Pharmaceutical Research, Vol. 3, No. 4, 1986

Review

The Applications of Synthetic Oligonucleotides to Molecular Biology

Ronald W. Ellis[1]

Over the last few years, the chemical synthesis of DNA *in vitro* has become a routine technique. The ready availability of synthetic DNA has revolutionized molecular biological research and has enabled new research approaches to be attempted that had not been possible previously. The most commonly used approaches can be divided into three broad areas: the cloning and manipulation of genes, the diagnosis of diseases by probing gene structure, and the specific *in vitro* mutagenesis of DNA for structure/function studies. In this review, each of these three areas is discussed through specific applications from published work. This review is not intended to be comprehensive either in its scope or in its documentation of published results. Rather, it is meant to present some of the most commonly used applications of oligonucleotides in molecular biology through representative examples.

KEY WORDS: oligonucleotides; recombinant DNA; molecular biology.

INTRODUCTION

Over a decade has passed since Khorana and coworkers utilized synthetic oligonucleotides to assemble the alanine transfer RNA (tRNA) gene (1). This technical tour de force required 20 man-years of work (2). The technology for the synthesis of DNA has changed remarkably over the intervening years. At present, commercially available semi- or fully automated or manual "gene machines" can synthesize several oligonucleotides simultaneously at the rate of one nucleotide added every 20 min, such that a DNA duplex the size of the tRNA gene now can be synthesized in only a few days. These gene machines are very reliable, being capable of synthesizing oligonucleotides of up to 100 residues with yields at each step of >99%. Several different chemical methods have been used for oligonucleotide synthesis. The technique used by Khorana was the phosphodiester method in solution. Major improvements followed with the development of the phosphodiester and phosphite triester chemistries as well as the adaptation of these chemical syntheses to solid-phase supports. The details of the chemistries of these techniques have been presented in several recent review articles (2–8) as well as in a comprehensive textbook (9), although the availability of dependable hardware has reduced the need for many molecular biologists to become nucleotide chemists!

The following article presents some of the more commonly used and versatile applications of synthetic oligonucleotides in molecular biology. In terms of the cloning and manipulation of genes, applications have included the isolation of genes and their complementary DNA (cDNA) derivatives from genomic DNA, the complete synthesis of functional genes, DNA sequence analysis, and the modification of DNA termini for the expression of recombinant DNA (rDNA) in transformed host cells. With respect to diagnostics, several different point mutations which form the molecular basis for inherited diseases now can be detected. Finally, the employment of oligonucleotides for the *in vitro* mutagenesis of DNA has been very useful for studies of gene function and for protein engineering. In each of these areas, specific examples are presented which exemplify the power of these different approaches.

CLONING AND MANIPULATION OF GENES

One of the central paths of discovery in modern molecular biology is the identification and sequence analysis of a gene encoding a particular protein. This is followed by the manipulation and expression of that gene or its mutated derivatives as rDNA in a host cell in order to express large quantities of that protein for ultimate structure–function studies. Oligonucleotides have proven extremely useful and, in many cases, indispensable to each of these steps.

Cloning of Genes and Their cDNA Products

The molecular cloning of genes as fragments of DNA has depended upon the ability of a known tagged nucleic acid probe (usually radiolabeled with ^{32}P) to hybridize specifically to the cDNA encoding the information necessary for synthesis of the protein of interest. In the middle to late 1970s, the development of techniques of rDNA technology led to the preparation of "libraries" of genomic DNA as fragments of total genomic DNA (10) and of cDNA clones as a complement of the messenger RNA (mRNA) present within cells (11). There were early constraints in the cloning of genomic DNA due to the limited availability of specific nucleic acid probes. Thus, genes cloned in the 1970s usually

[1] Department of Virus & Cell Biology, Merck Sharp & Dohme Research Laboratories, West Point, Pennsylvania 19486.

0724-8741/86/0800-0195$05.00/0 © 1986 Plenum Publishing Corporation

were limited to those encoding highly abundant proteins such as globin (12) or immunoglobulin (13) or protooncogenes for which viral oncogene probes were available (14,15).

The groundwork for the use of oligonucleotides as probes for DNA hybridization was laid in a number of early studies which demonstrated that defined oligonucleotides were capable of hybridizing to specific DNA (16–19) or mRNA (20) sequences. The key principle demonstrated by these studies was that oligonucleotides are able to form Watson–Crick base pairs (bp) which are thermodynamically stable under certain defined hybridization conditions. In addition to refinements in rDNA technology and oligonucleotide synthesis, two other significant advances in rDNA technology enabled oligonucleotides to become versatile probes for the cloning of genomic DNA and cDNA. These were the establishment of reliable techniques for the sequencing of DNA (21,22) and the development of methods for the hybridization on solid supports (e.g., nitrocellulose) of DNA fragments digested with restriction endonucleases and electrophoretically resolved on the basis of size (23). Most recent investigations have had the goal of cloning the complete cDNA from mRNA. The reason for this is that a cDNA clone, unlike a genomic DNA clone, affords a complete and perfect readout of the primary amino acid sequence of the encoded protein and also can direct expression of its gene product when cloned into an appropriate expression vector and introduced into a suitable prokaryotic or eukaryotic host cell.

The key requirement for the use of oligonucleotides as hybridization probes would seem to be a knowledge of the precise nucleotide sequence of the DNA to be detected. Clearly, this could be a circular argument, since the goal is to isolate the DNA; if the DNA sequence is known, then the oligonucleotide hybridization approach is passé. However, DNA sequence information is indirectly available through microsequencing of purified proteins. By reverse-reading the genetic code, one can impute the DNA sequences which can encode the newly discovered amino acid sequence of the protein of interest. Unfortunately, this analysis per se is insufficient to derive a precise nucleotide sequence of the oligonucleotide as a hybridization probe due to the degeneracy of the genetic code (each of the 20 primary amino acids is encoded by one to six codons in DNA).

Mixed-Sequence Oligonucleotides

Wallace and co-workers (24) first addressed this issue by suggesting that a mixture of synthetic oligonucleotides representing every possible set of coding sequences for a given amino acid sequence could be used for probing the DNA encoding that sequence. As also shown by others (25,26), a duplex containing a single bp mismatch between a synthetic oligonucleotide of 11–17 residues and a strand of the target DNA is thermodynamically less stable than a perfectly base-paired duplex. Experimentally, the stability of the duplex can be manipulated by adjusting the temperature at which the hybridization is performed (27), such that an appropriate (high) temperature can be chosen at which only the perfectly base-paired oligonucleotide will form a stable duplex. Empirical rules have been formulated (28) for the design of mixtures of oligonucleotides as probes for the identification of cloned genomic DNA or cDNA sequences (29). Such a mixture can be made by mixing approximately equal amounts of each nucleotide monomer at a position of redundancy in the chemical synthesis (30).

The strategy used to design a mixed sequence oligonucleotide pool as a hybridization probe based on available amino acid sequences is illustrated in Fig. 1. In general, an effort is made to focus upon a region for oligonucleotide design where most of the consecutive amino acid residues are encoded by one (Met or Trp) or two (Asn, Asp, Cys, Gln, Glu, His, Lys, Phe, or Tyr) codons (Fig. 2) in order to minimize codon degeneracy, thus maximizing the relative proportion of that oligonucleotide in the mixture capable of perfect base-pairing. Oligonucleotide mixtures typically end at residue +2 of a codon, since residue +3 is the source of almost all the codon degeneracy. In addition, for cloning from human DNA, a minimum of six amino acid codons should be represented (including the first two nucleotides of the last codon), since $(20)^6$, the number of theoretical permutations of a hexapeptide, well exceeds the number of genes in the complete genome of humans. This approach first was applied to the human β2-microglobulin gene (31). Total mRNA from a cell line was converted enzymatically to double-stranded cDNA, and this library was screened with sets of 8 or 16 mixed-sequence oligonucleotides corresponding to the known amino acid sequence of β2-microglobulin (32). In this way, the complete cDNA was cloned and sequenced. Since then, nearly 100 different cDNA clones have been isolated through the use of mixed oligonucleotide probes, some of which are mentioned below. The theoretical maximum redundancy of an oligonucleotide mixture that can be used as a probe is unknown; a mix of 384 oligomers of 23 base length has been used successfully (33).

In the last few years, some refinements have been introduced in the use of mixed oligonucleotides as hybridization probes. One is the use of primer extention for cDNA enrich-

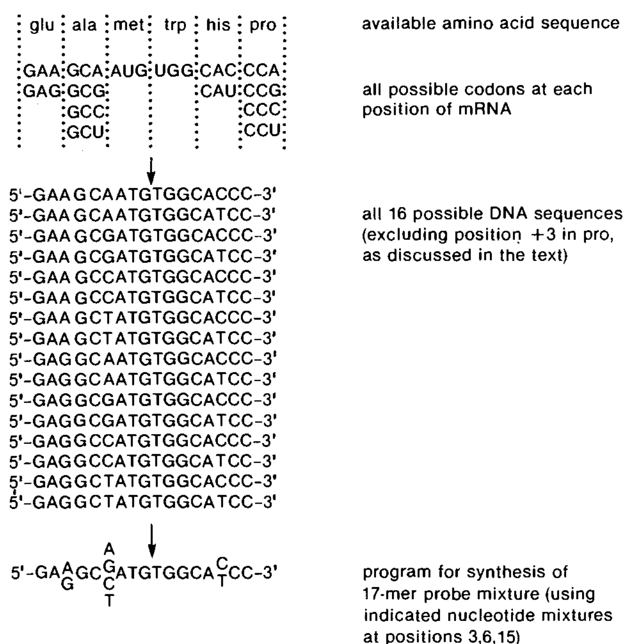

Fig. 1. The strategy for synthesizing mixed-sequence oligonucleotide probes.

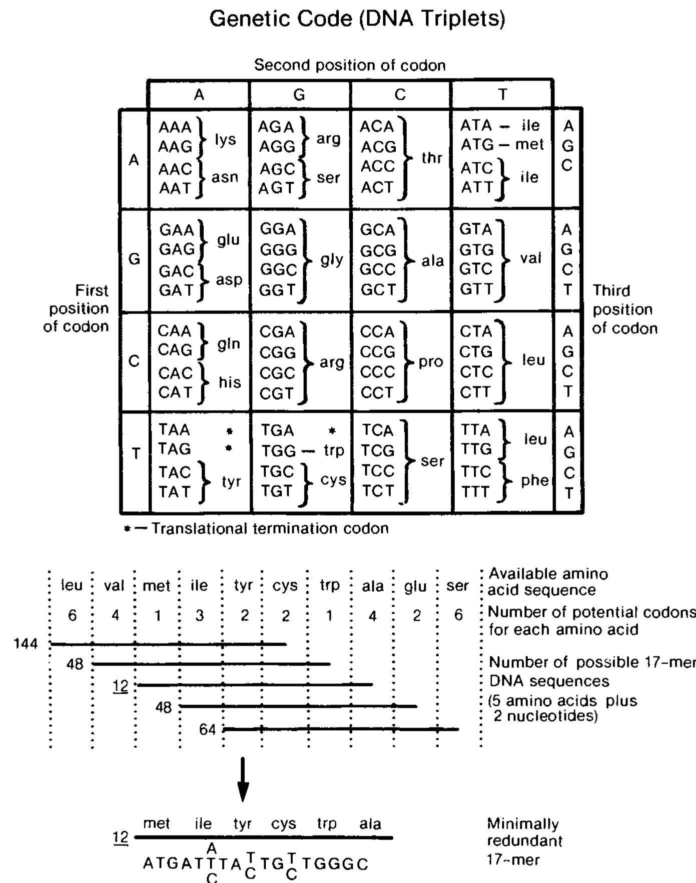

Fig. 2. The strategy for minimizing codon degeneracy in mixed-sequence oligonucleotide probes.

ment (see below). Another is the use of sets of mixed oligonucleotides based on nonoverlapping amino acid sequences from two distinct portions of the same protein; the ability of the different sets to hybridize to the same clone provides excellent confirmation that the clone does encode the protein of interest. This strategy was used to clone the cDNAs of human 3-phosphoglycerate kinase (34), bovine retinal transducin (γ subunit) (35), and chondroitin sulfate proteoglycan (36). A novel strategy was used by two independent groups to clone the cDNA for the human antihemolytic factor VIII (37,38). Mixed oligonucleotides were used to clone the human factor VIII gene. This gene then was used as a probe for identifying that tissue with the greatest amount of mRNA to be used as a source of cDNA synthesis. The final cDNA then was cloned from that library.

While workers have enjoyed clear successes with the use of mixed oligonucleotides as hybridization probes, concern has been raised over the different chemical coupling rates of deoxyadenosine (dA), deoxyguanosine (dG), deoxycytosine (dC), and deoxythymidine (dT) nucleotides at steps of mixed synthesis, such that a particular nucleotide might be underrepresented in a mixture. This concern has spawned the use of unique sequence oligonucleotides and other deoxynucleotides, as described below.

Unique-Sequence Oligonucleotides

As an alternative to the use of mixed oligonucleotide probes, Jaye and co-workers (39) introduced the concept of using a unique-sequence long oligonucleotide as a hybridization probe for the cloning of the antihemolytic factor IX cDNA. The design of the long oligonucleotide was based upon the following considerations: (i) a segment of the available amino acid sequence with the least codon degeneracy; (ii) codon usage, meaning the nonrandom distribution of nucleotide triplets which encode one particular amino acid, best corresponding to condons of other proteins secreted by the liver; (iii) the observation that a G:T mismatch at a given position in a duplex does not contribute to the instability of the duplex, while a G:A mismatch does (40); and (iv) the minimization of predictable secondary structure in the oligonucleotide. The resultant unique 52-base oligonucleotide was used successfully to detect and clone the cDNA; the authentic sequence proved to be 85% homologous to the oligonucleotide. Several other recent studies have utilized unique-sequence probes 50–90 nucleotides in length to clone several cDNAs, including pancreatic trypsin inhibitor (41), human insulin-like growth factor (42), cathepsin B (43), and human insulin receptor (44). Theoretical predictions for the design and use of unique sequence oligonucleotides as hybridization probes have been discussed (45).

Primer Extension

In conjunction with the use of oligonucleotides as hybridization probes for the cloning of DNA, primer extension

has proven a valuable tool. Primer extension is the use of an oligonucleotide hybridized to a DNA or RNA template as a template–primer combination for the synthesis of double-stranded DNA or cDNA with DNA polymerase or reverse transcriptase, respectively. Either mixed-sequence or unique-sequence oligonucleotides can be used, depending upon whether the sequence of the template is known. In addition, unique-sequence primers conforming to the rules of Jaye *et al.* (39) can be used. In the cloning of bovine adrenal preproenkaphalin (46), a mixture of two oligonucleotides based on a pentapeptide of the protein was used to prime cDNA synthesis from adrenal medullar mRNA, leading to the establishment of an enriched cDNA library which was screened with these oligonucleotides for clone isolation. Similarly, human β-nerve growth factor was cloned using a mixture of 16 oligonucleotides first as a primer and then as a screening probe (47). In the cloning of human tissue plasminogen activator cDNA (48), a clone was isolated through oligonucleotide screening that contained only a 3′ portion of the mRNA. Therefore, an oligonucleotide complementary to the 5′ end of the cDNA fragment was used to prime a second round of cDNA synthesis, thus leading to isolation of the complete cDNA. Primer extension experiments also are useful for increasing the specific activity of an oligonucleotide pool. In the cloning of the interleukin-2 receptor cDNA (49), the optimal available N-terminal amino acid sequence could be encoded by 64 possible 17-mer sequences. Therefore, the 64 were divided into 6 pools, only one of which gave a clear primer extension product when tested against T-cell mRNA. This oligonucleotide pool then was capable of detecting a cDNA clone which proved to be the receptor cDNA. Finally, perhaps the most fruitful uses of primer extension have been for DNA sequence analysis and for site-directed mutagenesis *in vitro*. These applications are discussed below.

Other Nucleotides

There have been some recent reports of the use of nucleotides other than dA, dG, dC, and dT as an alternative to the use of mixed oligonucleotides at positions where there is codon degeneracy for a given amino acid. Ohtsuka *et al.* (50) have reported the synthesis of oligonucleotides containing deoxyinosine (dI) at all ambiguous (third) positions of codons. The high dissociation temperature of the duplex of the dI-containing oligonucleotide with its homologous DNA suggested that dI neither stabilized nor destabilized the duplex, consistent with the ability of dI to hydrogen-bond, but not base-pair, with dA, dC, or dT. Therefore, long single-sequence oligonucleotides containing dI represent a useful alternative to mixed sequence oligonucleotides as hybridization probes, potentially reducing any concern regarding the differential coupling rates of oligonucleotides at positions of mixed synthesis. Huynh-Dinh *et al.* (51) describe the use of 2-amino-2-deoxyadenosine and of mixed deblocking of triazolodeoxynucleosides for certain codon degeneracies. They demonstrate that these represent valid alternatives to the use of mixed sequence oligonucleotides for the detection of human antithrombin III cDNA. Clearly, these approaches warrant further investigation.

Summary

The use of synthetic oligonucleotides for the cloning of DNA, and in particular cDNA, has extended technical capabilities to the point where essentially any gene for which the amino acid sequence of its purified protein product is available now can be cloned. A number of such genes has been described above. Some other human genes that have been cloned by this method and that encode products of pharmacological and physiological significance (chosen at random and not meant to exclude other reported clonings of equal merit) include sodium/potassium-dependent ATPase (52), calcium/magnesium-dependent ATPase (53), granulocyte-macrophage colony-stimulating factor (54), α-antitrypsin (55,56,57), α-galactosidase (58), apolipoprotein B (59), macrophage-specific colony-stimulating factor (60), erythropoietin (61), complement component C9 (62), antithrombin III (63), preprourokinase (64), and growth hormone-releasing factor (65). While mixed oligonucleotides have been used most frequently, the applications of long single-sequence oligonucleotides, other deoxynucleotides, and primer extension are contributing to the versatility of this general approach. Another recently developed technique, known as antibody screening (66), has been used alongside oligonucleotide screening. In this procedure, a cDNA library is constructed directly in a bacterial expression vector, and transformed colonies are screened for cDNA expression with monospecific (polyclonal or monoclonal) antibodies directed against the target protein. Many studies have utilized both antibodies and oligonucleotides to provide confirmatory data on individual clones, in addition to those studies using either technique independently.

Gene Synthesis

The seminal work of Khorana and co-workers in the synthesis of the alanine tRNA gene (1) has served as a prototype for the complete assembly of other genes from synthetic oligonucleotides. As recently reviewed (5), two different approaches have been used to assemble the oligonucleotides into complete functional genes. These two approaches as well as the applications of gene synthesis are discussed.

Assembly by DNA Ligase

This is the technique utilized for the synthesis of the alanine tRNA gene. In this approach, overlapping oligonucleotides that completely circumscribe both strands of the target gene (of known sequence) are synthesized (Fig. 3). Watson–Crick base-pairing stabilizes duplex formation across the oligomers following phosphorylation of the 5′-hydroxyl termini with T4 polynucleotide kinase; the single-stranded nicks are sealed by phosphodiester bond formation through the action of T4 DNA ligase. The resulting unnicked duplex, which is typically designed to leave the same cohesive termini (often a 4-bp overhang) as would be left by restriction endonuclease digestion, then is amplified by molecular cloning in a plasmid. Genes cloned in this manner include human growth hormone (66), a 585-bp duplex, human epidermal growth factor (67), somatomedin (68), complement fragment C5a (69), and interferon-α_1 (70).

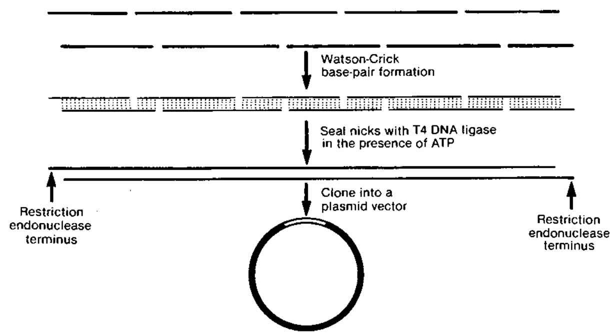

Fig. 3. The assembly of synthetic DNA duplexes using T4 DNA ligase.

Assembly by DNA Polymerase

The advances in technology that have enabled long oligonucleotides to be synthesized faithfully have led to the development of this alternate technique (71). Oligonucleotides are synthesized with only short, rather than total, stretches of complementarity between the two strands (Fig. 4). DNA polymerase I then catalyzes the incorporation of the four deoxynucleotide triphosphates onto this template–primer combination. Digestion with appropriate restriction endonucleases then generates the cohesive ends required for molecular cloning in a plasmid vector. With this technique, the human insulin A-chain and interferon-α2 genes have been synthesized (72). While less oligonucleotide synthesis is required for this approach compared to the ligase approach, it is not clear based upon the limited number of synthetic genes made to date which technique will prove more reliable in the long run.

Applications

There are two general areas where the synthesis of oligonucleotides for complete duplex DNA formation has proven useful. The first area is that of modifications to protein coding segments of DNA. It has been observed that in many species of host cell, the abundant proteins are encoded by a nonrandom distribution of codons, e.g., for an amino acid with four codons, one of these four might represent 80% of all codons found in the respective genes [see also Jaye *et al.* (39), above]. Codon distribution patterns for *Escherichia coli* (73) and *Saccharomyces cerevisiae* (74), two common hosts for the expression of rDNA-derived proteins, have been described. It has been reasoned that in a microbial host cell, which can express some foreign genes up to 10–20% of the total protein, the optimization of codon usage patterns of the rDNA-derived gene to the pattern of the host might optimize foreign protein expression. Such optimization has been performed for the somatomedin C gene for expression in *E. coli* (68) and for the first 21 amino acid codons of the hepatitis B surface antigen gene for expression in *S. cerevisiae* (75). Unfortunately, neither study reported a quantitative comparison of the expression of the "optimized" and natural gene, such that the above reasoning remains to be validated. Another useful modification to protein coding segments of DNA is the introduction of novel restriction endonuclease sites by "wobble" at the third base position of codons. The placement of sites at less than every 60–80 bp apart enables a segment to be removed and replaced by another synthetic fragment with altered coding capacity. This placement forms the basis for the *in vitro* modification technique known as "cassette mutagenesis," which is described below.

A second area for applying the chemical synthesis of duplex DNA is for noncoding DNA, especially promoter sequences which direct the position and strength of transcriptional initiation of genes. Not only can well-characterized promoters be synthesized for facilitating the design of versatile expression vectors, but novel promoters can be designed for optimization based on a concensus of which elements are thought to be most important for activity, and mutated promoters can be synthesized to test the importance of specific nucleotide residues in the promoter activity. Consequently,

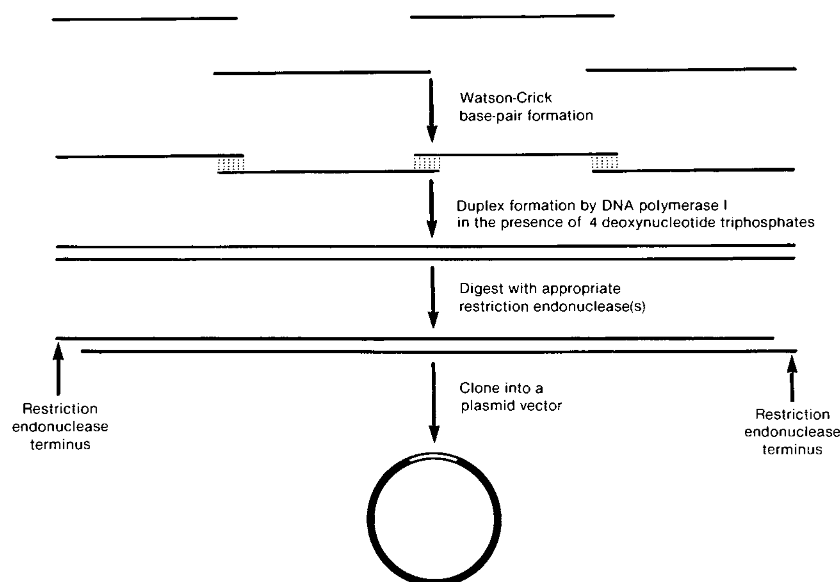

Fig. 4. The assembly of synthetic DNA duplexes using DNA polymerase I.

several natural (76,77) and artificial (78,79) *E. coli* promoters have been synthesized and tested. This approach also can be applied to the analysis of other genetic regulatory elements, including transcriptional termination, enhancers, and homeo-boxes. Also, the synthesis of complete protein coding sequences flanked by artificial restriction endonuclease sites facilitates the cloning of the synthetic gene in expression vectors (see also below; Gene Manipulation and Expression).

DNA Sequencing

One of the greatest breakthroughs in the history of molecular biology was the development of two different yet efficient procedures for rapid determination of DNA sequence (21,22), for which Walter Gilbert and Fred Sanger were awarded the Nobel Prize in Chemistry (with Paul Berg) in 1980. The contribution of their technologies to modern biology is incalculable, its most celebrated contribution being the definition of the amino acid sequence of the primary translational product of every gene that has been cloned. Each technique has its relative advantages and disadvantages which will not be discussed here, but essentially all molecular biology laboratories have adopted the use of either or both technologies. The method of Gilbert (Maxam–Gilbert) relies upon chemical degradation of DNA at specific nucleotide residues followed by gel electrophoresis of the degradation products. The Sanger method (chain termination) utilizes dideoxynucleotides (ddATP, ddGTP, ddCTP, ddTTP) to achieve premature termination of polymerization on a template–primer mediated by DNA polymerase, followed by gel electrophoresis of the products of synthesis. The development of M13 phage vectors for efficiently cloning DNA into single-stranded circles (80) has provided a convenient source of template, such template being a fragment of the gene of unknown sequence. A synthetic "universal" oligonucleotide primer which is complementary to the M13 sequence immediately 3′ to the target gene is annealed to the template, thus providing a template–primer for the ensuing polymerase reactions, as depicted in Fig. 5. Early applications of this technique utilized small inserts in the M13 vector, since individual sequencing experiments are limited to approximately 200 nucleotides each. The technique has become more efficient with the following additions (81): (i) utilize larger inserts in the M13 vector; (ii) sequence the first stretch of 200 nucleotides by means of the universal sequencing primer; (iii) synthesize as a second primer an oligonucleotide composed of the last 15–20 nucleotides in the first stretch; (iv) utilize the second sequencing primer for deriving the sequence of the next 200 nucleotides; and (v) synthesize another oligonucleotide primer, etc., until the entire large insert has been sequenced. This technology now has been used in hundreds of publications, and assay reagents are available from many different manufacturers in easy-to-use kits.

Gene Manipulation and Expression

The successful manipulation of rDNA for molecular cloning and for expression in heterologous hosts usually depends upon the availability of convenient restriction endonuclease sites for the isolation of protein coding sequences or genetic regulatory clements and their cloning in appropriate plasmid vectors, often for the purpose of expression of a gene. Sometimes such sites are naturally available, but usually life is not so straightforward! Synthetic oligonucleotides are highly useful in this regard, in that they can be used in conjunction with other techniques for the modification of existing restriction endonuclease sites or for the creation of novel sites not previously present. The creation of novel sites by wobble of the third base position of codons during gene synthesis has been described above. There are numerous and elegant published techniques for the use of synthetic oligonucleotides in the manipulation of restriction endonuclease sites; three of the common ones are outlined below (Fig. 6).

(i) DNA fragments, whose termini are generated by restriction endonuclease digestion or by random chemical/enzymatic cleavage, are rendered flush-ended by the action of T4 DNA polymerase or DNA polymerase I-Klenow fragment. These fragments are ligated by means of T4 DNA ligase to flush-ended (commercially available) synthetic oligonucleotide duplexes containing the new desired restriction endonuclease site (e.g., *Bam*HI), then the DNA is digested with the restriction endonuclease (*Bam*HI) such that the fragment can be cloned into a vector containing the new restriction endonuclease site. This general technique was applied by Maniatis *et al.* in the creation of the first described genomic DNA "library" (10) and has been utilized by other workers in subsequently described libraries.

(ii) With the increasing availability of reliable "gene machines," restriction endonuclease sites can be modified and created in a more versatile way by means of "custom-synthesized" oligonucleotides rather than by those commercially available. Two oligonucleotides are synthesized and annealed to create a duplex with one overhang (terminus) that can be ligated to the target DNA fragment and with a second terminus that is the overhang for a second restriction endonuclease site for the cloning of the fragment into a vector containing that second site. In the example shown, an *Eco*RI site in the target DNA (at amino acid codon +3) is modified to a *Bam*HI terminus by ligation to a synthetic duplex containing codons for amino acids 1–2, including the ATG translational initiation codon, as well as a "hidden" *Hind*III site which might be useful for future manipulations. This fragment now can be cloned into the *Bam*HI site of an expression vector adjacent to promoter sequences which will direct transcription of the gene in a heterologous host cell.

(iii) Where no restriction endonuclease site exists in the desired region of the rDNA molecule, the rDNA can be digested to the distal side of the site with an appropriate restriction endonuclease, then treated with an exonuclease such as Bal31 (82) back to the desired region of DNA (as diagnosed by DNA sequence analysis), and a synthetic oligonucleotide duplex ligated to the terminus for creation of a new restriction endonuclease site. This site then functions as a new terminus for the further manipulation of the rDNA fragment from the desired region. A most elegant application of this approach by McKnight and co-workers (83,84) was the creation of a series of clones containing deletions of varying size spanning a 100 nucleotide stretch of noncoding DNA which is important in transcriptional initiation of the

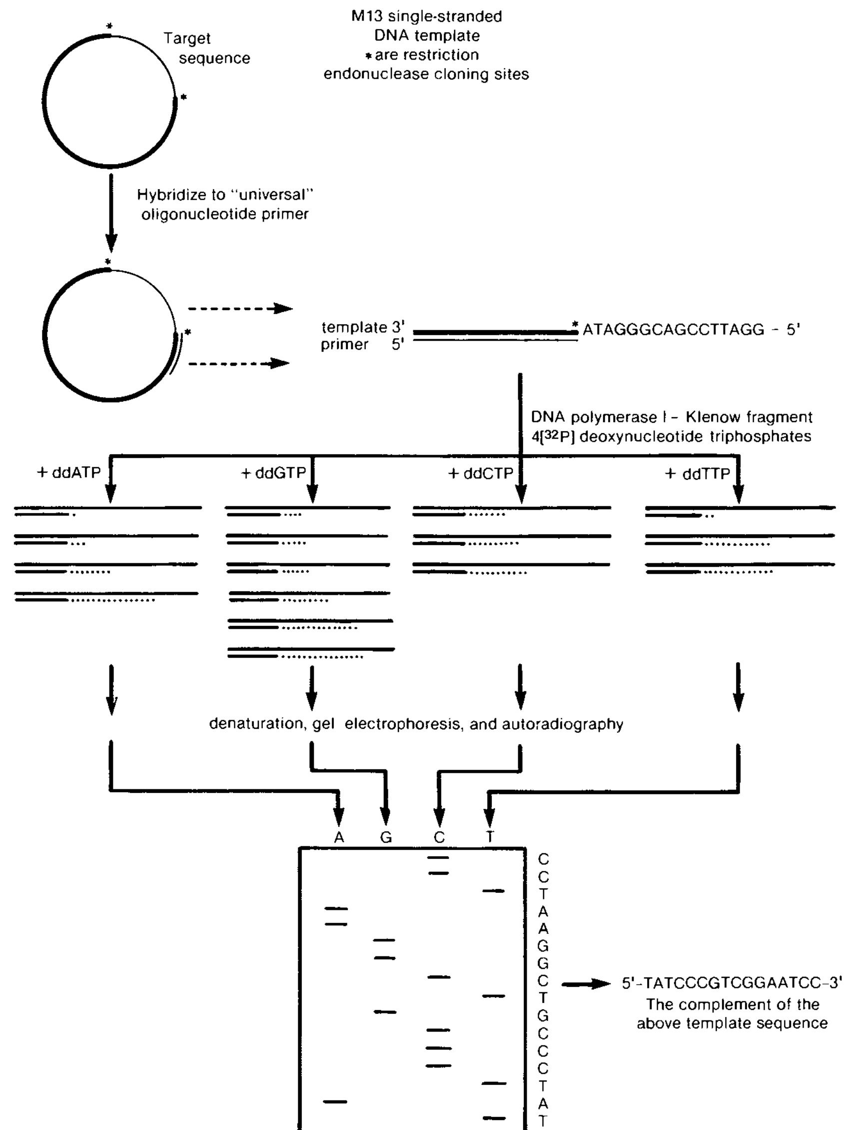

Fig. 5. The use of synthetic oligonucleotide primers in M13-dideoxynucleotide DNA sequencing.

thymidine kinase gene. These so-called "linker-scanning" mutants were used to elucidate three distinct elements (8–10 nucleotides each in length) of this gene promoter important in the regulation of transcriptional initiation.

DIAGNOSTICS

Specific oligonucleotides can be used as hybridization probes for the detection of point mutations in DNA. This technique is especially powerful when applied to particular human genes where single bp substitutions can cause certain genetic diseases (85–90). The molecular basis for this technique was first suggested from the work of Wallace and coworkers (24,29), who demonstrated that an oligonucleotide–DNA duplex with a single bp mismatch will not form stably under certain hybridization conditions under which a perfectly base-paired duplex will form. This means that the hybridization specificity of an oligonucleotide can discriminate between two different DNA molecules that differ by a point mutation (91). Thus, two oligonucleotides can discriminate by hybridization between two different gene copies allelic at a point mutation (Fig. 7).

Applications

Conner *et al.* (92) first applied this technique to the diagnosis of the point mutation ($Glu^6 \rightarrow Val^6$) in the sickle-cell β-globin gene (93) relative to the normal human β-globin gene. In addition, they demonstrated the utility of direct gel hybridization (94) of oligonucleotides as an alternative to the Southern blot method (23) of hybridization of electrophoretically resolved (on the basis of size) DNA transferred to nitrocellulose. In direct gel hybridization, an agarose gel containing electrophoretically resolved DNA is dried and hybridized directly with the radiolabelled probe, in this case the oligonucleotide, which readily diffuses into the gel. This technique is more efficient than Southern blot hybridization in terms of both shorter hybridization times as well as

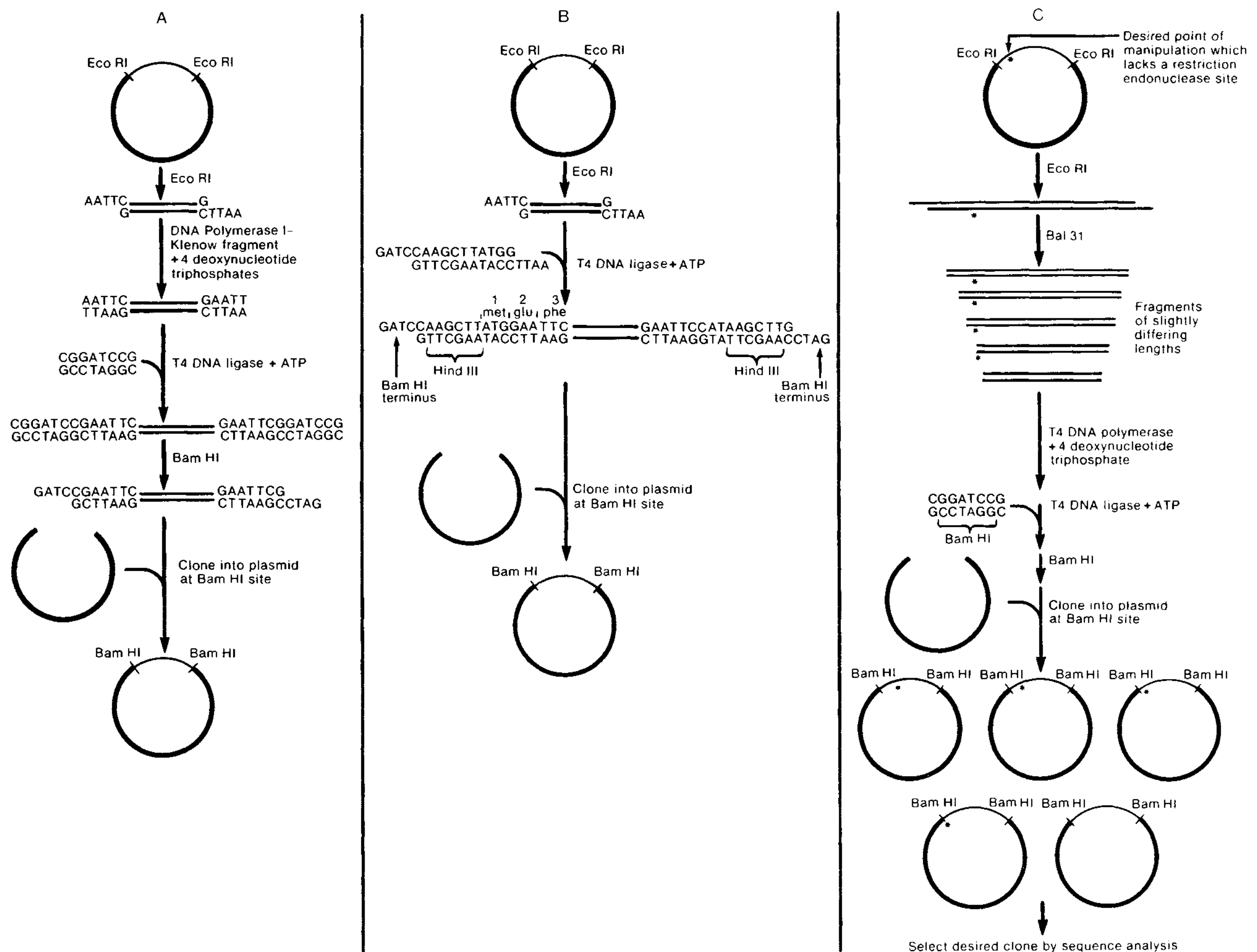

Fig. 6. The use of synthetic oligonucleotide duplexes for the modification of the termini of DNA fragments.

greater sensitivity because of reduced DNA losses. This group utilized two 19-mer oligonucleotides to resolve the β-globin gene alleles by this method. Subsequently, the diagnosis of point mutations in β-thalassemia (95,96) and α-antitrypsin deficiency (97) has been reported using these techniques.

Refinements

There have been recent technical refinements in the use of oligonucleotides as probes for DNA. Most oligonucleotide probes have utilized a single ^{32}P molecule enzymatically added to the 5′ end of the molecule as the detection signal. While such probes are sensitive in their detection level of DNA, these probes are stable for only a few weeks due to the radioactive decay of the ^{32}P. Moreover, the use of ^{32}P constitutes a biohazard risk to laboratory workers, thereby requiring special safety features. Thus, interest has turned to biotin as a tag on oligonucleotides and DNA duplexes. Biotin can bind to the glycoprotein avidin with a very high affinity (98) and then can be detected by biotinylated alkaline phosphatase in the presence of a substrate that develops a colored product upon digestion with an enzyme. The biotinylation of nucleotides first was reported utilizing RNA (99), then subsequently for duplex DNA through the use of a DNA polymerase reaction incorporating biotinylated nucleotides onto a template–primer (100,101). More recently, techniques have been described for the synthesis and use of biotinylated oligonucleotides as hybridization probes (102,103). These probes demonstrate a high specificity, longer useful half-lines, and greater safety, but less sensitivity, than ^{32}P-labeled oligonucleotide probes. Improvements in detection and labeling methods should increase the sensitivity of these probes.

A second technical advance has been in the increase in specific activity of oligonucleotide probes. As described above, probes classically are tagged with a single ^{32}P molecule at its 5′ end. The incorporation of all four ^{32}P-labeled deoxyribonucleoside triphosphates onto a DNA template–primer, followed by strand separation and purification of the ^{32}P-labeled primer strand, enables ^{32}P-labeled oligonucleotide probes to be prepared to a specific activity an order of magnitude higher than that attained by the earlier method (104). As a result, the amount of genomic DNA required for the detection of a single-copy gene by an oligonucleotide probe can be reduced from 10 to 1 μg per sample. This enables many more genetic analyses to be performed upon a single DNA sample.

The use of oligonucleotide probes is permitting the diagnosis of genetic diseases *in utero* (from DNA in amniotic

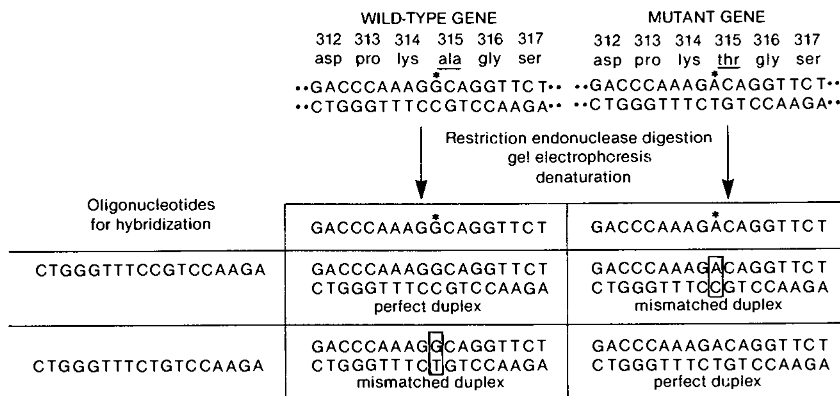

Fig. 7. The molecular basis for the use of synthetic oligonucleotides to diagnose point mutations in genomic DNA.

fluid) or for genetic analyses of adult tissues. One can envision a day when such analyses will be entirely automated by a computer utilizing hundreds of oligonucleotide probes for the rapid analysis of multiple genes. Some applications to adult genetic analysis would be relatively straightforward in terms of their ethical implications, such as forensics and paternity suits. However, the implications of other potential applications, e.g., a diagnosis of genetic predispositions to cancer, heart disease, or schizophrenia, are much more difficult to contemplate because of the great likelihood for discrimination were such information about an individual made widely known. Moreover, given the spectres of eugenics or abortion, it is far from clear as to how society will handle the information derived from the analysis of mutations in certain genes in fetal DNA.

IN VITRO MUTAGENESIS

With the development of rDNA cloning and sequencing techniques, the opportunity to define and direct genetic changes on the molecular level became available. In particular, the cloning and sequencing of wild-type genes and their available mutants coupled with a comparison of the function of the variant genes or gene products afford the potential to define structure/function relationships on the molecular level. While early analyses depended upon the availability of random, nonspecifically derived mutants, the availability of oligonucleotides has enabled scientists to create point mutations at any desired position in a DNA molecule. The techniques used for this site-specific oligonucleotide-directed mutagenesis as well as specific applications of these methods are discussed below. It is noteworthy that a wide range of other chemical and enzymatic techniques also has been used to create both site-directed and random point mutations as well as insertions and deletions in target genes. Several excellent review articles discuss these other approaches (105–107).

General Techniques

The general principle involved in the use of synthetic oligonucleotides to create insertions, deletions, or point mutations is outlined in Fig. 8. This approach is based upon the principle that an oligonucleotide that shares substantial but incomplete homology with a single-stranded DNA template is capable of hybridizing under appropriate conditions to the template and of priming DNA polymerase-mediated synthesis of a second strand. The incomplete homology can consist of a point mutation, insertion, or deletion with respect to the DNA template, as long as the oligonucleotide forms at least several perfect base pairs with the template on both sides of the nonhomology region. Following hybridization, the action of DNA polymerase I-Klenow fragment and T4 DNA ligase catalyzes the formation of covalently closed double-stranded DNA containing the mismatch. Upon the introduction of the DNA into *E. coli,* DNA replication of the plasmids will cause segregation of wild-type plasmids from mutant plasmids which contain the new sequence introduced by the oligonucleotide. The mutant and wild-type DNAs can be distinguished from one another by hybridization with the mutagenic oligonucleotide (24). The single-stranded template first used in this technique was ϕX174 (108–110). For routine use, the target gene can be introduced into M13 cloning vectors (80,81) which allow the isolation of single-stranded DNA containing the target gene as a template. Alternatively, the target gene can be cloned into a plasmid vector; the double-stranded DNA can be enzymatically nicked, and covalently closed circular single-stranded DNA isolated for use as a template (111). There have been several reported refinements to increase the efficiency of mutagenesis in plasmid (91,112) or M13 (113–118) vectors; the efficiency of mutagenesis generally has been higher in the M13 system.

Applications

Operationally, DNA sequences can be divided into two categories, protein coding sequences and noncoding sequences. The applications of site-directed *in vitro* mutagenesis have focused on both types of sequences as targets.

Protein Engineering

The ability to utilize oligonucleotides for *in vitro* mutagenesis has created great interest among protein chemists in the designing of new proteins. The creation of mutated proteins holds the prospect for an increased understanding of such areas of protein chemistry as the basis for enzymatic action, the mechanism and structural consequences of protein folding, and the mapping of active sites in a protein. A

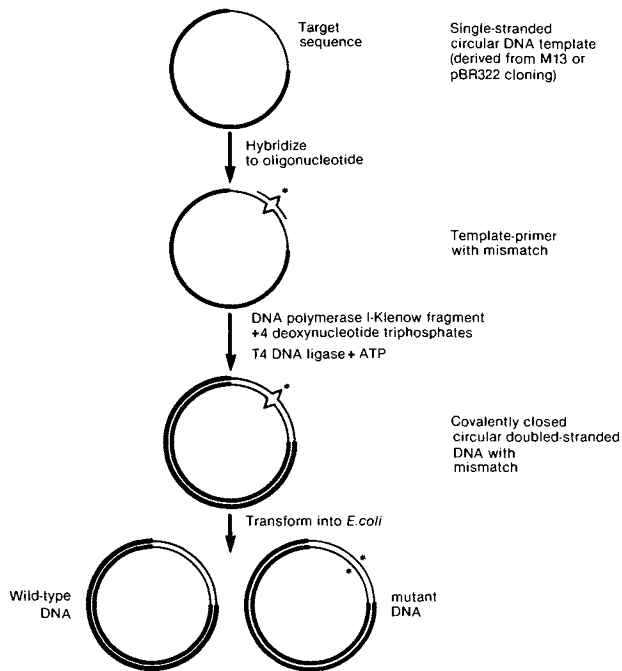

Fig. 8. The use of synthetic oligonucleotides in site-directed *in vitro* mutagenesis.

general review has been published recently (119) which discusses how amino acid modification generally can be used for understanding the biological function of proteins. Several examples of the results achieved by this technique follow. (i) The earliest applications of this technique were to probe the active site of β-lactamase, the bacterial protein which mediates ampicillin resistance. The Ser^{70}–Thr^{71} (meaning threonine at amino acid position 71) dyad in the presumptive active site was inverted to Thr^{70}–Ser^{71}; the resulting mutant no longer had enzymatic activity (120). In another study, it was shown that the replacement of Ser^{70} by Cys^{70} did not destroy enzymatic activity (121) but did alter substrate specificity (122). (ii) In the active site of the tyrosyl tRNA synthetase of *Bacillus stearothermophilus*, Cys^{35} in the active site was converted to Ser^{35} (123) and Thr^{51} was converted to either Ala^{51} or Pro^{51} (124). Among these mutants, enzyme activity was lowered 3-fold in one case and raised 25-fold in another case, apparently as a result of an altered K_m for ATP. (iii) Dihydrofolate reductase (DHFR), found in every known organism, plays a key metabolic role in the biosynthesis of several molecules. Three mutations in the enzyme were constructed at sites which, based upon X-ray crystal structural analyses of amino acid residues, were important in DHFR structure and function: $Asp^{27} \rightarrow Asn^{27}$, $Pro^{39} \rightarrow Cys^{39}$, and $Gly^{95} \rightarrow Ala^{95}$ (125). The mutant proteins displayed structural alterations consistent with roles for these residues in catalysis and three-dimensional protein structure. (iv) Ribulose-1,5-biphosphate carboxylase is an important enxyme in photosynthesis. The $Asp^{198} \rightarrow Glu^{198}$ mutation displayed altered kinetic properties, suggesting that the carboxyl group of Asp^{198} contributes to the formation of the divalent metal ion binding site (126). (v) The antiprotease α_1-antitrypsin inhibits hydrolysis of connective tissues, in particular lung. The active-site residue Met^{358} has been altered to either Val^{358} (57,127) or Arg^{358} (128). Changing the Met rendered the enzyme oxidation-resistant and also changed the antiprotease specificity of the enzyme. (vi) The Phe^{87} residue of yeast cytochrome *c*, having been suggested to be in a functionally important region, was changed to Ser^{87}, Tyr^{87}, or Gly^{87} (129). While the mutant proteins were equally active in electron transport as the wild type, their reduction potentials were lowered significantly. (vii) A functional analysis of the acetylcholine receptor was performed with respect to the contribution of several mutations in the α-subunit to overall receptor function (130). A total of 20 deletions (by means other than oligonucleotide mutagenesis) and 6 point mutations was constructed. In this way, several regions and individual residues were implicated in α-subunit function. (viii) Subtilisin, a serine protease of *Bacillus amyloliquefaciens*, can be inactivated oxidatively at Met^{222} (131). Nineteen mutant proteins at position 222 were created by substituting each of the 19 other amino acids for Met (132). These mutations were produced efficiently by means of the "cassette mutagenesis" technique (133). In this method, nearby restriction endonuclease sites are created that flank the condon on interest and that do not create any changes in the condons at the new restriction endonuclease sites. The wild-type DNA between the two sites is excised and replaced by synthetic DNA duplexes containing the desired mutation. In the case of subtilisin, 19 different 25-bp duplexes were utilized. It was found that some of the substitutions were highly resistant to oxidation. (ix) In another study of subtilisin from *Bacillus amyloliquefaciens*, based upon crystallographic data, the $Asp^{99} \rightarrow Ser^{99}$ mutation near the active site significantly changed the pH dependence of the enzyme (134). (x) Mutants of trypsin, $Gly^{216} \rightarrow Ala^{216}$ and $Gly^{226} \rightarrow Ala^{226}$, demonstrated reduced catalytic rates but enhanced substrate specificities (135). (xi) Analysis of the $Tyr^{248} \rightarrow Phe^{248}$ mutation of carboxypeptidase A demonstrated that this residue is not critical to catalysis but that it does participate in ligand binding (136).

As new genes for different proteins are cloned and sequenced, we can anticipate that the use of site-directed oligonucleotide-mediated *in vitro* mutagenesis will contribute to an understanding of the structure and function of these proteins. While the scope for analysis is very broad, this must be placed in perspective. More than one region of a protein can be involved in a particular aspect of structure and function; thus, analysis of mutants in only limited regions of a protein will yield only partial information. Nevertheless, the versatility of this technique makes it extremely useful for initiating such analyses.

Analysis of Noncoding Regions of Genes

DNA segments that do not encode mature protein products can have a large range of functions such as transcriptional initiation signals, transcriptional termination signals, transcriptional enhancement, and intracellular protein routing. *In vitro* mutagenesis has been very useful in dissecting the qualitative and quantitative nature of these functions.

In one of the first uses of this technique, Wallace *et al.* (137) created a 14-bp deletion in a region of the yeast tyrosine transfer RNA suppressor gene known as an intron, or intervening sequence. This is a region present in the initial RNA transcript of the gene that is removed (spliced out)

during processing to the final mature form of RNA. The altered gene was fully functional, thus demonstrating the nonessential nature of these intron sequences.

The function of particular individual nucleotide residues in transcriptional initiation regions (promoter sequences) has been studied. One such sequence in the promoter for chicken albumin received a single A → G mutation, which was shown to markedly reduce the efficiency of gene expression, i.e., transcriptional initiation (138). Recently, Kozak demonstrated that in a particular eukaryotic gene a single C → A mutation at nucleotide −3 to the 5′ side of the ATG translational initiation codon increased the efficiency of initiation at that ATG by a factor of 15 (139).

Another application of this approach is in the analysis of sequences responsible for signalling the entry of proteins into the secretory pathway of the cell. While these sequences initially are transcribed and translated into amino acids, these "signal" amino acids are cleaved proteolytically after the entry of the polypeptide chain into the secretory pathway, thus being absent in the mature protein product. While one study showed that a Ser → Arg mutation at the second amino acid residue of the signal sequence did not affect the secretion of β-lactamase (140), another study did identify key amino acid residues in signalling secretion. The first five amino acids of the natural secretion signal sequence of the major *E. coli* outer membrane lipoprotein have a net charge of +2. Mutants were created containing net charges of +2, +1, 0, and −1 in these residues. With increasing negative charge, less of the protein was secreted, thus implicating net charge as a contributing factor to the secretion signal (141).

Such approaches have great potential in defining the role of individual nucleotides in the regulation of the flow of genetic information DNA → RNA → protein.

CONCLUSION

Oligonucleotides rapidly are becoming routine laboratory reagents, and techniques for their use are increasingly easier to perform. Their potential applications and multiplicity of targets are essentially unlimited. We can anticipate fruitful applications not only in the above-mentioned areas but also in areas yet to be developed.

ACKNOWLEDGMENTS

I thank P. Kniskern for proofreading, R. Zeigler for the excellent preparation of artwork, and G. Albanesius for the careful typing of the manuscript.

REFERENCES

1. H. G. Khorana. *Science* 203:614–625 (1979).
2. M. H. Caruthers. *Science* 230:281–285 (1985).
3. S. A. Narang. *Tetrahedron* 39:3–22 (1983).
4. J. E. Davies and H. G. Gassen. *Angew. Chem. Int. Ed. Engl.* 22:13–31 (1983).
5. K. Itakura, J. J. Rossi, and R. B. Wallace. *Annu. Rev. Biochem.* 53:323–356 (1984).
6. C. B. Reese. *Tetrahedron* 34:3143–3179 (1980).
7. E. Ohtsuka, M. Ikehara, and D. Soll. *Nucleic Acids Res.* 10:6553–6570 (1982).
8. R. C. Sheppard. *Chem. Br.* 19:402–429 (1983).
9. M. Gait. (ed.) *Oligonucleotide Synthesis,* IRL Press, Oxford, 1984.
10. T. Maniatis, R. C. Hardbon, E. Lacy, J. Laver, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis. *Cell* 15:687–701 (1970).
11. F. Rudgeon, P. Kovrilsky, and B. Mach. *Nucleic Acids Res.* 2:2365–2378 (1975).
12. T. Maniatis, S. G. Kee, A. Efstratiadis, and F. C. Kafatos. *Cell* 8:163–182 (1976).
13. S. Tonegawa, C. Brack, N. Hozumi, and R. Schuller. *Proc. Natl. Acad. Sci. USA* 74:3518–3522 (1977).
14. M. Oskarrson, W. L. McClements, D. O. Blair, J. V. Maizel, and G. F. van de Woude. *Science* 207:1222–1224 (1980).
15. D. DeFeo, M. A. Gonda, H. A. Young, E. H. Chang, D. R. Lowy, E. M. Scolnick, and R. W. Ellis. *Proc. Natl. Acad. Sci. USA* 78:3328–3332 (1981).
16. S. K. Niyogi and C. A. Thomas. *J. Biol. Chem.* 243:1220–1223 (1968).
17. P. Besmer, R. C. Miller, M. H. Caruthers, A. Kumar, K. Minamoto, J. H. van de Sande, N. Sidarova, and H. G. Khorana. *J. Mol. Biol.* 72:503–522 (1972).
18. R. Wu. *Nature New Biol.* 236:198–200 (1972).
19. R. Wu, C. P. D. Tu, and R. Padmanabhan. *Biochem. Biophys. Res. Commun.* 55:1092–1099 (1973).
20. M. T. Doel and M. Smith. *FEBS Lett.* 34:99–102 (1973).
21. W. Gilbert. *Science* 214:1305–1312 (1981).
22. F. Sanger. *Science* 214:1205–1210 (1981).
23. E. M. Southern. *J. Mol. Biol.* 98:503–517 (1975).
24. R. B. Wallace, J. Schaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura. *Nucleic Acids Res.* 6:3543–3557 (1979).
25. J. B. Dodgson and R. D. Wells. *Biochemistry* 16:2367–2372 (1977).
26. J. W. Szostak, J. I. Stiles, B.-K. Tye, P. Chiu, F. Sherman, and R. C. Wu. *Methods Enzymol.* 68:419–428 (1979).
27. R. B. Wallace. In S. Woo (ed.), *DNA Recombinant Technology: Methods,* CRC Press, Cleveland, Ohio, 1984.
28. S. V. Suggs, T. Hirose, T. Miyake, E. H. Kawashima, M. J. Johnson, K. Itakura, and R. B. Wallace. In D. Brown (ed.), *Developmental Biology Using Purified Genes,* Academic Press, New York, 1982, pp. 683–693.
29. R. B. Wallace, M. J. Johnson, T. Hirose, T. Miyake, E. H. Kawashima, and K. Itakura. *Nucleic Acids Res.* 9:879–894 (1981).
30. Y. Ike, S. Ikuta, M. Sato, T. Huang, and K. Itakura. *Nucleic Acids Res.* 11:477–488 (1983).
31. S. V. Suggs, R. B. Wallace, T. Hirose, E. H. Kawashima, and K. Itakura. *Proc. Natl. Acad. Sci. USA* 78:6613–6617 (1981).
32. B. A. Cunningham, J. L. Wang, I. Berggard, and P. A. Peterson. *Biochemistry* 12:4811–4821 (1973).
33. A. S. Whitehead, G. Goldberger, D. E. Woods, A. F. Markham, and R. Colten. *Proc. Natl. Acad. Sci. USA* 80:5387–5391 (1983).
34. J. Singer-Sam, R. L. Simmer, D. H. Keith, L. Shively, M. Teplitz, K. Itakura, S. M. Gartler, and A. D. Riggs. *Proc. Natl. Acad. Sci. USA* 80:802–806 (1983).
35. J. B. Hurley, H. K. W. Fong, D. B. Teplow, W. J. Dreyer, and M. I. Simon. *Proc. Natl. Acad. Sci. USA* 81:6948–6952 (1984).
36. M. A. Bourdon, A. Oldberg, M. Pierschbacker, and E. Ruoslahti. *Proc. Natl. Acad. Sci. USA* 82:1321–1325 (1985).
37. W. I. Wood, D. J. Capon, C. C. Simonsen, D. C. Eaton, J. Gitschier, B. Keyt, P. H. Seeburg, D. H. Smith, P. Hollingshead, K. L. Wion, E. Delwart, E. G. D. Tuddenham, G. A. Vehar, and R. M. Lawn. *Nature* 312:330–337 (1984).
38. J. J. Toole, J. L. Knopf, J. M. Wozney, L. A. Sultzman, J. L. Buecker, D. D. Pittman, R. J. Kaufman, E. Brown, C. Shoemaker, E. C. Orr, G. W. Amphlett, W. B. Foster, M. L. Coe, G. J. Knutson, D. N. Fass, and R. M. Hewick. *Nature* 212:342–347 (1984).
39. M. Jaye, H. de la Salle, F. Schamber, A. Balland, V. Kohli, A. Findeli, P. Tolstoshev, and J.-P. Lecocq. *Nucleic Acids Res.* 11:2325–2335 (1983).
40. K. L. Agarwal, T. J. Brunsted, and B. E. Noyes. *J. Biol. Chem.* 256:1023–1028 (1981).

41. S. Anderson and I. B. Kingston. *Proc. Natl. Acad. Sci. USA* 80:6838–6842 (1983).
42. A. Ullrich, C. H. Berman, T. J. Dull, A. Gray, and J. M. Lee. *EMBO J.* 3:361–364 (1984).
43. B. S. Segundo, S. J. Chan, and D. F. Steiner. *Proc. Natl. Acad. Sci. USA* 82:2320–2324 (1985).
44. A. Ullrich, J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokura, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran. *Nature* 313:756–761 (1985).
45. R. Lathe. *J. Mol. Biol.* 183:1–12 (1985).
46. M. Noda, Y. Furutani, H. Takahashi, M. Toyosato, T. Hirose, S. Inayama, S. Nakanishi, and S. Numa. *Nature* 295:202–208 (1982).
47. A. Ullrich, A. Gray, C. Berman, and T. J. Dull. *Nature* 303:821–825 (1983).
48. D. Pennica, W. E. Holmes, W. J. Kohr, R. N. Harkins, G. A. Vehar, C. A. Ward, W. F. Bennett, E. Yelverton, P. H. Seeburg, H. L. Heyneker, and D. V. Goeddel. *Nature* 301:214–221 (1983).
49. D. Cosman, D. P. Cerretti, A. Larsen, L. Park, C. March, S. Dower, S. Gillis, and D. Urdal. *Nature* 312:768–771 (1984).
50. E. Ohtsuka, S. Matsuki, M. Ikehara, Y. Takahashi, and K. Matsubara. *J. Biol. Chem.* 260:2605–2608 (1985).
51. T. Huynh-Dinh, N. Duchange, M. M. Zakin, A. Lemarchand, and J. Igolen. *Proc. Natl. Acad. Sci. USA* 82:7510–7514 (1985).
52. G. E. Shull, A. Schwartz, and J. B. Lingrel *Nature* 316:691–695 (1985).
53. D. H. MacLennan, C. J. Brandl, B. Korczak, and N. M. Green. *Nature* 316:696–700 (1985).
54. G. G. Wong, J. S. Witek, P. A. Temple, K. M. Wilkens, A. C. Leary, D. P. Luxenberg, S. S. Jones, E. L. Brown, R. M. Kay, E. C. Orr, C. Shoemaker, D. W. Golde, R. J. Kaufman, R. M. Hewick, E. A. Wang, and S. C. Clark. *Science* 228:810–815 (1985).
55. A. Bollen, A. Herzog, A. Cravador, P. Herion, P. Chuchana, A. van der Straten, R. Loriau, P. Jacobs, and A. van Elsen. *DNA* 2:255–264 (1983).
56. M. Courtney, A. Buchwalder, L.-H. Tessier, M. Jaye, A. Benavente, A. Balland, C. Kohli, R. Lathe, P. Tolstoshev, and J.-P. Lecocq. *Proc. Natl. Acad. Sci. USA* 81:669–673 (1984).
57. S. Rosenberg, P. J. Barr, R. C. Najarian, and R. A. Hallewell. *Nature* 312:77–80 (1984).
58. D. H. Calhoun, D. F. Bishop, H. S. Bernstein, M. Quinn, P. Hantzopoulos, and R. J. Desnick. *Proc. Natl. Acad. Sci. USA* 82:7364–7368 (1985).
59. T. J. Knott, S. C. Rall, Jr., T. L. Innerarity, S. F. Jacobson, M. S. Urdea, B. Levy-Wilson, L. M. Powell, R. J. Pease, R. Eddy, H. Nakai, M. Byers, L. M. Priestley, E. Robertson, L. B. Rall, C. Betsholtz, T. V. Shows, R. W. Mahley, and J. Scott. *Science* 230:37–43 (1985).
60. E. S. Kawasaki, M. B. Ladner, A. M. Wang, J. van Arsdell, M. K. Warren, M. Y. Coyne, V. L. Schweickart, M.-T. Lee, K. J. Wilson, A. Boosman, E. R. Stanley, P. Ralph, and D. F. Mark. *Science* 230:291–296 (1985).
61. K. Jacobs, C. Shoemaker, R. Rudersdorf, S. D. Neill, R. J. Kaufman, A. Mufson, J. Sechra, S. S. Jones, R. Hewick, E. F. Fritsch, M. Kawakita, T. Shimizu, and T. Miyake. *Nature* 313:806–810 (1985).
62. R. G. DiScipio, M. R. Gehring, E. R. Podack, C. C. Kan, T. E. Hugli, and G. H. Fey. *Proc. Natl. Acad. Sci. USA* 81:7298–7302 (1984).
63. E. V. Prochownik, A. F. Markham, and S. H. Orkin. *J. Biol. Chem.* 258:8389–8394 (1983).
64. P. Jacobs, A. Cravador, R. Loriau, F. Brockly, B. Colau, P. Chuchana, A. van Elsen, A. Herzog, and A. Bollen. *DNA* 4:139–146 (1985).
65. K. E. Mayo, W. Vale, J. Rivier, M. G. Rosenfeld, and R. M. Evans. *Nature* 306:86–88 (1983).
66. E. Ohtsuka, Y. Taniyama, S. Iwai, K. Kitano, S. Miyamoto, T. Ohgi, Y. Sakuragawa, K. Fujiyama, T. Ikari, M. Kobayashi, T. Miyake, S. Shibahara, T. Tokunaga, M. Ikehara, A. Ono, T. Veda, T. Tanaka, H. Baba, T. Miki, A. Sakurai, and T. Oishi. *Nucleic Acid Res. Symp. Ser.* 12:79–82 (1983).
67. M. S. Urdea, J. P. Merryweather, G. T. Mullenbach, D. Cort, U. Heberlen, P. Valenzuela, and P. J. Barr. *Proc. Natl. Acad. Sci. USA* 80:7461–7467 (1983).
68. B. S. Sproat and M. J. Gait. *Nucleic Acids Res.* 13:2959–2977 (1985).
69. W. Mandecki, K. W. Mollison, T. J. Bolling, B. S. Powell, G. W. Carter, and J. L. Fox. *Proc. Natl. Acad. Sci. USA* 82:3543–3547 (1985).
70. M. D. Edge, A. R. Greene, G. R. Heathcliffe, P. A. Meacock, W. Schuch, D. B. Scanlon, T. C. Atkinson, C. R. Newton, and A. F. Markham. *Nature* 292:756–760 (1981).
71. J. J. Rossi, R. Kierzek, T. Huang, P. Walker, and K. Itakura. *J. Biol. Chem.* 257:9226–9229 (1982).
72. R. C. Scarpulla, S. A. Narang, and R. Wu. *Anal. Bio. Chem.* 121:353–365 (1982).
73. M. Gouy and C. Gautier. *Nucleic Acids Res.* 10:7055–7074 (1982).
74. J. L. Bennetzen and B. D. Hall. *J. Biol. Chem.* 257:3026–3031 (1982).
75. G. A. Bitter and K. M. Egan. *Gene* 32:263–274 (1984).
76. J. D. Windass, C. R. Newton, J. DeMaeyer-Guignard, V. E. Moore, and A. F. Markham. *Nucleic Acids Res.* 10:6639–6658 (1982).
77. P. L. deHaseth, R. A. Goldman, C. L. Cech, and M. H. Caruthers. *Nucleic Acids Res.* 11:773–787 (1983).
78. T. Miyake, T. Oka, S. Sumi, and M. Susuki. *Nucleic Acids Res. Symp.* 11:81–83 (1982).
79. J. J. Rossi, X. Soberon, Y. Marumoto, J. McMahon, and K. Itakura. *Proc. Natl. Acad. Sci. USA* 80:3203–3207 (1983).
80. J. Messing, B. Gronenborn, B. Muller-Hill, and P. H. Hofschneider. *Proc. Natl. Acad. Sci. USA* 74:3642–3646 (1977).
81. J. Messing, R. Crea, and P. H. Seeburg. *Nucleic Acids Res.* 9:309–321 (1981).
82. P. P. Lau and H. B. Gray. *Nucleic Acids Res.* 6:331–339 (1979).
83. S. L. McKnight and R. Kingsbury. *Science* 217:316–324 (1982).
84. S. McKnight. *Cell* 31:355–365 (1982).
85. R. A. Flavell, J. M. Kooter, E. DeBoer, P. F. R. Little, and R. Williamson. *Cell* 15:25–41 (1978).
86. R. F. Geever, L. B. Wilson, F. S. Nallaseth, P. F. Milner, M. Bittner, and J. T. Wilson. *Proc. Natl. Acad. Sci. USA* 78:5081–5082 (1981).
87. J. C. Chang and Y. W. Kan. *N. Engl. J. Med.* 307:30–32 (1982).
88. S. H. Orkin, P. F. R. Little, H. H. Kazazian, and C. D. Boehm. *N. Engl. J. Med.* 307:32–36 (1982).
89. M. Goossens, K. Y. Lee, S. A. Liebhaber, and Y. W. Kan. *Nature* 296:864–865 (1982).
90. K. Kurachi, T. Chandra, S. J. F. Degen, T. T. White, T. L. Marchioro, S. L. C. Woo, and E. W. Davie. *Proc. Natl. Acad. Sci. USA* 78:6826–6830 (1981).
91. R. B. Wallace, M. Schold, M. J. Johnson, P. Dembeck, and K. Itakura. *Nucleic Acids Res.* 9:3647–3656 (1981).
92. B. J. Conner, A. A. Reyes, C. Morin, K. Itakura, R. L. Teplitz, and R. B. Wallace. *Proc. Natl. Acad. Sci. USA* 80:278–282 (1983).
93. L. Pauling, H. A. Itano, S. J. Singer, and I. C. Wells. *Science* 25:543–548 (1949).
94. T. M. Shinnick, E. Lund, O. Smithies, and F. R. Blattner. *Nucleic Acids Res.* 2:1911–1929 (1975).
95. S. H. Orkin, A. F. Markham, and H. H. Kazazian. *J. Clin Invest.* 71:775–779 (1983).
96. M. Piratsu, Y. W. Kan, A. Cao, B. J. Conner, R. L. Teplitz, and R. B. Wallace. *N. Engl. J. Med.* 309:284–287 (1983).
97. V. J. Kidd, R. B. Wallace, K. Itakura, and S. L. C. Woo. *Nature* 304:230–234 (1983).
98. N. M. Green. *Adv. Protein Chem.* 29:85–133 (1975).
99. J. E. Manning, N. D. Hershey, T. R. Broker, M. Pellegrini, and N. Davidson. *Chromosoma* 53:107–117 (1975).
100. P. R. Langer, A. A. Waldrop, and D. C. Ward. *Proc. Natl. Acad. Sci. USA* 78:6633–6637 (1981).
101. J. J. Leary, D. J. Brigati, and D. C. Ward. *Proc. Natl. Acad. Sci. USA* 80:4045–4049 (1983).
102. A. Murasugi and R. B. Wallace. *DNA* 3:269–277 (1984).

103. T. Kempe, W. I. Sundquist, F. Chow, and S.-L. Hu. *Nucleic Acids Res.* 13:45–57 (1985).
104. A. B. Studencki and R. B. Wallace. *DNA* 3:7–15 (1984).
105. D. Shortle, D. DiMaio, and D. Nathans. *Annu. Rev. Genet.* 15:265–294 (1981).
106. D. Botstein and D. Shortle. *Science* 229:1193–1201 (1985).
107. M. Smith. *Annu. Rev. Genet.* 19:423–462 (1985).
108. A. Razin, T. Hirose, K. Itakura, and A. D. Riggs. *Proc. Natl. Acad. Sci. USA* 75:4268–4270 (1978).
109. S. Gilliam and M. Smith. *Gene* 8:81–97 (1979).
110. S. Gilliam and M. Smith. *Gene* 8:99–106 (1979).
111. R. B. Wallace, P. F. Johnson, S. Tanaka, M. Schold, K. Itakura, and J. Abelson. *Science* 209:1396–1400 (1980).
112. M. Schold, A. Colombero, A. A. Reyes, and R. B. Wallace. *DNA* 3:469–477 (1984).
113. M. J. Zoller and M. Smith. *Nucleic Acids Res.* 10:6487–6500 (1982).
114. M. J. Zoller and M. Smith. *Methods Enzymol.* 100:469–500 (1983).
115. M. J. Zoller and M. Smith. *DNA* 3:479–488 (1984).
116. K. Norris, F. Norris, L. Christiansen, and N. Fiil. *Nucleic Acids Res.* 11:5103–5112 (1983).
117. C. E. Bauer, S. D. Hesse, D. A. Waechter-Brulla, S. P. Lynn, R. I. Gumport, and J. F. Gardner. *Gene* 37:73–81 (1985).
118. P. Carter, H. Bedouelle, and G. Winter. *Nucleic Acids Res.* 13:4431–4442 (1985).
119. G. K. Ackers and F. R. Smith. *Annu. Rev. Biochem.* 54:597–629 (1985).
120. G. Dalbadie-McFarland, L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura, and J. H. Richards. *Proc. Natl. Acad. Sci. USA* 79:6409–6413 (1982).
121. I. S. Sigal, B. G. Harwood, and R. Arentzen. *Proc. Natl. Acad. Sci. USA* 79:7157–7160 (1982).
122. I. S. Sigal, W. F. DeGrado, B. J. Thomas, and S. R. Petteway. *J. Biol. Chem.* 259:5327–5332 (1984).
123. G. Winter, A. R. Fersht, A. J. Wilkinson, M. Zoller, and M. Smith. *Nature* 299:756–788 (1982).
124. A. J. Wilkinson, A. R. Fersht, D. M. Blow, P. Carter, and G. Winter. *Nature* 307:187–188 (1984).
125. J. E. Villafranca, E. E. Howell, D. H. Voet, M. S. Strobel, R. C. Ogden, J. N. Abelson, and J. Kraut. *Science* 222:782–788 (1983).
126. S. Gutteridge, I. Sigal, B. Thomas, R. Arentzen, A. Cordova, and G. Lorimer. *EMBO J.* 3:2737–2743 (1984).
127. J. Travis, M. Owen, P. George, R. Carrell, S. Rosenberg, R. A. Hallewell, and P. J. Barr. *J. Biol. Chem.* 260:4384–4369 (1985).
128. M. Courtney, S. Jallat, L.-H. Tessier, A. Benavente, R. G. Crystal, and J.-P. Lecocq. *Nature* 313:149–151 (1985).
129. G. J. Pielak, A. G. Mauk, and M. Smith. *Nature* 313:152–154 (1985).
130. M. Mishina, T. Tobimatsu, K. Imoto, K. Tanaka, Y. Fujita, K. Fukuda, M. Kurasaki, H. Takahashi, Y. Morimoto, T. Hirose, S. Inayama, T. Takahashi, M. Kuno, and S. Numa. *Nature* 313:364–369 (1985).
131. C. E. Stauffer and D. Etson. *J. Biol. Chem.* 244:5333–5338 (1969).
132. D. A. Estell, T. P. Graycar, and J. A. Wells. *J. Biol. Chem.* 260:6518–6521 (1985).
133. J. A. Wells, M. Vasser, and D. B. Powers. *Gene* 34:315–323 (1985).
134. P. G. Thomas, A. J. Russell, and A. R. Fersht. *Nature* 318:375–376 (1985).
135. C. S. Craik, C. Largman, T. Fletcher, S. Roczniak, P. J. Barr, R. Fletterick, and W. J. Rutter. *Science* 228:191–297 (1985).
136. S. J. Gardell, C. S. Craik, D. Hilvert, M. S. Urdea, and W. J. Rutter. *Nature* 317:551–555 (1985).
137. R. B. Wallace, P. F. Johnson, S. Tanaka, M. Schold, K. Itakura, and J. Abelson. *Science* 209:1396–1400 (1980).
138. T. Zarucki-Schulz, S. Y. Tsai, K. Itakura, X. Soberon, R. B. Wallace, M.-J. Tsai, S. L. C. Woo, and B. W. O'Malley. *J. Biol. Chem.* 257:11070–11077 (1982).
139. M. Kozak. *Nature* 308:241–246 (1984).
140. A. D. Charles, A. E. Gautier, M. D. Edge and J. R. Knowles. *J. Biol. Chem.* 14:7930–7932 (1982).
141. S. Inouye, X. Soberon, T. Franceschini, K. Nakamura, K. Irakura, and M. Inouye. *Proc. Natl. Acad. Sci. USA* 79:3438–3441 (1982).