GENETIC ENGINEERING, MICROBES

1. Introduction

Genetic engineering of microbes has undergone a major transformation during the last decade as many new tools have become available. Genomics, proteomics, fluxomics, and bioinformatics have changed the landscape of molecular biology by providing direct access to entire microbial genomes and insight concerning how global expression of these genomes changes in response to various environmental conditions. A comprehensive review of genetic engineering is a formidable task. Thus, this article is divided into three sections that focus on important aspects of genetic engineering of microbes. The first section will summarize some of the technologies that are routinely used for genetic engineering of microbes. The second section will focus on specific microbes that are important to industry for genetic engineering. The final section will review products and future opportunities that involve microbes.

2. Gene Technologies

The techniques for isolating and manipulating genes from microbes have become routine (1). Essentially, genes are cloned into plasmid- or phage-based vectors and introduced into the microbe of choice. A well-planned selection scheme or screening method is then necessary to identify host microbes that contain the cloned gene of interest. Most of the reagents that are needed for gene cloning are commercially available and are now frequently supplied as kits for routine tasks such as extraction of DNA, purification of DNA framents for cloning, and isolation of plasmids. In addition, construction and screening of gene libraries are performed by small biotechnology companies.

2.1. Gene-Transfer Methods. An efficient method of introducing engineered DNA into microbes is essential for gene manipulations. Generally, methods of transformation in which microbes take up DNA directly from the culture medium are the most useful for genetic engineering. Certain microbes such as *Bacillus subtilis* are naturally competent for DNA transformation during a particular growth phase. Other bacteria such as *Escherichia coli* are made competent for transformation by treatment with calcium chloride or rubidium chloride. Another approach involves removing the cell walls of bacteria and fungi and transforming the resulting spheroplasts with DNA in the presence of polyethylene glycol. However, since many bacteria cannot be made competent for transformation by any of these methods, electroporation and conjugation have become important techniques for inserting engineered DNA into industrially useful microbes (2,3).

When bacteria are exposed to an electric field, a number of physical and biochemical changes occur. The bacterial membrane becomes polarized at low electric field strength. When the membrane potential reaches a critical value of 200–300 mV, areas of reversible local disorganization and transient breakdown occur, resulting in a permeable membrane and molecular flux. The nature

of the membrane disturbance is not clearly understood, but bacteria, yeast and fungi are capable of DNA uptake following exposure to approriate electric field conditions (see YEASTS). This method, called electroporation, has been used to transform a variety of bacteria and yeast strains that are recalcitrant to other methods (2). The apparatus for electroporation is commercially available, and constant improvements in the design are being made.

Transformation and electroporation are not effective for all microbes. Conjugation is a practical method for introducing DNA into bacteria that can not be transformed. Conjugation is a process by which certain plasmids (conjugative or self-transmissible plasmids) are able to transfer copies of themselves from a host cell to recipient cells. Some plasmids are mobilizable. Mobilizable plasmids are not self-transmissible, ie, do not have the genes necessary for conjugation, but copies of a mobilizable plasmid can be moved to a recipient by a self-transmissible plasmid. Several mobilizable plasmid vectors have been developed that replicate in a broad range of bacterial species. These broad host range mobilizable vectors are transferred to recipient bacteria by use of a "helper" plasmid. Helper plasmids have all of the genes necessary for conjugation but have a very narrow host range for replication that is essentially limited to *E. coli*. The DNA can be cloned into a mobilizable plasmid vector and transformed into *E. coli* as the initial host. The cloned DNA can then be transferred into the desired nontransformable host in a triparental mating that involves mixing recipient cells with *E. coli* cells that contain the helper plasmid and *E. coli* cells that contain the mobilizable vector (3).

Two general types of mobilizable vectors are used for genetic manipulation of bacteria. The first type of mobilizable vector replicates in *E. coli* and in recipient cells. These vectors are used to maintain and express cloned genes in recipient bacteria on replicating plasmids. The second type of vector is usually called a "suicide" vector because it replicates in the *E. coli* donor but fails to do so in a recipient. Suicide vectors are useful because genes on a suicide vector can be used to select for homologous and nonhomologous recombination events that result in foreign DNA from the suicide vector being integrated into the chromosome of the recipient cell. Minitransposons can be used in combination with suicide vectors to isolate insertion mutations and to stably integrate cloned DNA into the chromosome (4).

2.2. Gene Amplification. The technique to amplify DNA *in vitro*, known as the polymerase chain reaction (PCR), has turned out to be extremely powerful and is widely used for gene manipulations (5,6). The PCR is highly efficient, and low amounts of target DNA are sufficient for amplification of target sequences. As a result, PCR is used in diagnostics and forensic science. PCR can also be used for site-directed mutagenesis, for domain swapping in proteins and for isolating homologous genes.

PCR is carried out in a series of automated cycles. Each cycle has three steps: denaturation, annealing, and elongation. Denaturation of double-stranded DNA at high temperature ($\sim$95°C) results in the template DNA (ie, the DNA that contains the target sequence that will be amplified) becoming single stranded. The annealing step allows primers to bind to the single stranded DNA. The primers are designed to anneal to complementry sequences that flank the target sequence so that the 3′ end of each primer is nearest the target sequence.

The stringency or specificity of the annealing reaction can be controlled by temperature. More mismatch between the primers and the template DNA can occur at lower annealing temperatures (below ~45°C) than at higher temperatures (~46–55°C). During the elongation step, DNA polymerase extends each primer from the 3′ end. The DNA polymerase from a thermophile allows the elongation to be carried out at a high temperature (72°C). Thermocyclers and the reagents needed for PCR are sold commercially by several companies.

2.3. Gene Alteration. A variety of procedures have been described for introducing specific mutations (base pair changes, deletions, or insertions) into a gene or to subject a specific gene to random mutagenesis. Among these techniques, PCR has proven to be particularly useful for site-directed mutagenesis (7). The basic approach to PCR based site directed mutagenesis involves using primers and templates that have mismatches. Mismatching a primer to the template allows introduction of specific base pair changes into the amplified sequence. Such mismatches can be used to construct convenient restriction sites for gene cloning experiments or to modify a protein's activity by altering its coding sequence. It is also possible to enhance the naturally high error rate of *Taq* polymerase to induce random mutations in a specific gene.

There has been an increase in the number of methods that generate variants of a protein by recombining similar genes with sequences that have diverged because of natural variation or random mutagenesis. One such technique is DNA shuffling, which mimics homologous recombination by using PCR to reassemble random fragments of genes (8–10). Another method involves a combinatorial approach to create a library of all possible single base pair deletions in a given piece of DNA by utilizing exonuclease III to incrementally truncate genes, gene fragments or gene libraries (11). For several variations on these methodologies see the review by Volkov and Arnold (12). Directed evolution of proteins for biocatalysis applications has been reviewed recently (13).

2.4. Genomics. Venter and Smith initiated a shotgun approach to sequencing microbial genomes that has transformed the landscape of microbial genomics (14). This random sequencing approach was initially used successfully to complete the 1.8 Mb genome of *Haemophilus influenza*. Over 100 complete microbial genomes are now available to the public in the NCBI database (www.ncbi.nlm.nih.gov) (15). Comparison of genome sequences is extremely useful in the manipulation of microbes for production of small molecules or proteins. Just as the sequence of the human genome will aid the pharmaceutical industry in discovery of new therapeutics, the sequences of microbial genomes and comparative genomics will accelerate production of industrial chemicals and proteins in a wide range of microbes that so far have not been amenable to engineering by classical methods (16,17). The genome sequences are also a rich source for a large number of gene sequences that can be engineered as biocatalysts (18).

2.5. Metabolic Engineering. Much of the early work on genetic engineering of microbes focused on manipulation of single genes with the goal of producing therapeutic proteins and industrial enzymes (insulin, growth hormone, subtilisin, etc). However, the development of a variety of new techniques has made it possible to alter entire biochemical pathways that are encoded by multiple genes to overproduce small molecules such as amino acids, vitamins or chemicals (19). Several small molecules are currently produced commercially by

biocatalysts that are the result of metabolic engineering. Development of new biocatalysts is being enhanced by coupling the traditional approaches of metabolic engineering with new molecular techniques. For example, DNA microarrays can be useful in identifying the rate limiting steps in regulatory networks (20).

2.6. DNA Microarray. DNA microarrays have two major applications, analysis of gene expression and detection of specific nucleotide sequences. The overwhelming advantage of DNA microarrays over other DNA hybridization technologies is that DNA microarrays can be used to analyze thousands of individual genes simultaneously. Hence, a DNA microarray can be used to analyze the regulation of gene expression for an entire genome or a large portion of a genome under different conditions. In addition, DNA microarrays can be used in high throughput approaches for gene discovery, measuring abundance of mRNAs and comparative genomics (21). A detailed description of the methodology can be found at www.gene-chips.com. Basically, a robot is used to spot a large number of samples of probe DNA onto a glass slide or a nylon membrane in an ordered array. The probe DNA represents a set or subset of genes (eg, all of the genes in a microorganism) and can be PCR products or oligonucleotides. The probe DNA is then hybridized with target DNAs. One typical application involves extracting messenger RNA (mRNA) from two cultures of the same organism that have been grown under different conditions (eg, rich growth medium versus minimal growth medium). The two samples of mRNA are transcribed into cDNA and labeled with fluorescent dyes using reverse transcriptase to produce two populations of target DNAs. The amount of the target DNA of interest can be detected based on hybridization (22,23). As a result, comparing separate hybridization of the two target DNA populations with the same probe DNA can reveal which genes are expressed at different levels under the two growth conditions (24–26). Gene expression profiles for several microbes such as *E. coli*, *B. subtilis*, *Saccharomyces cerevisae*, and *Corynebacterium glutamicum* have been studied. This method has been used also to determine the metabolic fluxes in *E. coli* and yeast (27,28).

Another elegant method to rapidly analyze levels of specific mRNAs in a population of transcripts has been developed by Brenner and co-workers (29). This method involves sequencing cDNA molecules that are tethered to microbeads to generate a unique sequence signature for each type of transcript. The novel sequencing procedure is performed in a flow cell so that physical separation of DNA fragments is not required to generate the sequence information. Powerful statistical methods are then used to count sequence signatures and determine functional relationships of genes.

2.7. Bioinformatics. Various computational tools have become essential for genetic engineering of microbes. The ability to genetically engineer microbes is greatly enhanced by the tremendous amount of sequence information that is now available. The number of nucleotide and protein sequences in public and private databases is rapidly increasing. There are a variety of software packages available that can be used to access, manipulate and analyze the sequences in these databases.

The National Center for Biotechnology Information (NCIB) maintains a variety of publicly accessible databases and bioinformatics tools that are

extremely useful (www.ncbi.nlm.nih.gov/Sitemap/index.html). GenBank contains nucleotide sequences from more than 130,000 organisms (www.ncbi.nlm.nih.gov:80/Genbank/GenbankOverview.html). Many of these sequences have been annotated and also translated into the corresponding amino acid sequences. Two portals are particularly useful for retrieving and analyzing GenBank sequences. *Entrez* is supported by the NCIB as a user friendly web page (www.ncbi.nlm.nih.gov:80/Entrez/index.html) so that Boolean searches of GenBank can be performed using key words such as the name for a particular organism or enzyme. The NCIB also supports a web page for the BLAST algorithm (www.ncbi.nlm.nih.gov:80/BLAST/). BLAST takes a user provided query sequence (nucleotide or amino acid) to search GenBank for related sequences and then aligns the query sequence with each of the related sequences.

2.8. Sources of Microbial Genes. Rapid sequencing methods and other molecular biology tools have greatly facilitated isolation and characterization of genes and microbes with relevant catalytic activities. With the development of PCR, any gene for which the sequence and species are available can be cloned. Even a gene of unknown sequence can be cloned by designing PCR primers that are based on conserved portions of a consensus sequence derived from related genes of known sequence from other species. The microbial populations in a variety of extreme environments have been explored as sources of novel biocatalysts. However, microbes with unique biochemical pathways have been isolated from more easily accessible sources such as wastewater bioreactors (30). In addition, several methods to isolate genes from mixed microbial consortia have been reported (31).

2.9. Expression Vectors. A vector that allows optimal expression of cloned genes is required for engineering a microbe to produce a protein or small molecule. Many useful general purpose and specialized purpose expression vectors are commercially available. Basic expression vectors typically contain an efficient, inducible promoter and convenient restriction sites that allow a coding sequence to be cloned in a proper translational reading frame with a start codon and Shine-Dalgarno sequence. Some expression vectors have additional specialized features. For example, several expression plasmids are designed so that translation of the cloned gene results in the expressed protein having a polyhistidine tail at either the amino terminus or the carboxy terminus. These tails are useful in affinity purification of the expressed proteins.

2.10. Display of Peptides and Proteins on Phages. The ability to express proteins on the surface of phage was reported in 1985 (32). Phage display of peptides and proteins is an extremely versatile method that allows large random libraries to be screened in a relatively short time for a specific function, eg, binding of a hormone (see HORMONES) to a receptor (33,34). Bacteriophages are viruses that specifically infect bacteria. A bacteriophage consists of a proteinaceous capsid surrounding genetic material, ie, either DNA or RNA. The proteinaceous capsid is made up of several proteins that have defined molecular shapes and architecture. In the phage display method, randomly synthesized oligonucleotides are cloned into a gene for a minor capsid protein. Expression of the modified capsid gene results in the minor capsid protein being fused to a polypeptide that is exposed to the outside of the capsid. Phages displaying peptide sequences that bind to a specific ligand are identified by a protocol called biopanning. Certain variations in the basic methodology have been developed. For example,

peptide sequences that bind receptors or mimic other nonproteinaceous compounds have been identified.

Phage display is a powerful technique because it is possible to isolate a peptide that has a specific binding affinity from a pool of 10^{12} random peptides in a relatively short period of time. This method has attracted the attention of several biotechnology companies because of the potential to provide much information for rational drug design and vaccine development. This technology has been applied predominantly for discovery of antibodies. However, phage display is having an increasing impact on a variety of applications that involve peptides and proteins. Phage display has accelerated protein engineering studies and potentially can be applied for de novo design of proteins along with computational studies (34). Phage display has been used for altering properties of an industrial enzyme (35). Peptides that bind to semiconductor materials have been identified by phage display (36). Such peptides may serve as scaffolds for controlled assembly of inorganic materials and small molecules.

2.11. Expression of Functional Antibodies. Monoclonal and polyclonal antibodies are used in both diagnostics and research. Kretzschmaer and von Ruden state that 30% of the human antibodies in clinical trial currently have been developed by phage display (37). Isolation of a specific monoclonal antibody having high affinity to antigen is expensive and laborious. A combination of gene amplification technology along with phage display provides an excellent method to express a repertoire of antibodies in a short time (38).

3. Host Systems for Gene Expression

One of the potentials of genetic engineering of microbes is production of large amounts of recombinant proteins and small molecules. A major challenge related to expression of functional proteins in heterologous hosts is that each protein is unique and the stability of the protein depends on the host. Thus, it is not feasible to have a single omnipotent microbial host for the production of all recombinant proteins. Rather, several microbial hosts have to be studied. Expression vectors have to be tailored to the microbe of choice which is dictated by the process economics and feed stock issues (39).

3.1. Escherichia Coli. A well-developed genetic system and a large base of biochemical information initially made *E. coli* the host organism of choice for gene manipulations in the 1970s. Since that time, the genetic engineering techniques for *E. coli* have undergone continuous refinement (1,39). The complete genome of *E. coli* has been sequenced and biochemical information concerning *E. coli* has greatly increased. Media and growth optimization for fermentation of recombinant *E. coli* strains have resulted in cell density yields corresponding to values of optical density at 550 nm (OD_{550}) of 100–150. As a result, *E. coli* continues to be a preferred host organism for the expression of heterlogous proteins and production of small molecules.

Some of the advances in high level expression of genes in *E. coli* have been reviewed by Makrides (40). Several expression vectors have been designed so that the recombinant protein can be produced as a fusion protein having an affinity tail at the C-terminus that can be used for purification. Improved methods

for assembly of oligonucleotides to form synthetic genes for the biosynthetic production of protein polymers has been reported (41). High levels of intracellular expression of proteins in *E. coli* often result in the formation of inclusion bodies that can be separated from debris and soluble cellular proteins, solubilized, and refolded to biological activity (39). However, the denaturation and refolding process is expensive. Therefore, vectors to produce properly folded proteins in *E. coli* have been developed. This is accomplished by targeting the recombinant protein to the periplasm. For example, active antibody fragments (1–2 g/L) have been expressed in the *E. coli* periplasm at high levels (42).

The genetics and biochemistry of *E. coli* make this organism particularly attractive for production of small molecules and metabolic engineering. Several amino acids have been overexpressed successfully in *E. coli* for commercial production (eg, phenylalanine). Genetic engineering of *E. coli* for production of small molecules such as indigo and 1,3-propanediol has also been reported (43,44).

3.2. Erwinia. *Erwinia* are Gram-negative bacteria that are significant as plant pathogens, causing soft-rot of vegetables mainly after harvest during storage. However, these bacteria have the potential to be used for production of specialty chemicals and enzymes. The principal interest in *Erwinia* has been in understanding the mechanism of plant pathogenesis, which primarily results from extracellular cell wall degrading enzymes such as pectinase, cellulases, and proteases. However, these enzymes can be used for biomass degradation. Vectors to express genes in *Erwinia* have been developed, and *Erwinia* has been engineered for the conversion of glucose to 2-ketol-gulonic acid, a precursor in vitamin C production (45).

3.3. Pseudomonas. Pseudomonads form a diverse group of Gram-negative bacteria that inhabit water and soil and colonize plants (46). Pseudomonads are notable for degradation of many xenobiotic compounds (ie, organic compounds that are generally difficult for other organisms to metabolize) and for carrying out diverse sets of biochemical reactions. As a result, the use of *Pseudomonas* strains in clean-up of chemical wastes and oil spills has drawn considerable attention. *Pseudomonas* is not commonly considered for production of recombinant proteins by fermentation. However, *Pseudomonas* species are used in the commercial production of some small molecules such as acrylamide (qv) (46). Plasmid vectors for regulatable gene expression are available for *Pseudomonas* strains, and the complete genome sequences of *P. aeruginosa* and *P. putida* are available (www.tigr.org).

Since certain *Pseudomonas* species colonize plant leaves and roots, much attention has focused on genetic engineering of *Pseudomonas* species for crop protection. For example, *P. syringea* has been engineered to have a trait that prevents ice nucleation (Ice^-). Ice^- bacteria prevent frost damage on certain vegetable plants by competing with Ice^+ bacteria that occur in Nature. Insecticidal endotoxin from *B. thurengeisis* has been expressed in *P. fluorescens*. The recombinant strain was capable of normal root colonization and showed toxicity toward corn root worms. It is evident from these examples that many of the potential applications involving *Pseudomonas* involve environmental release of genetically engineered microorganisms. Studies concerning the potential benefits and risks are being aided by vectors that allow the released microorganisms to be easily traced (47).

3.4. Bacillus. Members of the genus *Bacillus* represent a diverse group of Gram-positive, aerobic, endospore-forming bacteria that are primarily found in soil. Since *Bacillus* strains are generally easy to grow and low-cost growth media are readily available, an extensive fermentation technology has been developed for these bacteria (48). As a result, *Bacillus* species are particularly attractive for production of low-cost bulk enzymes and supply 58% of industrial enzymes. For example, proteases from *B. amyloliquefaciens*, amylases from *B. licheniformis* and glucose isomerase from *B. coagulans* are used in a variety of industrial processes (see ENZYME APPLICATIONS, INDUSTRIAL). The proteinaceous inclusions produced by *B. thuringiensis* are useful as insect toxins.

Bacillus subtilis remains the paradigm for *Bacillus* genetics and biochemistry (49). Since *B. subtilis* is naturally competent for genetic transformation and the entire sequence of the *B. subtilis* genome is available (http://genolist.pasteur.fr/SubtiList/), genetic manipulations are easily perfomed with this microorganism. In addition, electroporation has made it feasible to transform other industrially important strains such as *B. licheniformis, B. amyloliquefaciens*, and *B. thuringiensis*.

A variety of plasmids are available for cloning and gene expression in *B. subtilis*. Several specialized secretion vectors for overproduction of heterologous proteins in *B. subtilis* have been developed. High level (3–6 g/L) secretion of *Bacillus* proteins has been accomplished in *B. subtilis*, but secretion of certain non-*Bacillus* proteins has not reached commercially acceptable levels. Host cell proteases are a significant problem because they interfere with protein secretion in *B. subtilis*. Strains that lack most of the principal extracellular proteases are available and should help stabilize recombinant proteins. Although most attention to *B. subtilis* has focused on secretion of extracellular proteins, it is feasible to obtain high levels of intracellular protein expression.

Recombinant *B. subtilis* and *B. licheniformis* are currently used for commercial production of several recombinant proteins and small molecules. Genes for most of the *Bacillus* exoenzymes that are commercially important have been cloned. Thus, recombinant proteases and amylases should replace the traditional enzymes. Extensive protein engineering of subtilisin has resulted in creation of an enzyme that is tolerant of higher temperature and pH (48). *Bacillus subtilis* has been used to commercially produce riboflavin by metabolic engineering and is being developed for production of biotin (49). High levels of interleukin-3, secreted from *B. licheniformis* is in clinical trails in Europe. *Bacillus brevis* has been studied for the production of a variety of proteins in Japan (48).

3.5. Clostridium. Members of the genus *Clostridium* represent a diverse group of Gram-positive, anaerobic, endospore-forming bacteria that are primarily found in soil. *C. perfringes, C. tetani, C. botulinum* and *C. difficile* are important human pathogens. The availability of various toxin genes from these bacteria should both improve vaccines and help in human medicine and food diagnostics. Other members of the genus *Clostridium* yield interesting extracellular products. *Clostridium acetobutylicum* produces acetone, butanol, and ethanol during fermentation of various carbohydrates (50). The cellulases from *C. thermocellum* have been widely studied and the cellulase genes have been expressed in *E. coli* and *B. subtilis*. The complete genomes of *C. acetobutylicum* and *C. perfringes* have been sequenced (51). Plasmid vectors for gene

manipulations with *C. acetobutylicum* and *C. perfringes* are available, and gene transfer by electroporation is feasible for these bacteria.

3.6. Streptomyces. These Gram-positive, filamentous soil bacteria are the source of most clinically useful natural antibiotics. As a result, advanced fermentation technology has been developed for *Streptomyces*. *Streptomyces* species also secrete a variety of extracellular enzymes such as xylanases and lignases. The complete genome sequence of the model actinomycete *S. coelicolor* A3(2) has been determined (52). In addition, methods for direct transfer of large fragments of DNA between *Streptomyces* species have been devloped (53,54). Plasmids and cloning vectors for heterologous gene expression are available for certain *Streptomyces* species. Recently, genome shuffling of *S. fradiae* has resulted in improved production of tylosin (55). The impact of the genome sequence coupled with the genetic tools is likely to result in a pleothora of antibiotics and bioactive substances (56).

3.7. Rhodococcus. Bacteria belonging to the genus *Rhodococcus* are Gram-positive, aerobic, nonmotile actinomycetes. These bacteria are pleiomorphic in structure and range from cocci to filaments with short projections. Rhodococci are common throughout nature, and some are pathogenic to humans, plants, and animals. These bacteria exhibit a broad metabolic diversity and are capable of degrading hydrocarbons, halogenated phenols, aromatic amino acids, and halogenated alkanes. The genetics and molecular biology of *Rhodococcus* have not been well studied. Cloning vectors based on indigenous rhodoccocal plasmids have been constructed, and *E. coli–Rhodococcus* shuttle vectors are available. Protoplast transformation of several species of *Rhodococcus* have been accomplished. The development of cloning vectors and gene transformation make it feasible to engineer *Rhodococcus* species for a variety of purposes. The metabolic potential of this class of microbes can be exploited for both bioremediation and synthesis of novel chemicals. The nitrile hydratase from a strain of *Rhodococcus* is used in the commercial production of acrylamide in Japan (57).

3.8. Lactic Acid Bacteria. The lactic acid bacteria are ubiquitous in Nature, and are found in envrionments that range from plant surfaces to the gastrointestinal tracts of many animals. These Gram-positive facultative anaerobes convert carbohydrates (qv) to lactic acid and are used extensively in the food industry for the production of yogurt, cheese, sour dough bread, etc. In addition, certain species produce a variety of antibiotics.

Lactococcus and *Lactobacillus* are members of this group and gene-transfer methods are available, as are plasmid vectors that function in several different species of lactic acid bacteria. Some of the well-characterized staphylococcal plasmids are functional in these bacteria. Heterologous genes such as amylases and proteases have been expressed in certain lactic acid bacteria. The endogenous proteases produced by these bacteria act on caesin, resulting in small peptides. These peptides are thought to play a role in flavor generation. Food grade cloning vectors are available for industrial strains of *Lactococcus lactis*. The genes for some of the proteases have been cloned and expressed. Genome shuffling has been applied to *Lactobacillus* to improve the production of lactic acid (58).

3.9. Yeasts. Yeasts are the primary eukaryotic microbes used for the production of heterologous proteins (see YEASTS). *S. cerevisiae* is well-characterized biochemically and genetically and the entire genome sequence is available

(http://genome-www.stanford.edu/Saccharomyces/). However, heterologous protein expression seems to be better in some of the industrial strains of yeasts such as *Pichia pastoris*, *Hansenula polymorpha*, *Kluyveromyces lactis*, and *Yarrowia lipolytica* (59,60).

A set of regulatable gene expression vectors are available for *S. cerevisiae*. These vectors are based on autonomously replicating plasmids or genetic elements that integrate into the chromosome. Several therapeutic proteins such as insulin, growth hormone, and interleukins have been secreted from *S. cerevisiae*. A sizeable limitation of *S. cerevisiae* for the production of some therapeutic proteins is the extensive glycosylation of heterologous proteins. This hyperglycosylation, which may be responsible for the highly antigenic nature of some of the secreted proteins, may be advantageous for nontherapeutic purposes. For example, an increase in thermostability of *Aspergillus* glucose oxidase in *S. cerevisiae* was observed. High level (3 g/L) accumulation of a fungal glucose oxidase has also been accomplished in *S. cerevisiae* (60).

Methylotrophic yeasts *P. pastoris* and *H. polymorpha* have been used to secrete a variety of heterologous proteins. Stable multicopy integration of genes into the chromosome occurs in both *P. pastoris* and *H. polymorpha*. *Pichia pastoris* appears to be a promising host, and several heterologous proteins, eg, human serum albumin and human interleukin, have been secreted at high levels. The extent of glycosylation in *P. pastoris* is different than that in *S. cerevisiae*. *Kluyveromyces lactis* has also been engineered to secrete high levels of human serum albumin (61). A limitation of the methylotrophic yeasts is the lack of basic research on the mechanism of protein secretion and gene regulation. In addition, extensive clonal variation in the level of gene expression has been observed, making multiple clone screening a necessity. The clonal variation seems to be a result of the original transformation, and lack of understanding of this mechanism makes it hard to target genes to specific sites in the chromosome.

3.10. Filamentous Fungi. Fungi are highly versatile and generate a wide range of commercial products (62,63) including organic acids such as citric acid (qv), secondary metabolites such as antibiotics, and a variety of industrial enzymes. *Aspergillus nidulans* and *Neurospora crassa* have been the prototypes for genetics and biochemistry. However, some of the commercial strains that have been developed for heterologous protein expression are *A. niger* and *Trichoderma reesi* (64). *Aspergillus niger* is used in several commercial processes. *Trichoderma reesi* appears to have a very high capacity to secrete proteins to the level of 35 g/L. These filamentous fungi do not produce any pyrogens and thus do not pose any threat for the production of food processing enzymes and therapeutic proteins. However, the secreted proteins may be glycosylated and the glycosylation is different from that of mammalian cells which may present an increased antigenic problem for therapeutic proteins.

Vectors have been developed for regulated gene expression in both *A. niger* and *T. reesi*. Heterologous genes are maintained in these organisms by integration into the chromosome rather than on autonomous plasmids. Commercial levels of 1 g/L of chymosin (rennin) production have been accomplished in *A. niger* var. awamori (65). The recombinant chymosin has properties that are similar to the native calf rennin and can be used for cheese processing. The

recombinant chymosin has been approved by FDA for use in food processing (cheese). Chymosin from *E. coli* has been available from Pfizer since the early 1990s. Attempts to secrete high levels of chymosin in other microorganisms such as *B. subtilis* and yeast were not successful. Efficient production of antibody fragments (150 mg/L) in *T. reesi* has been reported. A mechanism of protein secretion and genetic tools needed for improving yields are not available in *A. niger* and *T. reesi*.

3.11. Other Microbial Systems. In addition to the systems described, gene cloning is routinely performed in several other bacterial strains including *Streptoccocus*, *Staphylococcus*, *Brevibacterium*, *Rhodobacter*, *Cornyebacterium*, *Glucanobacter*, *Acetobacter*, and *Zanthomonas*.

4. Products from Genetically Engineered Microbes

Genetically engineered microbes are used to produce many commercial products. These products range from therapeutic proteins to small molecules and serve a wide range of markets.

4.1. Therapeutics. It has been possible to produce a variety of therapeutic proteins such as insulin, human growth hormone, interleukins, interferon, and streptokinase in microbial systems (39). Therapeutic polypeptides are low volume, high value products. The production systems need not be very efficient, but the recombinant protein has to be extremely pure (66,67). Thus, high cost mammalian production systems can be tolerated if necessary.

4.2. Bulk Enzymes. Enzymes such as proteases, amylases, glucose isomerases, and rennin are used in food processing. Proteases and lipases are used in detergents. Cellulases and xylanases are used in the paper pulp industry. The estimated value of the industrial enzyme market in 2002 was $1.5 billion, primarily due to proteases used in the detergent industry. Application of new technologies such as protein engineering and directed evolution has resulted in the introduction of new products in a relatively short time. For example, proteases that can operate at lower temperatures in household washing machines have been recently introduced to reduce energy requirements without compromising the cleaning functionality. While some industrial enzyme market segments have stagnated, certain new enzyme activities have been introduced that have a positive environmental impact. One example is in the textile industry. The chemical process that was used in textile processing in the past consumed large amounts of water and alkali and was also a high temperature, energy consuming process (68). These problems have been reduced by the use of pectate lyase for removing cell-wall components from cellulose fibers.

4.3. Antibiotics. The genes involved in the synthesis of a variety of antibiotics have been isolated (69,70). Most of these antibiotics are synthesized by polyketide synthase or peptide synthetases. The architecture of these enzymes is modular in nature, thus allowing alteration of individual modules to synthesize novel antibiotics (71,72). In addition, a whole genome shuffling method has been used to increase the production of tylosin.

4.4. Chemicals. The use of microbes to produce chemicals has been reviewed recently (73,74). Biocatalysts have the potential of being regioselective,

chemoselective, and enantioselective. In addition, biocatalysts are active under milder conditions and frequently generate less waste than traditional chemical catalysts. As a result, there is much interest in developing biocatalysts for novel processes and as replacements for chemical catalysts in more established processes.

The use of nitrilases and nitrile hydratases in bioprocesses for enzymatic transformation of nitriles has been particularly successful. A nitrile hydratase based conversion of acrylonitrile to acrylamide is among the most important bioprocesses currently used for production of a commodity chemical (75). Mitsubishi Rayon produces ~20,000 metric tons/year of acrylamide using a third-generation biocatalyst, *Rhodococcus rhodochrous* J1. In another recently developed bioprocess, an immobilized microbial cell catalyst (*Acidovorax facilis* 72W), which contains a nitrilase, is used in the first step of a two part chemoenzymatic conversion of 2-methylglutaronitrile (MGN, a byproduct produced during the manufacture of adiponitrile for nylon-6,6) to 1,5-dimethyl-2-piperidone (1,5-DMPD, Xolvone) (76). The MGN byproduct is first hydrolyzed to 4-cyanopentanoic acid (4-CPA) ammonium salt using the 72W biocatalyst. The 4-CPA ammonium salt in the aqueous product mixture is then directly converted to 1,5-DMPD by low-pressure catalytic hydrogenation in the presence of added methylamine. When compared to an alternate chemical process in which MGN is directly converted via hydrogenation to a mixture of 1,3-DMPD and 1,5-DMPD, the chemoenzymatic process generates less waste, and produces a single lactam isomer in significantly higher yield.

Bioconversions using Baeyer-Villiger monooxygenases can be used to produce a wide range of compounds (77). The increase in sequenced genes from various bacterial sources has greatly expanded the number of putative monooxygenase genes present in GenBank. Conversion of cyclohexanone to caprolactone is among the chemical transformations mediated by a Baeyer-Villiger monooxygenase (78). In addition, the genes from the cyclododecanone pathway have been recently isolated and can be used to produce lauryl lactone or 12-hydroxydodecanoic acid, compounds that are not readily available by chemical synthesis (73).

Chemical transformations that are difficult to achieve by chemical catalysis are particularly attractive targets for biotechnology. Chemical routes for carboxylation of aromatics are difficult. Phenol carboxylase has been used to produce 4-HBA in an enzymatic process (79). However, a whole cell system would most likely be needed for large-scale production. Interestingly, biological carboxylation of phenol is similar to the Kolbe-Schmitt chemical process, which is used for commercial manufacture of 4-HBA. A biological carboxylation route to 4-HBA from phenol may offer higher yields with lower waste generation than the existing chemical process.

Oxidation of alkyl aromatics is another potential target for biotechnology. Industrial oxidation of alkyl aromatics is carried out extensively to produce compounds that have applications in films, fibers, paints, adhesives and softeners. Bacteria with xylene monooxygenases typically oxidize a single methyl group on *m*-xylene or *p*-xylene. Bacteria that oxidize both methyl groups on these xylenes have been described (80). In addition, it has been demonstrated that cloned xylene monooxygenases in *E. coli* can hydroxylate both methyl groups on *p*-xylene (81). Endogenous *E. coli* enzymes further oxidize the aromatic

alcohols to form aromatic carboxylates in a controllable, step-wise fashion. The industrial processes for complete oxidation of methyl groups on *p*-xylene to form polymer grade terephthalic acid are very efficient (82). However, it is difficult to produce partial oxidation derivatives of *p*-xylene and other starting compounds by chemical catalysis. Cloned xylene monooxygenase genes afford an alternative approach to producing partially oxidized derivatives such as carboxybenzyl alcohol.

4.5. Polymers. Poly-β-hydroxybutyrate (PHB) is a biodegradable thermoplastic that is produced by several microorganisms. PHB synthesis has been characterized in *Alcaligens eutrophus* and the operon involved in PHB production has been cloned. Recombinant *E. coli* strains that produce high levels of PHB have been constructed (83).

5. Future Prospects

Biotechnology offers a number of advantages to industry. However, a variety of novel biocatalysts must be developed for biotechnology to deliver on its potential for providing both new and improved processes. The basic tools are in place to take advantage of the rapidity with which recombinant DNA experiments can be performed. In addition, a large amount of DNA sequence is already available from a variety of organisms, with more sequence being deposited every day. Hence, we can expect the number of industrial applications based on genetically engineered microbes to rapidly increase.

BIBLIOGRAPHY

"Genetic Engineering" in *ECT* 3rd ed., Vol. 11, pp. 730–745, by A. M. Chakrabarty, University of Illinois at the Medical Center, Chicago; in Suppl. Vol., pp. 495–513, by E. Jaworski and D. Tiemeier, Monsanto Co.; in *ECT* 4th ed., Vol. 12, pp. 481–491, by Vasantha Nagarajan, E. I. du Pont de Nemours & Co., Inc.; "Genetic Engineering, Microbes" in *ECT* (online), posting date: December 4, 2000, by Vasantha Nagarajan, E. I. du Pont de Nemours & Co., Inc.

CITED PUBLICATIONS

1. J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning*, Vols. 1, 2, 3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
2. B. M. Chassy, A. Merciener, and J. Flickinger, *Trends. Biotechnol.* **6**, 303 (1988).
3. G. Ditta, S. Stanfield, D. Corbin, D. R. Helinski, *Proc. Natl. Acad. Sci. U. S. A.* **77**, 7347 (1980).
4. V. de Lorenzo and K. N. Timmis, *Methods Enzymol.* **235**, 386 (1994).
5. N. Arnheim and H. Erlich, *Ann. Rev. Biochem.* **61**, 131 (1992).
6. M. A. Innis and co-workers, *PCR Protocols: Guide to Methods and Applications*, Academic Press, Inc., New York, 1990.
7. M. M. Ling and B. H. Robinson, *Anal. Biochem.* **254**, 157 (1997).
8. W. P. Stemmer and co-workers, *Nature (London)* **370**, 324 (1994).

9. W. P. Stemmer, *Nature (London)* **391**, 288 (1998).
10. W. P. Stemmer, *Nature (London)* **370**, 389 (1994).
11. M. Ostermeier, A. E. Nixon, and S. J. Benkovic, *Bioorg. Med. Chem.* **7**, 2139 (1999).
12. A. A. Volkov and F. H. Arnold, *Methods Enzymol.* **328**, 447 (2000).
13. U. T. Bornschuer, *Biocatal. Biotrans.* **19**, 85 (2001).
14. R. D. Fleichmann and co-workers, *Science* **269**, 496 (1995).
15. E. V. Koonon and M. Y. Galperin *Computational approaches in comparartive genomics*, Kluwer Academic Publishers, Dordchet, The Netherlands, 2003.
16. C. M. Fraser and co-workers, *J. Bacteriol.* **184**, 6403 (2002).
17. C. Kurland, and G. Olsen, *Curr. Opin. Microbiol.* **5**, 497.
18. B. Marrs, S. Delagrave, and D. Murphy, *Curr. Opin. Microbiol.* **2**, 241 (1999).
19. G. N. Stephanopoulos, A. A. Aristidou, and J. Nielsen, *Metabolic Engineering, Principles and Methodologies*, Academic Press, New York, 1998.
20. J. F. Bailey, *Nature Biotechnol.* **19**, 503 (2001).
21. R. Ye and co-workers, *J. Microbiol. Methods* **47**, 257 (2001).
22. D. A. Lashkari and co-workers, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13057 (1997).
23. R. Ekins and F. W. Chu, *Trends Biotechnol.* **17**, 217 (1999).
24. A. Loos and co-workers, *Appl. Envir. Microbiol.* **67**, 2310 (2001).
25. A. Gaballa and co-workers, *J. Bacteriol.* **184**, 6508 (2002).
26. Y. Wei and co-workers, *J. Bacteriol.* **183**, 545 (2001).
27. L. M. Raamsdonk and co-workers, *Nature Biotechnol.* **19**, 45 (2001).
28. M. K. Oh and J. C. Liao, *Biotechnol. Prog.* **16**, 278 (2000).
29. S. Brenner and co-workers, *Nature Biotechnol.* **18**, 630 (2000).
30. M. G. Bramucci and V. Nagarajan, *Trends Biotechnol.* **18**, 501 (2000).
31. A. Majerník, G. Gottschalk, and R. Daniel, *J. Bacteriol.* **183**, 6645 (2001).
32. G. P. Smith, *Science* **228**, 1315 (1985).
33. J. J. Devin, L. C. Panganiban, and P. E. Devlin, *Science* **249**, 404 (1990).
34. R. H. Hoess, *Chem. Rev.* **101**, 3205 (2001).
35. R. M. Verhaert and co-workers, *J. Biotechnol.* **96**, 103 (2002).
36. S. R. Whaley and co-workers, *Nature (London)* **405**, 665 (2000).
37. T. Kretzschmaer and T. von Ruden, *Curr. Opinion Biotechnol.* **13**, 598 (2002).
38. R. A. Lerner, S. A. Benkovic, and P. G. Schultz, *Science* **252**, 659 (1991).
39. D. V. Goeddel, *Methods Enzymol.* 185 (1990).
40. S. C. Makrides, *Microbiol. Rev.* **60**, 512 (1996).
41. N. Goeden-Wood and co-workers, *Biomacromolecules* **3**, 874 (2002).
42. P. R. Carter and co-workers, *Biotechnology* **10**, 163 (1992).
43. U.S. Pat. 5,686,276 (1997), L. A. Laffend, V. Nagarajan, and C. Nakamura.
44. D. Murdock and co-workers, *Bio/Technol.* **11**, 381 (1993).
45. J. Robert-Baudoy, *Trends. Biotechnol.* **9**, 325 (1991).
46. T. Nakazawa, K. Furukawa, D. Haas, and S. Silver (eds.), *Molecular Biology of Psuedomonads*, American Society for Microbiology, Washington, D.C., 1996.
47. M. Wilson and S. E. Linsow, *Ann. Rev. Microbiol.* **47**, 913 (1993).
48. R. H. Doi and M. McGloughlin, *Biology of Bacilli Applications to Industry*, Butterworth-Heinemann, Stoneham, Mass., 1992.
49. A. L. Sonenshein, J. A. Hoch, and R. Losick, *Bacillus subtilis and other Gram-positive Microorganisms*, ASM, Washington, D.C., 1993.
50. P. Durre and co-workers, *J. Mol. Microbiol. Biotechnol.* **4**, 295 (2002).
51. J. Nölling and co-workers, *J. Bacteriol.* **183**, 4823 (2001).
52. S. D. Bentley and co-workers, *Nature (London)* **417**, 141 (2002).
53. Z. Hu, D. A. Hopwood, and C. Khosla, *Appl. Environ. Microbiol.* **66**, 2274 (2000).
54. J. Pigac and H. Schrempf, *Appl. Envir. Microbiol.* **61**, 352 (1995).
55. Y. Zhang and co-workers, *Nature (London)* **415**, 644 (2002).

56. S. Donadio, M. Sosio, and G. Lancini, *Appl. Microbiol. Biotechnol.* **60**, 377 (2002).
57. H. Yamada and M. Kobayashi, *Biosci. Biotech. Biochem.* **60**, 1391 (1996).
58. R. Patnaik and co-workers, *Nature Biotechnol.* **20**, 707 (2002).
59. K. Wolf, *Nonconventional yeasts in Biotechnology*, Springer-Verlag, New York, 1996.
60. A. De Baetselier and co-workers, *Biotechnology* **9**, 559 (1991).
61. R. Fleer and co-workers, *Biotechnology* **9**, 968 (1991).
62. P. J. Punt and co-workers, *Trends Biotechnol.* **20**, 200 (2002).
63. D. B. Finkelstein and C. Bell, *Biotechnology of Filamentous Fungi*, Butterworth-Heinemann, Stoneham, Mass., 1992.
64. R. W. Davies, in J. A. H. Murray, ed., *Transgenesis*, John Wiley & Sons, Inc., 1992, pp. 82–104.
65. N. S. Dunn-Coleman and co-workers, *Biotechnology* **9**, 976 (1991).
66. T. J. R. Harris, *Protein Production by Biotechnology*, Elsevier Science Publishers Ltd., New York, 1990.
67. R. H. Baltz, G. D. Hegeman, and P. L. Skatrud, *Industrial Microorganisms: Basic and Applied Molecular Genetics*, ASM, Washington, D.C., 1993.
68. O. Kirk, T. V. Borchert, and C. C. Fuglsang, *Curr. Opinion Biotech.* **13**, 345 (2002).
69. Y. Mao, M. Varoglu, and D. H. Sherman, *J. Bacteriol.* **181**, 2199 (1999).
70. P. L. Skatrud, *Trends. Biotechnol.* **10**, 324 (1992).
71. J. Stauton, *Curr. Opinion Chem. Biol.* **2**, 339 (1998).
72. T. A. Cropp and co-workers, *Biotechnol. Genetic Eng. Rev.* **19**, 159 (2002).
73. S. Thomas, R. DiCosimo, and V. Nagarajan, *Trends Biotech.* **20**, 238 (2002).
74. G. Chotani and co-workers, *Biochim. Biophys.* **1543**, 434 (2000).
75. A. Banerjee, Stereoselective microbial Baeyer-Villiger oxidations, in R. N. Patel, ed., *Stereoselective biocatalysis*, Marcel-Dekkar, Inc., New York, 2000, pp. 867–876.
76. M. Aresta and co-workers, *Tetrahedron.* **54**, 8841 (1998).
77. F. B. Cooling and co-workers, *J. Mol. Catal. B: Enzymatic* **11**, 295 (2001).
78. H. Yamada and M. Kobayashi, *Biosci. Biotech. Biochem.* **60**, 1391 (1996).
79. S. Breinig and co-workers, *J. Bacteriol.* **182**, 5849 (2000).
80. M. G. Bramucci and co-workers, *Appl. Microbiol. Biotechnol.* **58**, 255 (2002).
81. M. G. Bramucci, M. Singh, and V. Nagarajan, *Appl. Microbiol. Biotechnol.* **59**, 679 (2002).
82. K. Weissermel and H. -J. Arpe, *Industrial Organic Chemistry*, VCH, Weinhein, 1993.
83. A. Steinbuchel and S. Hein, *Adv. Biochem. Eng. Biotechnol.* **71**, 81 (2001).

VASANTHA NAGARAJAN
MICHAEL BRAMUCCI
Du Pont