNA (some genes code for functional RNA molecules, such as tRNA and rRNA). Genes are expressed by being first transcribed into RNA, and may then subsequently be translated into protein. A cell employs many different mechanisms to regulate gene expression. Gene

Figure 2.2 | **A general structural arrangement of the different components making up a eukaryotic gene**

Upstream regulatory elements (enhancers) and the promoter are required for regulation and initiation of transcription. Exons, which constitute the actual protein-coding regions, and interspersed introns are indicated. The 5′ and 3′ untranslated regions (UTRs) are mRNA sequences that do not encode protein, but are required for a correct translation process. Transcription start and termination sites are also indicated.



expression can be regulated at many different levels, from DNA transcription, pre-mRNA processing, mRNA stability and efficiency of translation up to protein modification and stability. Thus, a cell can precisely influence the expression level of every gene, and studying and predicting gene expression levels is a difficult task. Nevertheless, this is especially important for biotechnological applications, since it is desirable to precisely define the expression levels of introduced genes in transgenic organisms. In the following section, the major processes of gene expression will be introduced.

2.7.1 Transcription and translation

mRNA
Messenger RNA, the molecule that DNA is transcribed to and that is subsequently translated to protein.

The first step in gene expression is transcription, namely the production of a single-stranded RNA molecule known as mRNA in the case of protein-coding genes. The nucleotide sequence of the **mRNA** is complementary to the DNA from which it was transcribed. In other words, the genetic messages encoded in DNA are copied precisely into RNA. The DNA strand whose sequence matches that of the RNA is known as the *coding strand* and the complementary strand on which the RNA was synthesized is the *template strand*.

TRANSCRIPTION
The process of transferring genetic information from DNA to an RNA molecule. Performed by the enzyme RNA Polymerase.

Transcription is performed by an enzyme called RNA polymerase, which reads the template strand in 3′ to 5′ direction and synthesizes the RNA from 5′ to 3′ direction. To initiate transcription, the polymerase first recognizes and binds a promoter region of the gene. Thus a major regulatory mechanism of gene expression is the blocking or sequestering of the promoter region. This can be achieved either by tight binding of repressor molecules that physically block the RNA polymerase, or by spatially arranging the DNA so that the promoter region is not accessible (Thomas and Chiang, 2006).

In eukaryotes, transcription occurs in the nucleus, where the cell's DNA is sequestered. The initial RNA molecule produced by RNA polymerase is known as the primary transcript and must undergo post-transcriptional modification before being exported to the cytoplasm for translation. The splicing of introns present within the transcribed region is a modification unique to eukaryotes. The splicing reaction offers various possibilities for regulating and modulating gene expression in eukaryotic cells.

Following transcription and post-transcriptional mRNA processing, the mRNA molecule is ready for translation. In eukaryotes, the mRNA must first be transported from the nucleus to the cytoplasm, whereas in prokaryotes no nucleus exists and transcription and translation take place in the same compartment.

Translation is the process by which a mature mRNA molecule is used as a template for synthesizing a protein. Translation is carried out by the ribosome, a large macromolecular complex of several rRNA and protein molecules. Ribosomes are responsible for decoding the genetic code on the mRNA and translating it into the amino acid sequence of proteins. Likewise, they are catalysing the chemical reactions that add new amino acids to a growing polypeptide chain by the formation of peptide bonds (Ramakrishnan, 2002).

TRANSLATION
The process of using the genetic information from an mRNA molecule to synthesize a protein molecule. Performed by the ribosome.

The genetic code on the mRNA is read three nucleotides at a time, in units called codons, via interactions of the mRNA with specialized RNA molecules called transfer RNA (tRNA). Each tRNA has three unpaired bases, known as the anticodon, that are complementary to the codon it reads. The tRNA is also covalently attached to the amino acid specified by its anticodon. When the tRNA binds to its complementary codon in an mRNA strand, the ribosome ligates its amino acid cargo to the growing polypeptide chain. When the synthesis of the protein is finished, as encoded by a stop-codon on the mRNA, it is released from the ribosome. During and after its synthesis, the new protein must fold to its active three-dimensional structure before it can carry out its cellular function (Voet and Voet, 2002).

A single mRNA molecule can be translated several times and thus produce many identical proteins, depending on its half-life in the cell, i.e. the average time it remains within the cell before it is degraded.

2.8 REGULATION OF GENE TRANSCRIPTION

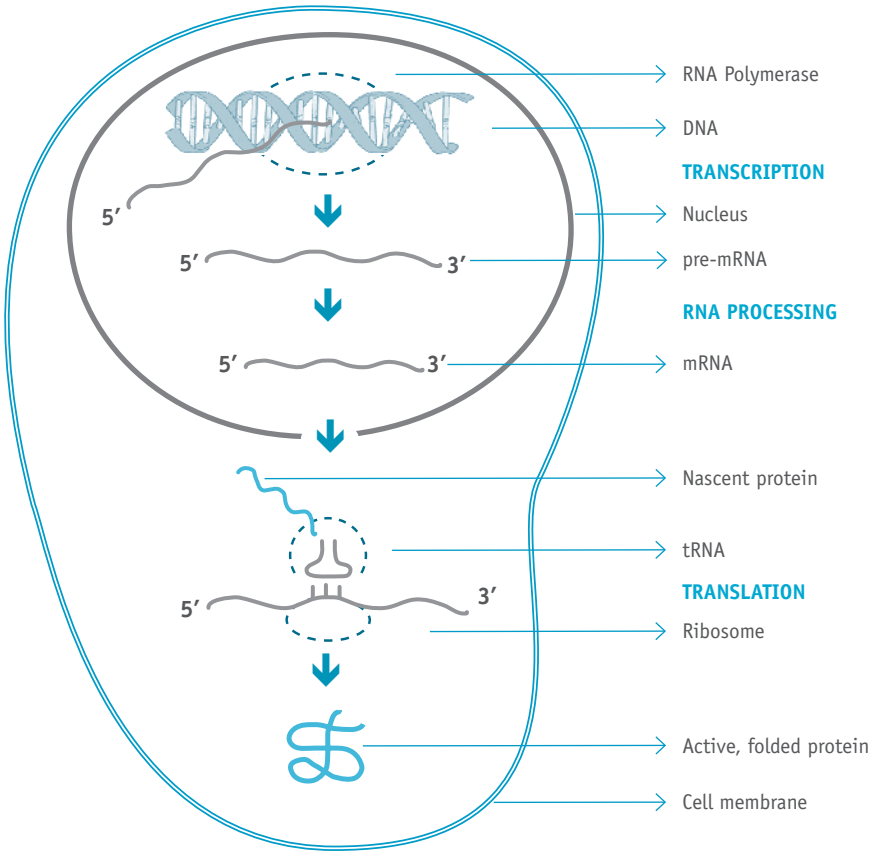
2.8.1 Promoters

PROMOTER
A DNA sequence associated with a gene that is responsible for recruiting the enzyme machinery required for the expression of that gene.

The promoter region of a gene is usually several hundred nucleotides long and immediately upstream from the transcription initiation site. The **promoter** constitutes the binding site for the enzyme machinery that is responsible for the transcription of DNA to RNA, the RNA polymerase. In eukaryotic cells several RNA polymerases

Figure 2.3 | Transcription and Translation

In the nucleus, DNA is transcribed to a pre-mRNA molecule by RNA polymerase. The pre-mRNA is processed, e.g. by intron excision, to the mature mRNA. The mRNA is exported to the cytoplasm and translated into protein, which is accomplished by ribosomes and tRNA that together decode the genetic code into amino acid sequence. Following translation, the synthesized protein adopts its correct 3-dimensional shape and is ready to perform its cellular function.



are present, the most prominent one, that is responsible for the transcription of protein-coding genes, being RNA polymerase II. There are different types of promoters for different RNA polymerases. Promoters for RNA polymerase II, the polymerase that transcribes protein-coding genes into mRNA, often contain the consensus sequence 5'-TATA-3', 30 to 50 bp upstream of the site at which transcription begins. Many eukaryotic promoters also have a so-called CAAT box with a GGNCAATCT consensus sequence centred about 75 bp upstream of the initiation start site (with N representing any of the four bases). RNA polymerases I and III are mainly responsible for the transcription of RNA molecules that possess an intrinsic function as catalytic or structural molecules, such as tRNA and rRNA, and that are not translated into proteins (Okkema and Krause, 2005). In general, the promoter region has a high importance for the regulation of expression of any gene. This concept will come up again later on in this module when the production of transgenic animals is introduced. Careful choice of promoters to drive gene expression in transgenic organisms is very important to ensure the transgenic organism possesses the desired characteristics.

2.8.2 Enhancers

ENHANCERS
DNA sequences that influence the expression of a gene, often over long distances of DNA sequence.

Enhancers were first described as sequences that increase transcription initiation but, unlike promoters, were not dependent on their orientation or the distance from the transcription start site. It is now apparent that **enhancers** are generally short sequences (less than 20 to 30 bp) that bind specific transcription factors, which then facilitate the assembly of an activated transcriptional complex (i.e. the RNA polymerase) at the promoter. Most enhancers function both on the coding and non-coding strand of the DNA (i.e. in either orientation), can act up to several thousand bps distant from their target promoter, and are a rather unspecific form of regulatory element (Visel *et al.*, 2007). This implies that an enhancer element may influence several, possibly very distant, promoters. Most enhancers are only active in specific cell types and therefore play a central role in regulating tissue specificity of gene expression. Some regulatory elements bind transcription factors that act to reduce

the efficiency of transcriptional initiation, and many genes contain a combination of both positive and negative upstream regulatory elements, which then act in concert on a single promoter. This allows gene expression to be controlled very precisely in a temporal and spatial manner with regard to cell type, developmental stage and environmental conditions. Mutations of promoters or enhancers can significantly alter the expression pattern, but not the structure of a particular gene product.

2.8.3 Operators

Operators are nucleotide sequences that are positioned between the promoter and the structural gene. They constitute the region of DNA to which repressor proteins bind and thereby prevent transcription. Repressor proteins have a very high affinity for operator sequences. Repression of transcription is accomplished by the repressor protein attaching to the operator sequence downstream of the promoter sequence (the point of attachment of the RNA polymerase). The enzyme must pass the operator sequence to reach the structural genes start site. The repressor protein bound to the operator physically prevents this passage and, as a result, transcription by the polymerase cannot occur (Reznikoff, 1992). Repressor proteins themselves can be affected by a variety of other proteins or small molecules, e.g. metabolites, that affect their affinity for the operator sequence. This allows a further level of gene expression regulation to be accomplished.

2.8.4 Attenuators

The attenuator sequences are found in bacterial gene clusters that code for enzymes involved in amino acid biosynthesis. Attenuators are located within so-called leader sequences, a unit of about 162 bp situated between the promoter-operator region and the start site of the first structural gene of the cluster. Attenuation decreases the level of transcription approximately 10-fold. As the concentration of an amino acid in the cell rises and falls, attenuation adjusts the level of transcription to accommodate the changing levels of the amino acid. High concentrations of the amino acid result in low levels of transcription of the structural genes, and low concentrations of

the amino acid result in high levels of transcription. Thus the biosynthesis of an amino acid can be linked to the actual concentration of that amino acid within the cell. Attenuation proceeds independently of repression, the two phenomena are not dependent on each other. Attenuation results in the premature termination of transcription of the structural genes (Yanofsky *et al.*, 1996).

Several other regulatory elements have been described that regulate gene expression at the level of transcription. In general, the interplay of all involved factors and sequences is, in most cases and especially in eukaryotes, very complex and not entirely understood. The **expression level** of a gene is therefore the net result of all stimulating and repressing factors acting on it (Watson, 2008). This combinatorial system of positive and negative influences allows the fine-tuning of gene expression and needs to be carefully considered when designing transgenic organisms.

EXPRESSION LEVEL
The frequency with which a given gene is transcribed and translated, i.e. how much of a given gene product is produced over time.

2.9 REGULATORY mRNA SEQUENCES

In the preceding paragraph, DNA sequences were described that regulate transcription of DNA to an mRNA transcript. This transcript, sometimes referred to as pre-mRNA, contains a variety of sequences in addition to the protein-coding sequences. This includes 5′ and 3′ untranslated sequences which are important in the regulation of translation, and introns (in the case of eukaryotes) which need to be excised before the process of translation can take place. In eukaryotes, processing of a pre-mRNA to a mature mRNA that is ready for translation takes place in the same compartment as transcription, the nucleus.

2.9.1 Introns and splice junctions

In eukaryotic pre-mRNA processing, intervening sequences (introns) that interrupt the coding regions are removed (spliced out), and the two flanking protein-coding exons are joined. This splicing reaction occurs in the nucleus and requires the intron to have a GU-dinucleotide sequence at its 5′-end, an AG-dinucleotide at its

3′-end, and a specific branch point sequence. In a two-step reaction, the intron is removed as a tailed circular molecule, or lariat, and is subsequently degraded. This splicing reaction is performed by RNA-protein complexes known as snRNPs (small nuclear ribonucleoproteins). The snRNPs bind to the conserved intron sequences to form a machinery called spliceosome, in which the cleavage and ligation reactions take place (Matthew *et al.*, 2008).

2.9.2 5′ Untranslated sequences

During the processing of precursor mRNA in the nucleus, the 3′ terminus as well as introns are removed. In addition, shortly after initiation of mRNA transcription, a methylguanylate residue is added to the 5′ end of the primary transcript. This 5′ "cap" is a characteristic feature of every mRNA molecule, and the transcriptional start or initiation site is also referred to as the capping site. The 5′ UTR extends from the capping site to the beginning of the protein coding sequence and can be up to several hundred bps in length. The **5′ UTRs** of most mRNAs contain the consensus sequence 5′ –CGAGCCAUC-3 involved in the intiation of protein synthesis (i.e. translation). In addition, some 5′ UTRs contain "upstream AUGs" that may affect the initiation of protein synthesis and thus could serve to control expression of selected genes at the translational level (Hughes, 2006).

5′UNTRANSLATED REGIONS
Sequence in mRNA, upstream of the coding region, that regulate initiation of translation.

2.9.3 3′ Untranslated sequences and transcriptional termination

The 3′ end of a mature mRNA molecule is created by cleavage of the primary precursor mRNA and the addition of a several hundred nucleotide long polyadenylic acid (poly-A) tail. The site for cleavage is marked by the sequence 5′ AAUAAA 3′ some 15 to 20 nucleotides upstream and by additional uncharacterized sequences 10 to 30 nucleotides downstream of the cleavage site. The region from the last protein codon to the poly-A addition site may contain up to several hundred nucleotides

of a **3′ UTR**, which includes signals that affect mRNA processing and stability. Many mRNAs that are known to have a short half life contain a 50 nucleotide long AU-rich sequence in the 3′ UTR. Removal or alteration of this sequence prolongs the half life of mRNA, suggesting that the presence of AU-rich sequences in the 3′ UTR may be a general feature of genes that rapidly alter the level of their expression. In general, the half-life of an mRNA indicates the average time that an mRNA molecule persists in the cell and thus can be translated before it is degraded. The mRNA half-life is therefore an important variable for the level of gene expression (Gray and Wickens, 1998).

3′ UNTRANSLATED REGION
Sequence in mRNA, downstream of the coding region, that regulates mRNA processing and stability.

2.9.4 Regulation of gene expression

Regulation of gene expression refers to the all processes that cells employ to convert the information carried by genes into gene products in a highly controlled manner. Although a functional gene product may be RNA or protein, the majority of known regulatory mechanisms affect the expression level of protein coding genes. As mentioned above, any step in the process of gene expression may be modulated, from transcription, to RNA processing, to translation, to post-translational modification of the protein. Highly sophisticated **gene expression regulatory systems** allow the cell to fine-tune its requirements in response to environmental stimuli, developmental stages, stress, nutrient availability etc. (Nestler and Hyman, 2002; Watson *et al.*, 2008).

REGULATION OF GENE EXPRESSION
All mechanisms employed by a cell/organism to regulate the expression level of its genes, in response to internal or external stimuli or developmental stages.

To conclude, this chapter has provided an overview of genes, gene expression and hereditary phenomena. Although this text offers only a brief introduction to the topic, it should have become clear that correct gene expression is based on a highly complicated network and interplay of numerous factors, and a complete comprehension of these networks is only beginning to emerge. However, a good understanding of the basic principles is required to follow and understand biotechnological applications and developments, as well as the associated current

limits and difficulties of this technology. The following chapter will introduce techniques and scientific concepts that are more specific to and highly important for modern, applied biotechnology, especially in the field of GMOs. The chapter is based on the principles of DNA structure, genes, and gene expression that have been described in this chapter.

CHAPTER 3

VECTORS AND PROMOTERS

3.1 RECOMBINANT DNA TECHNOLOGY – AN OVERVIEW

Following the elucidation of the DNA structure and the genetic code, it became clear that many biological secrets were hidden in the sequence of bases in DNA. Technical and biological discoveries in the 1970s led to a new era of DNA analysis and manipulation. Key among these was the discovery of two types of enzymes that made **DNA cloning** possible: cloning, in this sense, refers to the isolation and amplification of defined pieces of DNA. One enzyme type, called ***restriction enzymes,*** cut the DNA from any organism at specific sequences of a few nucleotides, generating a reproducible set of fragments. Restriction enzymes occur naturally in many bacteria, where they serve as defence mechanisms against bacteriophage (viruses infecting bacteria) infection by cutting the bacteriophages genome upon its entry into the cell. The other enzyme type, called *DNA ligases,* can covalently join DNA fragments at their termini that have been created by restriction enzymes. Thus, ligases can insert DNA restriction fragments into replicating DNA molecules such as plasmids (bacterial, circular DNA molecules), resulting in recombinant DNA molecules. The recombinant DNA molecules can then be introduced into appropriate host cells, most often bacterial cells. All descendants from such a single cell, called a clone, carry the same recombinant DNA molecule (Figure 3.1). Once a clone of

DNA CLONING
The isolation and amplification of defined sequences of DNA.

RESTRICTION ENZYME
Enzymes, naturally present in bacteria, that cut DNA at defined sequences.

cells bearing a desired segment of DNA has been isolated, unlimited quantities of this DNA sequence can be prepared (Allison, 2007). Furthermore, in case the DNA fragment contains protein-coding genes, the recombinant DNA molecule introduced into a suitable host can direct the expression of these genes, resulting in the production of the proteins within the host. These developments, DNA cloning and the production of **recombinant proteins**, were major breakthroughs in molecular biology and set the stage for modern biological research.

RECOMBINANT PROTEINS
Proteins produced with the aid of recombinant DNA technology.

3.2 VECTORS

A vector is a DNA molecule which can replicate in a suitable host organism, and into which a fragment of foreign DNA can be introduced. Most vectors used in molecular biology are based on bacterial plasmids and bacteriophages (bacteria-infecting viruses).

VECTOR
In molecular biology, a vector is a DNA molecule that can take up foreign DNA fragments and can be used to amplify and transfer this DNA to a suitable host.

Vectors need to have the following characteristics:

- Possess an origin of replication (ori), which renders the vector capable of autonomous replication independent of the host genome.
- Have a site (or sites) which can be cleaved by a restriction enzyme, where the foreign DNA fragment can be introduced.
- Contain convenient markers for identifying the host cell that contains the vector with the inserted DNA of interest. A common selection marker is an antibiotic resistance gene. If the host bacteria cells contain the vector then the bacteria will grow in the presence of that antibiotic, whereas growth of bacteria without the plasmid is restricted.

In addition to the above-listed features, the vector should be easily introducible into the host organism where it has to replicate and produce copies of itself and the foreign DNA. Furthermore, it should be feasible to easily isolate the vector from the host cell (Watson, 2008).

3.3 TYPES OF CLONING VECTORS

3.3.1 Plasmids

Plasmids are circular, double-stranded DNA molecules that are independent from a cell's chromosomal DNA. These extrachromosomal DNAs occur naturally in bacteria and in the nuclei of yeast and some higher eukaryotic cells, existing in a parasitic or symbiotic relationship with their host cell. Most naturally occurring **plasmids** contain genes that provide some benefit to the host cell, fulfilling the plasmid's portion of a symbiotic relationship. Some bacterial plasmids, for example, encode enzymes that inactivate antibiotics. Therefore, a bacterial cell containing such a plasmid is resistant to the antibiotic, whereas the same type of bacterium lacking the plasmid is killed. Plasmids range in size from a few thousand bps to more than 100 kilobases (kb).

PLASMIDS
Circular, extrachromosomal DNA molecules, usually derived from bacteria, capable of autonomous replication and maintenance within the cell.

The plasmids most frequently used in recombinant DNA technology are derived from and replicate in *E. coli* (Jana and Deb, 2005). In general, these plasmids have been modified to optimize their use as vectors in DNA cloning. One such modification, for example, is the reduction in size to approximately 3 kb, which is much smaller than that of naturally occurring *E. coli* plasmids. In addition, most plasmids contain a multiple cloning site (MCS), a short sequence of DNA containing many restriction enzyme sites close together. Thus, many different restriction enzymes can be used for the insertion of foreign DNA fragments. In addition to antibiotic resistance genes, many modern plasmid vectors also contain a system for detecting the presence of a recombinant insert, such as the blue/white β-galactosidase system that allows simple visual screening of bacterial clones.

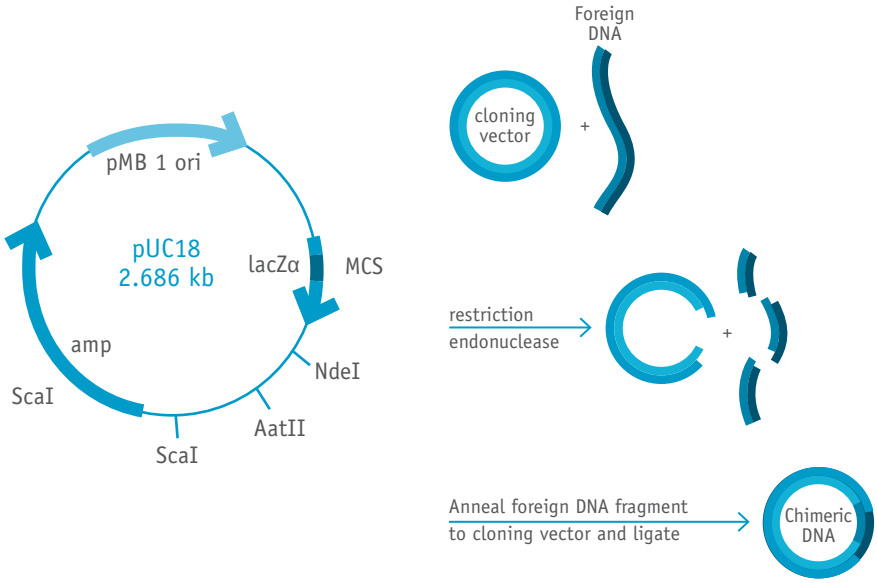
3.3.2 Bacteriophages

BACTERIOPHAGES
Viruses that infect bacteria. Bacteriophages can be manipulated and used as DNA cloning vectors.

Bacteriophages, or phages, are viruses that infect bacteria. They can display either lytic life cycles, leading to the death of the host bacterium and release of

Figure 3.1 | **A typical plasmid cloning vector and the principle of DNA cloning**

The pUC18 plasmid is a frequently used plasmid for DNA cloning. The plasmid size, ampicillin resistance gene (amp), origin of replication (ori) and multiple cloning site (MCS) are indicated. On the right hand side, the overall principle of DNA cloning is depicted.



new phage particles, or more complex lysogenic cycles during which the phage genome is integrated into the bacterial genome. One of the best studied phages is bacteriophage λ (Lambda) whose derivatives are commonly used as cloning vectors (Chauthaiwale *et al.*, 1992). The λ phage particle consists of a head containing the 48.5 kb double-stranded DNA genome, and a long flexible tail. During infection, the phage binds to certain receptors on the outer membrane of *E. coli* and subsequently injects its genome into the host cell through its tail. The phage genome is linear

and contains single-stranded ends that are complementary to each other (the so-called cos ends). Due to the complementarity, the cos ends rapidly bind to each other upon entry into the host cell, resulting in a nicked circular genome. The nicks are subsequently repaired by the cellular enzyme DNA ligase. A large part of the central region of the phage genome is dispensable for lytic infection, and can be replaced by unrelated DNA sequence. The limit to the size of DNA fragments which can be incorporated into a λ particle is 20 kb, which is significantly larger than fragments suitable for plasmids (around 10 kb maximum). A further advantage of λ-based vectors is that each phage particle containing recombinant DNA will infect a single cell. The infection process is about a thousand times more efficient than transformation of bacterial cells with plasmid vectors.

3.3.3 Cosmids

Both λ phage and *E. coli* plasmid vectors are useful for cloning only relatively small DNA fragments. However, several other vectors have been developed for cloning larger fragments of DNA. One common method for cloning large fragments makes use of elements of both plasmid and λ-phage cloning. In this method, called **cosmid** cloning, recombinant plasmids containing inserted fragments up to a length of 45 kb can be efficiently introduced into *E. coli* cells. A cosmid vector is produced by inserting the cos sequence from λ-phage DNA into a small *E. coli* plasmid vector about 5 kb long. Cosmid vectors contain all the essential components found in plasmids. The cosmid can incorporate foreign DNA inserts that are between 35 and 45 kb in length. Such recombinant molecules can be packaged and used to transform *E. coli*. Since the injected DNA does not encode any λ-phage proteins, no viral particles form in infected cells and likewise the cells are not killed. Rather, the injected DNA circularizes, forming in each host cell a large plasmid containing the cosmid vector and the inserted DNA fragment. Cells containing cosmid molecules can be selected using antibiotics as described for ordinary plasmid cloning.

COSMIDS
Artificial cloning plasmids, able to incorporate comparatively large DNA fragments of 35 to 45 kb.

A recently developed approach similar to cosmid cloning makes use of larger *E. coli* viruses such as bacteriophage P1. Recombinant plasmids containing DNA fragments of up to ≈100 kb can be packaged *in vitro* with the P1 system.

3.3.4 Yeast artificial chromosomes (YAC)

BACS AND YACS
Bacterial and yeast artificial chromosomes: Engineered chromosomes, that can act as vectors for very large DNA fragments.

YACs are constructed by ligating the components required for replication and segregation of natural yeast chromosomes to very large fragments of target DNA, which may be more than 1 Mb in length (Ramsay, 1994). **YAC** vectors contain two telomeric sequences (TEL), one centromere (CEN), one autonomously replicating sequence (ARS) and genes which act as selectable markers in yeast. YAC selectable markers usually do not confer resistance to antibiotic substances, as in *E. coli* plasmids, but instead enable growth of yeast on selective media lacking specific nutrients.

3.3.5 Bacterial artificial chromosomes (BAC)

BAC vectors were developed to avoid problems that were encountered with YACs to clone large genomic DNA fragments. Although YACs can accommodate very large DNA fragments they may be unstable, i.e. they often lose parts of the fragments during propagation in yeast.

In general, BACs can contain up to 300–350 kb of insert sequence. In addition, they are stably propagated and replicated in *E. coli*, are easily introduced into their host cell by transformation, large amounts can be produced in a short time due to the fast growth of *E. coli*, and they are simple to purify (Giraldo and Montoliu, 2001). The vectors are based on the naturally occurring plasmid F factor of *E. coli*, which encodes its own DNA polymerase and is maintained in the cell at a level of one or two copies. A BAC vector consists of the genes essential for replication and maintenance of the F factor, a selectable marker gene (SMG) and a cloning site for the insertion of target fragment DNA.

To summarize, cloning vectors are DNA molecules that can incorporate foreign DNA fragments and replicate in a suitable host, producing large quantities of the desired DNA fragment. Such methods are highly important for a variety of molecular biology applications and are the basis of recombinant DNA technology. However, for the production of transgenic organisms and related biotechnological applications, such vectors need to possess additional sequence elements and properties that allow targeted transfer of specific genes and controlled expression of these genes in a host organism. The necessary features to accomplish these tasks will be discussed in the following paragraphs.

3.4 PROMOTERS

As already introduced in Chapter 2, the promoter sequence is the key regulatory region of a gene that controls and regulates gene expression. More specifically, the promoter has a major importance in the regulation of transcription, i.e. the transfer of the information contained in a DNA coding region into an mRNA transcript. Promoters play an important role in the regulation of gene expression at different locations and times during the life cycle of an organism or in response to internal and external stimuli (Juven-Gershon *et al.*, 2008). Investigating and unravelling the precise function of promoter components and the additional factors associated with their performance revealed new possibilities of genetic engineering. Nowadays, it is feasible to modulate the expression of defined genes in an organism by combining them with a promoter of choice, resulting in the desired gene expression profile.

This approach can be used to modulate the expression of endogenous genes (i.e. genes that the organism possesses already) or to introduce foreign genes in combination with a foreign or endogenous promoter to create an organism with defined novel traits. Thus, promoters have a huge influence in follow-on research and development in biotechnology, and a more detailed understanding will certainly further influence the development of GMOs.

3.4.1 Types of promoters

PROMOTER TYPES
Promoters can be classified as constitutive, tissue-specific or inducible, according to their mode of regulating gene expression.

In general, promoters can be divided into different classes according to their function:

» **Constitutive promoters.** Constitutive promoters direct the expression of a gene in virtually all cells or tissues of an organism. The genes controlled by such promoters are often "housekeeping genes", i.e. genes whose products are constantly needed by the cell to survive and maintain its function. Constitutive promoters are to a large extent, or even entirely, insensitive to environmental or internal influences, thus the level of gene expression is always kept constant. Due to the insensitivity to external or internal stimuli and the high sequence conservation of such promoters between different species, constitutive promoters are in many cases active across species and even across kingdoms. An important example is the Cauliflower mosaic virus (CaMV) 35S promoter, which is frequently used to drive transgene expression in transgenic plants.

» **Tissue-specific promoters.** Tissue-specific promoters direct the expression of a gene in a specific tissue or cell type of an organism or during certain stages of development. Thus, the gene product is only found in those cells or tissues and is absent in others, where the promoter is inactive. In plants, promoter elements that specifically regulate the expression of genes in tubers, roots, vascular bundles, other vegetative organs or seeds and reproductive organs have been used for genetic engineering, both within a certain species and across different species. Frequently, such promoters rely on the presence or absence of endogenous factors to function, so in fact it is the presence or absence of these factors that defines the tissue-specificity of gene expression.

» **Inducible promoters.** Inducible promoters are of high interest to genetic engineering because their performance is dependent on certain endogenous or external factors or stimuli. In the ideal case, gene expression by an inducible

promoter can be controlled by the experimenter by simply adding a certain substance to the cell culture/the organism. This will result in expression of all genes controlled by this promoter – in the case of transgenic organisms, usually only the genes that have been specifically introduced (Padidam, 2003). Within the class of inducible promoters, one can find promoters controlled by abiotic factors such as light, oxygen level, heat, cold and wounding, while others are controlled by certain chemicals or metabolites. As it may be difficult to control some of these factors in the field, promoters that respond to chemical compounds, which are not found naturally in the organism of interest, are of particular interest. Substances that have been found to control certain promoters include rare metabolites, antibiotics, some metals, alcohol, steroids and herbicides, among other compounds. Once a promoter that responds to a certain compound has been identified it can be further engineered and adapted to induce gene expression in GMOs at will, independent of other factors encountered by the organism (Gurr and Rushton, 2005).

3.5 EXPRESSION VECTORS

Cloning a gene encoding a particular protein is only the first of many steps needed to produce a recombinant protein for agricultural, medical or industrial use. The next step is to transfer the DNA sequence containing the gene into the desired host cell for its expression and the production of the protein of interest. In order to allow expression of the gene of interest in the host cell or organism, it must be transferred into a vector that has several distinct sequence features. These features include all sequences that are required to drive and regulate expression of the gene, i.e. all components that are associated with a functional gene (see 2.8 and 2.9). Thus, in addition to the characteristics described for cloning vectors, an **expression vector** must carry a promoter, a polyadenylation site, and a transcription termination sequence. These sequences should have a correct orientation with regard to the multiple cloning site, where the foreign DNA is integrated. Inserting a coding

EXPRESSION VECTOR
A vector that, in addition to the properties of a cloning vector, contains sequences that direct expression of the inserted DNA sequence in an appropriate host organism.

sequence in proper orientation in between these expression control sequences will result in the expression of the gene in an appropriate host. A simplified version of an expression vector is depicted in Figure 3.2:

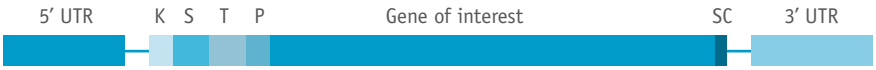
Figure 3.2 | **Generalized mammalian expression vector**

The multiple cloning site (MCS), where the foreign DNA can be inserted, and selectable marker gene (SMG) are under control of a eukaryotic promoter (p), polyadenylation (pa), and termination of transcription (TT) sequences. An intron (I) enhances the production of heterologous protein. Propagation of the vector in E. coli and mammalian cells depends on the origins of replication ori^{E} and ori^{euk}, respectively. The ampicillin gene (Amp^{r}) is used for selecting transformed E. coli cells.

In some cases, it is necessary and helpful to fuse some translation control and protein purification elements to the gene of interest (Figure 3.3) or to add them to the expression vector MCS. This is especially important if a recombinant protein is purified after its expression in a certain host cell or organism. For this purpose, short specific amino acid sequences, commonly referred to as tags, can be added to the protein by adding the sequence encoding them to the coding sequence of the protein. These tags can greatly facilitate protein purification, due to certain properties they possess and that are specific for each tag. If necessary, such tags can be removed from the final, purified protein by introducing a further specific amino acid sequence, which is recognized by a protease that cleaves the protein at this position and thus removes the tag. An example of a gene with such added sequences is given in Figure 3.3:

Figure 3.3 | **A gene of interest fitted with sequences that enhance translation and facilitate both secretion and purification of the produced protein**

These include the Kozak sequence (K) [5′-ACCAUGG-3′, its presence near the initiating AUG greatly increases the effectiveness of initiation], signal sequence (S) required for secretion, protein affinity tag (T), proteolytic cleavage site (P), and stop codon (SC). The 5′ and 3′ UTRs increase the efficiency of translation and contribute to mRNA stability.



In the case of **transgenic plant and animal** production, the general layout of an engineered gene as depicted in Figure 3.3 also holds true in most cases. However, other types of vectors to deliver the transgene to the plant or animal cells are frequently employed. Whereas cells in cell culture can be easily monitored for the presence of the desired expression vector and the expression vector is stably maintained within the cells, this is not necessarily the case for complex organisms. Therefore, the genes of interest are usually integrated in a vector that mediates integration of the transgene into the host organism's genome (i.e. into a chromosome). Thus, the transgene becomes an integral part of the organism's genome, and as such is present in all cells of an organism and is stably passed on to subsequent generations (Somers and Makarevich, 2004). This is usually not the case for plasmid vectors, which are maintained as extra-chromosomal entities and are frequently lost during cell divison and propagation.

TRANSGENIC PLANTS AND ANIMALS
Organisms, in which foreign DNA has been introduced by recombinant DNA technology.

The vectors and techniques that are employed to produce stable transgene integrations into the genome of a given organism are described in detail in the following chapters.

To summarize, this chapter has provided an introduction to the field of recombinant DNA technology. Specific DNA fragments can be cloned, by means of cloning vectors, and subsequently be isolated, investigated and further modified with great ease. Furthermore, specific DNA fragments containing protein-coding genes can be transferred to expression vectors, which will result in expression of the encoded proteins upon introduction of the vector into an appropriate host cell or organism. Thus, desired proteins can be produced in large quantities. Careful choice of the vector, the production host and promoter and other regulatory sequences is of high importance for the success of such approaches. Modern biotechnology offers the possibility to freely combine genes with promoters and other desired sequences, regardless of the original source of the genes and DNA sequences.

This technology also sets the basis for the creation of transgenic plants and animals, which are engineered to express new traits and properties by the specific introduction or modulation of genes and regulatory sequences. Plant and animal recombinant DNA techniques are introduced in the following two chapters.

CHAPTER 4

PLANT TRANSFORMATION AND SELECTION TECHNIQUES

4.1 PLANT TRANSFORMATION

In the last two chapters the molecular techniques, generally referred to as recombinant DNA technology, that allow isolation, manipulation and expression of specific genes were described. Furthermore, potential applications of this technology to produce specific proteins for medical or industrial use in cell culture were also discussed.

This chapter will provide the link between recombinant DNA technology and the creation of transgenic plants that possess novel traits of interest to agriculture, medicine or industry. This application is based on the techniques described so far, but in addition relies on novel techniques that are specific to and necessary for the creation of transgenic plants.

Genetic transformation is the (sometimes heritable) change in the genome of a cell or organism brought about by the uptake and incorporation of introduced, foreign DNA. Transformation encompasses a variety of gene transfer events, which can be characterized by the stability of transformation, the subcellular compartment transformed (nuclear, mitochondrial or plastid) and whether the transferred DNA is stably integrated into the host genome (Shewry *et al.*, 2008).

GENETIC TRANSFORMATION
The uptake of foreign DNA sequences into a cell. In some cases, this may take the form of stable incorporation of the DNA into the cell's/organism's genome.

Table 4.1 documents the generally accepted definitions of these alternative transformation events.

Table 4.1 | **Definitions of transformation**

Term	Definitions
Stable transformation	The transgene and novel genetic characteristics are stably maintained during the life of the cell culture or organism. The transgene is usually, but not necessarily, integrated into the host genome.
Transient expression	Expression of the transgene is detected in the first few days after its introduction into host cells. A subsequent decline in expression indicates that expression was based on non-integrated, extra-chromosomal DNA.
Integrative transformation	The transgene is covalently integrated into the genome of the host cell. In fertile plants (or animals) the transgene is inherited by the next generation (a form of stable integration).
Nuclear transformation	Gene transfer into the nuclear genome of the host cell, as confirmed by cellular fractionation, eukaryotic-type expression or mendelian inheritance.
Organellar transformation	Gene transfer into the plastid or mitochondrial genome of the host cell, as confirmed by cellular fractionation, prokaryotic-type expression or maternal inheritance.
Episomal transformation	Viral genomes or "mini-chromosomes" are introduced which replicate independently from the host genome. Stable over several generations in some cases.

4.2 PLANT TISSUE CULTURE

An important phenomenon that is a key determinant to plant transformation, and thus the generation of transgenic plants, is the finding that whole plants can be regenerated from single cells. Plant transformation thus depends on two events: successful introduction of foreign DNA into target plant cells, and subsequent development of a complete plant derived from the transformed cells.

PLANT REGENERATION
The possibility to regenerate complete plants from plant cell culture.

***In vitro* regeneration** is the technique of developing plant organs or plantlets from plant cells, tissues or organs isolated from the mother plant and cultivated on artificial media under laboratory conditions (Thorpe, 2007). Depending on different physical and physiological factors, in combination with various growth regulators, regeneration occurs via organogenesis (initiation of adventitious roots or shoots from plant cells or tissues) or embryogenesis (formation of plants

from somatic cells through a pathway resembling normal embryogenesis from the zygote). Both organogenesis and embryogenesis can be initiated either directly (from meristematic cells) or after formation of a callus (mass of undifferentiated parenchymatic cells induced by wounding or hormone treatment).

Transformed plants can thus be regenerated from calli or wounded plant tissues, such as leaf disks, into which foreign DNA has previously been introduced (Figure 4.1).

4.3 PLANT TRANSFORMATION TECHNIQUES

There is an expanding repertoire of **plant transformation** techniques available, ranging from established techniques to highly experimental methodologies (Newell, 2000). In Table 4.2 these alternative approaches to gene delivery are listed with brief comments on their application, efficiency and limitations. The most widely used techniques are the *Agrobacterium tumefaciens*-mediated transfer, microprojectile bombardment ("gene gun" or biolistic method) and direct gene transfer to protoplasts. The biolistic technique has proven especially useful in transforming monocotyledonous species like maize and rice, whereas transformation via *Agrobacterium* has been successfully practised in dicotyledonous species. Only recently has it also been effectively employed in monocotyledons. In general, the *Agrobacterium*-mediated method is considered preferable to the gene gun due to the higher frequency of single-site insertions of the foreign DNA into the host genome, making the transformation process easier to monitor. All available and currently employed transformation techniques are briefly described in the following sections.

PLANT TRANSFORMATION TECHNIQUES
All methods to achieve the uptake and introduction of foreign DNA into a plant's genome.

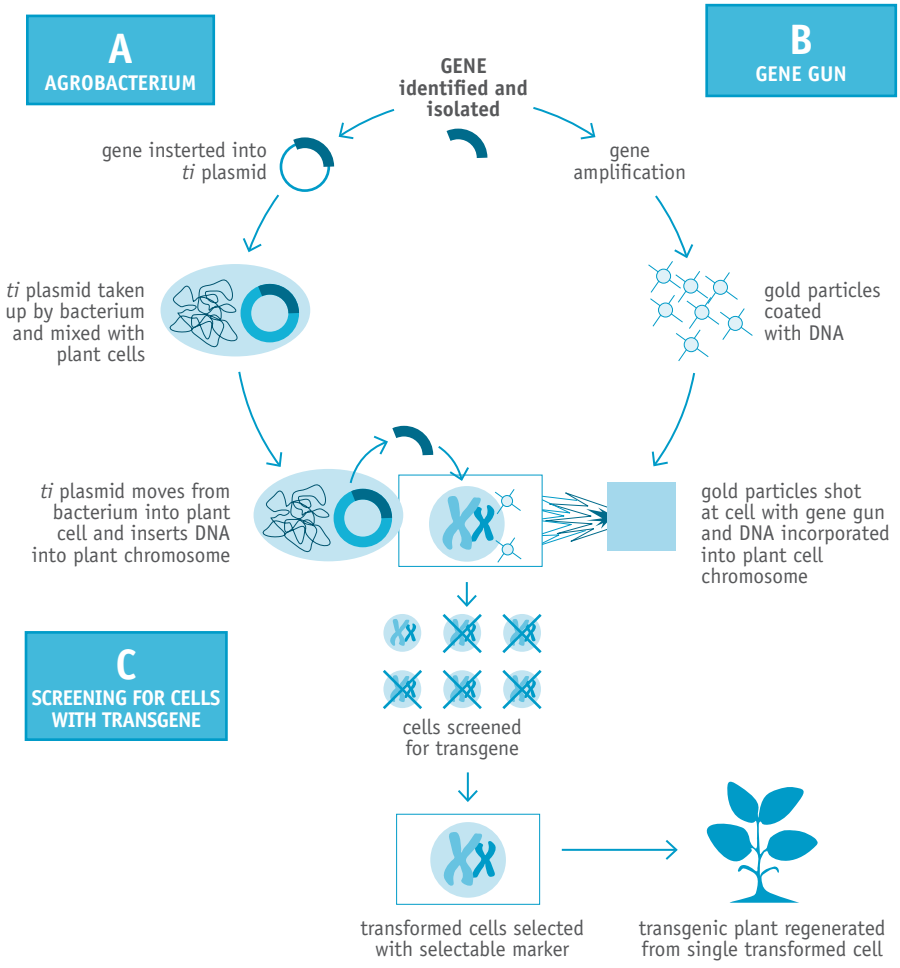
4.3.1 Microprojectile bombardment

This technique uses high velocity particles, or **microprojectiles**, that are coated with DNA and deliver exogenous genetic material into the target cell or tissue. Transformed cells are selected, cultured *in vitro* and regenerated to produce mature transformed plants (Kikkert *et al.*, 2005).

MICROPROJECTILE BOMBARDMENT
Plant cell transformation by shooting DNA-coated microparticles into plant material.

Figure 4.1 | **Steps involved in the generation of genetically transformed plants using either the *Agrobacterium tumefaciens* (*A. tumefaciens*) or microprojectile bombardment approaches**

Following introduction of foreign DNA into the plant cell, successfully transformed cells are selected and used to regenerate a transgenic plant (see text for details).



The particles, either tungsten or gold, are small (0.5-5 µm) but big enough to have the necessary mass to be sufficiently accelerated and penetrate the cell wall carrying the coated DNA on their surface. Once the foreign DNA is integrated into the plant genome in the cell nucleus, which is a somewhat spontaneous process, it can be expressed. Gold particles are chemically inert, although rather costly, and show a high uniformity. Tungsten particles, despite showing mild phytotoxicity and being more variable in size, are adequate for most studies. Furthermore, the chosen microprojectile should have good DNA binding affinity but, at the same time, be able to release the DNA once it has hit the target. DNA coating of surface-sterilized particles can be accomplished by defined DNA treatments using, for instance, the calcium chloride method, with the addition of certain chemicals to protect the DNA. However, a recent report describes the novel use of *Agrobacterium* as coating material for the microprojectiles, which are then shot into the target tissue. Once coated the particles are ready for shooting; the particles are accelerated and ultimatively collide with the target, usually plant cells or calli grown on a Petri dish. The DNA, delivered with this strategy, is expressed after reaching the nucleus and integrating randomly into the plant genome.

4.3.2 Agrobacterium-mediated plant transformation

A. tumefaciens are soil bacteria that have the ability to infect plant cells and transfer a defined sequence of their DNA to the plant cell in the infection process. Upon integration of the bacterial DNA into a plant chromosome, it directs the synthesis of several proteins, using the plant cellular machinery, that ensure the proliferation of the bacterial population within the infected plant. *Agrobacterium* infections result in crown gall disease (Gelvin, 2003).

In addition to its chromosomal genomic DNA, an ***A. tumefaciens*** cell contains a plasmid known as the Ti (tumour-inducing) plasmid. The Ti plasmid contains a series of *vir* (virulence) genes that direct the infection process, and a stretch of DNA termed

A. TUMEFACIENS
Pathogenic plant bacterium that has the ability to transfer a part of its DNA to the plant during the infection process.

T-DNA (transfer DNA), approximately 20 kb in length, that is transferred to the plant cell in the infection process. The T-DNA encodes proteins required for the maintenance of infection. These proteins include certain plant hormones that stimulate cell growth, resulting in the formation of galls, and proteins required for a certain metabolic pathway that secures the availability of nutrients for the bacteria (Figure 4.2).

Agrobacterium can only infect plants through wounds. When a plant root or stem is wounded it gives off certain chemical signals. In response to these signals, agrobacterial *vir* genes become activated and direct a series of events necessary for the transfer of the T-DNA from the Ti plasmid to the plant cell through the wound.

TI-PLASMID
The *A. tumefaciens* plasmid that is responsible for transferring DNA to the plant genome in the infection process. Engineered versions of the Ti-plasmid are used as transgene vectors for plant transformation.

To harness *A. tumefaciens* and the **Ti-plasmid** as a transgene vector, the tumor-inducing section of T-DNA is removed, while the T-DNA border regions and the *vir* genes are retained. The desired transgene is inserted between the T-DNA border regions, applying recombinant DNA technology. Thus, in the infection process, the transgene DNA is transferred to the plant cell and integrated into the plant's chromosomes (Lacroix *et al.*, 2006). To achieve transformation, *Agrobacterium* cells carrying an appropriately constituted Ti plasmid vector containing the desired transgene can be inoculated into plant stems, leaf disks etc., to allow infection and T-DNA transfer to the plant cells. The explants that have been co-cultivated with *Agrobacterium* are subsequently processed through various tissue culture steps resulting in the selection and production of transformed cells and plants.

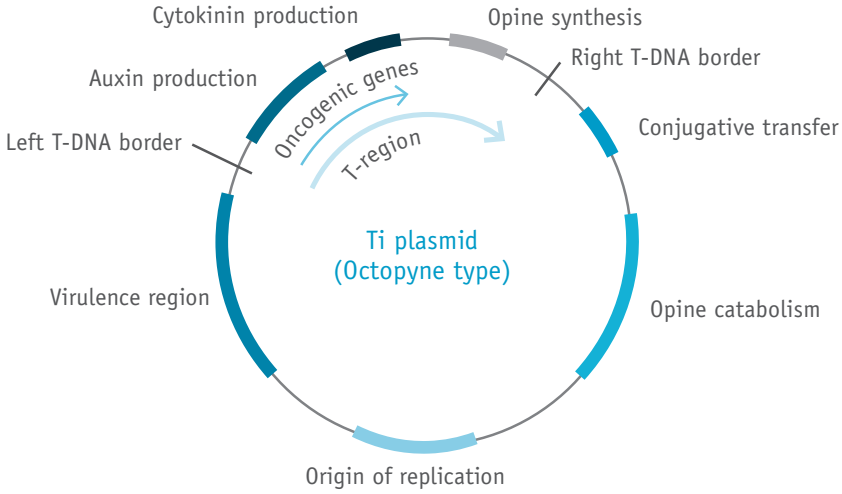
4.3.3 Protoplast transformation techniques

PROTOPLAST
Cultured plant cells whose cell wall has been removed.

One of the characteristic features of plant cells is that they are surrounded by a rigid, cellulose-based cell wall. **Protoplasts** are plant cells in which the cell wall has been removed (Davey *et al.*, 2005). Therefore protoplasts behave like animal cells, which naturally have no cell wall barrier. Plant regeneration from single protoplasts is possible due to the totipotency of plant cells, i.e. the potential of

Figure 4.2 | **Wild type Ti plasmid of *Agrobacterium tumefaciens* (*A. tumefaciens*)**

The T region, i.e. the region of the plasmid that can be exchanged and replaced with the transgene of interest, is highlighted in light blue.



a single cell to reconstitute a complete plant. Removal of the cell wall is achieved by treating the plant material (leaves, tissue cultures, suspended cells, etc.) with a cocktail of cell wall-destroying enzymes, including pectinases, cellulases, and/or hemicellulases in an appropriate incubation medium of the correct osmolality (i.e. the concentration of solutes in the medium). After removal of the cell wall, the protoplasts must be kept immersed in a solution of the appropriate solute concentration to prevent them from bursting. Thus, monitoring the correct osmolality of the culture medium until a new cell wall has formed is of high importance.

Different approaches exist for the delivery of transgene DNA into protoplasts through the plasma membrane. These include chemical treatments, electroporation and micro-injection techniques (Davey *et al.*, 2005).

4.3.3.1 Chemical techniques

The most commonly applied chemical protoplast transformation methods include polyethylene glycol (PEG) treatment, Ca^{2+}-DNA co-precipitation and liposomal DNA delivery. PEG treatment is the most widely used technique, employing solutions of 10-15 percent PEG in combination with high calcium content and a high pH. After mixing the isolated DNA and the protoplasts, followed by different washes, the DNA may be taken up by the protoplast. The role of PEG is to alter the plasma membrane properties, causing a reversible membrane permeabilization, thus enabling exogenous macromolecules to enter the cell cytoplasm.

Ca^{2+}-DNA co-precipitation depends on the formation of a co-precipitate of plasmid DNA and calcium phosphate. On contact with protoplasts under high pH conditions, the co-precipitate trespasses the cell's plasma membrane.

Liposomes, which are negatively-charged spheres of lipids, are also employed for DNA transfer and uptake into cells. DNA is first encapsulated into the liposomes which are subsequently fused with protoplasts, employing PEG as a fusogen.

4.3.3.2 Electroporation

Electrical pulses are applied to the DNA-protoplast mixture, provoking an increase in the protoplast membrane permeability to DNA. This technique is much simpler than the chemical method, providing satisfying results. However, the electrical pulses must be carefully controlled as cell death can occur above a certain threshold. The pulses induce the transient formation of micropores in the membrane lipid bilayer which persist for a few minutes, allowing DNA uptake to occur.

4.3.3.3 Micro-injection

This technique was originally designed to transform animal cells, and was later adapted for and gained importance in transforming plant cells. However, in plant cells the existence of a rigid cell wall, a natural barrier, prevents micro-injection. Furthermore, the presence of vacuoles that contain hydrolases and toxic metabolites that may lead to cell death after vacuole breakage presents a severe restriction to micro-injection. Therefore, protoplasts, rather than intact plant cells, are more suitable for micro-injection. This method is labour-intensive and requires special micro-equipment for the manipulation of host protoplasts and DNA. However, some success in transforming both monocotyledonous and dicotyledonous species has been achieved employing this technique.

4.3.4 Virus-mediated plant transformation/transduction

Virus-based vectors have been shown to be efficient tools for the transient, high-level expression of foreign proteins in plants (Chung *et al.*, 2006). These vectors are derived from plant viruses, e.g. Tobacco Mosaic Virus (TMV), and are manipulated to encode a protein of interest. Initial delivery of the virus-based vector to the plant can be achieved by *Agrobacterium* - the vector is encoded in the T-DNA, which is transferred to the plant. This method is applicable to whole plants, by the process of agroinfiltration, circumventing the need for labour-intensive tissue-culture.

Within a plant cell, the **virus-based vectors** are autonomously replicated, can spread from cell to cell and direct the synthesis of the encoded protein of interest. The advantages of this method are the applicability to whole plants and thus a much faster outcome than the establishment of a transgenic plant, and the high-level expression of the desired protein within a short time. The major disadvantage is that the process is transient: the expression level decreases over time, and the genetic change is not passed on to subsequent generations, i.e. it is not heritable.

VIRUS-BASED VECTORS
Vectors based on viral genomes, or parts thereof. Increasingly used for both stable and transient DNA transfer to cells or organisms.

The process of virus-mediated DNA transfer is referred to as transduction.

Several other plant transformation techniques, which have been reported but could not be reproduced or did not gain significant importance, are listed below in Table 4.2:

Table 4.2 | **Summary of plant transformation techniques**

Gene delivery method	Characteristics
Agrobacterium	Well-established transformation vector for many dicots and several monocots and a promising vector for gymnosperms. A wide range of disarmed Ti- or Ri-derived plasmid vectors are available. Additional value for the delivery of viral genomes to suitable hosts by agroinfiltration.
Direct DNA transfer to protoplasts	Well-established transformation technique with wide host range. Permeabilization of the plasma membrane to DNA by chemical agents or electroporation. Alternatively, genes can be delivered to protoplasts by injection or fusion with DNA in encapsulated liposomes.
Microprojectile bombardment	A widely used technique for introducing DNA via coated particles into plant cells. No host range limitation. Gene transfer to *in situ* chloroplasts has been documented.
Micro-injection	Effective gene delivery technique allowing visual DNA targeting to cell type and intracellular compartment. Labour-intensive and requiring specialist skills and equipment.
Macroinjection	Technically simple approach to deliver DNA to developing floral tissue by a hypodermic needle. Germline transformation not reproducibly reported.
Impregnation by whiskers	Suspensions of plant cells mixed with DNA and micron-sized whiskers. Both transient expression and stable transformation observed.
Laser perforations	Transient expression observed from cells targeted with a laser microbeam in DNA solution.
Impregnation of tissues	Transient and stable expression from tissue bathed in DNA solution or infiltrated under vacuum.
Floral dip	Stable DNA integration and expression following dipping of floral buds into DNA solution.
Pollen tube pathway	Claims of germ line transformation by treating pollen or carpels with DNA; remains controversial.
Ultrasonication	Stable transformation by ultrasonication of explants in the presence of DNA reported. Confirmation required.

4.4 SELECTION OF SUCCESSFULLY TRANSFORMED TISSUES

Following the transformation procedure, plant tissues are transferred to a selective medium containing a certain selective agent, depending on which SMG was used in the transgene expression cassette. Selectable markers are genes which allow the **selection** of transformed cells, or tissue explants, by enabling transformed cells to grow in the presence of a certain agent added to the medium (Miki and McHugh, 2004). One can differentiate between negative and positive selection: in positive selection, transformed cells possess a growth advantage over non-transformed cells, while in negative selection transformed cells survive whereas non-transformed cells are killed. Negative selection is the method of choice for most approaches. Thus, only cells/plants expressing the SMG will survive and it is assumed that these plants will also possess the transgene of interest. All subsequent steps in the plant regeneration process will only use the surviving cells/plants. In addition to selecting for transformants, marker genes can be used to follow the inheritance of a foreign gene in a segregating population of plants.

SELECTION
The process of selecting cells or organisms that have been successfully transformed with the desired transgene.

In some instances, transformation cassettes also include **marker/reporter genes** that encode gene products whose enzymatic activity can be easily assayed, allowing not only the detection of transformants but also an estimation of the level of foreign gene expression in the transgenic tissue. Markers such as β-glucuronidase (GUS), green fluorescent protein (GFP) and luciferase allow screening for enzymatic activity by histochemical staining or fluorimetric assays of individual cells and can be used to study cell-specific as well as developmentally regulated gene expression. These types of transgene constructs are usually used for optimizing transformation protocols and not for the development of commercial GM crops.

SELECTABLE MARKER/ REPORTER GENES
Genes that are incorporated into the transgene cassette and facilitate selection of transformed cells. Usually they confer resistance to certain antibiotic substances or allow visual detection of transformed cells.

In some cases, it may be desirable to produce a transgenic plant that does not contain the SMG used for the initial selection of transformed cells. Concerns have been raised about the release of transgenic plants containing antibiotic resistance or herbicide resistance genes, since the possibility of gene transfer to other

species cannot be ruled out. Therefore, techniques for producing marker-free plants have been developed, by either using markers not based on herbicide/antibiotic tolerance or by specifically deleting the SMG after selection of transformed cells (Darbani *et al.*, 2007).

4.5 SELECTABLE MARKER GENES (SMG)

The selectable portions on most transformation vectors are prokaryotic antibiotic resistance enzymes, which will also confer resistancy when they are expressed in plant cells. In some experiments, enzymes providing protection against specific herbicides have also been used successfully as marker genes (Miki and McHugh, 2004). The selective agent employed, i.e. the antibiotic or herbicide, must be able to exert stringent selection pressure on the plant tissue concerned, to ensure that only transformed cells survive. Below, some commonly used marker genes are briefly presented.

4.5.1 Neomycin phosphotransferase (npt-II) gene and hygromycin phosphotransferase (hpt) gene

Neomycin phosphotransferase-II (*npt*-II) is a small bacterial enzyme which catalyses the phosphorylation of a number of aminoglycoside antibiotics including neomycin and kanamycin. The reaction involves transfer of the γ-phosphate group of adenosine triphosphate (ATP) to the antibiotic molecule, which detoxifies the antibiotic by preventing its interaction with its target molecule - the ribosome. The hygromycin phosphotransferase (hpt) gene, conferring resistance to the antibiotic hygromycin, is also commonly used as selection marker.

4.5.2 Chloramphenicol acetyltransferase (CAT) gene

The chloramphenicol resistance (*cat)* gene encodes the enzyme chloramphenicol acetyltransferase (CAT) and was the first bacterial gene to be expressed in plants.

The enzyme specifically acetylates chloramphenicol antibiotics, resulting in the formation of the 1-, 3-, and 1,3-acetylated derivatives, which are inactive. Although not used as a selection system in plants, the gene is used frequently as a reporter gene in plant promoter studies.

4.5.3 Phosphinothricin acetyltransferase genes (bar and pat genes)

A commonly used herbicide is phosphinothricin (PPT, also known as Glufosinate). This compound binds to and inhibits glutamine synthethase, which is an important enzyme in the nitrogen metabolism and ammonium fixation pathways. PPT-induced glutamine synthetase inhibition results in elevated cellular ammonium levels and cell death. The enzyme phosphinothricin acetyltransferase (PAT), first identified in *Streptomyces hygroscopicus,* acetylates and thus detoxifies PPT. This allows transformed cells, or complete transgenic plants, to survive and grow in the presence of PPT.

4.5.4 β-Glucuronidase gene (GUS)

The *E. coli* β-glucoronidase gene has been adapted as a reporter gene for the transformation of plants. β-glucuronidase, encoded by the *uidA* locus, is a hydrolase that catalyses the cleavage of a wide variety of β-glucuronides, many of which are available commercially as spectrophotometric, fluorometric and histochemical substrates.

There are several features of the GUS gene which make it a useful reporter gene for plant studies. Firstly, many plants assayed to date lack detectable intrinsic glucuronidase activity, providing a null background in plants. Secondly, glucuronidase is easily, sensitively and cheaply assayed both *in vitro* and *in situ* and is sufficiently robust to withstand fixation, enabling histochemical localization in cells and tissue sections. The preferred histochemical substrate for tissue localization of GUS is 5-bromo-4chloro-3-indolyl-β-D-glucuronide (X-gluc). The advantage of

these substrates is that the indoxyl group produced upon enzymatic cleavage dimerizes to indigo which is virtually insoluble in an aqueous environment. The histochemical assay for GUS consists of soaking tissue in substrate solution and analysing the appearance of blue colour.

4.5.5 Luciferase gene

The luciferase (*luc)* gene isolated from *Photinus pyralis* (firefly) encodes the enzyme catalysing the ATP/oxygen-dependent oxidation of the substrate luciferin, resulting in the emission of light (bioluminescence). As a reporter, the gene is the basis of highly sensitive assays for promoter activity and for protein targeting sequences, involving the measurement of light emission using liquid scintillation counter photomultipliers, luminometers, X-ray film exposure or sensitive camera film.

4.5.6 Green fluorescent protein (GFP)

GFP is a widely used marker protein in modern biological research. The protein shows green fluorescence upon exposure to blue light. Originally, the protein was isolated from the jellyfish *Aequorea victoria*, but nowadays several other varieties from other marine organisms, as well as engineered versions (with different colour fluorescence), are available. GFP is widely applied for studies addressing gene expression or promoter efficiency as well as protein localization, stability and degradation.

4.6 MOLECULAR ANALYSIS OF TRANSGENIC PLANTS

After the successful transformation and selection of plant cells and the subsequent regeneration of a transgenic plant (see 4.2), it is desirable to monitor the presence of the transgene in the plant and to investigate the expression levels of the introduced genes encoding the protein(s) of interest (Stewart, 2005).

Analysis of transgenic plants at the molecular level is mainly performed by PCR (Box 7.1) and Southern blot analysis. PCR indicates the presence of the desired transgene within the plant, whereas stable integration of the transgene into the cellular genome is confirmed by Southern blot analysis. If plants are analysed that have been transformed using *A. tumefaciens*, it is important to prepare plant DNA from sterile tissue, as contamination with *A. tumefaciens* DNA will interfere with the interpretation of the results. Southern blot analysis using genomic DNA also yields information on the copy number of the integrated DNA sequences, whether any multiple inserts are tandemly linked or dispersed throughout the genome, and on the stability of the integrated DNA in the F_1 progeny of the transformed plants.

ANALYSIS OF TRANSGENIC PLANTS
To verify the presence of the desired transgene and the expression level, i.e. the accumulation of the desired protein(s).

Molecular analysis of the protein expression levels, including tissue-specific expression, developmental stage-specific expression, expression upon certain stimuli and so on, can be assayed by enzyme-linked immunosorbent assay ELISA (Box 7.2) or immunostaining of plant tissue. Expression of a gene of interest can also be assayed by determining the presence and quantity of the corresponding RNA transcript, e.g. by applying a modified PCR protocol (reverse transcriptase PCR) [RT-PCR]). All techniques will be described in detail in Chapter 7.

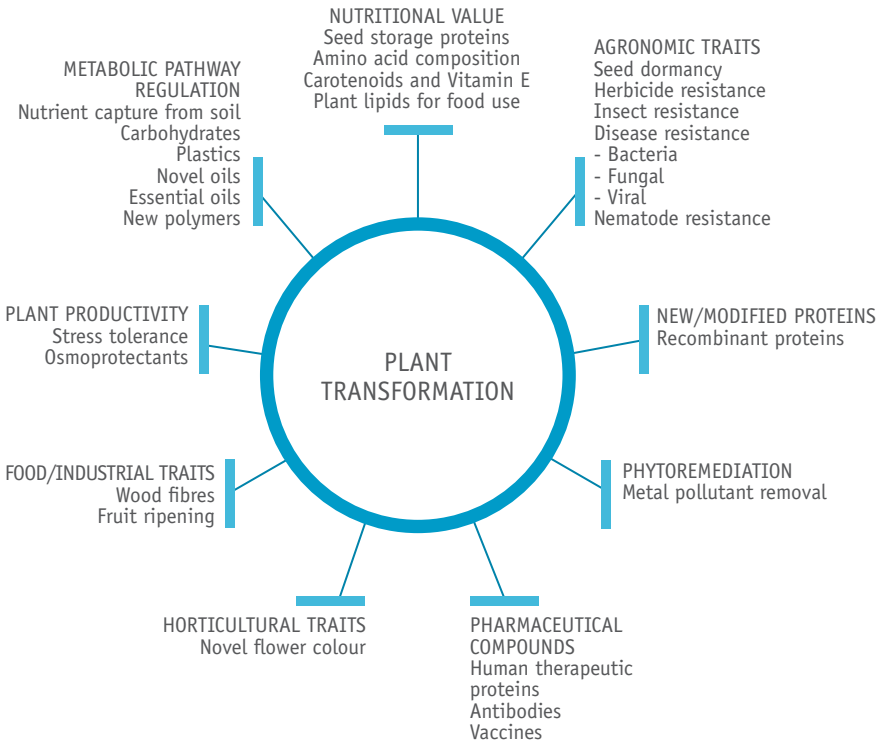
4.7 APPLICATION OF TRANSGENIC PLANTS

Numerous applications of transgenic plants are already reality or are envisaged and under investigation for the future; the main transgene targets being pest resistance and herbicide tolerance. In addition, resistances to abiotic stresses, such as drought, or improved nutrient profiles are increasingly investigated. Further possible applications that are under development are the production of medically valuable proteins or chemicals in plants (biopharmaceuticals), or the production of edible plants containing vaccines. In recent years, the technique of gene stacking, i.e. the introduction and targeting of several traits within one plant species, has also gained significant importance. Since the applications of transgenic plants are

PEST RESISTANCE AND HERBICIDE TOLERANCE
To date, these traits are the main targets for the development of transgenic plants.

diverse and numerous, no complete coverage of the field will be provided at this point. Please refer to the Annex for selected examples of transgenic plant applications explained in more detail and containing relevant literature references.

Figure 4.3 | **Applications of transgenic plants that are already available or envisaged for the future**



Adapted from: Newell, 2000.

CHAPTER 5

BIOTECHNOLOGY IN ANIMAL PRODUCTION

Modern biotechnology provides a number of possible applications in animal and livestock production. Research and development in the field focuses on improving animal growth, enhancing reproduction rates, enhancing breeding capacity and outcomes, improving animal health and developing new animal products (Basrur and King, 2005). The major techniques used to achieve these goals are the creation of transgenic animals, manipulation of animal reproduction, marker-assisted selection (MAS), molecular disease diagnostic and application of biotechnology to modify animal feed.

Once again, these technologies and applications are to a large extent based on the principles described in Chapters 2 and 3. This chapter looks at these biotechnological applications in detail and how they are impacting on animal improvement and production. Important to note is that to date no genetically modified agricultural animal has been approved for commercial release, in sharp contrast to the numerous commercial applications of GM plants. However, other biotechnological applications, which are not based on GMO production, are successfully applied in the field of animal and livestock industry.

5.1 BIOTECHNOLOGY IN ANIMAL BREEDING AND REPRODUCTION

Animal breeding, nowadays, is a field that is influenced by a whole range of biotechnological applications and developments (Bazer and Spencer, 2005). The common goal of all efforts undertaken in this field is **genetic progress** within a population, i.e. the improvement of the genetic resources and, ultimately, the phenotypic outcome. Genetic progress is influenced by several factors, namely the accuracy of choosing candidates for breeding, the additive genetic variation within the population, the selection intensity (i.e. the proportion of the population selected for further breeding), and the generation interval (the age of breeding). Note that the first three factors need to be increased in order to increase genetic progress, whereas the last factor, being generation interval, needs to be decreased. All factors can be influenced, to a varying extent, by modern biotechnology.

GENETIC PROGRESS
The improvement of the genetic resources within a population.

The techniques that are currently available to reach this end can be divided into two different groups. The first group includes all technologies that interfere with reproduction efficiency: artificial insemination, embryo transfer (ET), embryo sexing, multiple ovulation, ova pick-up and cloning, amongst others. The outcome of these technologies is an increased breeding accuracy, selection intensity and, in some cases, a shortened generation interval.

The second group of applications is based on the molecular determination of genetic variability and the identification of genetically valuable traits and characteristics. This includes the identification and characterization of **quantitative trait loci** (QTL) and the use of molecular markers for improved selection procedures. Quantitative traits are phenotypic characteristics that show a distribution of expression degree within a population (usually represented by a normal distribution), and that are based on the interaction of at least two genes (also known as polygenic inheritance). A typical example in humans is skin colour, which is based on the interaction of several genes

QUANTITATIVE TRAIT LOCI
Genes, or associated sequences, that contribute to a trait that is based on the influence of many different genes.

resulting in a large variety of phenotypes. A QTL is a DNA sequence that is associated with a certain quantitative trait – not even necessarily a gene that contributes to the trait, but possibly a sequence that is close in space to involved gene(s). Knowledge of the loci responsible for a certain quantitative trait and the underlying genes can help to select individuals for further breeding, or to start genetic engineering of the trait in question. Below, the most frequently applied techniques in animal breeding and reproduction will be summarized and explained in more detail.

5.1.1 Artificial insemination (AI)

ARTIFICIAL INSEMINATION Collection of semen from a particular individual and using it for the insemination of several female animals.

A**rtificial insemination** (AI) is the process of collecting semen from a particular male (e.g. a bull) that is subsequently used for the fertilization of many females (e.g. cows) (Galli *et al.*, 2003). The semen can be diluted and preserved by freezing (cryopreservation). This technique can enable a single bull to be used for fertilization simultaneously in several countries for up to 100 000 inseminations a year. The high intensity and accuracy of selection arising from AI can lead to a four-fold increase in the rate of genetic improvement in dairy cattle relative to that from natural mating. Since its establishment in the 1950s, AI has proven to be a very successful biotechnology, greatly enhancing the efficiency of breeding programmes (Rege, 1994). Use of AI can reduce the transmission of venereal diseases in a population and the need for farmers to maintain their own breeding males. Furthermore, it facilitates more accurate recording of pedigree and minimizes the cost of introducing improved stock. AI has significant importance for the breeding of cattle, swine and poultry.

5.1.2 Embryo transfer (ET)

EMBRYO TRANSFER Recovery or *in vitro* production of embryos and subsequent transfer to a foster mother.

Although not economically feasible for commercial use on small farms at present, **embryo technology** can greatly contribute to research and genetic improvement in local breeds. There are two procedures presently available for the production of embryos from donor females (McEvoy *et al.*, 2006). One consists of superovulation

using a range of hormone implants and treatments, followed by AI and then flushing of the uterus to gather the embryos. The other, called *in vitro* fertilization (IVF) consists of recovery of eggs from the ovaries with the aid of the ultrasound-guided transvaginal oocyte pick-up (OPU) technique. When heifers reach puberty at 11-12 months of age, their oocytes may be retrieved weekly or even twice a week. These are matured and fertilized *in vitro* and kept until they are ready for implantation into foster females. In this way, high-value female calves can be used for breeding long before they reach their normal breeding age. IVF facilitates recovery of a large number of embryos from a single female at a reduced cost, thus making ET techniques economically feasible on a large scale. Additionally, IVF produces embryos suitable for cloning experiments. However, ET is still not widely used despite its potential benefits.

5.1.3 Embryo sexing

Technologies for rapid and reliable sexing of embryos allow the generation of the desired sex at specific points in a genetic improvement programme, markedly reducing the number of animals required and enabling increased breeding progress. A number of approaches to the sexing of semen have been attempted; however, the only method of semen sexing that has shown any promise has been the sorting of spermatozoa according to the DNA content by means of flow cytometry (Rath and Johnson, 2008). Embryo sexing has been attempted by a variety of methods, including cytogenetic analysis, assays for X-linked enzyme activity, analysis of differential development rates, detection of male-specific antigens, and the use of Y-chromosome specific DNA sequences.

ANIMAL CLONING
The process of producing organisms that are genetically identical.

5.1.4 Animal cloning

A**nimal cloning** is defined as the process of producing organisms that are genetically identical. The cloning of animals can be achieved by two strategies: embryo splitting

and somatic cell nuclear transfer (SCNT) (somatic cell cloning). Both techniques offer the possibility for creating clone families from selected superior genotypes and to produce commercial clone lines (Vajta and Gjerris, 2006).

Somatic cell cloning is based on the procedure of removing the DNA from an unfertilized oocyte and replacing it with the DNA obtained from a somatic cell. The somatic cell DNA can be obtained from any individual, preferably an individual with desirable traits. Once introduced to the oocyte, the somatic cell's DNA is reprogrammed by the oocyte and the unfertilized oocyte can develop as an embryo. The resulting animal will be genetically identical as the somatic cell donor. In theory it is possible to obtain practically unlimited numbers of somatic donor cells from an individual, which allows cloning technology to be applied for the production of many genetically identical individuals. In addition, this technique offers another advantage: the somatic cells genome can be subjected to genetic manipulation prior to the introduction into the oocyte, resulting in a transgenic organism (see 5.3.3).

Embryo splitting, the second cloning technology, is the process of dividing a developing embryo, typically at the 8-cell stage, into two equal parts that continue to develop. The procedure can be repeated several times, but usually only four viable embryos can be obtained from a founder embryo. The technology has no significant importance in research and development nowadays.

5.2 GENETIC MARKERS AND MARKER-ASSISTED SELECTION (MAS)

A genetic marker is defined as a DNA sequence that is associated with a particular trait, in terms of spatial proximity of sequence, and thus segregates in an almost identical and predictable pattern as the trait. This marker can include the gene (or a part thereof) which is responsible for the trait, or DNA sequences that are sufficiently close to the gene(s) so that co-segregation is ensured.

GENETIC MARKER
A gene, or a sequence that is closely associated with it, that is responsible for a particular trait and that can be used to follow the inheritance of that trait within a population.

Genetic markers facilitate the "tagging" of individual genes or small chromosome segments containing genes which influence the trait of interest. Availability of large numbers of such markers has raised the likelihood of detection of major genes influencing quantitative traits. The process of selection for a particular trait using genetic markers is called marker assisted selection (MAS). MAS can accelerate the rate of breeding progress by increasing the accuracy of selection and by reducing the generation interval. Marker identification and use should enhance future prospects for breeding for such traits as tolerance or resistance to environmental stresses, including diseases (Dekkers, 2004; FAO, 2007).

Two types of marker can be considered. First, markers that are sufficiently close to the trait gene on the chromosome so that, in most cases, alleles of the marker and the trait gene are inherited together. This type of marker is called a linked marker. At the population level, alleles at linked markers cannot be used to predict the phenotype until the association between alleles at the marker and alleles at the trait gene is known (called "phase"). To determine phase, inheritance of the marker and trait gene has to be studied in a family. However, information on phase is only valid within that family and may change in subsequent generations through recombination (Ron and Weller, 2007).

The second type of marker is a functional trait. These markers are called "direct" markers. Once the functional polymorphism is known it is possible to predict the effect of particular alleles in all animals in a population, without first having to determine the phase. Therefore, "direct" markers are more useful than "linked" markers for predicting the phenotypic variation of target traits within a population. A further complication is that the mechanisms of genetic control differ between traits. The variation seen in some traits is directly controlled by a single gene (monogenic traits), which may have a limited number of alleles. In the simplest situation a gene will have two alleles: one allele will be associated with one phenotype and another allele with a different phenotype. An example is black

versus brown coat colour in cattle: the brown coat colour occurs as a result of a mutation in the melanocyte hormone receptor gene, which results in the creation of a different allele with a different function.

However, the traits that are important in livestock production are generally more complex and have a very large range of variation in the observed phenotype, caused by the interaction of multiple genes (polygenic traits). Growth rate and milk yield are examples of two traits that exhibit a continuous phenotypic variation. Such traits are called quantitative traits. The variation in quantitative traits is controlled by several genetic loci (called quantitative trait loci [QTL]), each of which is responsible for a small amount of the overall variation (Rocha *et al.*, 2002). The behaviour of genes (including major genes) that control a trait is likely to be dependent on the genetic background.

The myostatin allele responsible for double muscling in Belgian Blue cattle is also found in other breeds; however, the phenotype associated with the allele is variable between the breeds. This suggests that there are genes at other loci in the genome that act to modify the phenotypic expression of the major gene. Thus, information is required not only on the major genes that control a trait, but also on the interactions between genes. It is therefore premature to start using DNA-based selection widely, without further knowledge of gene interaction networks. However, some DNA tests for specific polymorphisms are being offered commercially, e.g. the GeneSTAR test for tenderness (based on variations in the calpastation gene, Pfizer Animal Genetics) and marbling (based on variations in the thyfoglobulin gene), and the Igenity test for fat deposition (based on variations in the leptin gene, Merial). These tests can be used by breeders and evaluated in their populations.

MAS and gene mapping are also considered as important tools to investigate, maintain and conserve the genetic diversity and the genetic resources of agricultural species. During the last decades an increasing portion of breeds became extinct, mainly local breeds that are not used in a sustainable manner and are not covered by breeding programmes. However, these local breeds are of high importance since they are adapted to local conditions, contribute to local food security and represent a unique source of genes that can be used for the improvement of industrial breeds. Molecular marker techniques can play an important role in the characterization and protection of agricultural genetic resources (FAO, 2006).

5.3 TRANSGENIC ANIMALS

TRANSGENIC ANIMAL
An animal, into which foreign DNA has been introduced.

A transgenic animal is an animal that carries a specific and deliberate modification of its genome – analogous to a transgenic plant. To establish a **transgenic animal**, foreign DNA constructs need to be introduced into the animal's genome, using recombinant DNA technology, so that the construct is stably maintained, expressed and passed on to subsequent generations. The last point, heritability of the genetic modification, can be achieved by creating an animal that carries the modification in the genome of its germ line: all offspring derived from this animal will be completely transgenic, as they will carry this modification in all their somatic and germ line cells.

Transgenic animals can be created for a variety of different purposes: to gain knowledge of gene function and further decipher the genetic code, study gene control in complex organisms, build genetic disease models, improve animal production traits, and produce new animal products (Melo *et al.*, 2007). This chapter will focus on the last two points, which are most important with respect to agricultural applications.

In 1982, the first transgenic animal was produced: a mouse, obtained by micro-injection of a DNA construct into a fertilized, single-cell stage oocyte (Palmiter *et*

al., 1982). The transgene construct used was composed of the rat growth hormone gene, fused to the mouse metallothionein-I promoter. The study was published in *Nature* magazine, and the impressive outcome of the study was chosen as the cover photo: the produced transgenic mice were unnaturally large, approximately twice the size as non-transgenic control mice. The impact of this study on both the scientific and public community was huge, and raised speculations about the potential applications of this technology for animals of agricultural importance. Since the insertion of a single growth hormone gene was sufficient to have tremendous effects on mice, it was anticipated that this procedure would also be applicable for agricultural animals, resulting in highly increased growth rate, feed efficiency and reduced fat deposition. Many other possible applications were also subject of speculation, such as a manipulation of milk production or production of milk with novel ingredients, increased wool production or increased resistance of farm animals to diseases and parasites.

By 1985, transgenic pigs and sheep had been obtained, with cattle and chicken following somewhat later (Melo *et al.*, 2007). Since that time the development of transgenic animals and the exploration of agricultural applications has been a steady process, although at a slower rate than what was initially expected. Engineering a specific trait proved to be much more difficult than simply introducing the responsible gene, and technical limitations, the high costs of the process and insufficient knowledge about gene function and regulation of gene expression severely restricted progress. This is particularly true for agricultural species such as cattle, which proved to be much more complicated than mice.

Nevertheless, research in the field continues and several agricultural applications are envisaged, and the approval and market release of the first transgenic animals is expected to take place in the next few years. The knowledge of gene expression and regulation is constantly extending, facilitating genetic engineering in complex animals, such as mammals. Likewise, the repertoire of available techniques to

manipulate DNA and animals is constantly increasing. Therefore, it is likely that genetic engineering techniques applied to animals of agricultural importance will play an increasingly important role in the years to come. The techniques that are currently applied to produce transgenic animals are listed below.

5.3.1 Micro-injection

MICRO-INJECTION
The first approach developed for the production of transgenic animals, based on the injection of the transgene into fertilized oocytes.

Micro-injection, the first successful approach for the creation of transgenic animals, has already been described in the preceding paragraphs. Briefly, it is based on the injection of a foreign DNA construct into a fertilized oocyte (Figure 5.1). The construct integrates randomly into the host oocyte genome, subsequently the zygote continues embryonic development, the embryo is transferred to a foster mother and eventually develops to a transgenic animal. However, this method has strong limitations: on average, less than 1 percent of embryos injected and 10 percent of animals born are transgenic, genes can only be added, not replaced or deleted, and multiple copies of the transgene are inserted at random, hindering the correct regulation of gene expression and possibly interfering with endogenous gene function (Robl *et al.*, 2007). This requires large amounts of oocytes to be injected, as the overall efficiency of the process is very low and basically a trial-and-error process, whose outcome can only be influenced to a small extent.

5.3.2 ES cell based cloning and transgenesis

EMBRYONIC STEM CELLS
Cells, derived from an early embryo, that possess the capability to differentiate into any of the cells of the adult animal (pluripotency).

To overcome the problems associated with micro-injection techniques, **embryonic stem cell** (ES cell) technology has been developed (Denning and Priddle, 2003). Embryonic stem cells, as the name suggests, are derived from embryos at a very early stage (the blastula), and possess the important characteristic of pluripotency. Pluripotency is the ability of these cells to differentiate to any of the cell types and tissues found in the adult organism. ES cells can be grown in culture for many passages and can be subjected to transformation with transgene constructs, resulting in modifications

of their genome. The constructs used not only permit the selection of successfully transformed cells, but also allow **gene targeting** to be accomplished (see 5.3.3) Thus, genes can be specifically introduced, replaced or deleted (so-called knock-ins and knock-outs). Transformed ES cells are re-introduced into the blastocoel cavity of an embryo, where they integrate and produce a mosaic (chimaeric) animal, i.e. an animal that is made up of transformed and non-transformed cells. Possibly, the chimaeric animal carries the transgene in the germ line; in this case, it is possible to obtain completely homozygous transgenic animals through selective breeding (Figure 5.1).

GENE TARGETING
The possibility to use gene constructs that integrate or replace DNA sequences at specific, determined sites in the host genome. Usually performed with ES cells.

This technique, mainly through the feature of gene targeting, allows a broad variety of genetic modifications to be introduced. For many years, several laboratories worldwide have tried to produce ES cells from farm animals, and although some success has been claimed, no robust and reproducible method has been published. Indeed, even in mice the production of ES cells is a costly and labour-intensive technology (Melo *et al.*, 2007).

5.3.3 Somatic cell nuclear transfer (SCNT)

The method of choice nowadays for the production of transgenic animals is **somatic cell nuclear transfer (SCNT)**. This method, also known as somatic cell cloning, initially gained importance for the possibility to clone animals in theoretically unlimited numbers (see 5.2.4). However, it can also be adapted to produce transgenic animals, with the additional benefit of targeted genetic manipulation (Heyman, 2005).

SOMATIC CELL NUCLEAR TRANSFER
A certain cloning procedure, based on the removal of a nucleus from a body cell of an animal and its introduction to a enucleated zygote, which subsequently develops into an animal with the identic genome as the donor animal.

The insertion of a transgene construct into a specific, pre-determined DNA site of the host genome is called gene targeting. The process and the construction of the transgene is more complex than random gene insertion, as is the case during micro-injection. Nevertheless, gene targeting is a powerful and widely used technique due to the ability to insert the transgene into a specific site (knock-in), inactivate

specific genes (knock-out) or replace the endogenous version of a gene with a modified version. This helps to overcome many of the problems and limitations that are associated with random transgene insertion.

The usual procedure is to produce a series of transgenic founder animals of both sexes, which are subjected to breeding, with the aim of producing homozygous offspring. The entire method is based on the following protocol: oocytes from a donor animal are enucleated, i.e. their nucleus containing the genome is removed. Subsequently a donor nucleus is injected into the enucleated oocyte, and the cells are fused by electrofusion. Following fusion, the oocyte is activated by chemical or mechanical means to initiate embryonic development, and the resulting embryo is transferred to a foster mother (Hodges and Stice, 2003). The donor nuclei can be derived from either somatic cells or ES cells that have been subjected to targeted genetic manipulation prior to injection into the oocyst. Thus, a large number of identical animals with targeted genetic modifications can be obtained (Figure 5.1).

Homozygous transgenic animals may also be obtained by a slightly modified approach: by targeting a transgene to one member of a pair of chromosomes, and subsequently target the same site on the other chromosome with the same transgene (Robl *et al.*, 2007).

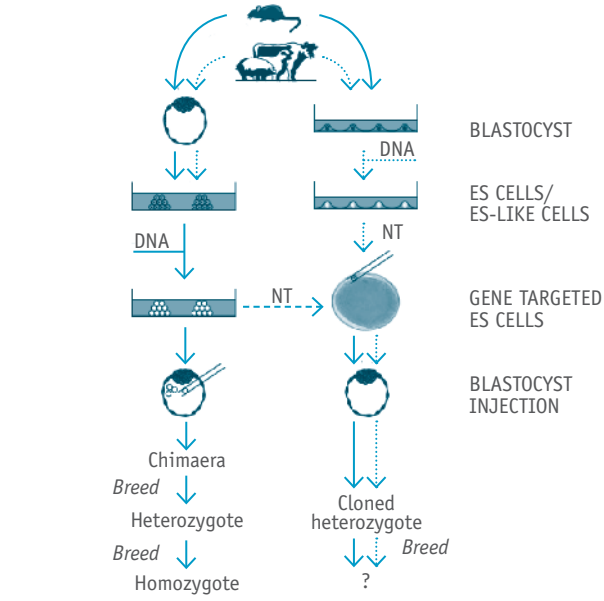
Another approach based on the techniques outlined above makes use of a rejuvenation system for bovine fibroblast (connective tissue) cells (Kuroiwa *et al.*, 2004). A bovine fibroblast cell line is derived from a bovine foetus, and subjected to genetic manipulation. Such primary cell lines grow for only a limited number of cell divisions in culture, allowing only a limited number of genetic manipulations to be introduced (usually only one) before the cells stop dividing and eventually die (Robl *et al.*, 2007). After the genetic manipulation, the cells are used in a cloning procedure to obtain cloned foetuses. These foetuses can be subjected to

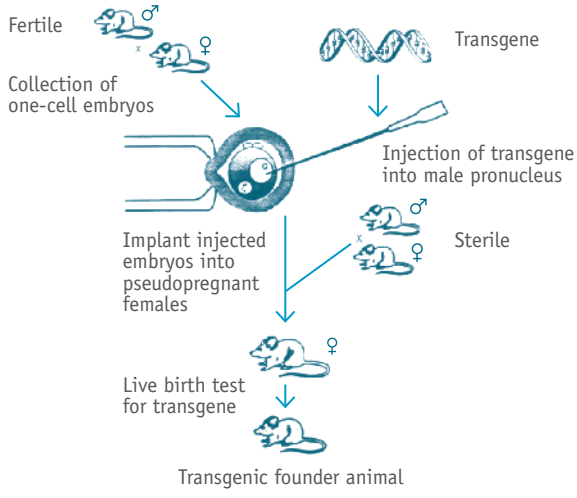
Figure 5.1 | **Comparison of micro-injection, ES cell techniques and somatic cell nuclear transfer (SCNT) for the creation of transgenic animals**

(A) ES cells are obtained from an early embryo, can be subjected to genetic modification in culture and are subsequently re-injected into an embryo or are used for nuclear transfer (NT) to an enucleated oocyte. Transgenic animals can be obtained by breeding in both cases.

(B) In SCNT, the nucleus from a somatic donor cell is removed and injected into an enucleated oocyte, resulting in a cloned animal.

(C) During micro-injection, the DNA construct is injected into a fertilized oocyte, resulting in random DNA integration and the production of a transgenic founder animal.





a second round of fibroblast isolation, manipulation and cloning. Once the genetic manipulations are completed, the final cell line can be used in a cloning procedure to produce transgenic offspring.

5.3.4 Artificial chromosome transfer

ARTIFICIAL CHROMOSOMES Chromosomes, designed by recombinant DNA technology, that can be used to transfer large pieces of foreign DNA to target organisms, where they are stably maintained and expressed.

Artificial chromosomes are a relatively recent development in animal transgenics (Robl *et al.*, 2003). One outstanding characteristic is their ability to carry very large fragments of DNA, up to several Mb (compared with 5-30 kb on a typical plasmid vector). **Artificial chromosomes** possess a centromere, telomeres and origins of replication, sequences that are responsible for their stable maintenance within the cell as autonomous, self-replicating chromosomes. This eliminates the need for integration into the host genome. Due to these properties, artificial chromosomes can be used to transfer either very large, complex genes or many small genes and regulatory elements to a target animal. The actual process of chromosome transfer and subsequent cloning of animals is similar to the SCNT approach.

The feasibility of this technique has been proven by the transfer of a human artificial chromosome, encoding the human antibody genes of 10 Mb in size to cattle (Kuroiwa *et al.*, 2002). The transferred chromosome was stable in the adult transgenic animals, and the encoded antibody genes were expressed to a certain extent.

5.3.5 Sperm-mediated DNA transfer

Several reports describe the use of sperm as a vector to deliver transgene DNA to the oocyte during the process of fertilization. The efficiency of this process varies considerably between species, and several approaches are under investigation to improve uptake and incorporation of foreign DNA. These include, among others, intracytoplasmic injection of DNA-coated sperm into the oocyte, or liposome

treatment of sperm to facilitate DNA uptake (Robl *et al.*, 2007). Nevertheless, the entire process is not completely understood and far from being used routinely.

5.3.6 Viral-vector mediated DNA transfer

Transgenesis may also be accomplished by employing virus-derived vectors, namely vectors based on the retrovirus-class of lentiviruses (Whitelaw *et al.*, 2008). Genes that are essential for viral replication are deleted from the viral genome, maintaining only the capacity for integration of the viral genome into the host genome. Parts of the vector that were occupied by viral genes can then be replaced by the transgene of interest – an approach analogous to the modification of the *A. tumefaciens* Ti-plasmid. Viruses carrying the modified vector are then produced *in vitro* and subsequently injected into the perivitelline space of the zygote (or an unfertilized oocyte), resulting in infection of the zygote and integration of the viral genome into the host genome. Transgenesis rates reaching up to 100 percent of injected embryos have been described (Park, 2007).

Major drawbacks of this method are a limited transgene size and random transgene integration. The maximal transgene size is 8 kb, which is rather low compared with other techniques. Random and possibly multiple transgene integration may lead to position effects, disturbance of the host genome and dose effects, as is the case with pronuclear injection. Solving these problems holds great promise for the further development and application of lentiviral vectors.

Other, less frequently used, methods include biolistics, liposome-mediated DNA transfer to cells and embryos, or DNA transfer to cells and embryos by electroporation. As mentioned in the introduction, constant progress is being made in the field of animal transgenesis, although no approval for commercial release has been obtained so far. Some of the envisaged applications of GM animals are given in the following paragraph.

5.4 APPLICATIONS FOR TRANSGENIC ANIMALS

Since the production of the first transgenic mice almost three decades ago much work has been performed on the development of technologies for efficient transgenesis. Many initial problems, such as low efficiencies, random transgene insertions and unexpected and undesirable behaviour of transgenic animals, have been overcome or are at least understood in more detail. Furthermore, an ever-increasing knowledge of genes, gene function and regulation of gene expression facilitates the planning and creation of transgenic animals with desired traits.

The main interest of modern agricultural research with regard to transgenic animals can be divided into two broad categories: production of animals with improved intrinsic traits, such as higher growth rates, improved milk production, disease resistance etc. The other is the production of animals that produce novel products, such as pharmaceuticals, proteins of medical relevance, vaccines etc. (Wheeler *et al.*, 2003; Niemann *et al.*, 2005). Examples of both categories will be given in the following sections.

5.4.1 Transgenic animals for food production

Engineering transgenic animals with an application in food production focuses mainly on improved meat production, improved carcass quality and enhanced milk production.

Milk is a complex biological fluid and has a high importance for contributing to the nutrition of many societies (Melo *et al.*, 2007). The major goals for transgenic animal development concerning milk production are increased milk production, higher nutrient content or milk containing novel substances.

Most milk proteins (circa 80 percent) belong to the caseins, and transgenic cattle were created that contain extra copies of casein genes. This resulted in elevated casein protein levels in milk (Brophy *et al.* 2003). Another milk application that is being investigated is the production of milk with no lactose (milk sugar) present,

since approximately 70 percent of the world population cannot metabolize lactose and thus cannot consume dairy products. Engineering milk with novel properties, e.g. milk containing the immune-stimulating human protein lactoferrin, is a further approach. Many other additives, e.g. different growth hormones or substances that stimulate health and development, have been proposed for overexpression in milk and thus possibly contribute to growth and health of developing offspring. In pigs, the transfer of the bovine a-lactalbumin gene led to increased milk production, resulting in faster piglet growth and survival rate.

One of the first reports with relevance for enhanced meat production was the article about the first transgenic mice, expressing rat growth hormone and showing increased body size and mass. However, transferring this approach to pigs initially did not yield promising results. Nevertheless, pigs showing increased muscle weight gain and feed efficiency by introducing porcine growth hormone or human insulin-like growth factor have been created (Niemann *et al.*, 2005). Furthermore, pigs expressing the enzyme phytase in their salivary glands have been created: these animals can metabolise the phosphor present as phytic acid in corn and soy products, thus needing less phosphor as feed additives and releasing less phosphor with their manure, reducing the environmental impact of pig farming (Haefner *et al.*, 2005).

Experiments in cattle are focusing on the myostatin gene, a negative regulator of muscle mass, resulting in a high increase in muscle mass in animals with a myostatin mutation or deletion.

Transgenesis is also employed for fish; injection of embryos with constructs containing either the bovine or Chinook salmon growth hormone has been reported, with the aim of improving fish growth in general and especially under adverse conditions, e.g. low water temperatures. This has resulted in an up to 5-11 fold increase in weight after one year of growth for transgenic salmon and 30-40 percent increased growth of transgenic catfish (Wheeler, 2007).

All these studies demonstrate the fundamental feasibility of applying transgenesis to agricultural animals for improved food production, but so far no transgenic food producing animal has been released for commercial use. In addition to the research and development necessary for the establishment of a transgenic animal, there are several other factors that strongly influence the use of transgenic animals for food production. Among these are considerations concerning the economic practicability, social acceptance of transgenic food and, possibly most important, regulations concerning the approval of GMOs and derived products.

Regulatory authorities need to consider three factors:

- safety of the food product for human consumption;
- environmental impact of the genetically modified animals;
- welfare of the animals.

These factors need to be considered on a case-to-case approach for every new transgenic animal or product that has been obtained using GMOs. In principle, this safety investigation is identical to the safety regulations and procedures that apply for transgenic plants. A detailed description of the safety evaluation procedures and the underlying regulatory documents and treaties is provided in the accompanying modules of this compendium.

5.4.2 Transgenic animals for production of human therapeutics

One major application of animal transgenesis nowadays is the production of pharmaceutical products, also known as animal pharming. The costs for producing transgenic animals are high, but since the pharmaceutical industry is a billion-dollar market the input is likely to be a feasible and economically worthwhile investment (Sullivan *et al.*, 2008). Since many human proteins cannot be produced in micro-organisms and production in cell culture is often labour-intensive with low yields, the

production of **biopharmaceuticals** in transgenic animal bioreactors is an attractive alternative (Kind and Schnieke, 2008). Furthermore, many human proteins cannot be produced in micro-organisms, since they lack post-translational modification mechanisms that are essential for the correct function of many human proteins.

BIOPHARMACEUTICAL Pharmaceuticals, produced in transgenic organisms (bioreactors) or in cell culture.

Pharmaceutical proteins or other compounds can be produced in a variety of body fluids, including milk, urine, blood, saliva, chicken egg white and seminal fluid, depending on the use of tissue-specific promoters (Houdebine, 2009). Nevertheless, milk is the preferred medium due to its large production volume. Furthermore, it has been shown that the mammary glands can produce up to 2 g of recombinant protein per litre of milk; assuming average protein expression and purification levels, only relatively small herds of transgenic animals would be required to supply the world market with a specific recombinant protein (e.g. 100 transgenic goats for the production of 100 kg monoclonal antibodies required per year [Melo *et al.*, 2007]). In Table 5.1, biomolecules expressed in mammary glands and their anticipated applications are listed:

Table 5.1 | **Pharmaceuticals produced by transgenic animals**

Pharmaceutical	Bioreactorspecies	Application/treatment	Company
Antithrombin III	goat	thrombosis, pulmonary embolism	GTC Biotherapeutics (USA)
tPA	goat	thrombosis	PPL Therapeutics (UK)
α-antitrypsin	sheep	emphysema and cirrhosis	PPL Therapeutics (UK)
Factor IX	sheep	hemophilia b	PPL Therapeutics (UK)
Factor VIII	sheep	hemophilia a	PPL Therapeutics (UK)
Polyclonal antibodies	cattle	vaccines	Hematech (USA)
Lactoferrin	cattle	bactericide	Pharming Group (NED)
C1 inhibitor	rabbit	hereditary angioedema	Pharming Group (NED)
Calcitonin	rabbit	osteoporosis and hypercalcemia	PPL Therapeutics (UK)

Adapted from: Melo *et al.*, 2007.

Another advantage of biopharmaceutical production in transgenic animals is the reduced risk of transmitting diseases, compared with human-derived material. Several cases are known where hundreds of patients were infected with HIV, Hepatitis C or Creutzfeld-Jakob-Disease following treatment with human-derived pharmaceuticals. Of course, animal-derived material needs to be subjected to a thorough purification procedure to exclude transmission of animal diseases (zoonoses) or contamination with animal DNA or protein that might induce an immune reaction.

Nevertheless, the development of transgenic animals that secrete high contents of the desired product in their milk, and the subsequent development of an effective and high-yield purification protocol to get rid of contaminating proteins, requires a lot of knowledge and financial and intellectual input. So far, only GTC Biotherapeutics Antithrombin III has been approved for the United States market and is sold under the name of ATryn (FDA, 2009). Furthermore, many potential target proteins as well as the technologies to develop a transgenic animal are covered by patents and intellectual property rights, thus only a small number of proteins are being investigated by a small number of pharmaceutical companies at the moment (Kind and Schnieke, 2007).

ANTIBODIES
Proteins, produced by an organism in response to a pathogen or foreign substance, that neutralize that substance/ pathogen and help the immune system to eliminate the infection.

A particularly promising approach is the development of transgenic animals that express human polyclonal antibodies. **Antibodies** are the fastest growing set of new biopharmaceuticals, for therapeutic use in cancer, autoimmune diseases, infections, transplantations, biodefence and immune deficiencies. Currently all approved therapeutic antibodies are produced by cell culture techniques.

The possibilities for the production of polyclonal human antibodies in transgenic cattle are currently being investigated; such antibodies would mimic the natural human immune response to a pathogen. Cattle would be especially suited for this purpose, since the total amount of antibodies in an adult animal is approximately 1 kg. One approach towards this end is the use of artificial chromosomes to

transfer the human antibody genes to the target animal (Kuroiwa *et al.*, 2002). Concomitantly, the endogenous antibody genes of the animal are knocked out to prevent their expression and thus allow purification of human antibodies without contaminating bovine antibodies. To obtain human polyclonal antibody sera from the animal, the animal would need to be immunized with a vaccine containing the pathogen of interest, e.g. a bacterium or a virus. Subsequently, the animal would build up an immune response and express the human antibodies directed against that pathogen. These antibodies could subsequently be extracted and purified from the animal's blood plasma and used to treat humans suffering from an infection with that particular pathogen. This perspective for a quick availability of large amounts of human antibody sera targeted against a certain pathogen or disease agent has raised speculations about a transformation of medicine similar to the introduction of antibiotics in the 1940s and 50s (Kind and Schnieke, 2007). Similar approaches, based on the same methodology, are being pursued for the use of plants as bioreactors for the production of medically valuable proteins and small-molecule drugs (Twyman *et al.*, 2005).

5.4.3 Transgenic animals for improved disease resistance

Resistance or susceptibility to diseases and the immune response typically depend on a variety of genes, but identification of some key genes has brought up the possibility of gene transfer to target important and specific aspects of the immune system (Niemann *et al.*, 2005). Diseases that are under investigation, by either introducing resistance genes or removing susceptibility genes, include bovine spongiform encephalopathy (BSE), brucellosis, other viral or bacterial infections, parasitic organisms, and intrinsic genetic disorders.

One often-cited example is resistance against mastitis: mastitis is a bacterial infection of the bovine mammary gland, leading to decreased productivity and milk contamination. Transgenic cattle have been produced that secrete the small

protein lyostaphin in their milk, which is a potent inhibitor of *Staphylococcus aureus (S. aureus),* the bacterium responsible for the majority of mastitis cases. According to first trials, the transgenic cows are resistant to *S. aureus* – mediated mastitis (Donovan *et al.*, 2005).

Further approaches of animal transgenics target animal reproductive performance and prolificacy, development of organs for transplantations (xenotransplantation) that do not evoke a rejection response, or improvement of animal fibre and wool.

5.5 BIOTECHNOLOGY IN ANIMAL HEALTH

Apart from the aforementioned possibilities to generate transgenic animals with enhanced resistances to diseases, biotechnology offers a variety of other techniques that contribute to improved animal health. These include the production of vaccines to immunize animals against diseases, and the development of improved disease diagnostic tools.

5.5.1 Vaccines

VACCINES
A vaccine is a substance, derived from a pathogen, that is administered to an organism and stimulates the organism's immune system to prevent infection from that pathogen.

Vaccines are substances, derived from a pathogen, that are used to stimulate an animal's immune system to produce the antibodies needed to prevent infection from that particular pathogen. Vaccination is therefore the main approach to protect animals from infectious diseases. The majority of vaccines are based on material directly derived from inactivated bacteria or viruses, which potentially revert to their virulent (disease-causing) form. Modern biotechnology offers possibilities to engineer specific vaccines that are free from pathogen-derived material and are more effective and safe in stimulating the immune response (Rogan and Babiuk, 2005).

One approach is based on recombinant protein technology: once a protein from a pathogen that serves as antigen (i.e. a molecule that stimulates an immune response) has been identified, this protein can be safely expressed in cell culture,

e.g. in *E. coli* or mammalian cells, using recombinant DNA technology. Subsequently, this protein can be harvested, purified and used as a vaccine (also known as subunit vaccines). In addition, it has also become possible to create fusions of several pathogen proteins, so that one final protein stimulates a variety of immune responses (Meeusen *et al.*, 2007).

A second approach consists of using DNA-based vaccines. This methodology is based on the delivery of plasmid DNA to the cells of a host animal that encodes pathogenic proteins. Once expressed within the cell, the proteins stimulate the animal's immune response in the same way as if the proteins were delivered from outside; thus the animal serves as its own bioreactor for vaccine production (Rogan and Babiuk, 2005). The efficiency of this method is largely dependent on effective plasmid delivery to the animal cells; methods for delivery include chemical transformation, electroporation, injection and the gene gun.

A third approach is the delivery of pathogen-derived antigens by live recombinant vectors. Bacteria, viruses or even parasites can be engineered to express foreign proteins from the pathogen of interest that act as antigens. The engineered organism is then delivered to the animal, where it induces a limited infection and presents the foreign pathogenic protein, thus stimulating an immune response against that pathogen.

Recently, a very interesting combination of transgenic plant technology and animal vaccination has emerged: plants are engineered to express an antigenic protein from a pathogen at high levels in their tissues or storage organs. Subsequently these plants can be fed to animals and the vaccine is presented to and taken up by the mucosal surfaces in the intestine, thus providing a direct feed-vaccination (Floss *et al.*, 2007).

In addition to the vaccine itself, substances that stimulate vaccine uptake and activity (so-called adjuvants) and the route of vaccine delivery (injection, inhalation, feed, etc.) are factors that are strongly investigated and further developed by biotechnological methods.

5.5.2 Diagnosis of disease and genetic defects

Successful control of a disease requires accurate diagnosis. Modern biotechnology offers many applications to diagnose diseases caused by pathogens as well as diseases caused by intrinsic genetic disorders of an organism. The currently available and deployed techniques are outlined below.

The ability to generate highly specific antigens by recombinant DNA techniques has significantly raised the number of ELISAs that have the capacity to differentiate between immune responses generated by vaccination from those due to infection. This has made it possible to overcome one of the major drawbacks of antibody detection tests: the fact that, because antibodies can persist in animals for long periods, their presence may not indicate a current infection (Rege, 1996).

The advent of PCR has enhanced the sensitivity of DNA detection tests considerably. For example, PCR used in combination with DNA hybridization analysis has been shown to provide a sensitive diagnostic assay to detect bovine leukosis virus. This holds true for many other pathogenic organisms that are difficult to detect by serological methods (Schmitt and Henderson, 2005).

Other diagnostic techniques include nucleic acid hybridization assays and restriction endonuclease mapping. A good example of the specificity of nucleic acid hybridization is its application in distinguishing infections caused by *peste des petits ruminants* (PPR) virus from *rinderpest*, diseases whose symptoms are clinically identical and which cannot be distinguished with available serological reagents. This technique also allows comparison of virus isolates from different geographical locations.

Molecular epidemiology is a fast growing discipline that enables characterization of pathogen isolates (virus, bacteria, parasites) by nucleotide sequencing, allowing the tracing of their origin. This is particularly important for epidemic diseases, where

the possibility of pinpointing the source of infection can significantly contribute to improved disease control. Furthermore, the development of genetic probes, which allow the detection of pathogen DNA/RNA (rather than host antibodies) in livestock, and the advances in accurate, pen-side diagnostic kits can considerably enhance animal health programmes (FAO, 2001).

DNA testing is also being used to diagnose hereditary weaknesses of livestock. One available test identifies the gene which is responsible for Porcine Stress Syndrome in pigs. Animals that carry this gene tend to produce pale, low-quality meat when subjected to the stress of transport or slaughter. The identification of pigs that carry this gene excludes them from breeding programmes, resulting in an overall decrease in the frequency of that gene within a population (Madan, 2005).

DNA TESTING
Testing an animal for the occurrence of specific gene versions, and thus preferably use or exclude that animal from breeding programmes.

Another example of DNA analysis is the diagnosis of a mutation of Holstein cattle that causes leucoyte adhesion deficiency. Cattle with this condition suffer diseases of the gum, tooth loss and stunted growth. The disease is fatal, and animals usually die before reaching one year of age. The available test identifies carriers of the defective gene, allowing the elimination of such animals from breeding herds. Ideally, all animals used for breeding should be tested to exclude any carriers of the gene (Madan, 2005).

5.6 DNA TECHNOLOGIES IN ANIMAL NUTRITION AND GROWTH

5.6.1 Nutritional physiology

Applications are being developed for improving the performance of animals through better nutrition. Specific enzymes can chemically modify feedstuffs and thus improve the nutrient availability and uptake by the animal. This lowers feed costs and reduces output of waste into the environment. Prebiotics (substances that stimulate microbial growth) and probiotics (live micro-organisms) as feed

additives or immune supplements can either stimulate growth of beneficial micro-organisms in the digestive system, or inhibit pathogenic gut micro-organisms and render the animal more resistant to them. Administration of the recombinantly produced growth hormone somatotropin (ST) results in accelerated growth and leaner carcasses in meat animals and increased milk production in dairy cows. Immunomodulation, i.e. administration of substances that stimulate or repress immune system function, can be used for enhancing the activity of endogenous anabolic hormones (FAO, 2001).

In poultry nutrition, possibilities for improvement include the use of feed enzymes, probiotics and antibiotic feed additives. The production of tailor-made plant products for use as feeds that are free from anti-nutritional factors through recombinant DNA technology is also a possibility.

Plant biotechnology may produce forages with improved nutritional value or incorporate vaccines or antibodies into feeds that may protect the animals against diseases (see 5.5.1).

5.6.2 Rumen biology

Rumen biology has the potential to improve the nutritive value of ruminant feedstuffs that are fibrous, low in nitrogen and of limited value for other animal species. Biotechnology can alter the amount and availability of carbohydrate and protein in plants as well as the rate and extent of fermentation and metabolism of these nutrients in the rumen (FAO, 2001).

Methods for improving rumen digestion in ruminants include the use of probiotics, which is the supplementation of animal feed with beneficial live micro-organisms, to improve the intestinal microbial balance for better utilization of feed and for good health (Weimer, 1998). The added bacteria may improve digestion of feed

and absorption of nutrients, stimulate immunity to diseases, or inhibit growth of harmful micro-organisms. Transgenic rumen micro-organisms (see Chapter 6) could also play a role in the detoxification of plant poisons or inactivation of antinutritional factors. Successful introduction of a caprine rumen inoculum into the bovine rumen to detoxify 3-hydroxy 4(IH) pyridine (3,4 DHP), a breakdown product of the non-protein amino acid mimosine found in *Leucaena* forage is an example (Rege, 1996).

To conclude this chapter, it should be noted that many biotechnological applications are already available in the field of animal production and utilization. However, all techniques that have been successfully adopted so far are based on conventional biological methodologies, such as assisted reproduction and MAS. On the contrary, the approval and commercialization of techniques based on the creation of GM animals is only beginning to emerge. This is in sharp contrast with the field of transgenic plants, which have been in commercial use since the mid-1990s. Nevertheless, research in the field of GM animals is actively searching for solutions to the problems that are still linked to the production and application of GM animals. The approval of the first drug that is produced in a transgenic organism is a positive sign in this respect, and many other applications of GM animals, both in agriculture and medicine, are envisaged to follow in the near future.

CHAPTER 6

GENETIC ENGINEERING OF MICRO-ORGANISMS OF INTEREST TO AGRICULTURE

6.1 INTRODUCTION

MICRO-ORGANISM
All organisms that are not visible to the naked eye; including bacteria, archae, fungi, protists, green algae and small animals.

Micro-organism is a term employed to cover all organisms that are not visible to the naked eye; this includes bacteria, archae, fungi, protists, green algae and small animals, such as plankton. The development of genetically modified micro-organisms of interest to agriculture is of significant importance. These micro-organisms may be used as gene transfer systems or donors and recipients of desirable genes. Micro-organisms functioning as gene transfer systems and as donors of genes have already been discussed (see previous chapters). The focus of this chapter is therefore on microbial recipients of transgenes to obtain organisms with novel traits and properties.

Micro-organisms play important roles in different sectors of agriculture, food processing, pharmaceutical industries and environmental management. This development already started early in the history of humankind with the use of micro-organisms for the fermentation process. In the early 1970s, micro-organisms, notably *E. coli*, were used at the forefront of molecular biology research, resulting in the advent of recombinant DNA technology. The first recombinant protein, produced in a micro-organism and approved as a drug by the FDA in 1982, was human insulin. Since then hundreds of recombinant proteins have been engineered and expressed in micro-organisms

and approved for use as pharmaceuticals. Nowadays, many microbial processes and pathways are understood and deciphered at the genetic level and can thus be subjected to specific and targeted genetic manipulation (Bull *et al.*, 2000). Traditionally this approach largely depended on the identification and selection of random mutants with desirable characteristics; recombinant DNA technology presents a significant advance in this respect, since specific metabolic pathways can be manipulated with high precision and completely new functions can be introduced into an organism. The following sections give some examples of micro-organisms of economic importance that have been genetically modified through recombinant DNA technology.

The techniques underlying the production of genetically modified micro-organisms are basically the same that we encountered throughout the previous chapters. However, genetic engineering of micro-organisms is in many respects much easier compared to plants and animals (Demain and Adrio, 2008). Since many micro-organisms are single-cell organisms, they can be easily grown in cell culture in the laboratory in large quantities. Furthermore, DNA can be easily introduced using a variety of techniques, and it is obviously not required to reconstitute a complete transgenic organism from a transformed cell, as is the case with plants and animals. Many micro-organisms can grow under a variety of different conditions and with different nutrient sources and survive periods of unfavourable growth conditions. Furthermore, the genetic makeup and function of many micro-organisms are known in detail and are in general less complicated compared with multicellular organisms, facilitating targeted genetic manipulations.

6.2 GENETICALLY MODIFIED MICRO-ORGANISMS AS BIOPESTICIDES AND BIOFERTILIZERS

BIOPESTICIDES
Substances derived from organic material that exhibit pesticidal activity.

Biopesticides are defined as all substances derived from natural materials, including plants, animals and micro-organisms, that exhibit pesticidal activity. Such biological control agents are increasingly targeted for genetic enhancement due to a rising

recognition of their potential benefits to modern agriculture (Rizvi *et al*., 2009). Biological control represents an alternative to chemical pesticides which have been subjected to much criticism due to their adverse impacts on the environment and human health. Therefore, there is a strong requirement to develop safer and environmentally amenable pest control using existing organisms in their natural habitats. Several such organisms, referred to as biological control agents, are available that offer protection against a wide range of plant pests and pathogenic microbial agents without damaging the ecosystem.

If biological control agents are to be effective in plant disease management, they must be efficacious, reliable and economical (Fravel, 2005). To meet these conditions superior strains are often required that are not found in nature. In this case the existing attributes of the biocontrol agents can be genetically manipulated to enhance their biocontrol activity and expand their impact spectrum.

The foreign genes used for transforming biological control agents can be integrated into the host genome or a plasmid. To express a heterologous gene in fungi or bacteria, the regulatory region of this gene must be modulated in its promoter and terminator regions in order to optimize the expression of the inserted gene in the new host. The addition of specific genes that are known to confer biocontrol activity may enhance or improve biocontrol capacity of organisms that do not naturally possess these genes.

Free-living bacteria associated with plants have been targeted to enhance their capacity either as soil inoculants or as biocontrol agents of plant pathogens. Studies on micro-organisms capable of enhancing plant growth have concentrated on the rhizosphere (root zone) whereas those on biocontrol target both the rhizosphere and phylloplane (leaf zone). Several important rhizobacteria including *Sinorhizobium meliloti* and *Pseudomonas putidrii,* both of which are excellent root colonizers, lack the ability to synthesize chitinases. Chitinases are enzymes that destroy chitin, a

major component of fungi cells (Dahiya *et al.*, 2006). Introducing genes encoding chitinases into their genome have enabled them to provide protection against plant pathogenic fungi. These two bacteria are good targets because of the unique beneficial characteristics they confer. *Sinorhizobium* is a symbiotic bacterium which stimulates formation of root nodules in legumes involved in fixing atmospheric nitrogen. Many *Pseudomonas* species in the rhizosphere environment produce siderophores which chelate iron ions, thereby increasing iron uptake by plants. The genetically modified commercial strain (RMBPC-2) of *Sinorhizobium meliloti* has added genes that regulate the nitrogenase enzyme involved in nitrogen fixation (Scupham *et al.*, 1996).

The *Trichoderma* species are widely present in soils and are antagonistic to other fungi. *T. harzianum,* in particular, is a strong rhizosphere colonizer which is also able to parasitize plant pathogenic fungi. It establishes tight physical contact with hyphae of target fungi with the aid of binding lectins. Several extracellular enzymes, including chitinases, glucanases, lipases and proteases, are produced by the *Trichoderma* species, which has been improved further with the transfer of chitinase genes, notably from *Serratia marcescens* (Benitez *et al.*, 2004).

The *Agrobacterium radiobacter* strain k84 protects plants against crown galls caused by *A. tumefaciens* strains carrying Ti-plasmids of the nopaline type. Protection conferred by *A. radiobacter* strain k84 is due to agrocin 84, an A nucleotide derivative. When taken up by *A. tumefaciens,* it inhibits DNA synthesis, resulting in cell death (Vicedo *et al.*, 1993). *A. radiobacter* has an additional negative effect on soil pathogens by being a very effective rhizosphere colonizer. Although *A. radiobacter* strain k84 has been widely used commercially for a long time, there was concern about its long-term effectiveness as a biocontrol agent. This is because the gene encoding agrocin is carried on a transmissible plasmid, which can be transferred by conjugation to *A. tumefaciens.* In the event of agrocin-encoding plasmid transfer, recipient *A. tumefaciens* strains would no longer be subjected to biocontrol by *A. radiobacter* strain k84. This concern was addressed by modification of the agrocin-encoding plasmid to prevent its transfer

to *A. tumefaciens*. The ensuing genetically engineered strain, known as *A. radiobacter* strain K1026, is a transgenic organism approved for use as a pesticide (EPA).

BACILLUS THURINGIENSIS
One of the best-known and adapted biopesticides, active against a wide range of insect pests.

Bacillus thuringiensis (Bt) has been used as a biopesticide for many years. The insecticidal activity of *B. thuringiensis* is based on the production of crystalline protein inclusions during sporulation. The crystal proteins are encoded by different cry genes and are also known as delta-endotoxins. The protein crystals are highly toxic to a variety of important agricultural insect pests; when the proteins are taken up by susceptible insect larvae they induce lysis of gut cells, resulting in death of the larvae by starvation and sepsis (Roh *et al.*, 2007). The toxin can be applied to plants as a spray consisting of a mixture of spores and protein crystals. However, the toxin has the disadvantage of fast degradation in sunlight. To overcome this limitation, different cry genes encoding the Bt toxin have been cloned and introduced into another bacterium, *Pseudomonas flourescens*. The transgenic *P. flourescens* strains are killed and used as a more stable and persistent biopesticide compared to the *B. thuringiensis* sprays (Herrera *et al.*, 1994). Furthermore, cry genes are widely used to create transgenic plants that directly express the toxin and are thus protected from susceptible insect pests (see Annex 1.2.2).

Baculoviruses (although, per definition, viruses are not micro-organisms) are also being manipulated to be effective biopesticides against insect pests such as corn borer, potato beetle and aphids (Szewczyk *et al.*, 2006).

6.3 MICRO-ORGANISMS FOR ENHANCING THE USE OF ANIMAL FEEDS

Animal digestive tracts harbour beneficial microflora that aid in the digestibility of various feeds. However, the function of these micro-organisms is easily affected by the unfavourable conditions within the gut, such as acidity and antibiotics used to treat pathogenic micro-organisms. Examples of gut micro-organisms that have been genetically modified include *Prevotella ruminicola* with a tetracycline

resistance gene, cellulolytic rumen bacteria with acid tolerance, hind gut bacteria with cellulose activity, rumen bacteria transformed with genes to improve protein yield and yeast (*Saccharomyces cerevisiae)* containing a transgene from the closely related *Saccharomyces diastaticus,* allowing it to increase the digestibility of low-quality roughage in conventional feeds (Weimer, 1998). The major limitation to the use of these engineered organisms has been their establishment in the appropriate regions of the gut. Some organisms are being used as beneficial supplements in animal feeds. These are called **probiotics** and their use aims at improving digestion of feed and absorption of nutrients, stimulate immunity to diseases and inhibit growth of harmful micro-organisms (Gomez-Gil *et al.*, 1998). For the improvement of silage, strains of the bacterium *Lactobillus planetarium* are being developed with the aim of increasing the lactate content and reduce the pH and ammonia content.

PROBIOTICS
Live micro-organisms as feed supplies to improve digestion, nutrient uptake and immune function.

Micro-organisms are being extensively used as bioreactors for the production of hormones and other substances that enhance animal size, productivity and growth rates. The recombinantly produced hormone bST (bovine somatotropin) was among the first recombinant hormones commercially available. It can increase milk yield by as much as 10 to 15 percent when administered to lactating cows (Etherton and Bauman, 1998). Current development efforts are looking at a wide spectrum of genes that affect growth and productivity within the animal and which could be expressed in recombinant micro-organisms to obtain the respective protein in large quantities.

6.4 GENETICALLY MODIFIED MICRO-ORGANISMS IN FOOD PROCESSING

Many micro-organisms are being manipulated with the objective of improving process control, yields and efficiency as well as the quality, safety and consistency of bioprocessed products. Modifications target food enzymes, amino acids, peptides (sweeteners and pharmaceuticals), flavours, organic acids, polysaccharides and vitamins. A classical example is the production of the recombinant cheese making enzyme, chymosin, in

bacteria. Its use was approved in 1990 in the United States, and nowadays 80 percent of US cheese is produced using this product (Law and Mulholland, 1991).

6.5 GENETICALLY MODIFIED MICRO-ORGANISMS IN BIOREMEDIATION

Micro-organisms are widely used in cleaning up pollution such as oil spills or agricultural and industrial wastes by degrading them into less toxic compounds (Chatterjee *et al.*, 2008). Some bacteria are being used as "bioluminescensors" that give luminescence in response to chemical pollutants. An example is the mercury resistance gene *mer* that is expressed in some bacteria and can result in bioluminescence upon encountering the presence of even very low levels of mercury in the environment.

A modified bacterium, *Rhodopseudomonas capsulate,* has the ability to grow rapidly in simple synthetic media. It is being used in advanced swine waste treatment plants in both Japan and Republic of Korea. The concentration of short chain fatty acids, one of the main sources of the bad odour of swine wastes, decreased dramatically after treatment. The residue after treatment can be used as a safe organic fertilizer. Several other applications of micro-organisms or plants for the purpose of **bioremediation** are being investigated.

BIOREMEDIATION
The use of living organisms to detoxify or remove pollutants from the environment.

To conclude this chapter, micro-organisms have always been at the forefront of research and development in the field of recombinant DNA methodology and biotechnology. As mentioned in the introduction to this chapter, this can be largely attributed to the comparative ease of culturing, analysing and manipulating many micro-organisms. Nevertheless, many micro-organisms and their potential benefits remain unexplored and new species are being discovered regularly; therefore, research and development of biotechnological applications for micro-organisms in the field of agriculture and nutrition holds great promise for the future (Bull *et al.*, 2000).

CHAPTER 7

GMO DETECTION, IDENTIFICATION AND QUANTIFICATION METHODS

7.1 INTRODUCTION

The precise and accurate detection of GMOs with high sensitivity in a given biological sample is of significant importance. This need for exact **GMO detection** methods will become increasingly clear in the following modules, when concepts for GMO surveillance, monitoring, biosafety measures and the implementation of relevant regulations are introduced.

GMO DETECTION
The process of detecting GMO material in a given sample.

Different stakeholders involved in the development, use and regulation of GMOs do at some point need to monitor and verify the presence and the amount of GMO material in agricultural products. Furthermore, comprehensive GMO monitoring also includes the analysis of biological samples, such as material derived from plant species that are related to an introduced GMO, to check for horizontal transfer of the transgene. This need has generated a demand for analytical methods capable of detecting, identifying and quantifying either the unique DNA sequences introduced or the protein(s) expressed in transgenic plants and animals. Thus, comprehensive GMO analysis techniques consist of three steps: detection, identification and quantification of GMO material (Anklam *et al.*, 2002).

- **Detection (screening for GMOs).** The objective of this first step is to determine if a product contains GMO material or not. For this purpose, a screening method can be used. The result is a qualitative positive/negative statement. Analytical methods for detection must be sensitive and reliable enough to obtain accurate and precise results and reliably identify small amounts of GMO material within a sample.
- **Identification.** The purpose of the identification step is to reveal how many different GMOs are present in a sample, to precisely identify each single one and determine if they are authorized or not. Specific information (i.e. details on the molecular make-up of the GMOs) has to be available for the identification of GMOs.
- **Quantification.** If a food product has been shown to contain one or more authorized GMOs, it becomes necessary to assess compliance of the set threshold level regulations for the product in question. This is achieved by determining the exact amount of each GMO that has been found in the sample.

GMO TESTING FRAMEWORK
Comprehensive GMO analysis consists of GMO detection, GMO identification and GMO quantifiifcation.

This **testing framework** is depicted in Figure 7.1, with labelling regulation thresholds of the European Union (adapted from Anklam *et al.*, 2002).

In general, the range of sample types that need to be tested for GMO content is extensive and covers raw commodities as well as highly processed food. Furthermore, the number and variety of worldwide commercially grown GMOs is constantly increasing. Therefore, it is necessary to carefully approach each sample on a case-by-case basis and thus determine the most appropriate testing method (Jasbeer *et al.*, 2008)

Every method developed for the detection of GMOs that is considered for routine use by official testing authorities and laboratories has to undergo several testing procedures to verify the analytical performance of the method (Michelini *et al.*, 2008). The performance requirements of each method include applicability (if it is suited for the detection purpose), practicability (costs, material and machine

requirements), specifity, dynamic range (range of different concentrations that can be detected), accuracy, limits of detection and quantification, and robustness (reproducibility of results) (Lipp *et al.*, 2005).

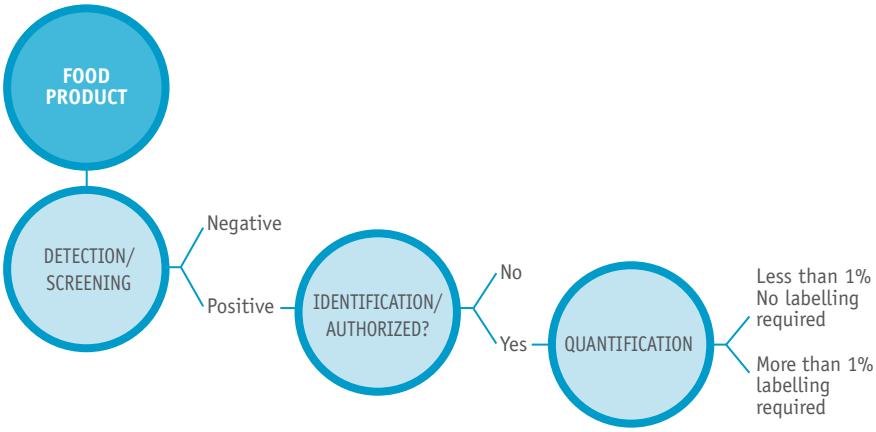
7.2 SAMPLING PROCEDURES

SAMPLING PROCEDURES
Prior to the actual analysis, the sampling procedure has major importance for the outcome and statistical value of a GMO analysis procedure.

Irrespective of the analytical method selected for GMO detection, correct **sampling procedures** are critically important for reliable and reproducible GMO analysis. An insufficient sampling plan can have strong effects on the reliability of the detected GMO level. In fact, the variance associated with the sampling procedure likely represents the major contribution to the overall variance of the detection procedure (Michelini *et al.*, 2008). Furthermore, GMO material usually shows a heterogeneous distribution within the bulk of a product, additionally contributing to sampling-dependent variance. Raw materials, in particular, may show a significant

Figure 7.1 | **GMO detection framework**

A comprehensive testing scheme consists of GMO detection/screening, GMO identification and GMO quantification.



heterogeneity, whereas processed materials and food usually display a more uniform distribution. The influence of the sampling strategy is more relevant when the overall GMO concentration is low.

Samples must therefore be taken in a manner that ensures that they are statistically representative of the larger lot volume or quantum of material. The sample size has to be adjusted to the required sensitivity and allow reliable GMO detection; the smaller the sample, the weaker the statistic significance of the testing procedure. So far, no generally accepted sampling guidelines have been established, and different control authorities employ different sampling schemes (Anklam *et al.*, 2002). The major parameters that influence the sampling plan are lot size, lot heterogeneity, the defined tolerance level and the applied testing methods. Furthermore, parameters that are specific for each event, i.e. the size of the host genome, the copy number of the transgene event involved, and the amount of material that can be analysed in a single test, need to be taken into consideration (Lipp *et al.*, 2005). Efforts are underway to define and internationally harmonize sampling plans, based on sound statistical requirements and analyses (Miraglia *et al.*, 2004).

An example of a sampling plan, based on kernels, is calculated by Grothaus *et al.*, 2006:

- To detect a lot concentration of 0.01 percent GMO material with 99 percent probability, 46 050 particles are required.
- To detect a lot concentration of 0.1 percent GMO material with the same confidence of 99 percent, 4 603 particles are required.
- If the confidence level for the detection of 0.1 percent GMO material is decreased to 95 percent, 2 995 articles are required.

Other calculations based on kernels state that at least 3 500 particles are required to detect a 1 percent contamination with a confidence level of 95 percent (Ovesna *et al.*, 2008). The International Organization for Standardization (ISO) has issued

a brochure on sampling procedures for GMO testing (ISO, 2006); a handbook from the International Seed Testing Agency (ISTA) on this topic is also available (ISTA, 2004). However, it should be noted that in any case the sampling strategy is highly dependent on the material analysed (raw, processed ingredients and processed food) and the required sensitivity, and it should be revised on a case-by-case basis. The establishment of a sampling plan that takes into account all relevant parameters and factors is a complex statistical procedure (refer to Remund *et al.*, 2001 for further information). The reduction of sampling errors and thus more reliable test results are important for all involved parties: for consumers, the probability of consuming food that has been accepted although containing GMO above set **threshold limits** is reduced, and for producers the probability of lot rejection although the GMO content is below the set threshold limit is reduced as well. Therefore, the adoption and implementation of standardized sampling procedures should be of interest to all parties involved in GMO production, trading and consummation (Miraglia *et al.*, 2004).

THRESHOLD LIMITS
Defined limits of GMO content that are allowed in a given product, and that needs to be labeled accordingly.

The actual sampling procedure consists of various steps: (1) sampling the lot of seed, grain or other material to obtain the bulk sample; (2) sampling the bulk sample to obtain the laboratory sample; (3) subsampling the laboratory sample to obtain the test sample; (4) homogenization (grinding etc.) of the test sample and sampling of the resulting meal to obtain the analytical sample; (5) extracting the analyte of interest (DNA, protein) from the analytical sample and using subsamples of it as final test portions (Lipp *et al.*, 2005). The final test portion, for example in the case of PCR analysis of DNA, is typically around 100-200 ng of DNA which can be used in a single PCR.

7.3 SAMPLE PREPARATION PROCEDURES

The next step in GMO detection and quantification analyses, following the sampling procedure, is sample preparation for subsequent analytical procedures. Since all

officially approved detection techniques rely on either DNA or protein-based assays, this section will focus on sample preparation and extraction techniques for these two compounds.

SAMPLE PREPARATION Extracting the analyte, usually DNA or protein, from a sample for subsequent analyses.

The ultimate aim of **sample preparation** is the isolation of DNA or protein with sufficient integrity, purity and quantity to allow reliable detection and quantification analyses. The choice of extraction procedure depends on the sample matrix, the target analyte and the type of analysis to be performed (GMO screening, identification or quantification). Different sample matrixes in combination with different extraction procedures have been shown to strongly influence the outcome of subsequent analyses (Cankar *et al.*, 2006), therefore the appropriate extraction method needs to be determined for each individual sample (Jasbeer *et al.*, 2008).

A further complication is the fact that samples often consist of highly processed food, i.e. the original plant or animal material has undergone several manufacturing steps. This might include simple mechanical procedures, such as milling, or complex chemical or enzyme-catalysed modifications. Since proteins and DNA are likely to be degraded during such processing steps, the detection of these compounds in highly processed food requires sensitive and reliable detection methods (Michelini *et al.*, 2004).

7.3.1 DNA extraction procedures

Compared with protein, DNA is a relatively stable molecule that can still be identified when it is partially degraded or denatured, contributing to its prime importance for GMO detection. It is possible to obtain DNA suitable for subsequent analyses from highly processed and refined food matrices; examples of failures to isolate DNA, to date, include refined soybean oil, soybean sauce and refined sugar (Jasbeer *et al.*, 2008). DNA can be isolated as intact, high molecular weight DNA from fresh material, or as fragmented DNA from processed, old material (Ovesna *et al.*, 2008).

Three parameters are characteristic for DNA extraction procedures:

- The DNA quantity: the overall amount of extracted DNA.
- The DNA quality: as mentioned, food processing has a negative effect on DNA quality. Heat exposure, enzymatic degradation or unfavourable chemical conditions contribute to DNA fragmentation or damage. Target sequences for subsequent analyses, therefore, often do not exceed 100-400 bp in length.
- DNA purity: DNA in food matrices might be severely contaminated, by substances such as polysaccharides, lipids or polyphenols. Obtaining DNA of high purity is important to avoid complications or misleading results during subsequent analyses.

The key steps in sample preparation include homogenization of the material, chemical or enzymatic pretreatment, extraction and purification (Jasbeer *et al.*, 2008). Concerning plant material, small aliquots of 100-350 mg are sufficient for DNA isolation, given that this laboratory sample is representative of the field sample and has been correctly homogenized (Anklam *et al.*, 2002).

Five DNA extraction methods are commonly used, depending on the food matrix to be analysed. These are the DNeasy Plant Mini Kit (Qiagen), Wizard extraction (Promega), GENESpin Kit (GeneScan), cetyl trimethylammonium bromide (CTAB) based extraction, or a combination of CTAB-extraction with DNA-binding silica columns (Michelini *et al.*, 2008). It is important to carefully determine the extraction method that is most suited for the food matrix in question in order to obtain reliable and reproducible extraction and analysis results.

7.3.2 Protein extraction procedures

In contrast to DNA, proteins are very heat-labile molecules. Furthermore, they are easily affected by chemical treatments or enzymatic degradation. The detection of a specific protein depends on the recognition of this protein by an antibody

directed against that protein. If the target protein is degraded or denatured (i.e. loses its specific 3-dimensional shape), this antibody-mediated detection can no longer be performed. Therefore, it is not possible to reliably and reproducibly detect and quantify proteins in complex food matrices, such as processed agricultural material and food products, that have been subjected to mechanical, thermal, enzymatic or chemical processing (Anklam *et al.*, 2002).

Due to these limitations, protein analysis is only applicable for materials in their raw state (Jasbeer *et al.*, 2008). However, the basic steps in sample preparation are the same as in DNA extraction: material homogenization, pretreatment, extraction and purification.

7.4 GMO DETECTION BY PHENOTYPIC CHARACTERIZATION

Phenotypic characterization is possible if the inroduced transgene(s) result in the absence or presence of a specific trait that can be screened by analysing the phenotype of the organism. Detection methods using this approach are referred to as bioassays. This approach can be used, for example, to test for the presence or absence of herbicide resistance transgenes. One such test is based on the germination of seeds in the presence of the herbicide of interest and subsequent analysis of germination capacity. Herbicide assays are considered to be accurate and inexpensive. Controls, including seeds with or without the trait targeted, should be included in all samples tested. Typically, a test sample consists of 400 seeds. The test accuracy is dependent on the overall germination efficiency of the seeds: the higher the germination efficiency, the higher the confidence level of the test. Obviously, only viable seed or grain can be tested (no processed products), and each test requires several days to complete. Furthermore, bioassays require separate tests for each trait in question and at present such tests will not detect non-herbicide tolerance traits. Therefore, the tests are only of limited value for inspection authorities.

7.5 MOLECULAR DETECTION AND QUANTIFICATION OF GMOs – DNA-BASED METHODS

As stated above, the methods of choice for detecting and quantifying GMO material on a molecular level are based on detecting either the inserted, foreign DNA fragments or the novel proteins that are expressed from this DNA. Methods for the detection of foreign DNA rely mainly on PCR (Box 7.1), that allows amplification and detection of specific DNA fragments from the entire genome. Another advantage of **DNA-based detection** is the finding that there is usually a linear relationship between quantity of GMO present in a sample and quantity of transgenic DNA, thus it can be used to accurately quantify the amount of GMO material present in a sample. Finally, the stability of DNA and the extractability of suitable DNA even from highly processed food matrices contribute to its prime importance for GMO analysis.

DNA-BASED GMO DETECTION
Detection of GMOs via DNA is the most commonly used and most reliable technique. The modified DNA is detected by PCR.

7.5.1 PCR-based GMO detection

As evident from the name, GMOs are the result of genetic modification. Therefore, the most suitable GMO detection methods are those that directly target the modification itself – the modified DNA.

Polymerase chain reaction (PCR), including variants of the technique such as competitive PCR and real-time-PCR, is the method of choice for DNA-based GMO detection, identification and quantification (Lipp *et al.*, 2005). Due to its very high sensitivity, PCR is well suited for the analysis of processed food matrices containing degraded DNA or material that has only low GMO content.

PCR-BASED GMO DETECTION
Detecting GMOs by amplifying sequences of the introduced transgene by PCR. Requires knowledge of the molecular makeup of a GMO.

PCR-based GMO detection is dependent on detailed knowledge of the molecular makeup of a GMO, i.e. the sequence of the transgene and, optimally, the transgene integration site in the host genome. For authorized and commercially released GMOs, such information is available in public databases such as AGBIOS

(Ovesna *et al.*, 2008). In general, a typical gene construct for the production of a GMO consists of at least three elements (refer to Chapter 1): a promoter to drive expression of the inserted gene(s), the inserted/altered gene(s), and a terminator as a stop signal behind these genes. Such sequences can be specifically detected in a PCR analysis.

If no detailed sequence information about a GMO is available, PCR-based methods rely on the detection of commonly used genetic elements. Such frequently used elements are, for example, the CaMV 35S promoter, the *A. tumefaciens* nopaline synthase terminator (nos3'), or the kanamycin resistance marker gene (nptII) (Michelini *et al.*, 2008). Focusing on such sequences for routine GMO screening purposes is promising, since many commercially available GMOs contain these elements, or varieties thereof, and can thus be detected in standard screening procedures.

GMO detection is frequently based on the detection of the P-35S and nos3' genetic elements; however, several approved GMOs do not contain the P-35S or nos3' sequences and additional target sequences are needed to detect their presence. Furthermore, to detect as many variants of a GMO marker as possible (there are at least eight variants of P-35S used in GM crops), a careful choice of primers (see Box 7.1) is required. In addition, it should be noted that the detection of a common GMO marker solely indicates the presence of material derived from a GMO within a sample, but does not provide any information about the species or the engineered trait (Jasbeer *et al.*, 2008).

Most PCR-based GMO detection methods include a positive control primer set for the amplification of a reference gene. This is often a so-called housekeeping gene, which is present in (and unique to) all varieties of the investigated species (Miraglia *et al.*, 2004). Examples include the lectin gene in soybean or the invertase gene in maize. If a strong signal cannot be obtained with the positive control primer set, then there may be problems with the integrity or purity of the extracted DNA.

BOX 7.1

THE POLYMERASE CHAIN REACTION (PCR)

The **Polymerase chain reaction (PCR),** developed in the early 1980s, allows the million-fold amplification of a specific target DNA sequence. This is achieved by framing the target sequence with two primers, synthetic oligonucleotides of 20 to 30 bp, that are complementary to either one end of the two strands of the target sequence. For primer design, exact knowledge of the target sequence is required. Amplification of the target sequence is achieved by elongation of these primers by an enzyme (a DNA polymerase) using the target sequence as template. Repeating this reaction several times results in an exponential accumulation of the target sequence, since the amount of target sequence is doubled during each reaction cycle.

In principle, the PCR is a multiple-step process with consecutive cycles of three different temperatures, where the number of amplified target sequence grows exponentially according to the number of cycles. In each cycle the three temperatures correspond to three different steps in the reaction.

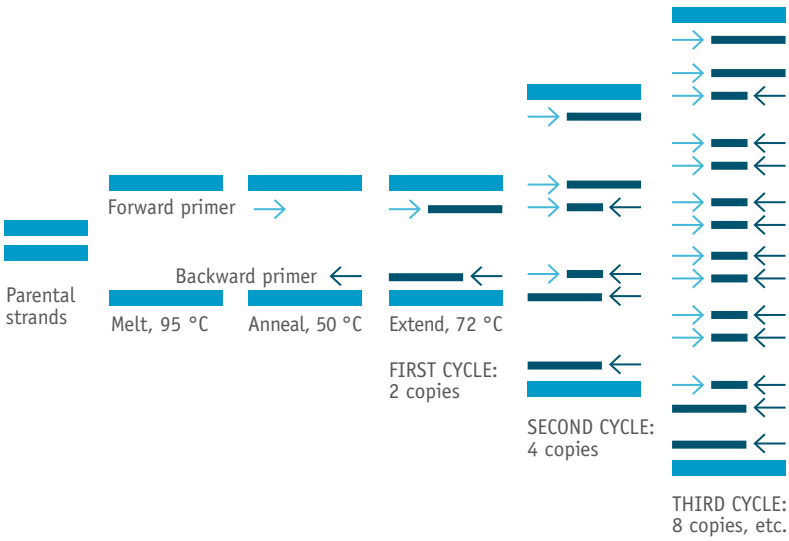
In the first step, the template DNA, i.e. the DNA serving as master-copy for the DNA fragment to be synthesized by DNA polymerase, is transformed from a double helix into single strands by heat denaturation at ~94 °C.

In the second step, the reaction mixture is cooled down to a temperature of 50-65 °C (depending on the primers used) to allow the annealing of the primers to both ends of the target sequence. Primer hybridization is favoured over DNA-DNA hybridization because of a high excess of primer molecules compared with template DNA in the reaction mixture. However, the annealing process is uncontrolled and can give rise to a large number of mismatched DNA duplexes.

POLYMERASE CHAIN REACTION
The process of rapidly amplifying a defined piece of DNA by repeated cycles of an enzymatic reaction.

Figure 7.2 | **The polymerase chain reaction**

See text for details.



In the third step, the annealed primers are extended using the target DNA strands as templates. This is usually performed by the enzyme DNA polymerase from the archaebacterium *Thermus aquaticus* (Taq) at its optimum temperature of 72 °C. With the elongation of the primers, a copy of the target sequence is generated.

The cycle is then repeated 20 to 50 times, depending on the initial amount of DNA present and the length of the amplicon (i.e. the amplified DNA fragment). The reaction results in an exponential amplification of the initially present target DNA, that can be subjected to subsequent analyses such as Southern blot, restriction digests or sequencing to verify its identity. In principle, PCR can be performed with as little as one template DNA initially present, but usually samples are adjusted to contain 25 to 100 template molecules. Due to this high sensitivity, PCR is very susceptible to contamination with undesirable DNA, that might produce false results.

Negative controls, for example samples with all necessary PCR ingredients but without template DNA, should also be included routinely to test for contamination with undesired DNA.

The **outcome of a PCR** can be evaluated by a variety of methods. Most frequently, amplified DNA fragments are subjected to agarose gel electrophoresis, a method to separate and visualize DNA fragments according to size. Since the expected size of a given target sequence is known, the presence of a fragment of that size indicates the presence of that target sequence in the original sample. If no fragment of the expected size is obtained, the sample did not contain the target sequence (given that the PCR worked well). To further verify the identity of an amplified fragment, it can be subjected to hybridization experiments with a complementary sequence, to analytical restriction enzyme digest, or to sequencing (Michelini *et al.*, 2008).

PCR EVALUATION
The outcome of a PCR can be most easily assayed by determining the size of the product. Other techniques are restriction enzyme digest, hybridization assays, or sequencing.

7.5.2 PCR-based GMO identification

Following a positive result from a GMO screening procedure, the next step is the unequivocal **identification of the GMO**(s) contained in a sample and the genetic modification event(s) involved. This can be achieved by PCR as well; however, compared with GMO detection, GMO identification is even more dependent on detailed information about the exact genetic modification of a GMO. In fact, this is a major limitation of PCR-based GMO detection and identification: if no such information is available, the GMO will not be detected or identified. Several approaches for GMO identification by PCR exist, and they are summarized below:

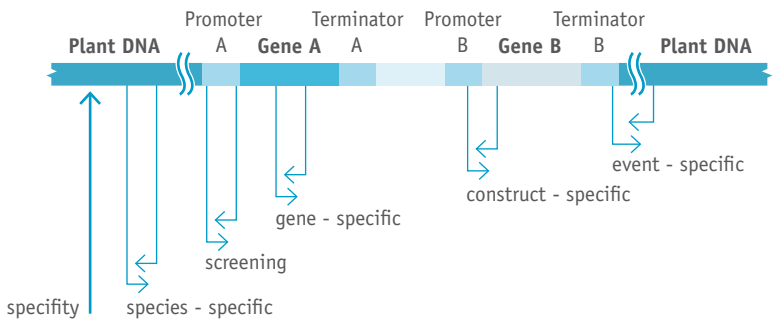
GMO IDENTIFICATION
Following GMO detection, the exact origin and GM species needs to be identified.

» *Gene-specific PCR:* In a gene-specific PCR, primers are used that lead to the amplification of a fragment from one gene of the transgenic element. This is rather unspecific, since many GMOs are engineered to contain the same, favourable genes. Thus, this method will fail to distinguish between these GMOs. This approach is therefore only useful if the target gene is present in only one GMO within a sample.

- *Construct-specific PCR:* This approach is more specific than gene-specific PCR. It is based on primers that target the junctions between different elements of the transgene insert, e.g. between the promoter and the gene or between different genes of the insert. Many GMOs contain identical genes, but the exact layout of their transgenes may differ, for example by a different arrangement of the genes or by the use of different promoters and terminators. By using construct-specific PCR, these different constructs, and thus GMOs, can be distinguished and identified.
- *Event-specific PCR:* Event-specific PCR is the most specific GMO identification strategy. Event, in this case, refers to the insertion of a transgene cassette into the host genome. The integration site is usually specific for each GMO. PCR primers, in this case, target the junction between the transgenic insert and the adjacent host genomic DNA. In most cases, this allows GMO identification with high certainty.

The different PCR layouts are also depicted in Figure 7.3. PCR evaluation is performed in the same way as described in 7.4.1, i.e. by visualization of amplified DNA fragments and subsequent sequence verification.

Figure 7.3 | **Different PCR strategies with increasing specificity**



Adapted from: Ovesna *et al.*, 2008.

Due to the dependency of PCR on detailed genetic information about GMOs there is a strong need for a continuous survey of all data available on GMOs – especially the introduced genetic elements and their integration sites. This applies not only for GM products approved for market release but also for any other GMO released for field trials worldwide. Only complete and accessible GMO information can guarantee comprehensive monitoring, detection and identification of GMOs.

7.5.3 PCR-based GMO quantification

The third step in GMO analysis, following detection and identification, is **GMO quantification**. Quantifying the GMO content in a sample is important to assess compliance with specific threshold levels for GMOs established by biosafety regulations. The typical approach to quantification utilizes one or more of the broad-spectrum primer sets that target common transgenic elements in GMOs. However, since different GMOs possibly contain these common elements in different numbers, accurate determination of GMO content cannot rely on the use of these common sequence elements alone. Quantification based on event-specific primers is therefore the most accurate means of obtaining quantitative results on GM content.

GMO QUANTIFICATION
Following GMO detection and identification, the exact amount of GMO material within a sample needs to be determined to check compliance with threshold and labeling regulations.

In general, two quantification approaches can be distinguished: absolute quantification and relative quantification. Absolute quantification, as the name suggests, yields absolute values of an analyte within a sample, e.g. how many milligrams of DNA could be extracted from a sample? This quantification is dependent on the sample size. The second approach is relative quantification: this is a measure of the amount of a substance compared to another substance, e.g. how many copies of transgene DNA per total DNA, or how many copy numbers of a gene per genome? Importantly, the final value obtained is a percentage, and the measurement is independent of the analysed sample size. Relative quantification is required for all GMO-related questions, such as compliance with labelling regulations (Jasbeer *et al.*, 2008).

7.5.3.1 Use of conventional PCR quantification

COMPETITIVE PCR
A varietation of PCR to quantifiy the amount of template DNA in a sample. Based on the co-amplification of a so-called competitor sequence.

One possibility for DNA quantification based on conventional PCR is double competitive PCR (DC-PCR). In **competitive PCR**, one primer pair is used to amplify both the target GMO template DNA and a synthetic template DNA fragment that is added to the same reaction mixture. The second fragment, which has a different size from the GMO target DNA (≤ 40 bp), is called the competitor. By conducting a series of experiments with varying amounts of the added synthetic DNA, it is possible to determine the amount of target GMO DNA in the sample. The competitor DNA serves as internal standard, and is added in different concentrations to the reaction mixture (an experimental setup known as titration). Following PCR amplification, the amplified fragments are visualized by agarose gel electrophoresis. The ratio of the two amplification products then represents the ratio of the initial two template sequences in the PCR mix. In other words, when the two products show equal amplification intensities, the amounts of initial template DNAs were the same. Since the amounts of added competitor DNA are known, this allows quantification of the target DNA in the sample.

Competitive and double-competitive PCR methods are semi-quantitative as a standard is required for comparison. In these cases the standard is the known amount of synthetic DNA. Consequently, the results will only indicate a value below, equal to or above a defined concentration of the standard.

7.5.3.2 Real-time PCR for GMO quantification

REAL-TIME PCR
A PCR format that allows quantification of the amounts of a specific DNA sequence within a sample.

Another strategy that improves accuracy, specificity and throughput of quantitative PCR is **real-time PCR**. This technique was originally developed in 1992 and is rapidly gaining popularity due to the introduction of several complete real-time PCR instruments and easy-to-use PCR assays. A unique feature of this PCR technique is that the amplification of the target DNA sequence can be followed during the

entire reaction by indirect monitoring of product formation. To this end, the conventional PCR reaction has been modified in order to generate a constantly measurable signal, whose intensity is directly related to the amount of amplified product. This signal is usually fluorescence, which is produced by an interaction between newly amplified DNA with certain added fluorophores. The increases in fluorescence during the reaction, that correspond to increasing concentrations of target DNA, are automatically measured, displayed on a computer screen, and can be analysed using suitable software.

DNA quantification by real-time PCR is based on the following principle: the PCR reaction mixture is submitted to several cycles of the reaction, until a fluorescent signal is encountered that is statistically significant above the noise level. The number of PCR cycles necessary to reach this threshold is recorded and referred to as Ct (cycle threshold) value. It is important to measure the Ct value in the exponential phase of the amplification procedure. During this stage, the Ct value is inversely proportional to the initial amount of template DNA molecules. In other words, a sample with many template molecules will reach a certain fluorescence threshold level faster than a sample with fewer molecules. For example, if a sample contains twice as many template molecules as a second sample, it will reach the threshold one cycle before the second sample since the amount of DNA is doubled during each reaction cycle. Thus, a low Ct value corresponds to a high initial concentration of target DNA.

Quantification of GMO DNA in a sample by RT-PCR is based on a combination of two absolute quantification values; one for the GMO target transgenic DNA and one for a species-specific reference gene. The GMO content in a sample can be calculated as a percentage using these two absolute values (Michelini *et al.*, 2008). Careful choice of suitable reference material is therefore of crucial importance for determining exact ratios of GMO to non-GMO material. Furthermore, it is important to know the copy number of the inserted transgenic sequences.

DETECTION LIMIT
Using real-time PCR, as little as 0.01 percent of GMO material can be reliably identified and quantified.

The **detection limit** of real-time PCR is very high; for corn, a detection of 0.01 percent GM corn versus non-GM corn has been demonstrated (Anklam *et al.*, 2002).

Several types of fluorescent probes for quantification of DNA using real-time PCR are currently available. One can discriminate between two classes of fluorophores: general DNA-binding dyes and fluorescent reporter probes. The first ones, a prominent example being SYBR Green, bind to double-stranded DNA in an unspecific manner and the resulting dye-DNA complex shows fluorescence. Since the overall amount of dsDNA in a PCR reaction increases, so does the intensity of fluorescence. The second type of probe consists of an oligonucleotide that is complementary to the target sequence, and a fluorophore and a quencher dye attached to it (e.g. the Taqman system). In the intact probe, the fluorophores' fluorescence is inhibited by the proximity of the quencher dye. During the annealing step of the PCR cycle, the oligonucleotide anneals to the target sequence between the two primers. Upon passage of the DNA polymerase during the elongation step, the oligonucleotide is cleaved and the fluorophore is liberated from the quencher dye. Thus, with increasing PCR cycles, the intensity of fluorescence increases as well. The latter, reporter-probe based method has the advantage that only the amplification of the desired target sequence is measured, while non-specific DNA binding dyes also react with non-specific PCR amplification products or other DNA hybrids (Miraglia *et al.*, 2004).

7.5.4 Confirmatory assays

CONFIRMATORY ASSAYS
All GMO detection, identification and quantififcation steps need to be verified by confirmatory assays to ensure the correctness of the obtained results.

Following PCR analysis, the identity of the amplicon needs to be confirmed and verified to ensure that the amplified sequence indeed represents the target sequence and is not an unspecific PCR artifact. Several **confirmatory assays** are available and commonly applied. Agarose gel electrophoresis, the simplest technique, can be applied to check if the amplicon is the expected size. However, it cannot be excluded that a PCR artifact, by coincidence, has the same size as the target sequence. To

further verify amplicon identity, it can therefore be subjected to restriction enzyme digest, since every DNA sequence has specific restriction profile. A further assay is Southern blotting, where the target amplicon is subjected to gel electrophoresis, transferred from the gel to a membrane, and hybridized with a complementary, labelled DNA probe; only the correct target sequence will yield a signal from binding of the complementary probe. A further possibility is nested PCR, where two primer pairs and two rounds of amplification are used: the second primer pair anneals within the target region of the first amplification, thus only the correct first amplification product will yield a second amplification product. The ultimate confirmatory assay is sequencing of the amplicon; however, this is rather expensive and requires special equipment that is not available in standard laboratories.

As stated above, PCR is able to amplify and thus identify very small amounts of initial target DNA. This implies that PCR is very sensitive to contamination with undesired DNA, possibly yielding false results in subsequent analyses. Therefore, high caution must be taken during all steps of PCR sample preparation and reaction setup to avoid cross-contamination. This already begins at sampling and sample preparation: it might already be sufficient to use the same grinding device for homogenization of two samples to produce contamination, even if no visible traces were left. Therefore it is of major importance to thoroughly clean and monitor all devices that come in contact with samples and that could potentially contribute to cross-contamination.

7.6 MOLECULAR DETECTION AND QUANTIFICATION OF GMOs – PROTEIN-BASED METHODS

A GMO is typically characterized by the introduction of novel genes, which direct the expression of novel proteins. Therefore, the second approach to detect GMOs is not based on detection of the modified DNA, but on the novel and newly expressed proteins. However, whereas modified DNA can be detected in all parts

of a transgenic organism at all times, this may not be the case for proteins: the genetic modification might not be directed at the production of novel proteins, protein expression levels might be too low to be detected, and proteins might only be expressed in certain parts of a plant or during certain stages of development (Jasbeer *et al.*, 2008).

PROTEIN-BASED GMO DETECTION
Detection and quantification of GMO material via the newly introduced proteins.

A further limitation for **protein-based GMO detection** is the susceptibility of proteins to heat denaturation and to chemical, enzymatic or mechanical degradation. Since protein detection requires intact, correctly folded protein molecules, it is only possible to reliably detect proteins in raw, non-processed commodities (Miraglia *et al.*, 2004).

Protein-based methods rely on a specific binding between the protein of interest and an antibody against that protein. The antibody recognizes the protein molecule, binds to it, and the resulting complex can be detected, for example by a chromogenic (colour) reaction. This type of assay is referred to as immunoassay, since antibodies are the molecules that are produced during an immune reaction to recognize and eliminate foreign (pathogenic) molecules. The main technique applying this procedure is called ELISA (enzyme-linked immunosorbent assay, Figure 7.4). The antibody required to detect the protein can only be developed with prior access to the purified protein; the protein can be purified from the GMO itself, or it can be synthesized in a laboratory if the composition of the protein is known in detail. Immunoassays can be applied both for detection and quantification of protein, over a wide range of protein concentration. Such assays are available for many proteins that are expressed in commercially released GMOs (Michelini *et al.*, 2008).

ELISA
Acronym for Enzyme-linked Immunosorbent Assay. Method to detect and quantify specific proteins.

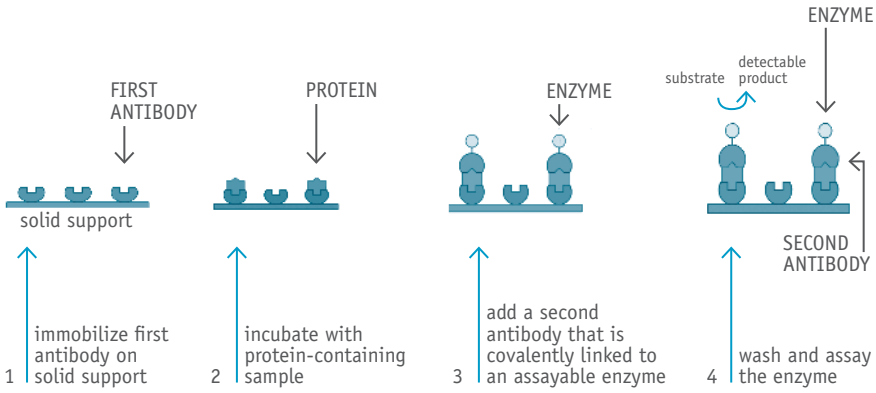
7.6.1 Enzyme-linked immunosorbent assay (ELISA)

In **ELISA**, a protein-antibody reaction takes place in solution on a solid support (plastic plates) and a protein-antibody complex is formed. This complex is usually

visualized by adding a second antibody that binds to the first antibody, and that is linked to a certain enzyme. This enzyme can catalyse the reaction of a specific substrate, which is added to the solution, to a coloured product (chromogenic detection). The intensity of the colour can be measured photometrically and used for quantitative assessments of protein concentration. ELISAs are available for several frequently engineered proteins in GM plants, including neomycin phosphotransferase (nptII), 5-enolpyruvyl-shikimate 3-phosphate synthase (EPSPS), the Bt insecticide Cry1Ab and phosphinotricin acetyltransferase (PAT) (Jasbeer *et al.*, 2008).

Figure 7.4 | **Enzyme-linked Immunosorbent Assay**

In this case a Sandwich-ELISA is depicted. A first antibody is immobilized on a solid support, followed by incubation with the target-protein containing solution. After a washing step (not shown), the second antibody, coupled to an assayable enzyme, is added and binds to the immobilized target protein. Finally the amount of bound secondary antibody, and thus target protein, can be assayed using the attached enzyme, which is usually done colorimetrically. In an easier approach, the target protein can be immobilized directly onto the plate, without a primary antibody.



Some ELISA plates are supplied with a calibration of known concentration of target protein in solution and a negative control defined by the absence of the target. These standards will exhibit distinctively different intensities of a given colour at the different concentrations of target molecules provided. By comparing the intensity of colour of the sample tested for GMO target molecules with that of the standards, it is possible to work out the concentration range of the target. These immunoassay measurements are semi-quantitative. Quantitative measurements can, however, be obtained by using a microplate reader which measures the absorbance of all samples and standards at the same time. This results in a very high precision of data acquisition and subsequently a precise calculation of target protein concentration in the test samples.

A major advantage of ELISA is the high specificity of the protein-antibody recognition, which allows accurate identification of proteins. Furthermore, they are fast, require only low work input, can be performed automatically to a large extent, and require only small investments in equipment and personnel.

However, ELISA may be around 100 times less sensitive than DNA-based methods, although detection of 0.01 percent of GM material has been described (Grothaus *et al.*, 2006). Furthermore, initial development and validation of a test for a specific protein is more time-consuming, and the supply of antibodies, which are derived from laboratory animals, is a limiting factor (Jasbeer *et al.*, 2008). Furthermore, protein detection and antibody affinity might be affected by the individual matrix under examination (Anklam *et al.*, 2002).

7.6.2 Lateral flow devices and dip sticks

LATERAL FLOW DEVICE
Rapid, on-site method for detecting GMO-derived protein, based on the same principle as ELISA.

Lateral flow devices and dip sticks are variations of the technology that ELISAs are based on; paper strips or plastic paddles on which antibody is captured on specific zones are used to detect protein targets derived from GMOs. The strip is dipped

into vials containing solutions of the sample to be tested. Each dip is followed by rinsing; the positive reaction is a colour change in a specific zone on the stick. Recent improvements of the dip stick have produced lateral flow strips in which reagents are transported through nylon membranes by capillary action. Antibodies specific to the target protein are coupled to a coloured reagent and are incorporated into the lateral flow strip. When the strip is brought into contact with a small amount of the sample containing the target protein, an antibody-antigen complex is formed with some of the antibody. The membrane contains two capture zones, one for the bound protein and the other for the coloured reagent. A coloured band appears in the capture zone corresponding to the bound antibody-protein complex and coloured reagent. Appearance of a single coloured band in the membrane is a negative test for the presence of the protein targeted. The presence of two bands represents detection of the target (Grothaus *et al.*, 2006).

These tests are available as kits and do not require major equipment or training, and thus represent a rapid GMO testing possibility. Sample preparation only involves homogenization of the sample and mixing with the reagents contained in the kit (Jasbeer *et al.*, 2008).

7.7 MOLECULAR DETECTION AND QUANTIFICATION OF GMOS – OTHER METHODS

Several other methods for the detection and quantification of GMOs have been proposed or are in developmental stages. Some of them are presented below – however, the main approved technologies for GMO analysis are PCR-based techniques and ELISA.

7.7.1 Chromatography and near infrared spectroscopy

If the chemical composition of a GMO has been altered, for example fatty acid or triglyceride content, chemical methods based on chromatography or near infrared

spectroscopy may be applied to detect these changes. These methods will detect differences in the chemical profile between GM organisms and conventional organisms. The applicability of such approaches has been demonstrated by investigating the triglyceride pattern of oils derived from GM canola by high performance liquid chromatography (HPLC). Triglyceride patterns and content can be compared between GM and non-GM samples. However, it should be noted that such techniques are only applicable when significant changes occur in the biochemical composition of GM plants or derived products. In addition, such methodologies only offer qualitative detection and no quantification (Anklam *et al.*, 2002). In particular, the addition of GM-derived products or raw material in small quantities to a larger lot of conventional material are probably not detectable given the sensitivity of the methods currently used.

7.7.2 Microarrays

MICROARRAYS
Comparatively new technology to detect GMOs. Possible to detect thousands of short DNA sequences in a single experimental setup.

Microarray technology (DNA-chip technology) has been developed in recent years for automated rapid screening of gene expression profiles and sequence variation of large numbers of samples. Microarray technology is based on the DNA hybridization principle, with the main difference that many (up to thousands) specific probes are attached to a solid surface and can be simultaneously detected. Different formats have been developed, including macroarrays, microarrays, high-density oligonucleotide arrays (gene chips or DNA chips) and microelectronic arrays.

GMO chip kits are designed to detect species-specific DNA of plants and viruses, frequently used transgene construction elements and specifically introduced genetic modifications, and thus allow the identification of approved and non-approved GMO varieties. One example of a GMO chip version that has been designed and tested for its applicability is capable of detecting species-specific DNA from soybean, maize, oilseed rape, rice, CaMV and several GMOs, including RR-soybean, Maximizer Bt 176 maize, Bt11 maize, Yieldgard Mon810 maize and Bt-Xtra maize. In addition,

GMO chips allow the detection of all GMOs that contain the widely used CaMV 35S promoter, Nos-terminator, *nptII*, *bar*, and *pat* genes (Leimanis *et al.*, 2006). Microarrays, in general, thus allow the detection, identification and quantification of a variety of GMOs in a single experimental setup.

7.8 SUMMARY OF GMO ANALYSIS: LIMITS AND OUTLOOK

As stated in the introduction, the field of GMO detection has a high relevance for all involved parties: research and development, producers, traders, consumers and legislation. Further progress in sampling and detection techniques and in traceability strategies needs to be made to enable adequate implementation and maintenance of GMO-relevant legislation and labelling requirements (Miraglia *et al.*, 2004). Promotion and implementation of reliable, international traceability strategies and agreements may also increase public trust in the transparency of GMOs and related products.

7.8.1 Summary of DNA and protein-based techniques

To summarize the previous sections, DNA and protein-based methods are currently the techniques of choice for GMO analysis. A PCR analysis can take between one to ten days and costs range from 100 to 400 euros. In comparison, an on-site ELISA takes two to eight hours and costs approximately 10 Euros; ELISA-based dipsticks take a few minutes to complete and cost around 3 euros (Miraglia *et al.*, 2004).

DNA-based analysis offers several advantages, including:

- a wide range of applications, from initial GMO screening to event-specific detection;
- the genome is the same in all cells of an organism, i.e. every part of an organism can be analysed;
- relative quantification, as required for labelling legislation, is possible;

- DNA is comparatively stable and can be isolated from a wide range of raw and processed matrices;
- a very high sensitivity.

Disadvantages of DNA-based methods include:

- the need for trained staff to operate high-end equipment;
- expensive, time-consuming and relatively unsuitable for on-site testing;
- DNA may be removed or degraded by certain processing procedures; certain food ingredients possibly interfere with DNA amplification and detection;
- PCR is very susceptible to cross-contamination;
- if no detailed sequence information of a GMO is available, DNA-based analysis is not possible.

Protein-based analysis offers the following advantages:

- comparatively cheap and less skilled personnel required;
- cheaper and less sophisticated equipment needed;
- fast conductance;
- quantification is possible;
- comparatively robust and simple assay formats;
- suitable for batch analysis of samples;
- possible to conduct on-site tests.

The disadvantages of protein-based analysis include:

- inferior sensitivity compared to DNA-based methods;
- the development of antibodies is difficult, expensive and requires skilled staff and equipment;
- only samples containing intact protein, i.e. fresh material, can be analysed;
- not possible to distinguish different events that produce the same protein (i.e. less specific than DNA-based methods);

- protein expression levels in a GM organism may vary significantly in a temporal and spatial manner;
- no relative, but absolute quantification;
- expression levels of target proteins may be too low to be detectable;
- reactivity of the antibody may be affected by other matrix components.

Thus, a careful evaluation of the most suitable analysis technique for a certain product should be performed to ensure that potential GMO contents are reliably, reproducibly and with high sensitivity detected and quantified. The choice of the technique may depend on a variety of factors, including the purpose (exact quantification for labelling legislation versus a simple yes/no result, GMO monitoring), the need for laboratory or on-site testing, financial background (including availability of personnel and equipment), exact GMO identification or just stating general GMO presence, the speed of analysis, composition of the food matrix to be analysed, etc. At present, however, PCR-based methods are the most widely applied and validated for GMO analysis purposes.

ANNEX 1

GENES OF INTEREST TO AGRICULTURE

A1.1 INTRODUCTION

Transgenic crops with novel agronomic and quality traits are grown in many developed and developing countries. A recent analysis of the current application of transgenic crops and the development over the last decade is provided by the International Service for the Acquisition of Agri-Biotech Applications (James, 2008). For a detailed account on the nature and extent of utilization of the various GM crops, one can consult online databases such as AGBIOS (http://www.agbios.com/dbase.php). The AGBIOS Web site includes details of the transgenes, the scientific background underpinning the traits and information on environmental and food safety issues of a variety of GM plants. A recent publication by the European Commission Joint Research Centre provides information about GM crops that are in the pipeline and expected to be marketed in the short to medium term, i.e. up to 2015 (Stein and Rodriguez-Cerezo, 2009). The database established by the authors is also available online at http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=2199. By surveying information in these and similar databases it is possible to get information on the

genes that have been used for the generation of transgenic crops, how these crops are commercially used and which additional crops are in developmental stages, in field trials or awaiting approval for commercial release. Each GMO is assigned a *Unique Identifier*, i.e. a code that allows allows direct identification of the GMO (Commission Regulation EC 65/2004).

A1.2 HERBICIDE TOLERANCE GENES

Glyphosate herbicide tolerance

The genetically modified glyphosate resistant crops contain a gene encoding the enzyme EPSPS, obtained from a strain of the soil inhabiting bacterium *Agrobacterium tumefaciens*. The EPSPS enzyme is an important part of the shikimate biochemical pathway which is required to produce aromatic amino acids, which plants need to grow and survive. EPSPS is also constitutively present in plants, but the enzyme is inhibited by binding of glyphosate. Conventional plants treated with glyphosate cannot produce the aromatic amino acids and die, whereas EPSPS from *A. tumefaciens* does not bind glyphosate and allows plants to survive the otherwise lethal effects of the herbicide (Tan *et al.*, 2006; Gianessi, 2008).

Glufosinate ammonium herbicide tolerance

Glufosinate ammonium is the active ingredient in the PPT herbicides. Glufosinate chemically resembles the amino acid glutamate and functions by inhibiting the enzyme glutamate synthase, which converts glutamate to glutamine. Glutamine synthesis is also involved in the ammonia detoxification of glufosinate resulting in reduced glutamine levels and increases in ammonia concentration. Elevated levels of ammonia damage cell membranes and impair photosynthesis. Glufosinate tolerance is the result of introducing a gene encoding the enzyme phosphinothricin-acetyl transferase (PAT). The gene was originally obtained from the soil actinomycete *Streptomyces hygroscopiens*. The PAT enzyme catalyses detoxification of phosphinothricin by acetylation (Duke, 2005; Tan *et al.*, 2006).

Sulfonylurea herbicide tolerance

Sulfonyl urea herbicides, such as triasulfuron and metsulfuron-methyl, target the enzyme acetolactate synthase (ALS), also called acetohydroxyacid synthase (AHAS), thereby inhibiting the biosynthesis of the branched chain animo acids valine, leucine and isoleucine (Tan *et al.*, 2005). This results in accumulation of toxic levels of the intermediate product alpha-ketoglutarate. In addition to the native ALS gene, herbicide tolerant crops contain the ALS gene from a tolerant line of *Arabadopsis thaliana*. This variant ALS gene differs from the wild type by one nucleotide and the resulting ALS enzyme differs by one amino acid from the wild type ALS enzyme. Still, this is sufficient to confer resistance to these herbicides, and provides an impressive example for the complexity and sensitivity of genes and proteins and the effects of mutations.

Oxynil herbicide tolerance

Oxynil herbicides and bromoxynil are effective against broad leaf weeds. Transgenic herbicide resistant crops contain a copy of the *bxn* gene isolated from the bacterium *Klebsiella pneumoniae*. The gene encodes a nitrilase which hydrolyses oxynil herbicides to non-phytotoxic compounds (Duke, 2005).

GENE STACKING
The development of transgenic plants that contain several transgenes, e.g. resistances to several different herbicides.

A recent development in herbicide tolerance is the development of plants containing several tolerance genes, allowing cocktails of different herbicides to be used (Green *et al.*, 2008). This technology is referred to as trait or **gene stacking**. Ideally, it will become possible to introduce not only herbicide tolerance traits, but also traits conferring insect resistance or quality traits (Halpin, 2005). One possible approach to this end is the development of artificial plant minichromosomes, capable of encoding many different, complex genes and regulatory sequences (Yu *et al.*, 2007).

A1.3 RESISTANCE TO BIOTIC STRESSES

Among insect pests, Lepidoptera (moths and butterflies) represent a diverse and important group. Most insect-resistant transgenic crop varieties developed so far

target the control of Lepidoptera, predominantly using transgene cassettes, including toxin-producing cry-type genes obtained from strains of the soil bacterium Bt. The Bt proteins bind to specific sites on the gut lining in susceptible insects (de Maagd *et al.*, 1999). The binding disrupts midgut ion balance which eventually leads to paralysis, bacterial sepsis and death. Important to note is that the original Bt cry-genes have been extensively modified, for example by deleting spurious splicing signals and optimizing the GC content, to improve the expression level in plants. Many cry genes exist that confer resistance to insects other than Lepidoptera. In addition to Bt cry genes, protease inhibitors, neuropeptides and peptide hormones that control and regulate the physiological processes of several insect pests have become candidates for developing insect-resistant crops. Other biocontrol toxins currently studied are chitinases, lectins, alpha-amylase inhibitors, cystatin and cholesterol-oxidase and glucosidase inhibitors (Christou *et al.*, 2006; Ranjekar *et al.*, 2003).

Among disease-causing organisms, viruses have received a lot of attention concerning the development of transgenic crops. This has been possible since the discovery of pathogen-derived resistance, where the expression of a viral protein (e.g. coat protein, replicase, helicase enzyme, etc.) in a transgenic plant renders that plant resistant to the virus (Prins *et al.*, 2008). As a result many viral genes have been cloned and used to transform crops. Genes encoding chitinases and glucanases have been used to generate plants resistant to fungal and bacterial pathogens, respectively. Other strategies for conferring resistance to pathogens in transgenic crops include genes for phytoalexine production pathways which are involved in pathogen-induced infection and defence, and R genes (resistance genes) which have been identified as responsible for additional defence mechanisms in plants (Campbell *et al.*, 2002).

A1.4 TOLERANCE TO ABIOTIC STRESSES

So far there are no commercialized transgenic crops with resistance to abiotic stresses such as drought, heat, salinity and frost. One possible explanation is that

the underlying genetic networks are rather complex, i.e. so far it has not been possible to identify single genes that would confer tolerance to these factors. However, a number of approaches are being developed to tackle these stress factors in crops (Bathnagar-Mathur *et al.*, 2008).

A1.5 QUALITY TRAITS

Modified flower colour

Many flowers including carnations, roses, lilies, chrysanthemums, roses and gerberas, which are important in the global flower trade, do not produce the blue pigment delphinidin. Transgenic carnation lines with unique violet/mauve colour have been developed. The genes of interest here include structural and regulatory genes of the flavanoid biosynthetic pathway.

Delayed fruit ripening and increased shelf life

Genes encoding an enzyme which degrades 1-aminocyclopropane 1-carboxylic acid (ACC), an ethylene precursor, and those encoding polygalacturonase (PG) have been suppressed in some transgenic plants. Suppression is accomplished by inserting a truncated or anti-sense version of the gene. Reduced ACC activity results in delayed fruit ripening while decreased activity of PG results in a lower level of cell wall breakdown and hence delays fruit softening and rotting (Prasanna *et al.*, 2007).

Modification of oil composition

Oilseed rape and soybean have been modified to increase the content of oleic acid in particular. The modified oils are lower in unsaturated fats and have greater heat stability than oils from the corresponding unmodified crops. In unmodified crops the FAD2 gene encodes a desaturase enzyme that converts C18:1 (oleic acid) to C18:2 and C18:3 acids. In the modified crop a mutant FAD2 gene prevents expression of the active desaturase, resulting in the accumulation of oleic acid (Kinney *et al.*, 2002).

Modified vitamin and mineral profiles

Vitamins and minerals are essential components of the human diet and dietary deficiencies of these nutrients can have severe effects on health and development. In addition to fortification and supplementation strategies for alleviating these deficiencies, transgenic crops with elevated and bio-available vitamins and minerals are being developed (Davies, 2007). Here the strategy is to express the genes responsible for the production or accumulation of the concerned nutrient in the edible parts of the plant. Thus promoters and other control sequences that target the expression of the gene(s) of interest to the correct part of the plant are highly important. In order to improve vitamin A production in rice the genes encoding phytoene synthase and phytoene desaturase have been expressed in the endosperm, resulting in the variety known as "Golden Rice". To improve iron accumulation and bio-availability in rice, genes such as ferritin synthase from soy (Fe storage), metallothionein (cystein-rich storage protein, improves Fe absorption) and a heat stable phytase gene (degrades phytic acid which inhibits Fe absorption) have been expressed in the rice endosperm.

A1.6 TRANSGENIC PLANTS AS BIOREACTORS FOR BIOPHARMACEUTICALS AND VACCINES

The first trials for the production of human proteins in plants dates back to the early 1990s; however, only in recent years has the use of transgenic plants as bioreactors for the production of small-molecule drugs or pharmaceutical proteins increasingly gained importance (Twyman *et al.*, 2005). The use of transgenic plants as a production platform presents a viable alternative to conventional production of such compounds, such as extraction from natural sources, various cell culture techniques or the use of animal bioreactors. In particular, plant-derived vaccines and antibodies are considered as promising (Tiwari *et al.*, 2009). Trials for the development of plants expressing vaccines in their edible parts, thus allowing cost-effective production and delivery of a vaccine, are a particularly intriguing option (Floss *et al.*, 2007).

REFERENCES

Allison, LA. 2007. *Fundamental molecular biology*. 1st edition. Malden, (MA), Wiley-Blackwell.

Anklam, E., Gadani, F., Heinze Hans Pijnenburg, P. & Van Den Eede, G. 2002. Analytical methods for detection and determination of genetically modified organisms in agricultural crops and plant-derived food products. *Eur. Food Res. Technol*. 214:3–26.

Basrur, P.K. & King, W.A. 2005. Genetics then and now: breeding the best and biotechnology. *Rev. Sci. Tech*. 24(1): 31-49.

Bazer, F.W. & Spencer, T.E. 2005. Reproductive biology in the era of genomics biology. *Theriogenology* 64(3): 442-56.

Benítez, T., Rincón, A.M., Limón, M.C. & Codón, A.C. 2004. Biocontrol mechanisms of Trichoderma strains. *Int. Microbiol*. 7(4): 249-60.

Bhatnagar-Mathur, P., Vadez, V. & Sharma, K.K. 2008. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. *Plant Cell. Rep*. 27(3): 411-24.

Brophy, B., Smolenski, G., Wheeler, T., Wells, D., L'Huillier, P. & Laible, G. 2003. Cloned transgenic cattle produce milk with higher levels of ß-casein and K-casein. *Nature Biotechnology* 21: 157– 62.

Bull, A.T., Ward, A.C. & Goodfellow, M. 2000. Search and discovery strategies for biotechnology: the paradigm shift. *Microbiol. Mol. Biol. Rev.* 64(3): 573-606.

Campbell, M.A., Fitzgerald, H.A. & Ronald, P.C. 2002. Engineering pathogen resistance in crop plants. *Transgenic Res*. 11(6): 599-613.

Cankar, K., Stebih, D., Dreo, T., Zel, J. & Gruden, K. 2006. Critical points of DNA quantification by real-time PCR--effects of DNA extraction method and sample matrix on quantification of genetically modified organisms. *BMC Biotechnol*. 6:37.

CBD (Convention on Biological Diversity). 5 June 1992. Rio de Janeiro. United Nations.

Chatterjee S., Chattopadhyay, P., Roy, S. & Sen, S.K. 2008. Bioremediation: a tool for cleaning polluted environments. *Journal of Applied Biosciences* 11: 594–601.

Chatterjee S., Chattopadhyay, P., Roy, S. & Sen, S.K. 2008. Bioremediation: a tool for cleaning polluted environments. *Journal of Applied Biosciences* 11: 594–601.

Chauthaiwale, V.M., Therwath, A. & Deshpandei., V.V. 1992**.** Bacteriophage lambda as a cloning vector. *Microbiological Reviews* pp. 577-591.

Christon, P., Capell, T., Kohli, A., Gatehouse, J.A. & Gataehouse, A.M. 2006. Recent developments and future prospects in insect pest control in transgenic crops. *Trends Plant Sci*. 11(6): 302-8.

Chung, S.M., Vaidya, M. & Tzafira, T. 2006. Agrobacterium is not alone: gene transfer to plants by viruses and other bacteria. *Trends Plant Sci*. 11(1): 1-4.

Clive, J. 2008. Global Status of Commercialized Biotech/GM Crops: *ISAAA Brief* No 39. Ithaca, New York, ISAA.

Dahiya, N., Tewari, R. & Hoondal, G.S. 2006. Biotechnological aspects of chitinolytic enzymes: a review. *Appl. Microbiol. Biotechnol.* 71(6): 773-82.

Darbani, B., Eimanifar A., Stewart, C.N. Jr. & Camargo, W.N. 2007. Methods to produce marker-free transgenic plants. *Biotechnol. J.* 2(1): 83-90.

Davey, M.R., Anthony, P., Power, J.B. & Lowe, K.C. 2005. Plant protoplasts: Status and biotechnological perspectives. *Biotechnol. Adv.* 23(2): 131-71.

Davies, K.M. 2007. Genetic modification of plant metabolism for human health benefits. *Mutat. Res.* 622(1-2): 122-37.

de Maagd, R.A., Bosch, D. & Stiekema, W. 1999. Bacillus thuringiensis-toxin-mediated insect resistance in plants. *Trends Plant Sci.* 4(1): 9-13.

Dekkers, J.C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82 E-Suppl: E313-328.

Demain, A.L. & Adrio, J.L. 2008. Contributions of microorganisms to industrial biology. *Mol. Biotechnol.* 38(1): 41-55.

Denning, C. & Priddle, H. 2003. New frontiers in gene targeting and cloning: success, application and challenges in domestic animals and human embryonic stem cells. *Reproduction* 126(1): 1-11.

Donovan, D.M., Kerr, D.E. & Wall, R.J. 2005. Engineering disease-resistant cattle. *Transgenic Res.* 14(5): 563-7.

Duke, S.O. 2005. Taking stock of herbicide-resistant crops ten years after introduction. *Pest Manag. Sci.* 61(3): 211-8.

EPA (Environmental Protection Agency). Factsheet 006474 on Agrabacterium radiobacter strain K1026. Available online at: http://www.epa.gov/pesticides/biopesticides/ingredients/factsheets/factsheet_006474.htm

Etherton, T.D. & Bauman, D.E. 1998. Biology of somatotropin in growth and lactation of domestic animals. *Physiol. Rev.* 78(3): 745-61.

FAO. 2001: Agricultural biotechnology for developing countries. Results of an electronic forum.

FAO. 2004. The state of food and agriculture. Agricultural biotechnology: meeting the needs of the poor?

FAO. 2005. Status of research and application of crop biotechnologies in developing countries. Preliminary assessment

FAO. 2006. The role of biotechnology in exploring and protecting agricultural genetic resources.

FAO. 2007. Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish.

FDA (U.S. Food and Drug Administration). Press Release, 6 February 2009. Retrieved on 15.10.2009 at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm109074.htm

Floss, D.M., Falkenburg, D. & Conrad, U. 2007. Production of vaccines and therapeutic antibodies for veterinary applications in transgenic plants: an overview. *Transgenic Res.* 16(3): 315-32.

Fravel, D.R. 2005. Commercialisation and implementation of biocontrol. *Annu. Rev. Phytopathol.* 43: 337–59.

Galli, C., Duchi, R., Crotti, G., Turini, P., Ponderato, N., Colleoni, S., Lagutina, I. & Lazzari, G. 2003. Bovine embryo technologies. *Theriogenology* 59(2): 599-616.

Gavrilescu, M. & Chisti, Y. 2005**.** Biotechnology – a sustainable alternative for chemical industry. *Biotechnology Advances* 23: 471–499.

Gellisen, G. (Ed). 2005. Production of recombinant proteins. 1st edition. Weinheim, Wiley-VCH.

Gelvin, S.B. 2003**.** Agrobacterium-mediated plant transformation. The biology behind the gene-jockeying tool. *Microbiol. Mol. Biol. Rev.* 67(1): 16-37.

Gianessi, L.P. 2008. Economic impacts of glyphosate-resistant crops. *Pest Manag. Sci.* 64(4): 346-52

Giraldo, P. & Montoliu, L. 2001**.** Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Research* 10**:** 83–103.

Gomez-Gil, B., Roque, A., Turnbull, J.F. & Inglis, V. 1998. A review on the use of microorganisms as probiotics. *Rev. Latinoam. Microbiol.* 40(3-4): 166-72.

Gray, N.K. & Wickens, M. 1998. Control of translation initiation in animals. *Annu. Rev. Cell. Dev. Biol.* 4: 399-458.

Green, J.M., Hazel, C.B., Forney, D.R. & Pugh, L.M. 2008. New multiple-herbicide crop resistance and formulation technology to augment the utility of glyphosate. *Pest Manag. Sci.* 64(4): 332-9.

Griffiths, A.J.F., Wessler, S.R., Lewontin, R.C., & Carroll, S.B. 2007. Introduction to genetic analyses. 9th edition. Palgrave Macmillan.

Grothaus, G.D., Bandla, M., Currier, T., Giroux, R., Jenkins, G.R., Lipp, M., Shan, G., Stave, J.W. & Pantella, V. 2006. Immunoassay as an analytical tool in agricultural biotechnology. *J. AOAC Int.* 89(4): 913-28.

Gurr, S.J. & Rushton, P.J. 2005**.** Engineering plants with increased disease resistance: how are we going to express it? *Trends in Biotechnology* 23(6): 283-290.

Haefner, S., Knietsch, A., Scholten, E., Braun, J., Lohscheidt, M. & Zelder, O. 2005. Biotechnological production and applications of phytases. *Appl. Microbiol. Biotechnol.* 68(5): 588-97.

Halpin, C. 2005. Gene stacking in transgenic plants--the challenge for 21st century plant biotechnology. *Plant Biotechnol. J.* 3(2): 141-55.

Herrera, G., Snyman, S.J. Thomson, J.A. 1994. Construction of a bioinsecticidal strain of pseudomonas fluorescens active against the sugarcane borer, Eldana saccharina. *Appl. Environ. Microbiol.* 60(2): 682-690.

Heyman, Y. 2005. Nuclear transfer: a new tool for reproductive biotechnology in cattle. *Reprod. Nutr. Dev.* 45(3): 353-61.

Hodges, C.A. & Stice, S.L. 2003. Generation of bovine transgenics using somatic cell nuclear transfer. *Reprod. Biol. Endocrinol.* 1: 81.

Houdebine, L.M. 2009. Production of pharmaceutical proteins by transgenic animals. *Comp. Immunol. Microbiol. Infect Dis.* 32(2): 107-21.

Hughes, T.A. 2006**.** Regulation of gene expression by alternative untranslated region. *Trends in Genetics* 22(3): 199-122.

International human genome sequencing consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945.

ISO (International Organization for Standardization). 2006. EN/TS 21568:2006, Foodstuffs — Methods of analysis for the detection of genetically modified organisms and derived products — Sampling strategies 2006; European Committee for Standardization, Brussels, Belgium

ISTA (International Seed Testing Agency). 2004. Handbook on Seed Sampling. 2nd Edition. ISTA Bulking and Sampling Committee. M. Kruse (ed.).

Jana, S. & Deb, J.K. 2005. Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 67: 289–298.

Jasbeer, K., Ghazali, F.M., Cheah, Y.K. & Son, R. 2008. Application of DNA and Immunoassay Analytical Methods for GMO Testing in Agricultural Crops and Plant-Derived Products. *ASEAN Food Journal* 15(1): 1-25.

J**obling, M.A. & Gill, P.** 2004. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* 10: 739-51.

Juven-Gershon, T., Hsu, J-Y., Theisen, J.W.M. & Kadonaga, J.T. 2008. The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell. Biol.* Author manuscript; available in PMC.

Kikkert, J.R., Vidal, J.R. & Reisch, B.I. 2005. Stable transformation of plant cells by particle bombardment/biolistics. *Methods Mol. Biol.* 286: 61-78.

Kind, A. & Schnieke, A. 2008. Animal pharming, two decades on. *Transgenic Res.* 2008 17(6): 1025-33.

Kinney, A.J., Cahoon, E.B. & Hitz, W.D. 2002. Manipulating desaturase activities in transgenic crop plants. *Biochem. Soc. Trans.* 30(Pt 6): 1099-103.

Kuroiwa, Y., Kasinathan, P., Choi, Y.J., Naeem, R., Tomizuka, K., Sullivan, E.J., Knott, J.G. Duteau, A., Goldsby, R.A., Osborne, B.A., Ishida, I. & Robl, J.M. 2002. Cloned transchromosomic calves producing human immunoglobulin. *Nat. Biotechnol.* 20(9): 889-94.

Kuroiwa, Y., Kasinathan, P., Matsushita, H., Sathivaselan, J., Sullivan, E.J., Kakitani, M., Tomizuka, K., Ishida, I. & Robl, J.M. 2004. Sequential targeting of the genes encoding immunoglobulin-mu and prion protein in cattle. *Nat. Genet.* 36(7): 775-80.

Lacroix, B., Li, J., Tzifira, T. & Citovsky, V. 2006. Will you let me use your nucleus? How Agrobacterium gets its T-DNA expressed in the host plant cell. *Can. J. Physiol. Pharmacol.* 84(3-4): 333-45.

Law, B.A. & Mulholland, F. 1991. The influence of biotechnological developments on cheese manufacture. *Biotechnol. Genet. Eng. Rev.* 9: 369-409.

Lawrence, S. 2007. Billion dollar babies – biotech drugs as blockbusters. *Nature Biotechnology* 25: 380-382.

Leimanis, S., Hernández, M., Fernández, S., Boyer, F., Burns, M., Bruderer, S., Glouden, T., Harris, N., Kaeppeli, O., Philipp, P., Pia, M., Puigdomènech, P., Vaitilingom, M., Bertheau, Y. & Remacle, J. 2006. A microarray-based detection system for genetically modified (GM) food ingredients. *Plant Mol. Biol.* 61(1-2): 123-39.

Lipp, M., Shillito, R., Giroux, R., Spiegelhalter,F., Charlton, S., Pinero, D. & Song, P. 2005. Polymerase chain reaction technology as analytical tool in agricultural biotechnology. *J. AOAC Int.* 88(1): 136-55.

Madan, M.L. 2005. Animal biotechnology: applications and economic implications in developing countries. *Rev. Sci. Tech.* 24(1): 127-39.

Mattick, J.S. 1994**.** Introns: evolution and function. *Current Opinion in Genetics and Development* 4: 823-831.

McEvoy, T.G., Alink, F.M., Moreira, V.C., Watt, R.G. & Powell, K.A. 2006. Embryo technologies and animal health - consequences for the animal following ovum pick-up, in vitro embryo production and somatic cell nuclear transfer. *Theriogenology*. 65(5): 926-42.

Meeusen, E.N., Walker, J., Peters, A., Pastoret, P.P. & Jungersen, G. 2007. Current status of veterinary vaccines. *Clin. Microbiol. Rev.* 20(3): 489-510.

Melo, E.O., Canavessi, A.M., Franco, M.M. & Rumpf, R. 2007. Animal trangenesis: state of the art and applications. *J. Appl. Genet.* 48(1): 47-61.

Michelini, E., Simoni, P., Cevenini, L., Mezzanotte, L. & Roda, A. 2008. New trends in bioanalytical tools for the detection of genetically modified organisms: an update. *Anal. Bioanal. Chem.* 392(3): 355-67.

Miki, B. & McHugh, S. 2004. Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J. Biotechnol.* 107(3): 193-232.

Miller, F.P., Vandome, A.F., McBrewster, J. 2009. *Central dogma of molecular biology*: History of molecular biology, Primary structure, DNA replication, Transcription (genetics). Alphascript Publishing.

Miraglia, M., Berdal, K.G., Brera, C., Corbisier, P., Holst-Jensen, A., Kok, E.J., Marvin, H.J., Schimmel, H., Rentsch, J., van Rie, J.P. & Zagon, J. 2004. Detection and traceability of genetically modified organisms in the food production chain. *Food Chem. Toxicol.* 42(7): 1157-80.

Morange. M. & Cobb, M. 2000. A history of molecular biology. 1st edition. Cambridge (MA), Harvard University Press.

Nestler, E.J. & Hyman, S.E. 2002**.** Regulation of gene expression. *Neuropsychopharmacology: The Fifth Generation of Progress.* Edited by Kenneth L. Davis, Dennis Charney, Joseph T. Coyle & Charles Nemeroff. American College of Neuropsychopharmacology.

Newell, C.A. 2000. Plant transformation techniques: Development and Application. *Mol. Biotechnol.* 16(1): 53-65.

Niemann, H., Kues, W. & Carnwath, J.W. 2005. Transgenic farm animals: present and future. *Rev. Sci. Tech.* 24(1): 285-98.

Official Journal of the European Union. Commission Regulation (EC) No 65/2004 of 14 January 2004 establishing a system for the development and assignment of unique identifiers for genetically modified organisms.

Okkema, P.G. & Krause, M. 2005. Transcriptional regulation. *WormBook*, ed. The *C.elegans* Research Community, WormBook.

Ovesna, J., Demnerova, K. & Pouchova, V. 2008. The detection of genetically modified organisms: an overview. *In* F. Toldrá (ed.), *Meat Biotechnology*. Springer Science+Business Media, LLC.

Padidam, M. 2003. Chemically regulated gene expression in plants. *Curr. Opin. Plant Biol.* 6(2): 169-77.

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C. & Evans, R.M. 1982. Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature* 300(5893): 611-5.

Park, F. 2007. Lentiviral vectors: are they the future of animal transgenesis? *Physiol. Genomics* 31(2): 159-73.

Park, S.J. & Cochran, J.R. 2009**.** Protein engineering and design. 1st Edition. Boca Raton (FL), CRC Press, Taylor and Francis Group.

Patrushev, L.I. & Minkevich, I.G. 2008. The problem of the eukaryotic genome size. *Biochemistry (Moscow)* 73(13): 1519-1552.

Pearson, H. 2006 Genetics: What is a gene? *Nature* 441: 398-401.

Plasmid genome database: http://www.genomics.ceh.ac.uk/plasmiddb/index.html

Prasanna, V., Prabha, T.N. & Tharanathan, R.N. 2007. Fruit ripening phenomena--an overview. *Crit. Rev. Food Sci. Nutr.* 47(1): 1-19.

Prins, M., Laimer, M., Noris, E., Schubert, J., Wassenegger, M. & Tepfer, M. 2008. Strategies for antiviral resistance in transgenic plants. *Mol. Plant Pathol.* 9(1): 73-83.

Ramakrishnan, V. 2002. Ribosome structure and the mechanisms of translation. Cell, 108: 557–572.

Ramsay, M. 1994. Yeast artificial chromosome cloning. *Mol. Biotechnol.* 1(2): 181-201.

Ranjekar, P.K., Patankar, A., Gupta, V., Bhatnagar, R., Bentur, J. & Ananda Kumar, P. 2003 genetic engineering of crop plants for insect resistance. *Current Science* 84(3).

Rath, D. & Johnston, L.A. 2008. Application and commercialization of flow cytometrically sex-sorted semen. *Reprod. Domest. Anim.* 43 Suppl 2: 338-46.

Rege, J.E.O. 1996. Biotechnology options for improving livestock production in developing countries, with special reference to sub-Saharan Africa. *In* S.H.B. Lebbie and E. Kagwini. *Small Ruminant Research and Development in Africa. Proceedings of the Third Biennial Conference of the African Small Ruminant Research Network, UICC, Kampala, Uganda, 5-9 December 1994*. Nairobi, Kenya, ILRI (International Livestock Research Institute).

Remund, K.M., Dixon, D.A., Wright, D.L. & Holden, L.R. 2001. Statistical considerations in seed purity testing for transgenic traits. *Seed Science Research* 11:101–119. A tool for calculating sample plans, provided by ISTA, is available online at: http://www.seedtest.org/upload/cms/user/seedcalc6.zip.

Reznikoff, W.S. 1992**.** The lactose operon-controlling elements: a complex paradigm. *Molecular Microbiology* 6(17): 2419-2422.

Rizvi, P.Q., Choudhury, R.A. & Ali, A. 2009. Recent advances in biopesticides. *In* M.S. Khan, A. Zaidi & J. Musarrat (eds.), Microbial Strategies for Crop Improvement. Springer-Verlag Berling.

Robl, J.M., Kasinathan, P., Sullivan, E., Kuroiwa, Y., Tomizuka, K. & Ishida, J. 2003. Artificial chromosome vectors and expression of complex proteins in transgenic animals. *Theriogenology*. 59(1): 107-13.

Robl, J.M., Wang, Z., Kasisnathan, P. & Kuroiwa, Y. 2007. Transgenic animal production and animal biotechnology. *Theriogenology* 67(1): 127-33.

Rocha, J.L., Pomp, D. & Van Vleck, L.D. 2002. QTL analysis in livestock. *Methods Mol. Biol.* 195: 311-46.

Rogan, D. & Babiuk, L.A. 2005. Novel vaccines from biotechnology. *Rev. Sci. Tech.* 24(1): 159-74.

Roh, J.Y., Choi, J.Y., Li, M.S., Jin, B.R. & Je, Y.H. 2007. Bacillus thuringiensis as a specific, safe, and effective tool for insect pest control. *J. Microbiol. Biotechnol.* 17(4): 547-59.

Rolinson, G.N. 1998. 40 years of beta-lactam research. *Journal of Antimicrobial Chemotherapy* 41: 589–603.

Ron, M. & Weller, J.I. 2007. From QTL to QTN identification in livestock-winning by points rather than knock-out: a review. *Anim. Genet.* 38(5): 429-39.

Schellenberg, M.J., Ritchie, D.B., & MacMillan., A.M. 2008. Pre-mRNA splicing: a complex picture in higher definition. *Trends in Biochemical Sciences* 33(6): 243-246.

Schmitt, B. & Henderson, L. 2005. Diagnostic tools for animal diseases. *Rev. Sci. Tech.* 24(1): 243-50.

Scupham, A.J., Bosworth, A.H., Ellis, W.R., Wacek, T.J., Albrecht, K.A. & Triplett, E.W. 1996. Inoculation with Sinorhizobium meliloti RMBPC-2 Increases Alfalfa Yield Compared with Inoculation with a Nonengineered Wild-Type Strain. *Appl. Environ/ Microbiol.* 62(11): 4260-4262.

Shewry, P.R., Jones, H.D. & Halford, N.G. 2008. Plant biotechnology: transgenic crops. *Adv. Biochem. Eng. Biotechnol.* 111: 149–186j.

Soccol, C.R., Vandenberghe, L.P., Woiciechowski, A.L., Thomaz-Soccol, V., Correia, C.T. & Pandey, A. 2003. Bioremediation: an important alternative for soil and industrial wastes clean-up. *Indian J. Exp. Biol.* 41(9): 1030-45.

Somers, D.A. & Makarevitch, I. 2004. Transgene integration in plants: poking or patching holes in promiscuous genomes? *Curr. Opin. Biotechnol.* (2): 126-31.

Stein, J.A. & Rodriguez-Cerezo, E. 2009. The global pipeline of new GM crops. Luxembourg, European Commission Joint Research Centre.

Stewart, C.N. Jr. 2005. Monitoring the presence and expression of transgenes in living plants. *Trends Plant Sci.* 10(8): 390-6.

Sullivan, E.J., Pommer, J. & Robl, J.M. 2008. Commercialising genetically engineered animal biomedical products. *Reprod. Fertil. Dev.* 20(1): 61-6.

Sullivan, E.J., Pommer. J. & Robl, J.M. 2008. Commercialising genetically engineered animal biomedical products., *Reprod, Fertil. Dev.* 20(1): 61-6.

Szewczyk, B., Hoyos-Carvajal, L., Paluszek, I., Skrzecz, I. & Lobo de Souza, M. 2006. Baculoviruses--re-emerging biopesticides. *Biotechnol. Adv.* 24(2): 143-60.

Tacket, C.O. 2009. Plant-based oral vaccines: results of human trials. *Curr. Top. Microbiol. Immunol.* 332: 103-17.

Tan, S., Evans, R. & Singh, B. 2006. Herbicidal inhibitors of amino acid biosynthesis and herbicide-tolerant crops. *Amino Acids* 30: 195–204.

Tan, S., Evans, R.R., Dahmer, M.L., Singh, B.K. & Shaner, D.L. 2005. Imidazolinone-tolerant crops: history, current status and future. *Pest Manag. Sci.* 61(3): 246-57.

Thomas, M.C. & Chiang, C-M. 2006. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* 41(3): 105-78.

Thorpe, T,A. 2007**.** History of plant tissue culture. *Mol. Biotecbnol.* 37(2): 169-80.

Tiwari, S., Verma, P.C., Singh, P.K. & Tuli, R. 2009. Plants as bioreactors for the production of vaccine antigens. *Biotechnol. Adv.* 27(4): 449-67.

Twyman, R.M., Schillberg, S. & Fischer, R. 2005. Transgenic plants in the biopharmaceutical market. *Expert Opin. Emerg. Drugs* 10(1): 185-218.

Vajta, G. & Gjerris, M. 2006. Science and technology of farm animal cloning: state of the art. *Anim. Reprod. Sci.* 92(3-4): 211-30.

Vicedo, B., Peñalver, R., Asins, M.J. & López, M.M. 1993. Biological Control of Agrobacterium *tumefaciens*, Colonization, and pAgK84 Transfer with Agrobacterium radiobacter K84 and the Tra Mutant Strain K1026. *Appl. Environ. Microbiol.* 59(1): 309-315.

Visel, A., Bristow, J. & Pennacchio, L.A. 2007. Enhancer identification through comparative genomics. *Semin. Cell. Dev. Biol*. 18(1): 140–152.

Voet, D. & Voet, J.G. 2004**.** *Biochemistry*. 3rd edition. Wiley & Sons.

Watson, J., Baker, T., Bell, S., Gann, A., Levine, M. & Losick, R. 2008. *Molecular biology of the gene*. Amsterdam, Addison-Wesley Longman.

Weimer, P.J. 1998. Manipulating ruminal fermentation: a microbial ecological perspective. *J. Anim. Sci.* 76(12): 3114-22.

Wheeler, M.B. 2007. Agricultural applications for transgenic livestock. *Trends Biotechnol.* 25(5): 204-10.

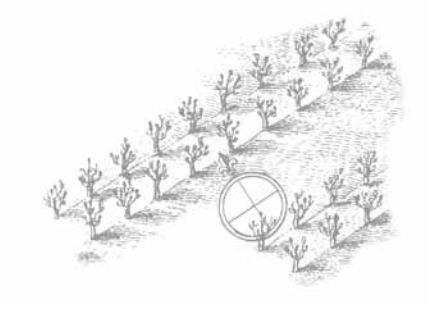
Wheeler, M.B., Walters, E.M. & Clark S.G. 2003. Transgenic animals in biomedicine and agriculture: outlook for the future. *Anim. Reprod. Sci.* 79(3-4): 265-89.

Whitelaw, C.B., Lillico, S.G. & King, T. 2008. Production of transgenic farm animals by viral vector-mediated gene transfer. *Reprod. Domest. Anim*. 43 Suppl 2: 355-8.

Williams, J.L 2005**.** The use of marker-assisted selection in animal breeding and biotechnology. *Rev. Sci. Tech*. 24(1): 379-91.

Yanofsky, C., Konan, K., Salsero, J. 1996. Some novel transcription attenuation mechanisms used by bacteria. *Biochemie* 78(11): 1017-1024(8).

Yu, W., Han, F. & Birchler, J.A. 2007. Engineered minichromosomes in plants. *Curr. Opin. Biotechnol.* 18(5): 425-31.



Design and layout: Pietro Bartoleschi and Arianna Guida (Studio Bartoleschi)
Cover illustrations elaborated from “l’Encyclopédie Diderot et d’Alembert”
Printed in Italy on ecological paper, Forest Stewardship Council (FSC) certified, May 2011

Biosafety Resource Book

MODULE A

MOLECULAR BIOLOGY AND GENETIC ENGINEERING

reviews the very basic scientific concepts and principles employed in producing GMOs, and provides a brief description of current and emerging uses of biotechnology in crops, livestock and fisheries.

For additional information please consult
www.fao.org/biotech
or contact
biotech-admin@fao.org