CHAPTER 14
Applications of recombinant DNA technology

Introduction

Biotechnology is not new. The making of beer, wine, bread, yoghurt and cheese was practised by ancient civilizations, such as the Babylonians, the Romans and the Chinese. Much, much later came vaccines, the production of basic chemicals (e.g. glycerol, citric acid, lactic acid) and the development of antibiotics. In each of these examples, existing properties of microorganisms were exploited. For example, *Penicillium* species naturally make penicillin. What the scientists have done is to increase the yield of penicillin by repeated rounds of mutation and selection, coupled with optimization of the growth medium. Similarly, sexual crosses between related plant species have created high-yielding and disease-resistant varieties of cereals. These improved cereals represent new combinations of genes and alleles already existing in wild strains.

With the development of gene manipulation techniques in the 1970s, there was a major paradigm shift. For the first time microorganisms could be made to synthesize compounds that they had never synthesized before, e.g. insulin production in *E. coli* (Johnson 1983). Soon all sorts of commercially or therapeutically useful proteins were being made in bacteria, principally *E. coli*, and thus the modern biotechnology industry was born. As the techniques developed for manipulating genes in bacteria were extended to plants and animals there was a concomitant expansion of the biotechnology industry to exploit the new opportunities being provided. Today there are many different facets to the commercial exploitation of gene manipulation techniques as shown in Fig. 14.1. Rather than discuss all these topics in detail, for that would take a book in itself, we have chosen to focus on six interdisciplinary themes that reflect both the successes achieved to date and the likely successes in the next decade.

Theme 1: Nucleic acid sequences as diagnostic tools

Introduction to theme 1

Nucleic acid sequences can be used diagnostically in two different ways. The first is to determine whether a particular, relatively long sequence is present in or absent from a test sample. A good example of such an application is the diagnosis of infectious disease. By choosing appropriate probes, one can ascertain in a single step which, if any, microorganisms are present in a sample. Alternatively, a search could be made for the presence of known antibiotic-resistance determinants so that an appropriate therapeutic regime can be instituted. In the second way in which sequences are used diagnostically, the objective is to determine the similarity of sequences from different individuals. Good examples of this approach are prenatal diagnosis of genetic disease and forensic profiling (‘DNA fingerprinting’).

Detection of sequences at the gross level

Imagine that a seriously ill individual has a disorder of the gastrointestinal tract. A likely cause is a microbial infection and there are a number of candidate organisms (Table 14.1). The question is, which organism is present and to which antibiotics is it susceptible? The sooner one has an answer to these questions, the sooner effective therapy can begin. Traditionally, in such a case, a stool specimen would be cultured on a variety of different media and would be examined microscopically and tested with various immunological reagents. A simpler approach is to test the sample with a battery of probes and determine which, if any, hybridize in a simple dot-blot assay. With such a simple format it is possible to vary the stringency of the hybridization reaction to accommodate any sequence differences that might
Applications of recombinant DNA technology

exist between the probe and the target. The downside of this approach is that if one wishes to test the sample with 10 different probes, then 10 different dot blots are required; otherwise there is no way of determining which probe has bound to the target.

Another disadvantage of this approach is that sufficient target DNA must be present in the sample to enable a signal to be detected on hybridization. Both of these problems can be overcome by using the polymerase chain reaction (PCR).

In a traditional dot-blot assay, the sample DNA is immobilized on a membrane and hybridized with a selection of labelled probes. An alternative is a ‘reverse dot blot’, where the probes are immobilized on the membrane and hybridized with the sample. In this way, only one hybridization step is required, because each probe occupies a unique position on the membrane (Fig. 14.2). For this approach to work, the sample DNA needs to be labelled and this can be achieved using the PCR. This has the added

**Table 14.1** Pathogens causing infection of the gastrointestinal tract.

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Viruses</th>
<th>Protozoa</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Salmonella</em> spp.</td>
<td>Rotavirus</td>
<td><em>Entamoeba</em> histolytica</td>
</tr>
<tr>
<td><em>Shigella</em> spp.</td>
<td>Enterovirus</td>
<td><em>Giardia</em> lamblia</td>
</tr>
<tr>
<td><em>Yersinia enterocolitica</em></td>
<td>Enterotoxigenic <em>E. coli</em></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium difficile</em></td>
<td><em>Clostridium welchii</em></td>
<td></td>
</tr>
<tr>
<td><em>Campylobacter</em> spp.</td>
<td><em>Aeromonas</em> spp.</td>
<td></td>
</tr>
<tr>
<td><em>Vibrio</em> spp.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 14.1** The different ways that recombinant DNA technology has been exploited.
benefit of amplifying the sample, thereby minimizing the amount of target DNA required. The downside of this approach is that the amplification step needs to be done with multiple pairs of primers (multiplex PCR), one for each target sequence. Although multiplex PCR is now an established method, considerable optimization is needed for each application (Elmfors et al. 2000). The major problems encountered are poor sensitivity or specificity and/or preferential amplification of certain specific targets. The presence of more than one primer pair in the reaction increases the chance of obtaining spurious amplification products because of the formation of primer dimers. Such non-specific products may be amplified more efficiently than the desired target. Clearly, the design of each primer pair is crucial to avoiding this problem. Another important feature is that all primer pairs should enable similar amplification efficiencies for their respective targets.

Despite the difficulties noted above, multiplex PCR has been used successfully in the diagnosis of infectious diseases. For example, Heredia et al. (1996) have used the method to simultaneously examine blood for the presence of HIV-1, HIV-2, human T-cell lymphotropic virus (HTLV)-I and HTLV-II. Similarly, Grondahl et al. (1999) have used the method to identify which of nine different organisms is responsible for respiratory infections. Once the clinical microbiologist knows the identity of a microorganism in a specimen, he/she can select those antibiotics that might be effective, provided that the organism in question does not carry multiple drug-resistance determinants. The presence of these determinants can also be established using multiplex PCR. Alternatively, if a virus causes the infection, one can monitor the progress of the infection by using quantitative PCR (see p. 23). This approach has been used to monitor cytomegalovirus infec-

---

**Fig. 14.2** Comparison of conventional dot blot assay with reverse dot blot method. Note that in the latter there is only a single hybridization reaction, regardless of the number of probes used, whereas in the former each probe has to be tested in a separate hybridization step.
Applications of recombinant DNA technology

It is worth noting that the development of multiplex PCR technology is being facilitated by the rapid progress in the sequencing of microbial genomes (for an up-to-date list, see http://igweb.integratedgenomics.com/GOLD/), since the data generated enable species-specific genes or sequences to be identified.

**Comparative sequence analysis: single-nucleotide polymorphisms (SNPs)**

In prenatal diagnosis of genetic disorders, there is a need to determine which alleles of a particular locus are being carried by the fetus, i.e. is the fetus homozygous for the normal or the deleterious allele or is it heterozygous? In forensic DNA profiling the requirement is to match DNA from the perpetrator of a crime with that of a suspect. In each case, a definitive answer could be obtained by sequencing the relevant samples of DNA. While this is possible, it is not practicable for mass screening. An alternative could be to detect hybridization of specific probes. This has been done for the detection of sickle-cell anaemia by Conner et al. (1983). They synthesized two 19-mer oligonucleotides, one of which was complementary to the amino-terminal region of the normal β-globin (β<sup>A</sup>) gene and the other of which was complementary to the sickle-cell β-globin gene (β<sup>S</sup>). These oligonucleotides were radiolabelled and used to probe Southern blots. Under appropriate conditions, the probes could distinguish the normal and mutant alleles. The DNA from normal homozygotes only hybridized with the β<sup>A</sup> probe and DNA from sickle-cell homozygotes only hybridized with the β<sup>S</sup> probe. DNA of heterozygotes hybridized with both probes. These experiments, therefore, showed that oligonucleotide hybridization probes can discriminate between a fully complementary DNA and one containing a single mismatched base. Similar results have been obtained (Fig. 14.3) with a point mutation in the α-antitrypsin gene, which is implicated in pulmonary emphysema (Cox et al. 1985).

The single-base changes that occur in the two clinical examples just quoted are examples of single-nucleotide polymorphisms (SNPs, pronounced ‘snips’). Many such polymorphisms occur throughout an entire genome and in humans the frequency is about once every 1000 bases. Their distribution is not entirely random. SNPs that alter amino acid

---

**Fig. 14.3** Schematic representation of the use of oligonucleotide probes to detect the normal α<sub>1</sub>-antitrypsin gene (M) and its Z variant. Human DNA obtained from normal (MM), heterozygous (MZ) and homozygous variant (ZZ) subjects is digested with a restriction endonuclease, electrophoresed and fragments Southern blotted on to a nitrocellulose membrane. The patterns shown were obtained on autoradiography of the filter following hybridization with either the normal (M-specific) or variant (Z-specific) probe.
sequences occur much less frequently than silent substitutions and SNPs in non-coding regions (Cargill et al. 1999). However, they are stably inherited. In some instances, these polymorphisms result in the creation or elimination of a restriction-enzyme site and this can be used diagnostically. A classical example is sickle-cell anaemia, where the 6th amino acid of β-globin is changed from glutamate to valine. (a) Location of recognition sequences for restriction endonuclease MstII in and around the β-globin gene. The change of A → T in codon 6 of the β-globin gene destroys the recognition site (CCTGAGG) for MstII as indicated by the asterisk. (b) Electrophoretic separation of MstII-generated fragments of human control DNAs (AA, AS, SS) and DNA from amniocytes (Amn). After Southern blotting and probing with a cloned β-globin DNA probe, the normal gene and the sickle gene can be clearly distinguished. Examination of the pattern for the amniocyte DNA indicates that the fetus has the genotype βAβS, i.e. it is heterozygous.

Fig. 14.4 Antenatal detection of sickle-cell genes. Normal individuals are homozygous for the βA allele, while sufferers from sickle-cell anaemia are homozygous for the βS allele. Heterozygous individuals have the genotype βAβS. In sickle-cell anaemia, the 6th amino acid of β-globin is changed from glutamate to valine. (a) Location of recognition sequences for restriction endonuclease MstII in and around the β-globin gene. The change of A → T in codon 6 of the β-globin gene destroys the recognition site (CCTGAGG) for MstII as indicated by the asterisk. (b) Electrophoretic separation of MstII-generated fragments of human control DNAs (AA, AS, SS) and DNA from amniocytes (Amn). After Southern blotting and probing with a cloned β-globin DNA probe, the normal gene and the sickle gene can be clearly distinguished. Examination of the pattern for the amniocyte DNA indicates that the fetus has the genotype βAβS, i.e. it is heterozygous.

Many of the SNPs that cause genetic diseases do not lie within a restriction-enzyme site, as is the case in sickle-cell anaemia. However, the restriction fragment length polymorphisms (RFLPs) caused by other SNPs can be used diagnostically, as shown in Fig. 14.5. In this case there is a close linkage between a polymorphic restriction site and the locus of interest and this can be used to trace the inheritance of the gene. When this approach was first developed, a major limitation was the availability of suitable polymorphic markers. Following the sequencing of the entire human genome, an encyclopedia of SNPs is being created (currently 1.4 million) and this will greatly facilitate association studies. Indeed, this approach is now being used to match patients with appropriate drugs (see p. 292). Again, when this approach was developed, RFLPs were detected by genomic Southern blotting. This laborious step can be bypassed by use of the PCR. Enough DNA can be synthesized in the PCR reaction so that after digestion with the restriction enzyme and electrophoresis, the DNA bands are directly visible following staining with ethidium bromide. Another advantage of PCR is that the gel step can be omitted altogether and the SNPs detected directly, using microarrays (‘DNA chips’) (p. 116).
Applications of recombinant DNA technology

Variable number tandem repeat (VNTR) polymorphisms

In higher eukaryotes, genes and their associated introns occupy only a small proportion of the genome. The intergenic DNA, which makes up the majority of the genome, is composed of a mixture of unique sequences and repetitive sequences. Many of the repetitive DNA sequence elements are arranged in tandem and are known as satellite DNA. Three types of satellite DNA can be distinguished on the basis of the level of repetition and the repeat-unit length (Table 14.2). Not only is satellite DNA dispersed throughout the genome, it is highly variable and provides a valuable tool for genetic individualization. An example of this is shown schematically in Fig. 14.6. In this case the RFLPs detected are due to variations in the number of repeat units (VNTR polymorphisms) between restriction sites, rather than changes in the location of the restriction sites, as discussed in the previous section.

VNTR polymorphisms are of importance to clinical geneticists because a number of important hereditary diseases are associated with alterations in the degree of repetition of microsatellites (for reviews see Bowater & Wells 2000, Gutekunst et al. 2000, Usdin & Grabczyck 2000). Probably the best example of such a disease is Huntington’s chorea, which is caused by an expansion of a CAG trinucleotide repeat in exon 1 of the gene coding for a protein of unknown function, which has been named huntingtin. Expansion beyond 40 repeat units correlates with the onset and progression of the disease (for review see Reddy et al. 1999).

Forensic applications of VNTRs

The existence of VNTR polymorphisms is of great utility in paternity testing and criminal investigations, since they allow ready comparison of DNA samples in the absence of detailed genetic information by the generation of a DNA profile or fingerprint. In principle, a multilocus DNA fingerprint can be generated either by the simultaneous application of several probes, each one specific for a particular locus, or by applying a single DNA probe that simultaneously detects several loci. When DNA profiling was first developed (Jeffreys et al. 1985a), multilocus probes were used and these were derived from a tandemly repeated sequence within an intron of the myoglobin gene (Fig. 14.7). These probes can hybridize to other autosomal loci – hence their utility. The first criminal court case to use DNA fingerprinting was in Bristol, UK, in 1987, when a link was shown between a burglary and a rape. In the following
year, DNA-fingerprinting evidence was used in the USA. It is worth noting that DNA evidence has been used to prove innocence as well as guilt (Gill & Werrett 1987).

Multilocus probes can also be used to prove or disprove paternity and a unique example, which was part of an immigration test case, is shown in Box 14.1. The technique also has application in many other areas, such as pedigree analysis in cats and dogs (Jeffreys & Morton 1987), confirming cell-line authenticity in animal cell cultures (Devor et al. 1988, Stacey et al. 1992) and monitoring the behaviour and breeding success of bird populations (Burke & Bruford 1987).

In criminal cases, a major disadvantage of multilocus probes is the complexity of the DNA fingerprint provided. Showing innocence is easy, but proving identity is fraught with problems. The issue boils down to calculations of the probability that two profiles match by chance as opposed to having come from the same person (Lewontin & Hartl 1991). For this reason forensic scientists have moved to the use of single-locus probes and an example is shown in Fig. 14.8. The latest variation of the technique, introduced in the UK in 1999, targets 10 distinct loci, and the likelihood of two people sharing the same profile is less than one in a billion (thousand million). Chance matches are even less likely in the USA, where the FBI routinely examines 13 VNTR loci. Another advantage of single-locus probes is

### Table 14.2 Classification of satellite DNA.

<table>
<thead>
<tr>
<th>Type of repeat</th>
<th>Degree of repetition per locus</th>
<th>Number of loci</th>
<th>Repeat-unit length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite</td>
<td>$10^3$–$10^7$</td>
<td>One to two</td>
<td>1000–3000</td>
</tr>
<tr>
<td>Minisatellite</td>
<td>$10^3$–$10^6$</td>
<td>Thousands per chromosome</td>
<td>9–100</td>
</tr>
<tr>
<td>Microsatellite</td>
<td>$10^2$–$10^4$</td>
<td>Up to $10^5$</td>
<td>1–6</td>
</tr>
</tbody>
</table>

Multilocus probes can also be used to prove or disprove paternity and a unique example, which was part of an immigration test case, is shown in Box 14.1. The technique also has application in many other areas, such as pedigree analysis in cats and dogs (Jeffreys & Morton 1987), confirming cell-line authenticity in animal cell cultures (Devor et al. 1988, Stacey et al. 1992) and monitoring the behaviour and breeding success of bird populations (Burke & Bruford 1987).

In criminal cases, a major disadvantage of multilocus probes is the complexity of the DNA fingerprint provided. Showing innocence is easy, but proving identity is fraught with problems. The issue boils down to calculations of the probability that two profiles match by chance as opposed to having come from the same person (Lewontin & Hartl 1991). For this reason forensic scientists have moved to the use of single-locus probes and an example is shown in Fig. 14.8. The latest variation of the technique, introduced in the UK in 1999, targets 10 distinct loci, and the likelihood of two people sharing the same profile is less than one in a billion (thousand million). Chance matches are even less likely in the USA, where the FBI routinely examines 13 VNTR loci. Another advantage of single-locus probes is

![Fig. 14.6 Restriction fragment length polymorphisms caused by a variable number of tandem repeats between the two HindIII restriction sites. The upper part of the diagram shows the DNA structure for three different individuals. The lower part of the diagram shows the pattern obtained on electrophoresis of HindIII cut DNA from the three individuals after hybridization with a probe complementary to the sequence shown in pink.](image)

![Fig. 14.7 Probes used for DNA fingerprinting.](image)

- **Core sequence:** A
- **Probe 33.6:** G G A G G T G G C A G G A G G
- **Probe 33.15:** [G G C T G G A G G]$_{18}$
- **Probe 33.5:** (A G A G G T G G C A G G T G G)$_{29}$
- **Probe 33.5:** (G G C G T G G C A G G A G G)$_{14}$
In 1984, a Ghanaian boy was refused entry into Britain because the immigration authorities were not satisfied that the woman claiming him as her son was in fact his mother. Analysis of serum proteins and erythrocyte antigens and enzymes showed that the alleged mother and son were related but could not determine whether the woman was the boy’s mother or aunt. To complicate matters, the father was not available for analysis nor was the mother certain of the boy’s paternity. DNA fingerprints from blood samples taken from the mother and three children who were undisputedly hers, as well as the alleged son, were prepared by Southern blot hybridization to two of the mini-satellite probes shown in Fig. B14.1. Although the father was absent, most of his DNA fingerprint could be reconstructed from paternal-specific DNA fragments present in at least one of the three undisputed siblings but absent from the mother. The DNA fingerprint of the alleged son contained 61 scorable fragments, all of which were present in the mother and/or at least one of the siblings. Analysis of the data showed the following.

- The probability that either the mother or the father by chance possess all 61 of the alleged son’s bands is $7 \times 10^{-22}$. Clearly the alleged son is part of the family.
- There were 25 maternal-specific fragments in the 61 identified in the alleged son and the chance probability of this is $2 \times 10^{-5}$. Thus the mother and alleged son are related.
- If the alleged mother of the boy in question is in fact a maternal aunt, the chance of her sharing the 25 maternal-specific fragments with her sister is $6 \times 10^{-6}$.

When presented with the above data (Jeffreys et al. 1985b), as well as results from conventional marker analysis, the immigration authorities allowed the boy residence in Britain. In a similar kind of investigation, a man originally charged with murder was shown to be innocent (Gill & Werrett 1987).

**Box 14.1 Use of DNA fingerprinting in an immigration test case**

![Fig. B14.1](image)
that it is possible to convert the DNA profile into a numerical format. This enables a database to be established and all new profiles can be matched to that database.

Detection of VNTR polymorphisms requires that an adequate amount of DNA be present in the test sample. This is not a problem in paternity disputes, but can be an issue in forensic testing. With the advent of single-locus probes, the amount of DNA required is much less of an issue, since the test loci in the sample can be amplified by PCR. As a result, it now is possible to type DNA from a face-mask worn by a bank robber, a cigarette-butt discarded at the scene of a crime or the back of a stamp on an envelope used to send a ‘poison-pen’ or blackmail letter.

**Historical genetics**

Just as multilocus probes have been used for many applications other than in crime testing, so too have single-locus probes. A good example is the determination of the parentage of grapevines used for wine making. Grapevines are propagated vegetatively, so that individual vines of a cultivar are genetically identical to each other and to the single original seedling from which the cultivar originated. Most of the cultivars in existence in north-eastern France are centuries old and their provenance was not known. However, using 17 microsatellite loci, Bowers et al. (1999) were able to show that 16 of the common cultivars have genotypes consistent with their being progeny of a single pair of grape cultivars that were widespread in the region in the Middle Ages.

VNTRs can be found in mitochondrial DNA, as well as nuclear DNA, and these have particular applications. The reasons for this are threefold. First, mitochondrial sequences are passed from mother to child in the egg. Thus, brothers and sisters have identical mitochondrial DNA. Secondly, the small size of mitochondrial DNA (16–20 kb) means that there is less scope for variability, but this is more than compensated for by the copy number (~10 000 copies per cell). That is, mitochondrial DNA is naturally amplified. Thirdly, in very old or degraded specimens, the nuclear DNA may be totally decomposed, but mitochondrial DNA can still be recovered. For example, mitochondrial-DNA analysis was used to confirm that skeletons found in Ekaterinburg, Russia, were the remains of the last tsar and his family (Gill et al. 1994). A similar analysis showed that an individual living in Cheddar Gorge in the UK was related to a Stone Age individual whose skull was found nearby. Since bones are more likely than soft tissue to survive in the event of major accidents that involve fire, mitochondrial DNA analysis will play an increasingly important role in identifying victims. Indeed, such an analysis was done in the UK following the 1999
Paddington train crash, in which one carriage was completely incinerated.

Just as the mitochondrion is transmitted maternally, the Y chromosome is transmitted only through male descendants. Because there is only a single copy of the Y chromosome in normal diploid cells, recombination between different Y chromosomes does not occur. Any changes that do occur in the Y chromosome from generation to generation must arise from DNA rearrangements or by accumulation of random mutations. That is, the Y chromosome should be highly conserved. Sykes and Irven (2000) obtained proof of this. They probed a randomly ascertained sample of males with the surname ‘Sykes’ with four Y-chromosome microsatellites and found that half of them had the same Y haplotype. This suggests that all those with the same haplotype have a common ancestor, even though conventional genealogical analysis suggests otherwise.

### Theme 2: New drugs and new therapies for genetic diseases

#### Introduction to theme 2: proteins as drugs

One of the earliest commercial applications of gene-manipulation techniques was the production in bacteria of human proteins with therapeutic applications. Not surprisingly, the first such products were recombinant versions of proteins already used as therapeutics: human growth hormone and insulin. Prior to the advent of genetic engineering, human growth hormone was produced from pituitary glands removed from cadavers. Not only did this limit the supply of the hormone but, in some cases, it resulted in recipients contracting Creutzfeld–Jakob syndrome. The recombinant approach resulted in unlimited supplies of safe material. This safety aspect has been extended to various clotting factors that were originally isolated from blood but now carry the risk of HIV infection. As the methods for cloning genes became more and more sophisticated, an increasing number of lymphokines and cytokines were identified and significant amounts of them produced for the first time. A number of these were shown to have therapeutic potential and found their way into clinical practice (Table 14.3).

The first generation of protein drugs were exact copies of the human molecules but protein engineering is now being used to develop second-generation molecules with improved properties (see theme 4, p. 299). More recently, macromodifications have been made to proteins, as exemplified by the recently approved drug (Table 14.3) for rheumatoid arthritis, which consists of the tumour necrosis factor receptor fused to the Fc portion of human IgG1.

#### Transgenic animals and plants as bioreactors: ‘pharming’

‘Pharming’ is the play on words that refers to the use of transgenic animals and plants to produce recombinant therapeutic proteins. As discussed in earlier chapters, recombinant-protein synthesis in animal cells has a number of advantages over microbial expression systems, the most important of which is the authentic post-translational modifications that are performed in animal cells. However, large-scale culture of animal cells is expensive, due to the amount of medium and serum required and the necessity for precise and constant growth conditions. The production of growth hormone in the serum of transgenic mice (Palmiter et al. 1982a) (see p. 209) provided the first evidence that recombinant proteins could be produced, continuously, in the body fluids of animals. Five years later, several groups reported the secretion of recombinant proteins in mouse milk. In each case, this was achieved by joining the transgene to a mammary-specific promoter, such as that from the casein gene. The first proteins produced in this way were sheep β-lactoglobulin (Simons et al. 1987) and human tissue-plasminogen activator (tPA) (Gordon et al. 1987, Pittius et al. 1988). There have been over 100 such reports since these early experiments, and a selection is listed in Table 14.4.

Although proteins can be produced at high concentrations in mouse milk (e.g. 50 ng/ml for tPA), the system is not ideal, due to the small volume of milk produced. Therefore, other animals, such as sheep and goats, have been investigated as possible bioreactors. Such animals not only produce large volumes of milk, but the regulatory practices regarding the use of their milk are more acceptable. An early success was Tracy, a transgenic ewe.
### Table 14.3  Some recombinant proteins that are used therapeutically.

<table>
<thead>
<tr>
<th>Year</th>
<th>Product</th>
<th>Clinical indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>Human insulin</td>
<td>Diabetes</td>
</tr>
<tr>
<td>1985</td>
<td>Human growth hormone</td>
<td>Pituitary dwarfism</td>
</tr>
<tr>
<td>1986</td>
<td>Hepatitis B vaccine</td>
<td>Prevention of hepatitis B infection</td>
</tr>
<tr>
<td>1986</td>
<td>Interferon-α&lt;sub&gt;2b&lt;/sub&gt;</td>
<td>Haemophillia B</td>
</tr>
<tr>
<td>1987</td>
<td>Tissue plasminogen activator</td>
<td>Acute myocardial infarction</td>
</tr>
<tr>
<td>1989</td>
<td>Erythropoietin</td>
<td>Anaemia associated with chronic renal failure</td>
</tr>
<tr>
<td>1990</td>
<td>Interferon-γ&lt;sub&gt;1b&lt;/sub&gt;</td>
<td>Chronic granulomatous disease</td>
</tr>
<tr>
<td>1991</td>
<td>Granulocyte–macrophage-colony-stimulating factor</td>
<td>Bone-marrow transplant</td>
</tr>
<tr>
<td>1991</td>
<td>Granulocyte-colony-stimulating factor</td>
<td>Chemotherapy-induced neutropenia</td>
</tr>
<tr>
<td>1992</td>
<td>Human interleukin-2</td>
<td>Renal-cell carcinoma</td>
</tr>
<tr>
<td>1993</td>
<td>Factor VIII</td>
<td>Haemophillia A</td>
</tr>
<tr>
<td>1994</td>
<td>Glucocerebrosidase</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>1996</td>
<td>Interferon-β&lt;sub&gt;1a&lt;/sub&gt;</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>1997</td>
<td>Factor IX</td>
<td>Haemophillia B</td>
</tr>
<tr>
<td>1997</td>
<td>Consensus interferon</td>
<td>Chronic HCV infection</td>
</tr>
<tr>
<td>1997</td>
<td>Platelet growth factor</td>
<td>Chemotherapy-induced thromboctopenia</td>
</tr>
<tr>
<td>1997</td>
<td>Platelet-derived growth factor β</td>
<td>Lower-extremity diabetic ulcers</td>
</tr>
<tr>
<td>1998</td>
<td>Tumour necrosis factor receptor linked to Fc portion of human IgG1</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>1999</td>
<td>Glucagon</td>
<td>Hypoglycaemia</td>
</tr>
<tr>
<td>1999</td>
<td>Factor VIIa</td>
<td>Haemophillia</td>
</tr>
</tbody>
</table>

### Table 14.4  Some recombinant proteins produced in the secretions of animal bioreactors.

<table>
<thead>
<tr>
<th>System</th>
<th>Species</th>
<th>Product</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>Mouse</td>
<td>Sheep β-lactoglobulin</td>
<td>Simons et al. 1987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human tissue-plasminogen activator</td>
<td>Gordon et al. 1987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human urokinase</td>
<td>Meade et al. 1990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human growth hormone</td>
<td>Deviney et al. 1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human fibrinogen</td>
<td>Prunkard et al. 1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human nerve growth factor</td>
<td>Coulibaly et al. 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spider silk</td>
<td>Karatzas et al. 1999</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Mouse</td>
<td>Human erythropoietin</td>
<td>Massoud et al. 1996</td>
</tr>
<tr>
<td>Sheep</td>
<td>Goat</td>
<td>Human α&lt;sub&gt;1&lt;/sub&gt;-antitrypsin</td>
<td>Wright et al. 1991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Human tissue-plasminogen activator</td>
<td>Ebert et al. 1991</td>
</tr>
<tr>
<td>Blood serum</td>
<td>Rabbit</td>
<td>Human α&lt;sub&gt;1&lt;/sub&gt;-antitrypsin</td>
<td>Massoud et al. 1991</td>
</tr>
<tr>
<td></td>
<td>Pig</td>
<td>Recombinant antibodies</td>
<td>Lo et al. 1991, Weidle et al. 1991</td>
</tr>
<tr>
<td>Urine</td>
<td>Mouse</td>
<td>Human growth hormone</td>
<td>Kerr et al. 1998</td>
</tr>
<tr>
<td>Semen</td>
<td>Mouse</td>
<td>Human growth hormone</td>
<td>Dyck et al. 1999</td>
</tr>
</tbody>
</table>
producing extremely high levels (30 g/l) of human α1-antitrypsin (AAT) in her milk (Wright et al. 1991). Artificially inseminated eggs were microinjected with a DNA construct containing an AAT gene fused to a β-lactoglobulin promoter. These eggs were implanted into surrogate mothers, of which 112 gave birth. Four females, including Tracy, and one male were found to have incorporated intact copies of the gene and all five developed normally. Over the lactation period, sheep can produce 250–800 l of milk, so the production potential is significant.

Using similar protocols, Ebert et al. (1991) have demonstrated the production of a variant of human tPA in goat milk. Of 29 offspring, one male and one female contained the transgene. The transgenic female underwent two pregnancies and one out of five offspring was transgenic. Milk collected over her first lactation contained only a few milligrams of tPA per litre, but improved expression constructs have since resulted in an animal generating several grams per litre of the protein. Recombinant human antithrombin III, which is used to prevent blood clots forming in patients that have undergone heart-bypass operations, was the first protein expressed in transgenic animal milk to reach commercial production, and is currently marketed by Genzyme Transgenics Corporation.

The production of foreign proteins in secreted body fluids has the obvious advantage that transgenic animals can be used as a renewable source of the desirable molecule. In addition to milk, other production systems have been investigated, including serum (Massoud et al. 1991), semen (Dyck et al. 1999) and urine (Kerr et al. 1998). In each case, an important consideration is whether the protein is stable and whether it folds and assembles correctly. The assembly of complex proteins comprising up to three separate polypeptides has been demonstrated in milk, e.g. fibrinogen (Prunkard et al. 1996), collagen (John et al. 1999) and various immunoglobulins (e.g. Castilla et al. 1998). Other abundantly secreted fluids that are likely to be exploited for recombinant-protein expression in the future include the albumen of hens’ eggs and silkworm cocoons. There has already been some success with the latter, using both microinjection (e.g. Nagaraju et al. 1996) and infection of silkworm larvae with baculovirus vectors (Tamura et al. 1999, Yamao et al. 1999) (see Chapter 10). The use of animals as bioreactors has been extensively reviewed (Clark 1998, Rudolph 1999, Wall 1999, Houdebine 2000).

**Plants as bioreactors**

Plants are a useful alternative to animals for recombinant-protein production because they are inexpensive to grow and scale-up from laboratory testing to commercial production is easy. Therefore, there is much interest in using plants as production systems for the synthesis of recombinant proteins and other specialty chemicals. There is some concern that therapeutic molecules produced in animal expression systems could be contaminated with small quantities of endogenous viruses or prions, a risk factor that is absent from plants. Furthermore, plants carry out very similar post-translational modification reactions to animal cells, with only minor differences in glycosylation patterns (Cabanes-Macheteau et al. 1999). Thus plants are quite suitable for the production of recombinant human proteins for therapeutic use.

A selection of therapeutic proteins that have been expressed in plants is listed in Table 14.5. The first such report was the expression of human growth hormone, as a fusion with the *Agrobacterium* nopaline synthase enzyme, in transgenic tobacco and sunflower (Barta et al. 1986). Tobacco has been the most frequently used host for recombinant-protein expression although edible crops, such as rice, are now becoming popular, since recombinant proteins produced in such crops could in principle be administered orally without purification. The expression of human antibodies in plants has particular relevance in this context, because the consumption of plant material containing recombinant antibodies could provide passive immunity (i.e. immunity brought about without stimulating the host immune system). Antibody production in plants was first demonstrated by Hiatt et al. (1989) and During et al. (1990), who expressed full-size immunoglobulins in tobacco leaves. Since then, many different types of antibody have been expressed in plants, predominantly tobacco, including full-size immunoglobulins, Fab fragments and single-chain Fv fragments (scFvs). For example, a fully humanized antibody against herpes simplex virus-2 (HSV-2) has been expressed...
in soybean (Zeitlin et al. 1998). Even secretory IgA (sIgA) antibodies, which have four separate polypeptide components, have been successfully expressed in plants. This experiment involved the generation of four separate transgenic tobacco lines, each expressing a single component, and the sequential crossing of these lines to generate plants in which all four transgenes were stacked (Ma et al. 1995). Plants producing recombinant sIgA against the oral pathogen Streptococcus mutans have been generated (Ma et al. 1998), and these plant-derived antibodies (‘plantibodies’) have recently been commercially produced as the drug CaroRx™, marketed by Planet Biotechnology Inc. A number of other biotechnology companies are bringing antibody-expressing transgenic plants into commercial production (see Fischer & Emans 2000).

The impact of genomics

Many of the drugs currently on the market treat the symptoms of the disease rather than the cause of the disease. This is analogous to reversing a mutant phenotype by selecting a mutation at a second site. Not surprisingly, many of these drugs have side-effects quite separate from those (e.g. toxicity) caused by their metabolism. Now that the first drafts of the human genome sequence are available, it will be possible to convert disease phenotypes into nucleotide changes in specific genes (Bailey et al. 2001). These genes will then become targets for new small-molecule drugs. Already drugs have been developed based on such genotype–phenotype correlations. For example, Wettereau et al. (1998) identified a molecule that normalizes atherogenic lipoprotein levels caused by a genetic deletion of a microsomal triglyceride transfer protein.

The identification of genetic changes associated with particular disease phenotypes offers a number of novel approaches to the development of therapies. As well as using such changes as novel targets for small-molecule drug design, there is an opportunity to use the techniques described in Chapter 11 to generate animals with the exact same genetic defect and which can be used as models to test new drug candidates (disease modelling). Furthermore, where drugs cannot be developed to treat a particular disorder, there might be an opportunity to correct the disease by further modification to the genome (gene therapy). Finally, it is likely that, in the near future, transgenic animals could be used to provide healthy organs for humans requiring transplants (xenotransplantation) (Box 14.2). These topics are discussed in more detail below.

Transgenic animals as models of human disease

Mammals have been used as models for human disease for many years, since they can be exploited to
Applications of recombinant DNA technology

Box 14.2 Xenotransplantation

Transplantation is widely used to treat organ failure but there is a shortage of organ donors resulting in long waiting times and many unnecessary deaths. In the future, transgenic animals could be used to supply functional organs to replace failing human ones. This process is termed xenotransplantation. There is vigorous debate concerning the ethics of xenotransplantation but, ethics aside, the technique remains limited by the phenomenon of hyperacute rejection, which is caused by the host immune system. Hyperacute rejection is dependent both on antibodies raised against the foreign organ and the activation of the host complement system. In both cases, the major trigger for rejection appears to be a disaccharide group (Gal-α(1,3)-Gal) which is present in pigs but not in primates (Cooper et al. 1994, Sandrin et al. 1994).

Transgenic strategies have been investigated to avoid hyperacute rejection, including the expression of complement-inactivating protein on the cell surface (Cozzi & White 1995, reviewed by Pease et al. 1997), the expression of antibodies against the disaccharide group (Vanhove et al. 1998) and attempts to inhibit the expression of α(1,3) galactosyltransferase, the enzyme that forms this particular carbohydrate linkage, an enzyme that is present in pigs but not in primates. In the latter case, the simplest strategy would be to knock out the gene by homologous recombination. Gene targeting has not yet been achieved in pigs, although the success of gene targeting/nuclear transfer in sheep (Chapter 11) suggests that knockout pigs could be produced in the next few years. Alternative procedures include introducing genes encoding other carbohydrate-metabolizing enzymes, so that the Gal-α(1,3)-Gal groups are modified into some other less immunogenic moiety (e.g. Sandrin et al. 1995, Cohn et al. 1997, Osman et al. 1997). Once hyperacute rejection has been overcome, there may be further problems, including delayed rejection, involving natural killer cells and macrophages (Bach et al. 1996), and the requirement for T-cell tolerance (Bracy et al. 1998, Kozlowski et al. 1998, Yang et al. 1998). There are also concerns that endogenous pig retroviruses could be activated following transplantation, perhaps even recombining with human retroviruses to produce potent new hybrids with unknown properties. Detailed studies have so far shown no evidence of such a phenomenon (e.g. Heneine et al. 1998, Patience et al. 1998). For recent reviews on the prospects of xenotransplantation (particularly porcine-to-human), the reader should consult Lambrigts et al. (1998), Sandrin and McKenzie (1999) and Logan (2000).

Transplantation is widely used to treat organ failure but there is a shortage of organ donors resulting in long waiting times and many unnecessary deaths. In the future, transgenic animals could be used to supply functional organs to replace failing human ones. This process is termed xenotransplantation. There is vigorous debate concerning the ethics of xenotransplantation but, ethics aside, the technique remains limited by the phenomenon of hyperacute rejection, which is caused by the host immune system. Hyperacute rejection is dependent both on antibodies raised against the foreign organ and the activation of the host complement system. In both cases, the major trigger for rejection appears to be a disaccharide group (Gal-α(1,3)-Gal) which is present in pigs but not in primates (Cooper et al. 1994, Sandrin et al. 1994).

Transgenic strategies have been investigated to avoid hyperacute rejection, including the expression of complement-inactivating protein on the cell surface (Cozzi & White 1995, reviewed by Pease et al. 1997), the expression of antibodies against the disaccharide group (Vanhove et al. 1998) and attempts to inhibit the expression of α(1,3) galactosyltransferase, the enzyme that forms this particular carbohydrate linkage, an enzyme that is present in pigs but not in primates. In the latter case, the simplest strategy would be to knock out the gene by homologous recombination. Gene targeting has not yet been achieved in pigs, although the success of gene targeting/nuclear transfer in sheep (Chapter 11) suggests that knockout pigs could be produced in the next few years. Alternative procedures include introducing genes encoding other carbohydrate-metabolizing enzymes, so that the Gal-α(1,3)-Gal groups are modified into some other less immunogenic moiety (e.g. Sandrin et al. 1995, Cohn et al. 1997, Osman et al. 1997). Once hyperacute rejection has been overcome, there may be further problems, including delayed rejection, involving natural killer cells and macrophages (Bach et al. 1996), and the requirement for T-cell tolerance (Bracy et al. 1998, Kozlowski et al. 1998, Yang et al. 1998). There are also concerns that endogenous pig retroviruses could be activated following transplantation, perhaps even recombining with human retroviruses to produce potent new hybrids with unknown properties. Detailed studies have so far shown no evidence of such a phenomenon (e.g. Heneine et al. 1998, Patience et al. 1998). For recent reviews on the prospects of xenotransplantation (particularly porcine-to-human), the reader should consult Lambrigts et al. (1998), Sandrin and McKenzie (1999) and Logan (2000).
gain-of-function disease models include Alzheimer’s disease, which was modelled by overexpression of the amyloid precursor protein (Quon et al. 1991), and the triplet-repeat disorder spinocerebellar ataxia type 1 (Burright et al. 1995). Simple transgene addition can also be used to model diseases caused by dominant negative alleles, as recently shown for the premature ageing disease, Werner’s syndrome (Wang et al. 2000).

Recessively inherited diseases are generally caused by loss of function, and these can be modelled by gene knockout. The earliest report of this strategy was a mouse model for hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency, generated by disrupting the gene for HRPT (Kuehn et al. 1987). A large number of genes have been modelled in this way, including those for cystic fibrosis (Dorin et al. 1992, Snouwaert et al. 1992), fragile-X syndrome (Dutch–Belgian Fragile X Consortium, 1994), β-thalassaemia (Skow et al. 1983, Ciavattia et al. 1995) and mitochondrial cardiomyopathy (Li et al. 2000). Gene targeting has been widely used to model human cancers caused by the inactivation of tumour suppressor genes, such as TP53 and RB1 (reviewed by Ghebranious & Donehower 1998, Macleod & Jacks 1999).

While the studies above provide models of single-gene defects in humans, attention is now shifting towards the modelling of more complex diseases, which involve multiple genes. This is a challenging area of research but there have been some encouraging early successes. In many cases, the crossing of different modified mouse lines has led to interesting discoveries. For example, undulated mutant mice lack the gene encoding the transcription factor Pax-1, and Patch mutant mice are heterozygous for a null allele of the platelet-derived growth-factor gene. Hybrid offspring from a mating between these two strains were shown to model the human birth defect spina bifida occulta (Helwig et al. 1995). In other cases, such crosses have pointed the way to possible novel therapies. For example, transgenic mice overexpressing human α-globin and a mutant form of the human β-globin gene that promotes polymerization provide good models of sickle-cell anaemia (Trudel et al. 1991). However, when these mice are crossed to those ectopically expressing human fetal haemoglobin in adulthood, the resulting transgenic hybrids show a remarkable reduction in disease symptoms (Blouin et al. 2000). Similarly, crossing transgenic mice overexpressing the anti-apoptotic protein Bcl-2 to rds mutants, which show inherited slow retinal degeneration, resulted in hybrid offspring in which retinal degeneration was strikingly reduced. This indicates that Bcl-2 could possibly be used in gene therapy to treat the equivalent human retinal-degeneration syndrome (Nir et al. 2000).

The most complex diseases involve many genes, and transgenic models would be difficult to create. However, it is often the case that such diseases can be reduced to a small number of ‘major genes’ with severe effects and a larger number of minor genes. Thus, it has been possible to create mouse models of Down’s syndrome, which in humans is generally caused by the presence of three copies of chromosome 21. Trisomy for the equivalent mouse chromosome 16 is a poor model because the two chromosomes do not contain all the same genes. However, a critical region for Down’s syndrome has been identified by studying Down’s patients with partial deletions of chromosome 21. The generation of yeast artificial chromosome (YAC) transgenic mice carrying this essential region provides a useful model of the disorder (Smith et al. 1997) and has identified increased dosage of the Dyrk1a (minibrain) gene as an important component of the learning defects accompanying the disease. Animal models of Down’s syndrome have been reviewed (Kola & Hertzog 1998, Reeves et al. 2001).

**Gene transfer to humans – gene therapy**

The scope of gene therapy

Gene therapy is any procedure used to treat disease by modifying the genetic information in the cells of the patient. In essence, gene therapy is the antithesis of the disease modelling discussed above. Whereas disease modelling takes a healthy animal and uses gene-manipulation techniques to induce a specific disease, gene therapy takes a diseased animal (or human) and uses gene-manipulation techniques in an attempt to correct the disorder and return the individual to good health. Gene transfer can be carried out in cultured cells, which are then reintroduced into the patient, or DNA can be transferred to the patient in vivo, directly or using viral vectors.
The *ex vivo* approach can be applied only to certain tissues, such as bone marrow, in which the cells are amenable to culture. Gene therapy can be used to treat diseases caused by mutations in the patient’s own DNA (inherited disorders, cancers), as well as infectious diseases, and is particularly valuable in cases where no conventional treatment exists or where that treatment is inherently risky. Strategies include the following (Fig. 14.9):

- Gene-augmentation therapy (GAT), where DNA is added to the genome with the aim of replacing a missing gene product.
- Gene targeting to correct mutant alleles.
- Gene-inhibition therapy, using techniques such as antisense RNA expression or the expression of intracellular antibodies to treat dominantly acting diseases.
- The targeted ablation of specific cells.

![Fig. 14.9](image_url) Overview of gene-therapy strategies.
Therapeutic gene transfer effectively generates transgenic human cell clones and, for this reason, only somatic cells can be used as targets. The prospect of germ-line transgenesis in humans raises serious ethical concerns and, with the rapid advances in technology allowing germ-line transformation and nuclear transfer in numerous mammals, these concerns will need to be addressed in the very near future (Johnson 1998). As an alternative to permanent gene transfer, transient gene therapy can be achieved using oligonucleotides, which can disrupt gene expression at many levels but do not permanently change the genetic material of the cell (Pollock & Gaken 1995).

The tools and techniques for gene therapy are essentially similar to those used for gene transfer to any animal cells. Transfection, direct delivery or transduction (see Chapter 11) can be used to introduce DNA into cells. Viral vectors are most popular because of their efficiency of gene transfer in vivo. However, extreme precautions need to be taken to ensure the safety of such vectors, avoiding potential problems, such as the production of infectious viruses by recombination and the pathological effects of viral replication. A number of viral vectors have been developed for gene therapy, including those based on oncoretroviruses, lentiviruses, adenovirus, adeno-associated virus, herpes virus and a number of hybrid vectors combining advantageous elements of different parental viruses (Robbins et al. 1998, Reynolds et al. 1999). The risks associated with viral vectors have promoted research into other delivery methods, the most popular of which include direct injection of DNA into tissues (e.g. muscle), the injection of liposome–DNA complexes into the blood and direct transfer by particle bombardment. Although inherently much safer than viruses, such procedures show a generally low efficiency (Scheule & Cheng 1996, Tseng & Huang 1998).

**Gene-augmentation therapy for recessive diseases**

The first human genetic-engineering experiment was one of gene marking, rather than gene therapy, and was designed to demonstrate that an exogenous gene could be safely transferred into a patient and that this gene could subsequently be detected in cells removed from the patient. Both objectives were met. Tumour-infiltrating lymphocytes (cells that naturally seek out cancer cells and then kill them by secreting proteins such as tumour necrosis factor (TNF)) were isolated from patients with advanced cancer. The cells were then genetically marked with a neomycin-resistance gene and injected back into the same patient (Rosenberg et al. 1990).

The first clinical trial using a therapeutic gene-transfer procedure involved a 4-year-old female patient, Ashanthi DeSilva, suffering from severe combined immune deficiency, resulting from the absence of the enzyme adenosine deaminase (ADA). This disease fitted many of the ideal criteria for gene-therapy experimentation. The disease was life-threatening (therefore making the possibility of unknown treatment-related side-effects ethically acceptable), but the corresponding gene had been cloned and the biochemical basis of the disease was understood. Importantly, since ADA functions in the salvage pathway of nucleotide biosynthesis (p. 177), cells in which the genetic lesion had been corrected had a selective growth advantage over mutant cells, allowing them to be identified and isolated in vitro. Conventional treatment for ADA deficiency involves bone-marrow transplantation from a matching donor. Essentially the same established procedure could be used for gene therapy, but the bone-marrow cells would be derived from the patient herself and would be genetically modified ex vivo (Fig. 14.10). Cells from the patient were subjected to leucapheresis and mononuclear cells were isolated. These were grown in culture under conditions that stimulated T-lymphocyte activation and growth and then transduced with a retroviral vector carrying a normal ADA gene as well as the neomycin-resistance gene. Following infusion of these modified cells, both this patient and a second, who began treatment in early 1991, showed an improvement in their clinical condition as well as in a battery of in vitro and in vivo immune-function studies (Anderson 1992). However, the production of recombinant ADA in these patients is transient, so each must undergo regular infusions of recombinant T lymphocytes. Research is ongoing into procedures for the transformation of bone-marrow stem cells, which would provide a permanent supply of corrected cells.

Gene-augmentation therapies for a small number of recessive single-gene diseases are now undergoing...
clinical trials. We consider cystic fibrosis (CF) as an example. CF is a disorder that predominantly affects the lungs, liver and pancreas. The disease is caused by the loss of a cAMP-regulated membrane-spanning chloride channel. This results in an electrolyte imbalance and the accumulation of mucus, often leading to respiratory failure. CF is a recessive disorder suggesting that the loss of function could be corrected by introducing a functional copy of the gene. Indeed, epithelial cells isolated from CF patients can be restored to normal by transfecting them with the cloned cystic-fibrosis transmembrane regulator (CFTR) cDNA. Unlike ADA deficiency, the cells principally affected by CF cannot be cultured and returned to the patient, so in vivo delivery strategies must be applied. Targeted delivery of the CFTR cDNA to affected cells has been achieved using adenoviral vectors, which have a natural tropism for the epithelial lining of the respiratory system. Recombinant viruses carrying the CFTR cDNA have been introduced into patients using an inhaler (Zabner et al. 1993, Hay et al. 1995, Knowles et al. 1995). The CFTR cDNA has also been introduced using liposomes (e.g. Caplen et al. 1995). While such treatments have resulted in CFTR transgene expression in the nasal epithelium, there were neither consistent changes in chloride transport nor reduction in the severity of CF symptoms, i.e. they have been largely ineffective.

Gene-therapy strategies for cancer

Cancer gene therapy was initially an extension of the early gene-marking experiments. The tumour-infiltrating leucocytes were transformed with a gene for TNF in addition to the neomycin-resistance gene, with the aim of improving the efficiency with which these cells kill tumours by increasing the amount of TNF they secrete. Although TNF is highly toxic to humans at levels as low as 10 µg/kg body weight, there have been no side-effects from the gene therapy and no apparent organ toxicity from secreted TNF (Hwu et al. 1993). One alternative strategy is to transform the tumour cells themselves, making them more susceptible to the immune system through the expression of cytokines or a foreign antigen. Another is to transform fibroblasts, which are easier to grow in culture, and then co-inject these together with tumour cells to provoke an immune response against the tumour. A number of such ‘assisted killing’ strategies have been approved for clinical trials (see review by Ockert et al. 1999).

Direct intervention to correct cancer-causing genes is also possible. Dominantly acting genes (oncogenes) have been targeted using antisense technology, either with antisense transgenes, oligonucleotides (see Carter & Lemoine 1993, Nellen & Lichtenstein 1993) or ribozymes (Welch et al. 1998, Muotri et al. 1999). An early report of cancer gene therapy
with antisense oligonucleotides was that of Szczylik et al. (1991) for the treatment of chronic myeloid leukaemia. They used two 18-mers specific for the BCR–ABL gene junction generated by the chromosomal translocation that causes this particular cancer, and showed that colony formation was suppressed in cells removed from cancer patients. Cancers caused by loss of tumour-suppressor gene function have been addressed by replacement strategies, in which a functional copy of the appropriate gene is delivered to affected cells (e.g. see Cai et al. 1993, Harper et al. 1993, Smith et al. 1993, Hahn et al. 1996). A further strategy, known as prodrug activation therapy, involves the activation of a particular enzyme specifically in cancer cells, which converts a non-toxic ‘prodrug’ into a toxic product, so killing the cancer cells. This can be achieved by driving the expression of a so-called ‘suicide gene’ selectively in cancer cells. An example is the HSV thymidine kinase gene, in combination with the prodrug ganciclovir. Thymidine kinase converts ganciclovir into a nucleoside analogue, which is incorporated into DNA and blocks replication by inhibiting the DNA polymerase. Activation of the enzyme specifically in cancer cells can be achieved by preferential delivery to dividing cells through the use of oncoretroviruses (e.g. Moolten 1986, Culver et al. 1992, Klatzmann et al. 1996). Another way is to use transcriptional regulatory elements that are active only in cancer cells (e.g. Harris et al. 1994, Su et al. 1996).

The importance of SNPs

As noted earlier (p. 277), SNPs are single-base variations that occur about once every 1000 bases along the human genome. With the availability of the human genome sequence, a high-resolution SNP map is being developed and this will facilitate disease therapy in a number of ways (Rothberg 2001). Two examples will be cited here: understanding polygenic disorders and pharmacogenomics.

The disease studies cited above have all focused on single-gene disorders. However, many of the common diseases of humans are polygenic in origin and attempts to map genes for these complex conditions have generally failed. The availability of SNP maps is providing a new tool for use in genetic association studies to identify genes for polygenic disorders, and success has already been achieved with hypertension (Geller et al. 2000), non-insulin-dependent diabetes (Deeb et al. 2000) and cardiovascular disease (Mason et al. 1999).

Pharmacogenomics is the term used to describe the identification and elucidation of genetic variations that will have an impact on the efficacy of drugs. SNPs are essential to pharmacogenomics (McCarthy & Hilfiker 2000), for they provide an easy means of determining the genotype of the patient. For example, the 2-adrenergic receptor agonists are the most widely used agents in the treatment of asthma and several polymorphisms have been described within the target genes. Several studies have shown associations between SNPs in these genes and response to therapy. One study (Buscher et al. 1999) found that homozygotes for one allele were up to 5.3 times more likely to respond to albuterol than homozygotes for the other allele. Another striking example is the response of Alzheimer’s patients to the drug tacrine. Approximately 80% of patients not carrying the ApoE4 allele improved after tacrine treatment, whereas 60% of patients with the allele deteriorated after treatment (Poirier et al. 1995). One scenario for the future is that, before prescribing a drug, patients will be genotyped by SNP analysis to determine the most effective therapy.

Another use for pharmacogenomics is the prevention of adverse drug reactions resulting from drug metabolism. For example, between 3 and 10% of the Caucasian population fail to metabolize the adrenergic-blocking drug debrisoquine and treatment results in severe hypotension. In Afro-Americans the frequency of this ‘poor metabolizer’ condition is 5% and in Asians it is only 1%. Affected individuals are homozygous for a mutant cytochrome P450 gene (CYP2D6) and they also fail to metabolize over 20% of all commonly prescribed drugs, including codeine (Gonzalez et al. 1988). The same gene also has alleles that cause an elevated-metabolizer phenotype, which has been correlated with increased susceptibility to cancer. Clearly, before selecting patients for clinical trials of new drugs, it would make sense to screen candidates for their drug-metabolizing phenotype and most large pharmaceutical companies now do this.
Theme 3: Combating infectious disease

Introduction to theme 3

The usual way of treating bacterial and infectious diseases is with antibiotics. As is well known, certain microbes quickly develop resistance to the antibiotics in current use and this means that new antibiotics are required. The traditional way of obtaining new antibiotics is the screening of new microbial isolates from nature. An alternative way will be described in theme 5: combinatorial biosynthesis (p. 306). Another way is to identify new cellular targets and screen chemical libraries for inhibitory activities. A number of methods for identifying key genes involved in pathogenesis have been developed and these are described later in this section. In contrast to bacteria and fungi, viruses are not sensitive to antibiotics and few therapies have been available. However, this could change with the development of antisense drugs, as described in the previous section (p. 291). Where suitable therapies do exist, it can be advantageous to know the identity of the pathogen as soon as possible. Conventional laboratory procedures take several days, but PCR methodology offers a more rapid identification, as was described on p. 274.

For many pathogens, prevention is much better than cure, and hence vaccines are of great value. Gene-manipulation techniques have greatly facilitated the development of new vaccines, as described below.

Novel routes to vaccines

An effective vaccine generates humoral and/or cell-mediated immunity, which prevents the development of disease upon exposure to the corresponding pathogen. This is accomplished by presenting pertinent antigenic determinants to the immune system in a fashion which mimics that in natural infections. Conventional viral vaccines consist of inactivated, virulent strains or live, attenuated strains, but they are not without their problems. For example, many viruses have not been adapted to grow to high titre in tissue culture, e.g. hepatitis B virus. There is a danger of vaccine-related disease when using inactivated virus, since replication-competent virus may remain in the inoculum. Outbreaks of foot-and-mouth disease in Europe have been attributed to this cause. Finally, attenuated virus strains have the potential to revert to a virulent phenotype upon replication in the vaccinee. This occurs about once or twice in every million people who receive live polio vaccine. Recombinant DNA technology offers some interesting solutions.

Given the ease with which heterologous genes can be expressed in various prokaryotic and eukaryotic systems, it is not difficult to produce large quantities of purified immunogenic material for use as a subunit vaccine. A whole series of immunologically pertinent genes have been cloned and expressed but, in general, the results have been disappointing. For example, of all the polypeptides of foot-and-mouth disease virus, only VP1 has been shown to have immunizing activity. However, polypeptide VP1 produced by recombinant means was an extremely poor immunogen (Kleid et al. 1981). Perhaps it is not too surprising that subunit vaccines produced in this way do not generate the desired immune response, for they lack authenticity. The hepatitis B vaccine, which is commercially available (Valenzuela et al. 1982), differs in this respect, for expression of the surface antigen in yeast results in the formation of virus-like particles. A similar phenomenon is seen with a yeast Ty vector carrying a gene for HIV coat protein (Adams et al. 1987). These subunit vaccines also have another disadvantage. Being inert, they do not multiply in the vaccinee and so they do not generate the effective cellular immune response essential for the recovery from infectious disease.

Recombinant bacterial vaccines

An alternative approach to the development of live vaccines is to start with the food-poisoning organism Salmonella typhimurium. This organism can be attenuated by the introduction of lesions in the aro genes, which encode enzymes involved in the biosynthesis of aromatic amino acids, p-aminobenzoic acid and enterochelins. Whereas doses of $10^4$ wild-type S. typhimurium reproducibly kill mice, aro mutants do not kill mice when fed orally, even when doses as high as $10^{10}$ organisms are used. However, the mutant strains can establish self-limiting infections in the mice and can be detected in low numbers in organs such as the liver and spleen. Such attenuated strains of S. typhimurium
are particularly attractive as carriers of heterologous antigens because they can be delivered orally and because they can stimulate humoral, secretory and cellular immune responses in the host (Charles & Dougan 1990). Already a wide range of heterologous antigens have been expressed in such vaccine strains (Hackett 1993).

The use of the BCG vaccine strain as an alternative vector has many advantages, including its known safety, low cost, widespread use as a childhood vaccine and ability of a single dose to induce long-lasting protection. Several recombinant BCG strains have been constructed that stably express foreign genes (Stover et al. 1991) and preliminary results from animal studies are very encouraging. An alternative vector is the human oral commensal *Streptococcus gordonii* (Fischetti et al. 1993).

A different procedure for attenuating a bacterial pathogen to be used as a vaccine has been proposed by Kaper et al. (1984). They attenuated a pathogenic strain of *Vibrio cholera* by deletion of DNA sequences encoding the A1 subunit of the cholera enterotoxin. A restriction-endonuclease fragment encoding the A1, but not the A2 or B, sequence was deleted in vitro from cloned cholera-toxin genes. The mutation was then recombined into the chromosome of a pathogenic strain. The resulting strain produces the immunogenic but non-toxic B subunit of cholera toxin but is incapable of producing the A subunit. This strain has been found to be safe and immunogenic in carefully controlled clinical trials in a number of countries (Cryz 1992).

Yet another approach to vaccine preparation has been developed by Pizza et al. (2000), based on whole-genome sequencing. They wanted to develop a vaccine against group B meningococci but were hampered by the fact that there was considerable sequence variation in surface-exposed proteins. Starting with the entire genome sequence of a group B strain of *Neisseria meningitidis*, they identified 350 proteins as potential protective antigens. All 350 candidate antigens were expressed in *E. coli*, purified and used to immunize mice. The sera from the mice allowed the identification of proteins that are surface-exposed in meningococci, that are conserved across a range of strains and that induce a bactericidal antibody response, a property known to correlate with vaccine efficacy in humans.

**Recombinant viruses as vaccines**

Recombinant viruses can be used as vectors to express heterologous antigens, and thus function as live vaccines. The first animal virus to be exploited in this way was vaccinia, which had been used previously as a non-recombinant vaccine providing cross-protection against variola virus, the causative agent of smallpox. Contributing to its success as a live vaccine were its stability as a freeze-dried preparation, its low production cost and the ability to administer the vaccine by simple dermal abrasion. Vaccinia remains the most widely explored recombinant viral vaccine, and many antigens have been expressed using this vector, essentially using the methods described on p. 197 (Ulabeto & Hruby 1994, Moss 1996). The most successful recombinant-virus vaccination campaign to date involved the use of recombinant vaccinia virus expressing rabies-virus glycoprotein. This was administered to the wild population of foxes in central Europe by providing a bait consisting of chicken heads spiked with the virus. The epidemiological effects of vaccination were most evident in eastern Switzerland, where two decades of rabies came to a sudden end after only three vaccination campaigns (Brochier et al. 1991, Flamand et al. 1992), and in Belgium, where the disease was all but eliminated (Brochier et al. 1995).

One disadvantage of this approach in humans is the unacceptably high risk of adverse reactions to the vaccine. For use in humans, vaccinia must be further attenuated to make it replication-deficient and to minimize the likelihood of replication-competent viruses arising by recombination (e.g. Cooney et al. 1991, Sutter & Moss 1992, Tartaglia et al. 1992). Highly attenuated vaccinia derivatives (Table 14.6) have been used to express a range of viral, bacterial and parasite antigens, with some reaching clinical trials (reviewed by Moss 1996, Paolelli 1996). Other poxviruses, such as canarypox, have also been developed as potential vaccines (e.g. see Perkins et al. 1995, Fries et al. 1996, Myagkikh et al. 1996).

As well as vaccinia, adenovirus and alphavirus vectors have also been investigated as potential recombinant vaccines. Adenovirus has been used to express various antigens, including HSV and rabies virus glycoproteins (Gallichan et al. 1993, Xiang et al. 1996). Among the alphaviruses, Semliki Forest
Applications of recombinant DNA technology

The virus has been used to express the gp160 protein of HIV (Berglund et al. 1997) and Sindbis virus has been used to express antigens from Japanese encephalitis virus (Pugachev et al. 1995).

Recombinant proteins expressed by viruses have been shown to be immunogenetic. However, stronger stimulation of the immune response can often be achieved by presenting the antigen on the surface of the virus attached to a host virus-derived carrier protein. The advantage of this strategy is that the recombinant antigen is presented as multiple copies. This strategy is borrowed from the surface display of foreign antigens on bacteria that are used as live vaccines (see above) but is much safer. Indeed, while a number of conventional viral vaccines, such as vaccinia, have been developed as surface-display systems (e.g. see Katz & Moss 1997, Katz et al. 1997), even plant viruses can be used for this purpose, as discussed below (Porta & Lomonossoff 1996, Johnson et al. 1997).

### Table 14.6 Immune response to heterologous antigens expressed by vaccinia virus recombinants.

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Neutralizing antibodies</th>
<th>Cellular immunity</th>
<th>Animal protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rabies virus glycoprotein</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Vesicular stomatitis glycoprotein</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Herpes simplex virus glycoprotein D</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Hepatitis B surface antigen</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Influenza virus haemagglutinin</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Human immunodeficiency virus envelope</td>
<td>+</td>
<td>+</td>
<td>not determined</td>
</tr>
</tbody>
</table>

Reproduced with permission from Tartaglia & Paoletti (1988).

### Table 14.7 A selection of recombinant vaccines against animal viruses produced in plants.

<table>
<thead>
<tr>
<th>Antigen</th>
<th>Host-plant system</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herpes virus B surface antigen</td>
<td>Tobacco</td>
<td>Mason et al. 1992</td>
</tr>
<tr>
<td>Rabies glycoprotein</td>
<td>Tomato</td>
<td>McGarvey et al. 1995</td>
</tr>
<tr>
<td>Norwalk virus coat protein</td>
<td>Tobacco, potato</td>
<td>Mason et al. 1996</td>
</tr>
<tr>
<td>Foot-and-mouth virus VP1</td>
<td>Arabidopsis</td>
<td>Carrillo et al. 1998</td>
</tr>
<tr>
<td>Cholera toxin B subunit</td>
<td>Potato</td>
<td>Arakawa et al. 1998</td>
</tr>
<tr>
<td>Human cytomegalovirus glycoprotein B</td>
<td>Tobacco</td>
<td>Tackaberry et al. 1999</td>
</tr>
</tbody>
</table>

Plants as edible vaccines

Plants have been explored as a cheap, safe and efficient production system for subunit vaccines (Table 14.7), with the added advantage that orally administered vaccines can be ingested by eating the plant, thereby eliminating the need for processing and purification (reviewed by Mason & Arntzen 1995, Walmsley & Arntzen 2000). The earliest demonstration was the expression of a surface antigen from the bacterium *S. mutans* in tobacco. This bacterium is the causative agent of dental caries, and it was envisaged that stimulation of a mucosal immune response would prevent the bacteria colonizing the teeth and therefore protect against tooth decay (Curtis & Cardineau, 1990).

A number of edible transgenic plants have been generated expressing antigens derived from animal viruses. For example, rabies glycoprotein has been expressed in tomato (McGarvey et al. 1995), hepatitis
Virus antigen in lettuce (Ehsani et al. 1997) and cholera antigen in potato (Arakawa et al. 1997). As well as animal virus antigens, autoantigens associated with diabetes have also been produced (Ma et al. 1997, Porceddu et al. 1999). Plants have also been infected with recombinant viruses expressing various antigen epitopes on their surfaces. Cowpea mosaic virus (CMV) has been extensively developed as a heterologous antigen-presenting system (Porta et al. 1994, Lomonossoff & Hamilton 1999). There have been some recent successes in vaccination trials using recombinant CMV vectors expressing epitopes of HIV gp41 (McLain et al. 1995, 1996a,b) and canine parvovirus (Dalsgaard et al. 1997). The first clinical trials using a plant-derived vaccine were conducted in 1997 and involved the ingestion of transgenic potatoes expressing the B subunit of the \textit{E. coli} heat-labile toxin, which causes diarrhea. This resulted in a successful elicitation of mucosal immunity in test subjects (Tacket et al. 1994).

**DNA vaccines**

The immune system generates antibodies in response to the recognition of proteins and other large molecules carried by pathogens. In each of the examples above, the functional component of the vaccine introduced into the host is the protein responsible for the elicitation of the immune response. The introduction of DNA into animals does not generate an immune response against the DNA molecule, but, if that DNA is expressed to yield a protein, that protein can stimulate the immune system. This is the basis of DNA vaccination, as first demonstrated by Ulmer et al. (1993). DNA vaccines generally comprise a bacterial plasmid carrying a gene encoding the appropriate antigen under the control of a strong promoter that is recognized by the host cell. The advantages of this method include its simplicity, its wide applicability and the ease with which large quantities of the vaccine can be produced. The DNA may be administered by injection, using liposomes or by particle bombardment. In the original demonstration, Ulmer and colleagues introduced DNA corresponding to the influenza virus nucleoprotein and achieved protection against influenza infection. Since then, many DNA vaccines have been used to target viruses (e.g. measles (Cardoso et al. 1996); HIV (Wang et al. 1993, Fuller et al. 1997, Hinkula et al. 1997); Ebola virus (Xu et al. 1998), other pathogens (e.g. tuberculosis (Huygen et al. 1996)) and even the human cellular prion protein in mice (Krasemann et al. 1996).

The DNA-vaccination approach has several additional advantages. These include the following:

- Certain bacterial DNA sequences have the innate ability to stimulate the immune system (see Klinman et al. 1997, Roman et al. 1997).
- Other genes encoding proteins influencing the function of the immune response can be co-introduced along with the vaccine (e.g. Kim et al. 1997).
- DNA vaccination can be used to treat diseases that are already established as a chronic infection (e.g. Mancini et al. 1996).

In principle, DNA vaccination has much in common with gene therapy (discussed above), since both processes involve DNA transfer to humans, using a similar selection of methods. However, while the aim of gene therapy is to alleviate disease, by either replacing a lost gene or blocking the expression of a dominantly acting gene, the aim of DNA vaccination is to prevent disease, by causing the expression of an antigen that stimulates the immune system.

**Selecting targets for new antimicrobial agents**

In attempting to develop new antimicrobial agents, including ones that are active against intractable pathogens such as the malarial parasite, it would be useful to know which genes are both essential for virulence and unique to the pathogen. Once these genes have been identified, chemical libraries can be screened for molecules that are active against the gene product. Two features of this approach deserve further comment. First, inhibition of the gene product could attenuate the organism’s virulence but would not result in death of the organism \textit{in vitro}: that is, no effect would be seen in whole-organism inhibition assays and so molecules only active \textit{in vivo} would be missed. Secondly, target genes can be selected on the basis that there are no human counterparts. Thus, active molecules will be less likely to be toxic to humans. A number of different approaches have been developed for identifying virulence
determinants, and details are presented below for three of these: *in vivo* expression technology (IVET), differential fluorescence induction and signature-tagged mutagenesis.

**In vivo expression technology (IVET)**

IVET was developed to positively select those genes that are induced specifically in a microorganism when it infects an animal or plant host (Mahan *et al.* 1993). The basis of the system is a plasmid carrying a promoterless operon fusion of the *lacYZ* genes fused to the *purA* or *thyA* genes downstream of a unique *BglII* cloning site (Fig. 14.11). This operon fusion was constructed in a suicide-delivery plasmid. Cloning of pathogen DNA into the *BglII* site results in the construction of a pool of transcriptional fusions driven by promoters present in the

![Diagram of IVET process](image)

**Fig. 14.11** The basic principle of IVET. See text for details.
cloned DNA. The pool of fusions is transferred to a strain of the pathogen carrying a purA or thyA deletion and selection made for integration into the chromosome. The recombinant pathogen is then used to infect a test animal. Fusions that contain a promoter that is active in the animal allow transcription of the purA or thyA gene and hence bacterial survival. When the surviving pathogens are reisolated from the test animal, they are tested in vitro for their levels of β-galactosidase. Clones that contain fusions to genes that are specifically induced in the test animal show little lacZ expression on laboratory media and DNA sequencing can be used to identify the genes to which these promoters belong.

The IVET system was initially developed for use in a murine typhoid model. Since it was first described, IVET has been used with a wide variety of Gram-positive and Gram-negative organisms, including ‘difficult’ organisms, such as Mycobacterium. One disadvantage of the original IVET system is that it inherently selects for promoters that are constitutively and highly expressed in vivo. An alternative IVET system uses tnpR operon fusions in place of purA or thyA. The tnpR locus encodes a site-specific resolvase and is used as the selection gene. Expression of this gene results in resolvase synthesis and the deletion of a resolvase-specific reporter. The advantage of this system is that it can be used to identify those genes that are expressed only transiently during animal infection (Camilli & Mekalanos 1995, Camilli 1996). Handfield and Levesque (1999) have described a number of other modifications to the IVET technology.

Differential fluorescence induction
This technique represents a different way of identifying environmentally controlled promoters and was originally developed to facilitate identification of Salmonella genes that are differentially expressed with macrophages (Valdivia & Falkow 1997). Fragments of DNA from Salmonella were fused to a promoterless gene for green fluorescent protein (GFP) and returned to the Salmonella, which was then used to infect macrophages. Cells that became fluorescent were recovered using a fluorescence-activated cell sorter (FACS) and grown on media in the absence of macrophages. Bacteria that were non-fluorescent in the extracellular environment were sorted and used for a second round of macrophage infection. Bacteria still capable of generating fluorescent macrophages were found to contain GFP fusions that were up-regulated by the macrophage’s intracellular environment.

**Signature-tagged mutagenesis**
This technique is a variation on the use of transposon mutagenesis that has been applied to the identification of bacterial virulence genes. The basic principle is to create a large number of different transposon-generated mutants of a pathogenic organism and to identify those mutants that can survive in vitro but not in vivo – that is, it is a whole-genome scan for habitat-specific genes. Although one could test each mutant individually, this would be very laborious and would use large numbers of animals. By using signature-tagged mutagenesis, one can test large numbers of different mutants simultaneously in the same animal. This is achieved by tagging each transposon mutant with a different DNA sequence or tag.

The use of sequence-tagged mutagenesis was first described for the identification of virulence genes from S. typhimurium in a mouse model of typhoid fever (Hensel et al. 1995), as shown in Fig. 14.12. The tags comprise different sequences of 40 bp [NK]_{20} where N = A, C, G or T and K = G or T. The arms were designed so that the amplification of the tags by PCR with specific primers would produce probes with 10 times more label in the central region than in each arm. The double-stranded tags were ligated into a Tn5 transposon and transferred from E. coli to S. typhimurium by conjugation. A library of over 1500 exconjugants resulting from transposition events was stored in microtitre dishes. Of these exconjugants, 1152 were selected and prepared as 12 pools of 96 mutants. Each pool was injected into the peritoneum of a different mouse and infection allowed to proceed. Bacteria were then recovered from each mouse by plating spleen homogenates on culture media. DNA was extracted from the recovered bacteria and the tags in this DNA were amplified by PCR. Those tags present in the initial pool of bacteria but missing from the recovered media represent mutations in genes essential for
Applications of recombinant DNA technology

virulence. In this way 28 different mutants with attenuated virulence were identified and some of these mutants were in genes not previously identified.

The principle of signature-tagged mutagenesis has been extended to the analysis of pathogenicity determinants in a wide range of bacteria (for a review, see Handfield & Levesque 1999) and to fungi (Brookman & Denning 2000).

Environomics

The techniques described above to identify bacterial virulence genes can be applied to other microbe–host interactions, e.g. protozoan infections of humans, bacterial or fungal pathogenesis of plants, or even beneficial rhizosphere–microbe interactions (Rainey 1999). Essentially, these techniques are methods for scanning the entire microbial genome for genes that are expressed under particular environmental conditions. This approach has been termed ‘environomics’.

Theme 4: Protein engineering

Introduction to theme 4

One of the most exciting aspects of recombinant DNA technology is that it permits the design, development and isolation of proteins with improved operating characteristics and even completely novel proteins. The simplest example of protein engineering involves site-directed mutagenesis to alter key residues, as originally shown by Winter and colleagues (Winter et al. 1982, Wilkinson et al. 1984). From a detailed knowledge of the enzyme tyrosyl-tRNA synthetase from Bacillus stearothermophilus, including its crystal structure, they were able to predict point mutations in the gene that should increase the enzyme’s affinity for the substrate ATP. These changes were introduced and, in one case, a single amino acid change improved the affinity for ATP by a factor of 100. Using a similar approach, the stability of an enzyme can be increased. Thus Perry and Wetzel (1984) were able to increase the thermostability of T4 lysozyme by the introduction of a disulphide bond. However, although new cysteine residues can be introduced at will, they will not necessarily lead to increased thermal stability (Wetzel et al. 1988).

Improving therapeutic proteins with single amino acid changes

As noted earlier (p. 283), many recombinant proteins are now being used therapeutically. With some of them, protein engineering has been used to generate second-generation variants with improved
pharmacokinetics, structure, stability and bioavailability (Bristow 1993). For example, in the neutral solutions used for therapy, insulin is mostly assembled as zinc-containing hexamers. This self-association may limit absorption. By making single amino acid substitutions, Brange et al. (1988) were able to generate insulins that are essentially monomeric at pharmaceutical concentrations. Not only have these insulins preserved their biological activity, but they are also absorbed two to three times faster. Similarly, replacing an asparagine residue with glutamine altered the glycosylation pattern of TPA. This in turn significantly increased the circulatory half-life, which in the native enzyme is only 5 min (Lau et al. 1987). Proteins can also be engineered to be resistant to oxidative stress, as has been shown with AAT (see Box 14.3).

**Box 14.3 Oxidation-resistant variants of \( \alpha_1 \)-antitrypsin (AAT)**

Cumulative damage to lung tissue is thought to be responsible for the development of emphysema, an irreversible disease characterized by loss of lung elasticity. The primary defence against elastase damage is AAT, a glycosylated serum protein of 394 amino acids. The function of AAT is known because its genetic deficiency leads to a premature breakdown of connective tissue. In healthy individuals there is an association between AAT and neutrophil elastase followed by cleavage of AAT between methionine residue 358 and serine residue 359 (see Fig. B14.2).

After cleavage, there is negligible dissociation of the complex. Smokers are more prone to emphysema, because smoking results in an increased concentration of leucocytes in the lung and consequently increased exposure to neutrophil elastase. In addition, leucocytes liberate oxygen free radicals and these can oxidize methionine-358 to methionine sulphonide. Since methionine sulphoxide is much bulkier than methionine, it does not fit into the active site of elastase. Hence oxidized AAT is a poor inhibitor. By means of site-directed mutagenesis, an oxidation-resistant mutant of AAT has been constructed by replacing methionine-358 with valine (Courtney et al. 1985). In a laboratory model of inflammation, the modified AAT was an effective inhibitor of elastase and was not inactivated by oxidation. Clinically, this could be important, since intravenous replacement therapy with plasma concentrates of AAT is already being tested on patients with a genetic deficiency in AAT production.

![Fig. B14.2](image) The cleavage of \( \alpha_1 \)-antitrypsin on binding to neutrophil elastase.
Many of the changes described above were made to improve the ability of subtilisin to hydrolyse protein when incorporated into detergents. However, serine proteases can be used to synthesize peptides and this approach has a number of advantages over conventional methods (Abrahmsen et al. 1991). A problem with the use of subtilisin for peptide synthesis is that hydrolysis is strongly favoured over aminolysis, unless the reaction is undertaken in organic solvents. Solvents, in turn, reduce the half-life of subtilisin. Using site-directed mutagenesis, a number of variants of subtilisin have been isolated with greatly enhanced solvent stability (Wong et al. 1990, Zhong et al. 1991). Changes introduced included the minimization of surface changes to reduce solvation energy, the enhancement of internal polar and hydrophobic interactions and the introduction of conformational restrictions to reduce the tendency of the protein to denature. Designing these changes requires an extensive knowledge of the enzyme's structure and function. Chen and Arnold (1991, 1993) have provided an alternative solution. They utilized random mutagenesis combined with screening for enhanced proteolysis in the presence of solvent (dimethyl formamide) and substrate (casein).

The engineering of subtilisin has now gone one step further, in that it has been modified such that aminolysis (synthesis) is favoured over hydrolysis, even in aqueous solvents. This was achieved by changing a serine residue in the active site to cysteine (Abrahmsen et al. 1991). The reasons for this enhancement derive mainly from the increased affinity and reactivity of the acyl intermediate for the amino nucleophile (Fig. 14.13). These engineered "peptide ligases" are in turn being used to synthesize novel glycopeptides. A glycosyl amino acid is used in peptide synthesis to form a glycosyl peptide ester, which will react with another C-protected peptide in the presence of the peptide ligase to form a larger glycosyl peptide.

**Methods for engineering proteins: the rational approach**

Given the range of modifications made to subtilisin, the question is not "what modifications are possible?" (for review, see Arnold 2001) but "how do I achieve the modifications that I wish to make?" There are two general techniques: the rational approach and "directed evolution". There are two variations on the rational approach. The first of these makes use of comparisons between related proteins. For example, barnase and binase are ribonucleases that have 85% sequence identity but differ in their thermal stability. Using site-directed mutagenesis, Serrano et al. (1993) tested the impact on thermostability of barnase of all 17 amino acid differences between barnase and binase. They observed effects between +1.1 and −1.1 kcal/mol and a multiple mutant combining six of the substitutions displayed a stability increase of 3.3 kcal/mol over the wild-type enzyme.

The second variation on the rational approach requires that the three-dimensional structure of the protein be known. Based on an analysis of this structure, particular residues are selected for mutagenesis and the biological effects then evaluated. That this approach works was clearly shown by the early work on lysozyme and subtilisin, but the chances of success are not predictable. One disadvantage of this approach is illustrated by the results of Spiller et al. (1999), who were investigating the thermostability of an esterase using directed evolution. When they evaluated the effect of certain mutations, they concluded that no amount of rational analysis of the crystal structure would have led them to predict the results obtained.

**Protein engineering through directed evolution**

Directed evolution involves repeated rounds of random mutagenesis, followed by selection for the improved property of interest. Any number of methods can be used to introduce the mutations, including mutagenic base analogues, chemical
mutagenesis, error-prone PCR and spiked synthetic oligonucleotides. The key element in the process is the ability to screen large numbers of mutants. An example is the isolation of a more thermostable subtilisin. Up to 1000 mutant clones are gridded out on replica plates and, once they are grown, one plate is incubated at an elevated temperature long enough to inactivate the wild-type enzyme. When an assay for hydrolytic activity is subsequently performed, only mutants with stability greater than that of the wild type will display measurable activity. Once stable mutants have been identified, the replicate colony can be grown to identify the mutation. Once stabilizing single amino acid changes have been identified, building a highly stable subtilisin can be accomplished by combining individual amino acids into the same molecule (Pantoliano et al. 1988, 1989, Zhao & Arnold 1999).

One of the disadvantages of the screening method described above is that it is labour-intensive and the maximum feasible number of mutants that can be examined in a single screen is $10^4-10^5$. By combining mutagenesis with phage display, up to $10^9$ different mutants can theoretically be screened in a single experiment. The constraint with this approach is that phage display is really best suited to selecting proteins with altered binding characteristics, rather than the other properties one might wish to engineer. Despite this, a number of groups (Atwell & Wells 1999, Demartis et al. 1999, Olsen et al. 2000) were able to select variant proteases with novel substrate specificities.

**Gene families as aids to protein engineering**

An alternative approach to directed evolution is ‘DNA shuffling’, which is also known as ‘molecular breeding’ (Minshull & Stemmer 1999, Ness et al. 2000). This method can only be adopted if the target protein belongs to a known protein family. If it does, the genes for the different family members are isolated and artificial hybrids created (Fig. 14.14). As an example of this approach, Ness et al. (1999)
Applications of recombinant DNA technology

**Theme 5: Metabolic engineering**

**Introduction to theme 5**

When the large-scale production of penicillin was begun in the 1940s, the yields of penicillin were measured in micrograms per litre of culture. Demand for the antibiotic was outstripping supply and higher-yielding strains were badly needed. Since nothing was known about the biosynthetic pathway, a programme of strain improvement was set in place that involved random mutation and screening. The best strain from each cycle of improvement then became the starting-point for the next round of selection. In this way the yield of penicillin was steadily increased until it reached the tens of grams per litre that can be achieved today. As each new antibiotic was discovered, the same process of strain improvement was applied. In every case, the biochemical and genetic basis of the beneficial mutations was not known. Only when the details of gene regulation and metabolic-pathway regulation (allosteric control) had been elucidated could we even begin to understand how antibiotic yields might have been improved.

Based on our current knowledge of metabolic regulation we can predict that the changes in the improved strains described above will involve all of the following:

- Removal of rate-limiting transcriptional and allosteric controls.
- Kinetic enhancement of rate-limiting enzymes.
- Genetic blockage of competing pathways.
- Enhanced carbon commitment to the primary metabolic pathway from central metabolism.
- Modification of secondary metabolic pathways to enhance energy metabolism and availability of enzymatic cofactors.
- Enhanced transport of the compound out of the cell.

Today, if one starts with a wild-type strain and wants to turn it into an overproducer, then the approach would be to use recombinant DNA technology to make these desired changes in a rational way, as exemplified below by phenylalanine. There are many other examples of rational ‘metabolic engineering’ and these have been reviewed by Chotani *et al.* (2000).

**Designed overproduction of phenylalanine**

Phenylalanine is a key raw material for the synthesis of the artificial sweetener aspartame. Phenylalanine can be synthesized chemically but is too expensive if made this way. In the 1980s bacterial strains that overproduced phenylalanine were developed, using the traditional mutation and selection method. At the same time, a programme of rational strain development was instituted at G.D. Searle, the company who owned the patent for aspartame. The starting-point for this programme was an analysis of the biosynthetic pathway (Fig. 14.15). Removing feedback inhibition of key steps is an essential first step. In the case of a phenylalanine producer, it is essential to knock out any feedback inhibition of the pathway from chorismate to phenylalanine and this was achieved by selecting strains resistant to phenylalanine analogues. The conversion of erythrose-4-phosphate (E4P) and phosphoenol pyruvate (PEP) to DAHP is also subject to feedback inhibition, but, since there are three different enzymes here, each
inhibited by a different aromatic end-product, all that is necessary for a phenylalanine overproducer is to clone the tryptophan-sensitive enzyme and have it overexpressed. To overcome repression of enzyme synthesis, the existing promoters were removed and replaced with one that could be controlled more easily in industrial-scale fermentations. The above changes removed the natural control circuits. The next step was to remove competing pathways, i.e. the synthesis of tyrosine and tryptophan. This was easily achieved by making a tyrosine and tryptophan double auxotroph. Note that stable (non-reverting) auxotrophs can best be made by deleting part or all of the relevant genes. This is a task that is easy using recombinant DNA technology. Once all the control circuits and competing pathways had been removed, attempts were made to increase the carbon flux through the biosynthetic pathway. Surprisingly, overexpressing all the genes in the pathway did not enhance the yield of phenylalanine. One explanation was that the supply of precursors (E4P and/or PEP) was rate-limiting. This was confirmed when cloning transketolase (to enhance E4P levels) and eliminating pyruvate kinase (to enhance PEP levels) enhanced yields.

**New routes to small molecules**

Recombinant DNA technology can be used to develop novel routes to small molecules. Good examples are the microbial synthesis of the blue dye indigo (Ensley *et al.* 1983) and the black pigment melanin (Della-Cioppa *et al.* 1990). Neither compound is produced in bacteria. The cloning of a single gene from *Pseudomonas putida* – that encoding naphthalene dioxygenase – resulted in the generation of an *E. coli* strain able to synthesize indigo in a medium containing tryptophan (Fig. 14.16). Similarly, cloning a tyrosinase gene in *E. coli* led to conversion of tyrosine to dopaquinone, which spontaneously converts to melanin in the presence of air. To overproduce these compounds, one generates a strain of *E. coli* that overproduces either tryptophan or tyrosine, rather than phenylalanine, as described.
above. With both indigo and melanin, yields are improved by increasing the levels of cofactors. Also, in the case of indigo biosynthesis, it is necessary to engineer the tryptophan synthase gene. The reason for this is that indole is an intermediate in the biosynthesis of tryptophan (Fig. 14.15). However, normally it is not free in the cytoplasm but remains trapped within the tryptophan synthase complex. By modifying the \( \text{trpB} \) gene, encoding the subunit of tryptophan synthase, it was possible for the indole to be released for conversion by the dioxygenase (Murdock et al. 1993).

One disadvantage of the new route to indigo is that one of the intermediates in its synthesis, indoxyl, can undergo an alternative spontaneous oxidation to isatin and indirubin. The latter compound is an isomer of indigo with similar dyeing properties, but instead of being blue it is a deep
burgundy colour. To make textile-quality indigo, there must be no indirubin present. Screening soil microorganisms with the capacity to degrade indole resulted in the identification of an enzyme, isatin hydrolase, that can degrade isatin to isatic acid. After cloning the gene for isatin hydrolase in the indigo overproducing strains, the indigo product obtained performed as well as chemically produced material.

A slightly different approach to that above has yielded a new route to vitamin C. The conventional process starts with glucose and comprises one microbiological and four chemical steps (Fig. 14.17). By cloning in Erwinia a single gene – that from Corynebacterium encoding 2,5-diketogluconic acid reductase – the process can be simplified to a single microbiological and a single chemical step (Anderson et al. 1985). After observations of unexpectedly low yields of 2-ketogulonic acid in the recombinant strain, it was found that 2-ketogulonic acid was converted to l-idonic acid by an endogenous 2-ketoaldonate reductase. Cloning, deletion mutagenesis and homologous recombination of the mutated reductase gene into the chromosome were some of the several steps taken to develop an organism capable of accumulating large amounts (120 g/l) of 2-ketogulonic acid (Lazarus et al. 1990). So far, attempts to manufacture vitamin C directly from glucose have been unsuccessful. However, enzymes that can convert 2-ketogulonic acid to ascorbic acid have been identified and the objective now is to clone these activities into Erwinia (Chotani et al. 2000).

**Combinatorial biosynthesis**

A number of widely used antibiotics and immunosuppressants belong to a class of molecules known
Applications of recombinant DNA technology

307

as polyketides. These molecules, which are synthesized by actinomycetes, have a fairly complex structure (Fig. 14.18). The genes involved in the biosynthesis of polyketides are clustered, thereby facilitating the cloning of all of the genes controlling the synthetic pathway. The first cluster (the \textit{act} genes) to be cloned was that for actinorhodin. When parts of the \textit{act} gene cluster were introduced into streptomycetes making related polyketides, completely new antibiotics were produced (Hopwood \textit{et al.} 1985). For example, introducing the \textit{actVA} gene from \textit{Streptomyces coelicolor} into a strain that makes medermycin leads to the synthesis of mederrhodin\textsubscript{A} (Fig. 14.19). This approach has been repeated many times since

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{polyketides.png}
\caption{Some examples of polyketides.}
\end{figure}
with other polyketides (for review, see Baltz 1998) and is known as *combinatorial biosynthesis*.

Once a number of polyketide biosynthetic gene clusters had been cloned and sequenced, new insights were gained on the mechanism of synthesis. In particular, two enzymic modes of synthesis were discovered. In particular, polyketide synthesis takes place on an enzyme complex in a manner analogous to fatty acid synthesis. Furthermore, there are two types of complex. In type II complexes, the different enzymic activities are encoded by separate subunits. In contrast, in type I synthesis all the different enzyme activities are encoded by a single, very large gene. Clearly, the polyketide synthases are prime candidates for DNA shuffling, as described in the previous theme (p. 302), and this approach has been widely adopted (Baltz 1998). However, novel polyketides can also be generated by simply changing the order of the different activities in type I synthases (McDaniel et al. 1999).

**Engineering metabolic control over recombinant pathways**

When a recombinant cell overproduces a protein or the components of a biosynthetic pathway, there is a marked reduction in metabolic activity, coupled with a retardation of growth (Kurland & Dong 1996). These phenotypic characteristics result from the fact that the constant demands of production placed upon the cell interfere with its changing requirements for growth. To alleviate this problem, Farmer and Liao (2000) designed a dynamic control circuit that is able to sense the metabolic state of the cell and thereby regulate the expression of a recombinant pathway. This approach is termed ‘metabolic control engineering’.

An essential component of a dynamic controller is a signal that will reflect the metabolic state of the cell. Acetyl phosphate was selected as the signal, since it is known to be a regulator of various operons influenced by nutrient availability. A component of the Ntr regulon in *E. coli*, NRI, is capable of sensing the acetyl phosphate level in the cell. When phosphorylated by acetyl phosphate, it is capable of binding to the *glnAp2* promoter and activating transcription (Fig. 14.20). To reconstruct this control module, the NRI-binding site and the *glnAp2* promoter were inserted into a plasmid vector upstream of a cloning site.

When the *lacZ* gene was placed under the control of the *glnAp2* promoter, there was no significant β-galactosidase synthesis until late in the exponential phase of growth, just as expected. A similar result was obtained when the *lacZ* gene was replaced with a construct encoding two different metabolic enzymes. Finally, as a real test of the system, an engineered construct encoding a pathway for the synthesis of the carotenoid lycopene was placed under the control of the *glnAp2* promoter. This was a particularly interesting test because one of the precursors of lycopene, pyruvate, is also an immediate precursor of acetyl phosphate. Once again, product
Applications of recombinant DNA technology

(lycopene) formation was controlled by the metabolic state of the cell. Although metabolic control engineering is still at an early stage, it represents a significant degree of sophistication of control compared with the use of simple controllable promoters.

**Metabolic engineering in plant cells**

Plants synthesize an incredibly diverse array of useful chemicals. Most are products of secondary metabolism – that is, biochemical pathways that are not involved in the synthesis of essential cellular components but which synthesize more complex molecules that provide additional functions. Examples of these functions are attraction of pollinators and resistance to pests and pathogens. In many cases, these secondary metabolites have specific and potent pharmacological properties in humans: well-known examples include caffeine, nicotine, morphine and cocaine.

Plants have long been exploited as a source of pharmaceutical compounds, and a number of species are cultivated specifically for the purpose of extracting drugs and other valuable molecules. We discussed above how gene transfer to bacteria and yeast can be used to produce novel chemicals, so in theory it would be possible to transfer the necessary components from these useful plants into microbes for large-scale production. However, the secondary metabolic pathways of plants are so extensive and complex that, in most cases, such a strategy would prove impossible. Fortunately, advances in plant transformation have made it possible to carry out metabolic engineering in plants themselves, and large-scale plant cell cultures can be used in the same manner as microbial cultures for the production of important phytochemicals (reviewed by Verpoorte 1998, Verpoorte et al. 2000).

The secondary metabolic pathways of most plants produce the same basic molecular skeletons, but these are ‘decorated’ with functional groups in a highly specific way, so that particular compounds may be found in only one or a few plant species. Furthermore, such molecules are often produced in extremely low amounts, so extraction and purification can be expensive. For example, the Madagascar periwinkle *Catharanthus roseus* is the source of two potent anti-cancer drugs called vinblastine and vincristine. These terpene indole alkaloids are too complex to synthesize in the laboratory and there are no alternative natural sources. In *C. roseus*, these molecules are produced in such low amounts that over 1 ha of plants must be harvested to produce a single gram of each drug, with a commercial value of over $1 million.

It would be much more convenient to produce such drugs in fermenters containing cultured plant cells, and this has been achieved for a number of
compounds, two of which (paclitaxel and shikonin) have reached commercial production (see Verpoorte et al. 2000). However, cell-suspension cultures often do not produce the downstream products made by the parent plant. This applies to vinblastine and vincristine from *C. roseus*, and also to other important drugs, such as morphine, codeine and hyoscyamine. Part of the reason for this is the complexity of secondary metabolism. The entire pathway is not completed in a single cell, but is often segregated into different cell types, with consequent shuttling of intermediates between cells. Within the cell, different stages of the pathway are also compartmentalized, so that intermediates must be transported between organelles. As in bacteria, knowledge of the target biosynthetic pathway is therefore essential for metabolic engineering, but in plants only a few secondary pathways are understood in sufficient detail. Examples include the phenylpropanoid and flavonoid pathways, which yield anthocyanins (plant pigments) and phytoalexins (antimicrobial compounds), and the terpene indole alkaloid biosynthetic pathway, which generates important alkaloids, such as vinblastine and vincristine. As more plant genomes are sequenced, we are likely to learn much more about such pathways.

All terpene indole alkaloids derive from a single universal precursor called strictosidine. This is formed by the convergence of two pathways, the iridoid pathway (culminating in secologanin) and the terpenoid pathway (culminating in tryptamine). Strictosidine is formed by the condensation of secologanin and tryptamine, catalysed by the enzyme strictosidine synthase, and is then further modified in later steps to produce the valuable downstream alkaloids (Fig. 14.21). The conversion of tryptophan...
to tryptamine is a rate-limiting step in the terpenoid pathway, and this has been addressed by overexpressing the enzyme tryptophan decarboxylase in *C. roseus* cell-suspension cultures. However, while transformed cultures produced much higher levels of tryptamine, no downstream alkaloids were synthesized (Goddijn *et al.* 1995, Canel *et al.* 1998). The simultaneous overexpression of the next enzyme in the pathway, strictosidine synthase, did increase the levels of the alkaloid ajmalicine, a useful sedative, in some cultures, but did not result in the synthesis of vinblastine or vincristine (Canel *et al.* 1998).

Thus it seems that single-step engineering may remove known bottlenecks only to reveal the position of the next. The limited success of single-gene approaches has resulted in the development of alternative strategies for the coordinated regulation of entire pathways using transcription factors. Using the yeast one-hybrid system (Vidal *et al.* 1996a), a transcription factor called ORCA2 has been identified that binds to response elements in the genes for tryptophan decarboxylase, strictosidine synthase and several other genes encoding enzymes in the same pathway. A related protein, ORCA3, has been identified using insertional vectors that activate genes adjacent to their integration site. By bringing the expression of such transcription factors under the control of the experimenter, entire metabolic pathways could be controlled externally (see review by Memelink *et al.* 2000).

Apart from the modification of endogenous metabolic pathways to produce more (or less) of a specific endogenous compound, plants can also be engineered to produce heterologous or entirely novel molecules. An example is the production of the alkaloid scopolamine, an anticholinergic drug, in *Atropa belladonna* (Hashimoto *et al.* 1993, Hashimoto & Yamada 1994). Scopolamine is produced in *Hyoscyamus niger* but not in *A. belladonna*, which accumulates the immediate precursor hyoscyamine. *H. niger* converts hyoscyamine into scopolamine using the enzyme hyoscyamine-6-hydroxylase (H6H), which is absent in *A. belladonna*. Hashimoto and colleagues isolated a complementary DNA (cDNA) encoding H6H from *H. niger* and expressed it in *A. belladonna*. The transgenic *A. belladonna* plants produced scopolamine because they were able to extend the metabolic pathway beyond its endogenous end-point. Another example of the production of novel chemicals in plants is the diversion of carbon backbones from fatty acid synthesis to the formation of polyhydroxyalkanoates, which form biodegradable thermoplastics (Steinbuchel & Fuchtenbusch 1998). In this case, the foreign genes derive not from other plants, but from bacteria.

**Theme 6: Plant breeding in the twenty-first century**

**Introduction to theme 6**

We have discussed several uses for transgenic plants earlier in this chapter, i.e. as bioreactors producing recombinant proteins (p. 285) and novel metabolites (see previous section). Transgenic plants can potentially express any foreign gene, whether that gene is derived from bacteria, yeast, other plants or even animals. The scope for exploitation and improvement is virtually limitless, and gene-manipulation techniques have therefore given the biotechnology industry a new lease of life. In the following sections, we consider the development of plant biotechnology in two key areas: improvement of agronomic traits and modification of production traits.

**Improving agronomic traits**

The initial focus of plant biotechnology was on improving agronomic traits, i.e. the protection of crops against pests, pathogens and weeds, and thus increasing yields. Major crop losses are caused every year by these so-called ‘biotic’ constraints, as well as physical (or ‘abiotic’) factors, such as flooding, drought, soil quality, etc. The aims of the biotechnology industry went hand in hand with those of conventional breeders, but offered the possibility of importing useful genes from distant species that could not be used for breeding. It has been found that, in many cases, single genes transferred from another organism can provide high levels of protection.

**Herbicide resistance**

Herbicides generally affect processes that are unique to plants, e.g. photosynthesis or amino acid biosynthesis (see Table 14.8). Both crops and weeds share
these processes, and developing herbicides that are selective for weeds is very difficult. An alternative approach is to modify crop plants so that they become resistant to broad-spectrum herbicides, i.e. incorporating selectivity into the plant itself rather than relying on the selectivity of the chemical. Two approaches to engineering herbicide resistance have been adopted. In the first, the target molecule in the cell either is rendered insensitive or is overproduced. In the second, a pathway that degrades or detoxifies the herbicide is introduced into the plant. An example of each strategy is considered below.

Glyphosate is a non-selective herbicide that inhibits 5-enol-pyruvylshikimate-3-phosphate (EPSP) synthase, a key enzyme in the biosynthesis of aromatic amino acids in plants and bacteria. A glyphosate-tolerant Petunia hybrida cell line obtained after selection for glyphosate resistance was found to overproduce the EPSP synthase as a result of gene amplification. A gene encoding the enzyme was subsequently isolated and introduced into petunia plants under the control of a cauliflower mosaic virus (CaMV) 35S promoter. Transgenic plants expressed increased levels of EPSP synthase in their chloroplasts and were significantly more tolerant to glyphosate (Shah et al. 1986). An alternative approach to glyphosate resistance has been to introduce a gene encoding a mutant EPSP synthase. This mutant enzyme retains its specific activity but has decreased affinity for the herbicide. Transgenic tomato plants expressing this gene under the control of an opine promoter were also glyphosate-tolerant (Comai et al. 1985). Following on from this early research, several companies have introduced glyphosate tolerance into a range of crop species, with soybean and cotton the first to reach commercialization (Nida et al. 1996, Padgette et al. 1996). Currently, nearly three-quarters of all transgenic plants in the world are resistant to glyphosate (James 2000).

Phosphinothricin (PPT) is an irreversible inhibitor of glutamine synthetase in plants and bacteria. Bialaphos, produced by Streptomyces hygroscopicus, consists of PPT and two alanine residues. When these residues are removed by peptidases the herbicidal component PPT is released. To prevent self-inhibition of growth, bialaphos-producing strains of S. hygroscopicus also produce the enzyme phosphinothricin acetyltransferase (PAT), which inactivates PPT by acetylation. The bar gene that encodes the acetylase has been introduced into potato, tobacco and tomato cells using Agrobacterium-mediated transformation. The resultant plants were

<table>
<thead>
<tr>
<th>Herbicide</th>
<th>Pathway inhibited</th>
<th>Target enzyme</th>
<th>Basis of engineered resistance to herbicide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyphosate</td>
<td>Aromatic amino acid biosynthesis</td>
<td>5-Enol-pyruvyl shikimate-3-phosphate (EPSP) synthase</td>
<td>Overexpression of plant EPSP gene or introduction of bacterial glyphosate-resistant aroA gene</td>
</tr>
<tr>
<td>Sulphonylurea</td>
<td>Branched-chain amino acid biosynthesis</td>
<td>Acetolactate synthase (ALS)</td>
<td>Introduction of resistant ALS gene</td>
</tr>
<tr>
<td>Imidazolinones</td>
<td>Branched-chain amino acid biosynthesis</td>
<td>ALS</td>
<td>Introduction of mutant ALS gene</td>
</tr>
<tr>
<td>Phosphinothricin</td>
<td>Glutamine biosynthesis</td>
<td>Glutamine synthetase</td>
<td>Overexpression of glutamine synthetase or introduction of the bar gene, which detoxifies the herbicide</td>
</tr>
<tr>
<td>Atrazine</td>
<td>Photosystem II</td>
<td>Qb</td>
<td>Introduction of mutant gene for Qb protein or introduction of gene for glutathione-S-transferase, which can detoxify atrazines</td>
</tr>
<tr>
<td>Bromoxynil</td>
<td>Photosynthesis</td>
<td></td>
<td>Introduction of nitrate transferase, which detoxifies bromoxynil</td>
</tr>
</tbody>
</table>
resistant to commercial formulations of PPT and bialaphos in the laboratory (De Block et al. 1987) and in the field (De Greef et al. 1989) (Fig. 14.22). More recently, it has been shown that bialaphos-resistant transgenic rice plants that were inoculated with the fungi causing sheath blight disease and subsequently treated with the herbicide were completely protected from infection (Uchimiya et al. 1993). This agronomically important result depends on the observation that bialaphos is toxic to fungi as well as being a herbicide. PPT resistance is widely used in plants as a selectable marker (see p. 231); however, it has also been introduced into a number of different crops for weed control, including sugar cane and rice (Gallo-Meagher & Irvine 1996, Oard et al. 1996). Thus far, PPT-resistant transgenic plants have not been commercially released.

The benefit of herbicide-resistant transgenic crops is the increased yield and seed quality as competing weed species are eliminated. However, there was initial concern that this would come with an associated penalty of increased herbicide use, which could have a serious impact on the environment. Contrary to these predictions, the introduction of herbicide-resistant plants has actually reduced chemical use by up to 80% in many areas, as farmers adopt better weed-control policies and switch to herbicides with low use rates. A further risk is that transgenes for herbicide tolerance could spread to weed species, resulting in a new breed of ‘superweeds’ (Kling 1996). It is too early to say whether this will be a problem. Although a range of herbicide-resistant transgenic crops are being tested, only crops resistant to glyphosate or bromoxynil are currently grown on a commercial scale. The benefits and risks of herbicide-resistant crops and strategies to combat gene transfer to weeds are discussed in recent reviews (Gressel 1999, 2000).

**Virus resistance**

Major crop losses occur every year as a result of viral infections, e.g. tobacco mosaic virus (TMV) causes losses of over $50 million per annum in the tomato industry. There is a useful phenomenon known as cross-protection, in which infection of a plant with one strain of virus protects against superinfection with a second, related strain. The mechanism of cross-protection is not fully understood, but it is believed that the viral coat protein is important. Powell-Abel et al. (1986) developed transgenic plants that express the TMV coat protein and which had greatly reduced disease symptoms following virus infection. Since that observation, the principle of heterologous coat-protein expression has been extended to many different plants and viruses (reviewed by Beachy et al. 1990). In the case of resistance to TMV, the coat protein must be expressed in the epidermis and in the vascular tissue, through which the virus spreads systemically (Clark et al. 1990). Transgenic squash containing multiple viral coat-protein genes and demonstrating resistance to cucumber mosaic virus, water-melon mosaic virus 2 and zucchini yellow mosaic virus was the first
virus-resistant transgenic crop to reach commercial production (Tricoli et al. 1995).

While the heterologous coat-protein approach can be successful, it has been demonstrated that, in many cases, the effect of transgene expression is mediated at the RNA rather than the protein level. This can be proved by generating transgenic plants carrying coat-protein genes that cannot be translated to yield functional protein, as first shown by Lindbo and Dougherty (1992) using the tobacco etch virus coat-protein gene. The transgene RNA apparently interferes with viral replication (a phenomenon called RNA-mediated viral resistance (RMVR)). This requires homology between the transgene and the target virus, and involves high-level transgene transcription but low-level accumulation of the transcript. This has much in common with post-transcriptional gene silencing (discussed in more detail in Chapter 13), which can lead to transgene silencing and the cosuppression of homologous endogenous genes, as well as viral resistance. The interested reader can consult several excellent reviews covering this and related phenomena (Waterhouse et al. 1998, Grant 1999, Plasterk & Ketting 2000, Hammond et al. 2001).

A different method of minimizing the effects of plant virus infection was developed by Gehrlach et al. (1987). They generated plants that expressed the satellite RNA of tobacco ringspot virus and such plants were resistant to infection with tobacco ringspot virus itself. Another potential method of inducing resistance to viruses is the production of antiviral proteins in transgenic plants. American pokeweed produces an antiviral protein called dianthrin that functions as a ribosome-inactivating protein. The cDNA for this protein has been cloned (Lin et al. 1991) and expressed in Nicotiana benthamiana (a relative of tobacco), providing resistance against African cassava mosaic virus (ACMV) (Hong et al. 1996). Interestingly in this experiment, the dianthrin gene was expressed under the control of an ACMV promoter, such that the antiviral protein was expressed only upon viral infection. In this manner, the toxic effects of constitutive transgene expression were avoided.

Antibodies specific for virion proteins have also been used to protect plants from viruses. In the first demonstration of this approach, Tavladoraki et al. (1993) expressed a single-chain Fv fragment (scFv) specific for ACMV in transgenic N. benthamiana, and demonstrated resistance to viral infection. Other groups have generated transgenic tobacco plants expressing antibodies specific for TMV, resulting in reduced infectivity. Voss et al. (1995) expressed full-size IgGs, while Zimmermann et al. (1998) expressed scFv fragments. Targeting scFv fragments to the plasma membrane also provides protection against virus infection (Schillberg et al. 2001).

**Resistance to microbial pathogens**

Progress has also been made in developing resistance to plant-pathogenic fungi which are traditionally controlled by appropriate farming practices (e.g. crop rotation) and the application of expensive and environmentally harmful fungicides. A straightforward approach is to engineer plants with antifungal proteins from heterologous species. This was first demonstrated by Broglie et al. (1991) who showed that expression of bean chitinase can protect tobacco and oil-seed rape from post-emergent damping off caused by Rhizoctonia solani. Plants synthesize a wide range of so-called ‘pathogenesis-related proteins’ (PR proteins), such as chitinase, which are induced by microbial infection. They also synthesize antifungal peptides called defensins and other antifungal proteins. As the genes for more of these proteins have been cloned and characterized, the number of transgenic plants constitutively expressing such proteins continues to rise. For example, tobacco osmotin has been expressed in transgenic potato, providing resistance to Phytophthora infestans (Liu et al. 1994), and in transgenic rice, providing resistance to R. solani (Lin et al. 1995). Instead of using a protein to provide direct protection, a metabolic-engineering strategy can be utilized. Phytoalexins are alkaloids with antifungal activity, and transforming plants with genes encoding the appropriate biosynthetic enzymes can increase their synthesis. Hain et al. (1993) generated tomato plants expressing the grapevine gene for stilbene synthase, and these plants demonstrated increased resistance to infection by Botrytis cinerea. Similarly, Anzai et al. (1989) have used a bacterial gene facilitating tabtoxin detoxification to protect tomato plants against Pseudomonas syringae infection.
An alternative to the use of antifungal proteins or metabolites is to manipulate the hypersensitive response, which is a physiological defence mechanism used by plants to repel attacking pathogens. Resistance occurs only in plants carrying a resistance gene \((R)\) that corresponds to an avirulence \((avr)\) gene in the pathogen. Elicitors (signalling molecules) released by pathogens are detected by the plant and activate a range of defence responses, including cell death, PR-gene expression, phytoalexin synthesis and the deposition of cellulose at the site of invasion, forming a physical barrier. Importantly, the hypersensitive response is systemic, so that neighbouring cells can pre-empt pathogen invasion. A recently developed strategy is to transfer avirulence genes from the pathogen to the plant, under the control of a pathogen-inducible promoter. This has been demonstrated in tomato plants transformed with the \(avr9\) gene from \(Chladosporium fulvum\), resulting in resistance to a range of fungal, bacterial and viral diseases (Keller et al. 1999, reviewed by Melchers & Stuiver 2000).

**Resistance to insects and other pests**

Insect pests represent one of the most serious biotic constraints to crop production. For example, more than a quarter of all the rice grown in the world is lost to insect pests, at an estimated cost of nearly US$50 billion. This is despite an annual expenditure of approximately US$1.5 billion on insecticides for this crop alone. Insect-resistant plants are therefore desirable not only because of the potential increased yields, but also because the need for insecticides is eliminated and, following on from this, the undesirable accumulation of such chemicals in the environment is avoided. Typical insecticides are non-selective, so they kill harmless and beneficial insects as well as pests. For these reasons, transgenic plants have been generated expressing toxins that are selective for particular insect species.

Research is being carried out on a wide range of insecticidal proteins from diverse sources. However, all commercially produced insect-resistant transgenic crops express toxin proteins from the Gram-positive bacterium *Bacillus thuringiensis* (Bt). Unlike other *Bacillus* species, Bt produces crystals during sporulation, comprising one or a small number of ~130 kDa protoxins called crystal proteins. These proteins are potent and highly specific insecticides. The specificity reflects interactions between the crystal proteins and receptors in the insect midgut. In susceptible species, ingested crystals dissolve in the alkaline conditions of the gut and the protoxins are activated by gut proteases. The active toxins bind to receptors on midgut epithelial cells, become inserted into the plasma membrane and form pores that lead to cell death (and eventual insect death) through osmotic lysis. Approximately 150 distinct Bt toxins have been identified and each shows a unique spectrum of activity.

Bt toxins have been used as topical insecticides since the 1930s, but never gained widespread use, because they are rapidly broken down on exposure to daylight and thus have to be applied several times during a growing season. Additionally, only insects infesting the exposed surface of sprayed plants are killed. These problems have been addressed by the expression of crystal proteins in transgenic plants. Bt genes were initially introduced into tomato (Fischhoff et al. 1987) and tobacco (Barton et al. 1987, Vaeck et al. 1987) and later into cotton (Perlak et al. 1990), resulting in the production of insecticidal proteins that protected the plants from insect infestation. However, field tests of these plants revealed that higher levels of the toxin in the plant tissue would be required to obtain commercially useful plants (Delannay et al. 1989). Attempts to increase the expression of the toxin gene in plants by use of different promoters, fusion proteins and leader sequences were not successful. However, examination of the bacterial *cry1Ab* and *cry1Ac* genes indicated that they differed significantly from plant genes in a number of ways (Perlak et al. 1991). For example, localized AT-rich regions resembling plant introns, potential plant polyadenylation signal sequences, ATTTA sequences that can destabilize mRNA and rare plant codons were all found. The elimination of undesirable sequences and modifications to bring codon usage into line with the host species resulted in greatly enhanced expression of the insecticidal toxin and strong insect resistance of the transgenic plants in field tests (Kostiel et al. 1993). By carrying out such enhancements, Perlak and colleagues expressed a modified *cry3A* gene in potato to provide resistance against Colorado beetle
In 1995, this crop became the first transgenic insect-resistant crop to reach commercial production, as NewLeaf™ potato marketed by Monsanto. The same company also released the first commercial transgenic, insect-resistant varieties of cotton (Bollgard™, expressing cry1Ac and protected against tobacco bollworm) and maize (YieldGard™, expressing cry1Ab and resistant to the European corn-borer). Many other biotechnology companies have now produced Bt-transgenic crop plants resistant to a range of insects.

Although Bt-transgenic plants currently dominate the market, there are many alternative insecticidal proteins under investigation. Two types of protein are being studied in particular: proteins that inhibit digestive enzymes in the insect gut (proteinase and amylase inhibitors) and lectins (carbohydrate-binding proteins). Research into these alternatives is driven in part by the fact that some insects are not affected by any of the known Bt crystal proteins. Homopteran insects, mostly sap-sucking pests such as planthoppers, fall into this category, but have been shown to be susceptible to lectins such as Galanthus nivalis agglutinin (GNA). This lectin has been expressed in many crops, including potato (Shi et al. 1994, Gatehouse et al. 1996), rice (Bano-Maqbool & Christou 1999), tomato and tobacco (reviewed by Schuler et al. 1998).

**Modification of production traits**

Production traits can include many characteristics, such as the colour and smell of flowers and the taste, consistency and nutritional composition of food. While modifying flower colour would seem a harmless pursuit (experiments in this area are discussed on p. 261), the genetic modification of food for human consumption is the subject of intensive current debate. Opinions as to the benefits and risks of the latter differ widely according to region, ethical standpoint and financial/food security of the nation. There is no doubt, however, that gene manipulation has the potential to greatly increase the quality of food. Examples include modifying the starch or oil content, facilitating the accumulation of vitamins and minerals in foods that are normally poor in such nutrients, and altering the properties of fruits and vegetables to delay ripening, increase the levels of soluble solids and facilitate processing. Several examples are discussed below and Box 14.4 describes the recently developed terminator technology, which provides a mechanism by which producers of transgenic crops could dictate where and when their modified plants are grown.

**Delayed ripening**

The earliest example of quality improvement by gene transfer was the introduction of antisense constructs targeting the polygalacturonase (pg) gene in tomato to delay ripening (Sheehy et al. 1988, Smith et al. 1988), as discussed in Chapter 13 (p. 260). Other strategies to delay ripening include: cosuppression of the pg gene using sense constructs (Smith et al. 1990), the use of antisense RNA to suppress the expression of two key enzymes required for ethylene synthesis (Hamilton et al. 1990, Oeller et al. 1991) and the degradation of a key intermediate in the ethylene biosynthetic pathway (Klee et al. 1991: see review by Fray & Grierson 1993).

**Starch and oil modification**

Much research has been carried out into strategies for the modification of the starch and oil content of seeds. In each case, it is essential to have a thorough understanding of the normal biosynthetic pathways. Starch is the major storage carbohydrate in higher plants. It consists of two components: amylose, which is a linear polymer, and amylopectin, which is a branched polymer. The physicochemical, nutritional and textural properties of food are significantly influenced by the nature of the starch. For example, cooked amylose produces fibre-like structures that are resistant to digestion, serve as dietary fibre and require less insulin for metabolism. Amylopectin, in contrast, is waxy and viscous. The food and other industries use a wide range of different starches. These are obtained by sourcing the starch from different plant varieties coupled with chemical and enzymatic modification. The genetic modification of plants offers a new approach to creating novel starches with new functional properties.

The enzyme ADP-glucose pyrophosphorylase (GP) catalyses the first step in starch biosynthesis in plants. Stark et al. (1992) cloned the gene for an E.
The production of any marketable commodity requires a great deal of investment in terms of research and development. Most commodities are not self-renewing, so manufacturers can recover their investments through long-term sales. Transgenic plants, however, are capable of reproduction. Farmers have to make an initial outlay for seed purchase, but, by saving a proportion of seeds from each harvest, they need never return to the manufacturer again. For producers who have invested heavily in development, this is unacceptable. So how can the producers protect their investment?

Until recently, this was achieved through the use of contracts that obliged farmers not to save any seeds. However, such contracts are difficult to enforce, particularly in developing countries. A novel strategy is to modify the plants so that the seeds are sterile, thereby forcing the farmer to return to the manufacturer year after year for fresh supplies of seeds. This strategy, which has been termed technology protection by the biotechnology companies and terminator technology by disgruntled consumers and opponents, is perhaps the most controversial potential use of genetically modified plants. In essence, the technology works rather like the anti-copying devices now incorporated into video recorders – the industry is protecting its investment by preventing unauthorized duplication of its products. The principle of terminator technology is the expression of a toxic transgene, the terminator gene, at a critical stage of embryonic development, thus killing the embryo and rendering the seeds incapable of germination. Several variants on this technology were described in a patent application, granted in March 1998, to Pine Lands Corporation, which has since been bought by Monsanto. One of the variations is discussed below.

In this example, the terminator gene encodes a ribosomal-inactivating protein and is expressed under the control of a promoter that is active in late embryogenesis, such as the promoter of the *lea* (late embryogenesis abundant) gene. Seeds produced from such plants are sterile, but, since seed storage products (e.g. starch and oil) accumulate in early and mid-embryonic development, the seeds are nutritionally unimpaired. In order to maintain a stock of fertile plants, the terminator gene is controlled by inducible site-specific recombination, as described in detail in Chapter 13. The gene is rendered inactive by inserting a blocking element between the gene and its promoter. This blocking element is flanked by *loxP* sites. The transgenic plant also contains a cre recombinase gene, which is controlled by a tetracycline-inducible promoter, and a third transgene encoding the tetracycline repressor protein, which is constitutively expressed. In the absence of tetracycline, the repressor prevents cre expression, the terminator gene is not activated and the plants can be grown as normal, allowing the producer to grow plants and produce seeds in unlimited quantity. Before distribution, however, the seeds are soaked in tetracycline, which causes the repressor to release the cre promoter, thus inducing the cre gene, leading to excision of the blocking fragment and activation of the terminator gene. Since the terminator gene is controlled by an embryonic promoter, it is not switched on until the following generation of developing seeds.

The advent of terminator technology was met by a public outcry, particularly from farmers and from environmentalists concerned that terminator genes could spread from the transgenic crops into wild plants, with unknown consequences. Perhaps in response to this, Monsanto has since pledged not to implement the technology now at its disposal.

---

**Box 14.4 Protecting your investment: terminator technology**

The production of any marketable commodity requires a great deal of investment in terms of research and development. Most commodities are not self-renewing, so manufacturers can recover their investments through long-term sales. Transgenic plants, however, are capable of reproduction. Farmers have to make an initial outlay for seed purchase, but, by saving a proportion of seeds from each harvest, they need never return to the manufacturer again. For producers who have invested heavily in development, this is unacceptable. So how can the producers protect their investment?

Until recently, this was achieved through the use of contracts that obliged farmers not to save any seeds. However, such contracts are difficult to enforce, particularly in developing countries. A novel strategy is to modify the plants so that the seeds are sterile, thereby forcing the farmer to return to the manufacturer year after year for fresh supplies of seeds. This strategy, which has been termed technology protection by the biotechnology companies and terminator technology by disgruntled consumers and opponents, is perhaps the most controversial potential use of genetically modified plants. In essence, the technology works rather like the anti-copying devices now incorporated into video recorders – the industry is protecting its investment by preventing unauthorized duplication of its products. The principle of terminator technology is the expression of a toxic transgene, the terminator gene, at a critical stage of embryonic development, thus killing the embryo and rendering the seeds incapable of germination. Several variants on this technology were described in a patent application, granted in March 1998, to Pine Lands Corporation, which has since been bought by Monsanto. One of the variations is discussed below.

In this example, the terminator gene encodes a ribosomal-inactivating protein and is expressed under the control of a promoter that is active in late embryogenesis, such as the promoter of the *lea* (late embryogenesis abundant) gene. Seeds produced from such plants are sterile, but, since seed storage products (e.g. starch and oil) accumulate in early and mid-embryonic development, the seeds are nutritionally unimpaired. In order to maintain a stock of fertile plants, the terminator gene is controlled by inducible site-specific recombination, as described in detail in Chapter 13. The gene is rendered inactive by inserting a blocking element between the gene and its promoter. This blocking element is flanked by *loxP* sites. The transgenic plant also contains a cre recombinase gene, which is controlled by a tetracycline-inducible promoter, and a third transgene encoding the tetracycline repressor protein, which is constitutively expressed. In the absence of tetracycline, the repressor prevents cre expression, the terminator gene is not activated and the plants can be grown as normal, allowing the producer to grow plants and produce seeds in unlimited quantity. Before distribution, however, the seeds are soaked in tetracycline, which causes the repressor to release the cre promoter, thus inducing the cre gene, leading to excision of the blocking fragment and activation of the terminator gene. Since the terminator gene is controlled by an embryonic promoter, it is not switched on until the following generation of developing seeds.

The advent of terminator technology was met by a public outcry, particularly from farmers and from environmentalists concerned that terminator genes could spread from the transgenic crops into wild plants, with unknown consequences. Perhaps in response to this, Monsanto has since pledged not to implement the technology now at its disposal.

---

(a) No tetracycline

(b) Tetracycline present

coli mutant GP which was deficient in allosteric regulation. When this gene was inserted into potato plants, the tubers accumulated higher levels of starch but had normal starch composition and granule size. Since transgenic plants expressing the wild-type E. coli GP had normal levels of starch, the allosteric regulation of GP, and not its absolute level, must influence starch levels. A different method of modulating GP activity has been used by Muller-Rober et al. (1992). They used an antisense construct to reduce the level of plant GP to 2–5% of wild type. These plants had very low levels of starch (also 2–5% of wild type) but had a very much higher number of tubers. These tubers had six- to eightfold more glucose and sucrose and much lower levels of patatin and other storage proteins. Visser et al. (1991) have modified the starch composition of potato plants. They demonstrated that potato plants expressing an antisense gene to the granule-bound starch synthase have little or no amylase compared with wild-type potato, where it can be as much as 20%.

Whereas some plants accumulate starch as a carbon reserve, others high levels of triacylglycerols in seeds or mesocarp tissues. Higher plants produce over 200 kinds of fatty acids, some of which are of food value. However, many are likely to have industrial (non-food) uses of higher value than edible fatty acids (Kishore & Somerville 1993, Murphy 1999). Thus there is considerable interest in using gene-manipulation techniques to modify the fatty acid content of plants. As a first step towards this goal, a number of genes encoding desaturases have been cloned, e.g. the MS desaturase, which converts linoleic acid to α-linolenic acid (Arondel et al. 1992).

In most plants the Δ9 stearoyl acyl carrier protein (ACP) desaturase catalyses the first desaturation step in seed-oil biosynthesis, converting stearoyl-ACP to oleoyl-ACP. When antisense constructs were used to reduce the levels of the enzyme in developing rape-seed embryos, the seeds had greatly increased stearate levels (Knutzon et al. 1992). The significance of this result is that high stearate content is of value in margarine and confectionery fats.

Most oil-seed crops accumulate triacylglycerols with C_{16} to C_{18} acyl chains. However, plants that accumulate medium-chain triacylglycerols (MCT), with C_{8} to C_{12} acyl chains, would be very useful, for these lipids are used in detergent synthesis. Some plants, such as the California bay, accumulate MCT because of the presence of a medium-chain specific thioesterase. When the gene for this thioesterase was expressed in rape-seed and Arabidopsis, the transgenic plants accumulated high levels of MCT (Voelker et al. 1992).

**Improving vitamin and mineral content**

The vitamin and mineral content of plants varies from species to species, and from tissue to tissue in a particular plant. Millled cereal grains – basically the endosperm component of the seed – are particularly deficient in essential vitamins and minerals, and yet they represent the staple food for much of the world’s population. In developed countries, where diet is varied, this is seldom a problem. However, in poorer nations, where cereal grains are often the only food available, this can cause major health problems. Iron deficiency is the most widespread nutrient deficiency in the world, reflecting the combination of low iron in cereal grains and the high level of phytic acid, which reduces the efficiency of iron absorption from the gut. Many cereal grains also lack β-carotene, which is required in the diet for vitamin A biosynthesis. Vitamin A deficiency in early childhood leads to blindness, an avoidable consequence of poor diet in many parts of Asia and Africa.

A number of attempts to increase the nutritional value of cereals have been reported, and we discuss two examples of genetically engineered rice below, since this cereal is the staple diet of at least two-thirds of the world’s population. Lucca et al. (2001) described three different routes to improving the iron content of rice grain, which could be used singly or in combination. The first route was to transform rice with a gene encoding ferritin, an iron-storage protein. The overexpression of bean ferritin in rice endosperm led to the accumulation of iron in a bioavailable form, resulting in transgenic rice grains with twice the normal levels of iron. The second route was to transform rice with a gene from the fungus Aspergillus fumigatus encoding the enzyme phytase, which metabolizes phytic acid. Since the transgenic grain has lower levels of phytic acid, iron absorption from the gut should be more efficient. The third approach was to transform rice with a
Applications of recombinant DNA technology

construct causing the overexpression of an endogenous cysteine-rich metallothionein-like protein, since cysteine peptides have been shown to have the opposite effect to phytic acid, i.e. they enhance iron absorption. The resulting rice grains had over five times the normal level of cysteine residues and a 130-fold increase in phytase activity, theoretically sufficient to eliminate all phytic acid from the grains.

Burkhardt et al. (1997) described rice plants transformed with the phytoene synthase gene from the daffodil (Narcissus pseudonarcissus). The enzyme encoded by this gene represents the first of four steps in the \( \beta \)-carotene biosynthesis pathway, and the plants were shown to accumulate the immediate product of the recombinant enzyme, phytoene. Further work by the same group (Ye et al. 2000) resulted in rice plants expressing several daffodil enzymes simultaneously in the endosperm, and recapitulated the entire heterologous pathway. Hence, the transgenic rice grains were highly enriched for \( \beta \)-carotene (provitamin A) and were consequently golden in colour.

Epilogue: from genes to genomes

Principles of Gene Manipulation tells the story of the development of genetic-engineering technology, a story that began in the mid-1970s. When the first edition of this book was published in 1980, basic techniques, such as cloning, sequencing and expression analysis, were all in their infancy and many of today’s most widely used procedures, including the PCR, had not been invented. In those days, the cloning and characterization of a single gene was an elaborate task that required a great deal of inventiveness and hard work. The sequencing and analysis of entire cellular genomes was an impossible goal, as far-fetched as space travel was to people 100 years ago.

Over the last 25 years, however, the technology of gene manipulation has rapidly increased in sophistication. Through the sharing of information between laboratories, standard molecular-biology procedures have become more robust and reliable, new techniques have evolved, and many of the more laborious procedures that scientists had to carry out in the early years have been streamlined or automated. Biotechnology companies now take onthe burden of developing and refining the tools and protocols, producing high-quality enzymes, specialist molecular-biology kits and premade custom genomic or cDNA libraries. Most importantly, there has been a drive to develop high-throughput versions of many of the techniques discussed in this book, allowing more and more samples to be processed in parallel, producing more data in less time than ever before. Sequencing has been at the forefront of this paradigm shift, with the result that the sequencing of entire genomes is now commonplace. As this sixth edition of Principles of Gene Manipulation goes to press, over 40 bacterial genomes plus those of four model eukaryotes have been completely sequenced and the first draft of the entire human genome sequence has been published (International Human Genome Sequencing Consortium 2001, Venter et al. 2001).

These extraordinary achievements have ushered in a new era in molecular biology, which is analysis at the whole-genome level. More data are emerging from genome-sequencing projects and the random cloning of partial cDNAs (known as expressed sequence tags (ESTs)) than can be dealt with using traditional single-gene analysis methods. A new breed of high-throughput functional-analysis techniques has been developed, some of which are discussed briefly in Chapters 6, 9 and 13 of this book. Simple northern blots and reverse transcription (RT)-PCR experiments are being replaced by the use of DNA chips containing hundreds or thousands of individual sequences that can be hybridized with a suitable probe simultaneously. Instead of analysing single proteins by western blots, two-dimensional gel electrophoresis is being used to simultaneously study every single protein expressed under various conditions. The creation and functional analysis of individual mutants is being superseded by systematic approaches to mutating and characterizing every single gene in the genome. Techniques are also being developed for the analysis of all possible protein–protein interactions in the cell.

While the impact of the new genomics is discussed in this book, as least as far as it affects trends in gene-manipulation technology, a full account of its development, application and potential requires a book of its own. The story continues in our sister text, Principles of Genome Analysis and Genomics (third edition).