Introduction

While a number of methods are available for site-directed mutagenesis, the use of a properly designed synthetic gene offers many advantages, particularly where extensive mutagenesis is planned. A synthetic gene used for heterologous expression codes for the correct amino acid sequence, but contains an artificial nucleotide sequence. The nucleotide sequence is generally designed to contain a large number of unique restriction endonuclease cleavage sites. The degeneracy of the genetic code and the large number of available restriction enzymes allows on the order of 40 unique restriction sites to be introduced for a synthetic gene of 1-kb pairs in length.

Mutagenesis using a synthetic gene involves the replacement of a restriction fragment within the gene by a synthetic DNA duplex counterpart that contains the desired codon alteration(s). Cassette mutagenesis can be used to introduce point mutations, or extensive changes such as deletions, insertions, and multiple nucleotide substitutions as desired. Random mutagenesis techniques such as combinatorial cassette mutagenesis are also facilitated by using a synthetic gene. Many of these site- or domain-directed mutagenesis procedures are difficult to accomplish by mismatch primer methods.

Gene synthesis has been particularly useful in the study of G protein-coupled receptors. The superfamily of G protein-coupled receptors consists of a large number of related seven-transmembrane segment plasma membrane proteins (1). In studying receptor function by heterologous expression of mutant receptors, it is often useful to create chimeric receptors in which putative functional domains are exchanged. For example, the transfer of a cytoplasmic loop sequence from one pharmacological receptor subtype to another is one approach to study the specificity of ligand-dependent G protein activation. A synthetic gene for one receptor subtype can be engineered readily to produce either a chimeric construct or the mutant gene for a functionally different receptor subtype. The use of synthetic genes allows domain exchanges without the potential limitations of naturally occurring restriction endonuclease cleavage sites.
In this chapter, a summary of the mutagenesis methods facilitated by the use of synthetic genes is presented. The leading gene synthesis strategies are compared and contrasted. A general method for designing a synthetic gene is discussed in detail. Detailed procedures for the various steps in gene synthesis, from oligonucleotide preparation to cassette mutagenesis, are also given. Specific examples of the use of synthetic genes in the study of receptor structure and function are presented.

Mutagenesis Methods Facilitated by the Use of Synthetic Genes

The use of a synthetic gene facilitates mutagenesis methods that rely on restriction fragment replacement (cassette mutagenesis). Site-directed point mutations are easily introduced into cloned DNA by a variety of mismatch primer methods, some of which employ the polymerase chain reaction (PCR). However, site-directed cassette mutagenesis can be successfully employed to introduce extensive alterations of the nucleotide sequence within a particular gene segment. Such an approach may be useful, for example, in structure–function studies of discrete receptor domains, such as the cytoplasmic loops of a G protein-coupled receptor (2–4). A long stretch of amino acid residues easily can be replaced by a random sequence or by a homologous sequence from a related protein.

In addition, a number of random mutagenesis methods have developed based on cassette mutagenesis. For example, the classical approach to combinatorial mutagenesis involves the synthesis and cloning of a "random" DNA duplex. The utility of random mutagenesis ultimately depends on the existence of an adequate functional screening method. Various mutagenesis methods that are facilitated by the use of a synthetic gene are presented below.

Mutagenesis by Restriction Fragment Replacement
(Cassette Mutagenesis)

Mutagenesis by restriction fragment replacement was first demonstrated in the naturally occurring gene of bacteriorhodopsin (5). This mutagenesis strategy was possible because of the fortuitous natural placement of unique restriction sites. Cassette mutagenesis using a synthetic gene also involves replacement of a restriction fragment by a synthetic duplex counterpart that contains the desired codon alteration(s). However, the presence of a
relatively large number of unique and evenly spaced restriction sites in a carefully designed synthetic gene allows the general use of the method. Because all of the restriction sites within the gene are unique, cloning is “directional” and screening is not generally necessary for identification of a desired recombinant transformant.

Cassette mutagenesis is also of more general utility than most forms of site-directed mismatch primer mutagenesis because of the ease of producing defined mutations at multiple sites within a domain to yield deletions, extensive substitutions, domain swaps, or the construction of chimeric genes.

One beautiful example of the creative use of cassette mutagenesis was reported in a disulfide cross-linking study of the *Escherichia coli* Tar receptor (6). Individual cysteine replacements were introduced at every position in a transmembrane domain. Interactions between the four transmembrane domains that comprise the Tar receptor dimer could then be assessed by measuring the tendency of cysteine mutants to form disulfide cross-links with one another. To produce cysteine mutants at 24 separate positions with minimal synthesis, oligonucleotides were synthesized simultaneously on two columns, one producing the native sequence and the second producing repeated cysteine codons. After each codon synthesis, the resins containing the elongating chains were exchanged in a defined proportion. The resulting mutagenic duplexes were cloned and the recombinants were characterized.

Cassette mutagenesis is also the basis of a series of random mutagenesis techniques collectively known as combinatorial cassette mutagenesis as described below.

**Combinatorial Cassette Mutagenesis**

The approach of combinatorial cassette mutagenesis has been used to determine the “informational content” of individual amino acid residues in a protein. Previous reports using this approach suggest that proteins tolerate a range of amino acid substitutions without significant changes in biochemical properties (7–9). For example, the effects of approximately 1500 single amino acid substitutions at 142 positions in the *lac* repressor were studied (10). About one-half of all of the mutations were phenotypically silent. At some amino acid positions many different nonconservative substitutions were tolerated. These positions must play little or no role in structure and function and, therefore, were deemed to have “low informational content.” At other positions, no substitutions or only conservative substitutions were tolerated without effects on function. These positions are the most important for *lac* repressor activity and were deemed to have “high informational content.”
These positions were generally hydrophobic residues involved in core amino acid packing and were crucial for protein folding or stability.

In transmembrane receptor molecules, where little if any structural information is known beyond primary structure, initial mutagenesis experiments that target a particular amino acid may not prove informative. It may be appropriate to prepare large numbers of mutations in a particular putative domain if an adequate functional screening method can be devised. Combinatorial cassette mutagenesis can be employed in such situations.

The general strategy of combinatorial cassette mutagenesis is to perform restriction fragment replacement with a set of synthetic duplexes that should provide codons for each of the 20 amino acids at one or more positions within the duplex (8). This can be accomplished by synthesizing the noncoding (top) strand of the duplex with equal mixtures of all four bases in the first two positions of a codon and with equal mixtures of guanine and cytosine at the third position. Inosine is inserted at each of the randomized base positions in the bottom strand because it is able to pair with each of the four natural bases. The heterogeneous top-strand oligonucleotides and the bottom-strand oligonucleotide are annealed and the resulting duplex is ligated into an appropriate vector. Bacterial transformation essentially produces a library of mutants which can be cloned or studied batchwise depending on particular circumstances.

Another method has been described to increase the chances of sampling large numbers of amino acid changes at all positions while removing wild-type background (11). This method involves contaminating the individual strands of the mutagenic duplex to different degrees. A method has been developed for combinatorial mutagenesis that does not rely on cassette mutagenesis. This strategy uses degenerate mismatch primer mutagenesis and can be applied to a gene with a selectable phenotype (12).

**Combinatorial Cassette Mutagenesis with a Limited Alphabet**

Because of the large number of possibilities in a random sequence, combinatorial mutagenesis with a limited alphabet can be carried out (13). In studies of the leucine zipper domain of the *Saccharomyces cerevisiae* transcription factor GCN4, cassette mutagenesis was used to produce substitutions of four codons at eight positions. The codons of interest were those for lysine, glutamic acid, and alanine. Threonine codons were unavoidably introduced as well. Even with this limited alphabet approach, randomization of eight positions with four codons should yield $4^8$ (65,536) unique sequences. A statistical analysis was carried out on a set of 121 mutants to quantify the contributions of specific sites on leucine zipper stability.
Binomial Cassette Mutagenesis

An additional method to limit the number of mutants produced while achieving the desired information has been called binomial mutagenesis (14). The method was designed to allow the study of the effect of a mutation in the context of other mutations. The goal is to identify amino acid residue pairs that may interact functionally despite being distant in primary or even tertiary structure. Alanine is substituted simultaneously at \( N \) positions with a probability of 0.5 to give a set of \( 2^N \) mutant proteins. The additivity of the effect of mutations is evaluated. The method was employed to study the helix-turn-helix of \( \lambda \) repressor.

Recursive Ensemble Mutagenesis

Recursive ensemble mutagenesis has been developed to allow successive rounds of combinatorial cassette mutagenesis. This method relies on an algorithm that allows information gained in one round of mutagenesis experiments to control the nucleotide composition of DNA synthesized for the next round. The result is that much more information can be gained because the relative number of "positive" mutants is greatly increased (15, 16).

Gene Synthesis Strategies

The general strategies discussed below relate to the construction of synthetic genes. Three main strategies are presented: (i) the Khorana method, (ii) the clamp method, and (iii) PCR methods. Other strategies such as stepwise synthesis with multiple cloning and amplification steps have been reported as well (17, 18). The actual design of synthetic genes, which in general involves the introduction of multiple restriction sites to facilitate mutagenesis by restriction fragment replacement, is presented later since the design process needs to be coordinated with the method chosen.

Khorana Method

The classical method of gene synthesis evolved prior to the development of automated oligonucleotide synthesis. This method involves the complete synthesis of both duplex strands. In 1976, Khorana and co-workers reported the total synthesis of the 126-bp gene for the precursor of a tyrosine suppressor transfer RNA from \( E. \) coli (19-21). The synthetic strategy involved the chemical synthesis of 26 oligodeoxynucleotides, the joining of several 5' end phosphorylated oligos to form four DNA duplexes, and finally the enzymatic
ligation of the four duplexes to form the full-length gene. The most important factor in this method is the high-efficiency DNA ligase-catalyzed joining of the oligonucleotides (22).

The Khorana method has evolved with automated DNA synthesis and a number of relatively long synthetic genes have been constructed by this approach. The first gene synthesis that exploited the use of solid-phase automated DNA synthesis in combination with efficient purification techniques such as reverse-phase HPLC was the synthesis of a 514-bp gene coding for human interferon-α1 (23). Nucleotide changes were introduced in this gene to eliminate complementary and repeated sequences and to optimize the sequence for E. coli codon usage. Subsequently, the desire to introduce restriction sites for cassette mutagenesis and to optimize codon usage for E. coli expression of eukaryotic genes has led to the synthesis of numerous genes based on these early methods (24–56).

The obvious advantage of the Khorana method is the high fidelity of synthesis and high general success rate since most of the synthesis is chemical and not enzymatic. The disadvantage is the time and expense involved in preparing oligonucleotides for complete upper and lower strands. However, these disadvantages can be somewhat overcome by economies of scale, the use of batchwise 5' end phosphorylation reactions of oligonucleotides, and ligation of up to 22 oligonucleotides in a single reaction. Another minor disadvantage is that gene design involves planning the duplex joining sites as well as the restriction sites. Although not foolproof, planning the duplex joining sites is relatively straightforward.

The Khorana method is presented schematically in Fig. 1 for the synthesis of the 1464-bp rat glucagon receptor gene (57). Other long synthetic genes synthesized by this method are listed in Table 1 (58–69).

**Clamp ("Fill-in") Method**

The basic principle of the clamp method of DNA synthesis was first proposed by Khorana (70). This method involves the synthesis of relatively long oligonucleotides from alternating DNA strands. The neighboring oligos have complementary 3' ends of 10 to 15 bp. After annealing of the oligos, the single-stranded portions are repaired by the action of DNA polymerase I (Klenow fragment or Sequenase) in the presence of excess deoxynucleotide 5'-triphosphates (dNTP's). This method is useful for short genes or fragments (200–300 bp) and may be justified by the reduction of synthesis required (58). The clamp method is presented schematically in Fig. 2.
1. Final synthetic gene design

2. Design of joining sites and synthesis of oligonucleotides

3. Ligation and isolation of fragments

4. Cloning and amplification of fragments

5. Assembly of full-length gene into vector

FIG. 1 Gene synthesis by the Khorana method. This method involves the total synthesis of both upper and lower DNA strands followed by ligation reactions to assemble synthetic duplexes into gene fragments. The fragments are then cloned, amplified, and subcloned to assemble the entire gene. A schematic is presented for the synthesis of the rat glucagon receptor gene as an example. The gene was designed with a large number of unique restriction sites (*). Three gene fragments were prepared in vitro: fragment A, EcoRI to MluI; fragment B, MluI to BamHI; fragment C, BamHI to NorI. Detailed procedures for gene synthesis by this method are presented in the text.

Polymerase Chain Reaction Methods

Several methods have been reported for the synthesis of genes using a combination of automated DNA synthesis and the polymerase chain reaction (66, 71, 72). One PCR method is presented schematically in Fig. 3 (66). It involves the synthesis of the full-length top and bottom DNA strands in single automated synthesizer runs. The crude oligonucleotide mixes are combined and full-length duplexes are amplified by PCR. This method was demonstrated by the synthesis of HIV-1 rev and nef genes of 393 and 655 bp, respectively.
Table I  Selected Synthetic Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length (base pair)</th>
<th>Synthetic method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucagon receptor</td>
<td>1472</td>
<td>Khorana method</td>
<td>Carruthers et al. (57)</td>
</tr>
<tr>
<td>Red cone pigment</td>
<td>1130</td>
<td>Khorana method</td>
<td>Oprian et al. (59)</td>
</tr>
<tr>
<td>Green cone pigment</td>
<td>1130</td>
<td>Khorana method</td>
<td>Oprian et al. (59)</td>
</tr>
<tr>
<td>Blue cone pigment</td>
<td>1080</td>
<td>Khorana method</td>
<td>Oprian et al. (59)</td>
</tr>
<tr>
<td>Transducin α subunit</td>
<td>1076</td>
<td>Khorana method</td>
<td>Sakmar and Khorana (60)</td>
</tr>
<tr>
<td>Rhodopsin</td>
<td>1048</td>
<td>Khorana method</td>
<td>Ferretti et al. (61)</td>
</tr>
<tr>
<td>Thymidylate synthase</td>
<td>978</td>
<td>Khorana method</td>
<td>Clinie and Santi (62)</td>
</tr>
<tr>
<td>Mirabilis antiviral</td>
<td>759</td>
<td>Khorana method</td>
<td>Habuka et al. (63)</td>
</tr>
<tr>
<td>protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriorhodopsin</td>
<td>757</td>
<td>Khorana method</td>
<td>Nassal et al. (64)</td>
</tr>
<tr>
<td>Sensory rhodopsin-I</td>
<td>720</td>
<td>Khorana method</td>
<td>Krebs et al. (65)</td>
</tr>
<tr>
<td>HIV-1 nef</td>
<td>655</td>
<td>PCR method</td>
<td>Ciccarelli et al. (66)</td>
</tr>
<tr>
<td>HBV core protein</td>
<td>560</td>
<td>Khorana method</td>
<td>Nassal (67)</td>
</tr>
<tr>
<td>Human interleukin-1α</td>
<td>484</td>
<td>Khorana method</td>
<td>Yanofsky and Zurawski (68)</td>
</tr>
<tr>
<td>HIV-1 rev</td>
<td>393</td>
<td>PCR method</td>
<td>Ciccarelli et al. (66)</td>
</tr>
<tr>
<td>Charybdoxin</td>
<td>151</td>
<td>Khorana method</td>
<td>Park et al. (69)</td>
</tr>
</tbody>
</table>

*Selected genes synthesized since 1985 are listed for illustrative purposes. Genes are listed in descending order of size. A previous review contains a detailed discussion of earlier work (58).*

Synthetic Gene Design

An overview of synthetic gene design is presented below that includes the choice of restriction endonuclease cleavage sites and the choice of an appropriate nucleotide sequence. The discussion applies to all methods for gene synthesis. However, the Khorana method involves the design of joining sites in addition to other considerations.

Choice of Restriction Endonuclease Cleavage Sites

The aim of gene synthesis is to prepare a DNA duplex which encodes the amino acid sequence of a particular protein while containing an optimal number of unique restriction sites. The choice of restriction endonuclease sites to be considered in the design of a synthetic gene includes the following criteria: (i) reliable availability, (ii) high activity and freedom from any exonuclease activity, (iii) a recognition sequence of five or more nucleotides, and (iv) the generation of staggered rather than blunt ends. Because of the degeneracy of the universal genetic code, a very large number of potential
1. Final synthetic gene design

2. Design and synthesis of overlapping oligonucleotides

3. Extension with DNA polymerase and D-NTP's

4. Cloning and amplification of fragments
5. Assembly of full-length gene into vector

Fig. 2 The synthesis by the clamp (fill-in) method. This method involves the partial synthesis of upper and lower DNA strands followed by DNA polymerase reactions to fill in single-stranded portions. This method can be used for small- to medium-sized genes, or for fragments, which can then be assembled as in Fig. 1.

nucleotide sequences can encode a given amino acid sequence. This potential variability in nucleotide sequence generates a large number of potential restriction maps.

The traditional approach for synthetic gene design was to begin with the native DNA sequence and restriction map, retain all potentially useful restriction sites, and then attempt to add new sites in the intervening sequences (61). This approach, however, is not general. Manual approaches were used to reverse translate restriction endonuclease recognition sequences in order to consider the locations of all possible restriction sites in a particular amino acid sequence (60). More recently, this general approach has been greatly facilitated by the availability of sequence analysis software packages that allow the identification of all of the potential restriction sites within a putative gene. This can be accomplished by starting with the amino acid sequence and using a reverse translation algorithm. A fully degenerative nucleotide sequence is thus created from the initial amino acid sequence. A restriction map can then be created from the degenerate DNA sequence. Once the
1. Final synthetic gene design

2. Synthesis of full-length top and bottom strands

3. Combine crude top and bottom synthesis products with PCR primers

4. PCR amplify and clone full-length gene or fragment into expression vector

Fig. 3 Gene synthesis by very long oligonucleotide synthesis followed by polymerase chain reaction (PCR). This method involves the total synthesis of both upper and lower DNA strands as two oligonucleotides. The yield of the full-length product is exceedingly low. However, the crude oligonucleotides are combined and subjected to PCR to amplify the full-length product. This method can be used for small- to medium-sized genes, or for fragments, which can then be assembled as in Fig. 1.

restriction map of the degenerate gene sequence is available, it is straightforward to choose a maximum number of evenly spaced sites.

In order to reduce the number of restriction sites identified in the initial map, it is preferable to limit the file of restriction enzymes to those of use for cassette mutagenesis as discussed above. These restriction endonucleases would include those with unambiguous 6- or 8-base recognition sequences (palindromic or interrupted palindromic) that generate cohesive ends of two or more nucleotides. Methylase-sensitive enzymes should be avoided, but in some cases nucleotides outside of the endonuclease recognition sequence can be altered to remove the methylase recognition sequence. Sites for enzymes generating blunt ends can be used if long gaps are present after all enzymes generating staggered ends are considered. However, blunt-end cutters should not be juxtaposed in the restriction map. Convenient cloning sites should be chosen for each end of the gene. For example, a number of
genes have been synthesized with an EcoRI site at the 5' end and a NotI site at the 3' end (57, 59, 60).

Some care must be taken so that enzymes that generate identical cohesive overhangs are not juxtaposed. For example, the juxtaposition of BclI and BglII would not be advisable. The restriction sites chosen define the corresponding DNA sequence and remove nucleotide ambiguity. Once the position for a particular enzyme is chosen, other potential sites for the same enzyme must be removed. Finally, the user-defined DNA sequence can be translated and the resulting amino acid sequence compared to the original to assure that the correct amino acid sequence results. This is important because the amino acid sequence translated from a degenerate codon sequence will not generally match the original amino acid sequence. This is because certain degenerate codons code for more than one amino acid. For example, serine has six codons, four of the form TCN and two of the form AGY (see Table II). The two forms reduce to the single degenerate codon WSN. If this degenerate codon is translated, it will be assigned an unknown amino acid X, since WSN can expand to any of the following: TCN (Ser), ACN (Thr), AGY (Ser), AGR (Arg), TGY (Cys), TGA (Ter), or TGG (Trp).

Creation of Appropriate Nucleotide Sequence

After defining the nucleotide sequence that corresponds to the desired restriction map, a majority of the gene sequence is still undefined. In general, the natural sequence can be retained. However, other considerations such as codon usage, potential secondary structure, and guanine-cytosine content can be considered as discussed below.

Codon Usage Considerations

In cases where a particular expression system is established, codon usage bias can be considered to attempt to optimize the level of expression (73). In some cases, codon usage can be used as well to slow translation of signal sequences so that cellular membrane translocation systems are not saturated. This was the case for the expression in E. coli of the gene for the light-driven proton pump bacteriorhodopsin from Halobacterium halobium (64). Some investigators have also placed a mammalian translation initiation consensus sequence immediately preceding the initiation methionine codon (60, 61, 74).

A dramatic example of using gene synthesis to affect expression level by altering codon bias was demonstrated in the case of the synthetic gene for human interleukin-2 (IL-2) (75). In designing the synthetic gene, the percentage of preferred codons for E. coli was increased from 43 to 85%. While message levels and stability were similar for the cDNA and synthetic
genes expressed in *E. coli*, the synthetic gene produced up to 16 times more IL-2 than the cDNA.

**Secondary Structure Considerations**

Computer algorithms can also be employed to identify areas of significant secondary structure. In many cases, nucleotide sequences can be modified to remove potential hairpin loops that might affect DNA sequencing or mRNA stability.

**Guanine–Cytosine Content Considerations**

The synthetic gene design process allows for the reduction of G-C content if desired. Stretches of four or more guanines or cytosines can be avoided where possible to minimize potential difficulties in oligonucleotide synthesis or DNA sequencing.

**Choice of Sites for Enzymatic Ligation of Synthetic Duplexes (Khorana Method)**

Once a full-length synthetic gene has been designed, both strands must be assembled from individual oligonucleotides. The lengths of the oligonucleotides are designed to allow adequate purification. Generally, 70 to 90 bases
are the maximum length that can be purified by HPLC or acrylamide gel electrophoresis.

One strategy of joining synthetic duplexes is based on the classical principles developed by Khorana and co-workers (20, 21). In this strategy, complementary oligonucleotides are synthesized. However, to avoid low-efficiency ambiguous blunt-end to blunt-end ligation of synthetic duplexes, the complementary oligonucleotides are designed with short 5’ overhangs. The optimal length of overhangs is 4 or 5 bases. This length is optimal for DNA ligase activity and allows enough unique sequences so that a relatively large number of duplexes can be ligated batchwise. No cohesive overhang within a gene fragment should be self-complementary, and the overhangs should obviously be unique within a gene fragment. When evaluating for self-complementarity, guanine and thymine should be considered to form stable base pairs. Restriction site overhangs are not useful for enzymatic joining of synthetic duplexes because they are generally palindromic and self-ligation will occur.

Synthetic Gene Construction (Khorana Method)

Methods are presented below for the construction of synthetic genes by the Khorana method where both upper and lower strands are synthesized. Oligonucleotides of moderate length are synthesized and purified. Phosphorylation of the 5’ ends of oligonucleotides with noncomplementary cohesive overhangs is carried out. Sets of oligonucleotides are annealed and joined with DNA ligase. The full-length gene fragment is isolated and cloned. Detailed methods relevant to other synthetic gene construction approaches are presented elsewhere as discussed above. In addition, detailed experimental methods for various mutagenesis procedures, such as combinatorial cassette mutagenesis, are presented in original sources.

The methods presented in this chapter for the design and synthesis of synthetic genes by the Khorana method were recently employed in the synthesis of a 1472-bp gene for the rat glucagon receptor (57). Glucagon is a 29 amino acid pancreatic hormone. Glucagon binds to a seven transmembrane domain G protein-coupled receptor on cells of target tissues such as liver and fat resulting in the activation of glycogenolysis and gluconeogenesis. The molecular mechanism of hormone–receptor interaction and receptor activation is not known. In order to facilitate these studies by site-directed mutagenesis, we have designed and synthetized a gene for the rat glucagon receptor (76). The gene codes for the native 485 amino acid protein, but contains 47 unique restriction sites. The full-length gene was assembled from three fragments of 362, 546, and 564 bp. The fragments consisted of 8, 14, and 14 oligonucleotides, respectively, ranging in length from 74 to 91 bases (Fig. 4).
Experimental Procedures

Oligonucleotide Synthesis

Automated oligonucleotide synthesis can be easily carried out on oligonucleotide synthesizers with commercially available solvents and reagents. The most commonly used chemistry involves the phosphite triester approach using protected β-cyanoethyl phosphoramidite nucleosides. The fully protected 3′ terminal phosphoramidite of the oligonucleotide is coupled to a...
solid support such as control pore glass or polystyrene (77). After protic acid treatment to remove the 5' protecting group, the fully protected incoming phosphoramidite is activated by tetrazole so that a phosphite triester bond is formed at high efficiency. The small amount of unreacted 5' hydroxyl of the first nucleoside is capped by a quantitative reaction with acetic anhydride in the presence of 1-methylimidazole. Finally, the newly formed internucleotide linkage is converted from a phosphite triester to a more stable phosphotriester by oxidation with iodine where water is the oxygen donor. The 5' hydroxyl of the dinucleotide can now be deprotected with acid treatment to complete a cycle. The cycle is repeated until the full-length oligonucleotide is obtained. Thus, the oligonucleotide is elongated from 3' to 5'. Cleavage from the support and removal of phosphate and exocyclic amine protecting groups are achieved by treatment with concentrated ammonium hydroxide.

For synthesis of the glucagon receptor gene, automated oligonucleotide synthesis is performed on an Applied Biosystems Model 392 DNA synthesizer. Phosphoramidite chemistry is employed using 0.2 μmol or 40 pmol synthesis scales and standard cycle routines. Each synthetic oligonucleotide is automatically cleaved from the solid support after removal of the terminal 5' hydroxyl protecting group. Each oligonucleotide solution is transferred into a screw top vial. After the addition of 2 ml of fresh concentrated ammonium hydroxide, the vial is tightly capped and heated at 55°C for at least 8 hr. Each fully deprotected oligonucleotide is dried by vacuum centrifugation in a polystyrene tube and the pellets are dissolved in 50 μl of TE (10 mM Tris–HCl, 1 mM EDTA, pH 7.4). An ultraviolet spectrum is measured from 310 to 210 nm after the proper dilution in TE is made. The yield in total absorbance units at 260 nm is calculated.

**Oligonucleotide Purification and Characterization**

The synthetic oligonucleotides are mixed with denaturing polyacrylamide load buffer containing formamide, boiled for 3 min, quick-chilled on ice, and subjected to electrophoresis on a 40-cm-long, 3-mm-thick 6% polyacrylamide gel. The gel is run until the product band is estimated to be at least half-way down the gel. The gel is removed from the glass plates and covered with plastic wrap. The DNA is displayed by ultraviolet shadowing with a handheld lamp, and the full-length synthetic oligonucleotide band is outlined with a marker and cut out with a razor blade. The gel fragment is transferred into a polypropylene tube, crushed thoroughly with a glass rod, and eluted from the gel into 10 ml of buffer (1 M triethylammonium bicarbonate, pH 7.5) by gently mixing for 12 to 24 hr at room temperature.

Triethylammonium bicarbonate (TEAB) buffer is most conveniently made by transferring 1 mol of liquid triethylamine (TEA) by weight into a 2-liter
flask and adding 850 ml of distilled, deionized water. Carbon dioxide (CO₂) gas is then bubbled through the liquid (in a fume hood if possible) until the TEA goes into solution. The bubbling is continued until the pH of the buffer drops to about 7.5. The buffer can then be brought to a final volume of 1 liter. The TEAB should be stored tightly sealed at 4°C. The pH should be checked periodically and can be adjusted by additional CO₂ treatment.

Oligonucleotides are purified from gel contaminants and desalted using C18 reverse-phase cartridges (SEP-PAK, Waters) (5). Cartridges are affixed to 20-ml syringes and are primed by successive washes with acetonitrile (10 ml), 17% acetonitrile in 100 mM TEAB (10 ml), and 25 mM TEAB (10 ml). The TEAB oligonucleotide solutions are applied to the cartridges after spinning out gel fragments in a table top centrifuge. The cartridges are then washed with 25 mM TEAB (10 ml), and the oligonucleotides are eluted with 17% acetonitrile in 100 mM TEAB (4.5 ml). The eluted volume is dried by vacuum centrifugation and the pellet dissolved in TE. An ultraviolet spectrum is measured from 310 to 210 nm after the proper dilution in TE is made. The concentration of purified oligonucleotide is determined using the estimate that a solution with an absorbance at 260 nm of 1 has a concentration of 40 µg/ml.

5' Terminal Phosphorylation
The synthetic oligonucleotides involved in joining reactions are 5’ end-phosphorylated in order to serve as substrates for DNA ligase. Phosphorylation reactions can be carried out individually (61). However, it is convenient to carry out the phosphorylations of all of the oligonucleotides within a fragment batchwise. The 5’ oligonucleotide on the upper strand and the 5’ oligonucleotide on the lower strand of a particular fragment are not phosphorylated. A batchwise reaction mixture is prepared with 100 pmol of each oligonucleotide in 50 mM Tris-HCl (pH 8.0). The mixture is heated for 3 min at 90°C and quick-chilled on ice. Concentrated stock solutions are added to give final concentrations of 10 mM MgCl₂, 2 mM spermidine, 10 mM DTT, 1 mM ATP, and 4 units of T4 polynucleotide kinase (New England Biolabs). The mixture is incubated for 45 min at 37°C, then heated 3 min at 90°C, and quick-chilled on ice. The phosphorylation reaction is repeated a second time at 37°C for 45 min with the addition of DTT and 4 units of T4 polynucleotide kinase. The reaction is stopped by adding EDTA to a final concentration of 25 mM. Phenol chloroform extraction is carried out followed by application to a prewashed spun-filter microconcentrator (Microcon-30, Amicon) to desalt and remove AMP and excess ATP from the sample. The concentration step is repeated three times after the addition of TE and the sample is recovered in a total volume of 50 µl of TE.
Annealing and Ligation Reaction

Complementary upper and lower strand oligonucleotides for an entire gene fragment are annealed by adding 100 pmol each of the 5' upper strand and 5' lower strand oligonucleotides to the phosphorylated oligonucleotide mix in a final volume of 100 μl in 50 mM Tris--HCl (pH 8.0), 10 mM MgCl₂. The mixture is overlayed with 50 μl of mineral oil, heated to 90°C for 3 min, and slow-cooled to room temperature over 2 to 3 hr. An aliquot is saved for analysis on an agarose gel. The remaining solution is stored at −20°C or immediately prepared for the ligation reaction.

The annealed oligonucleotides are joined by incubation for 16 hr at 14°C in 66 mM Tris--HCl (pH 7.5), 5 mM MgCl₂, 1 mM DTT, 1 mM ATP, and 25 units of T4 DNA ligase (Boehringer-Mannheim Biochemical) in a final volume of 100 μl. Phenol chloroform extraction followed by microconcentration on a Microcon-30 unit removes protein and desalts the solution. An aliquot is saved for agarose gel analysis.

Recovery, Cloning, and Characterization of Ligation Product

The products of the ligation reaction are separated on a preparative 1.5% agarose gel and visualized by ethidium bromide staining. The full-length synthetic DNA fragment band is cut out of the gel and purified from the agarose with the Qiaex Gel Extraction Kit (Qiagen). The DNA fragment is cloned into an appropriate cloning vector or expression vector that has been linearized by restriction endonuclease digestion to produce the two overhangs corresponding to the fragment ends. Ligation is carried out on 100 fmol of linear vector and 300 fmol of synthetic DNA fragment in a reaction mixture as previously described using 5 units of T4 DNA ligase for 16 hr at 14°C. As a control, 100 fmol of linearized vector are also subjected to the same treatment. The ligation mixtures are ethanol-precipitated and resuspended in 10 μl of TE.

Transformation of E. coli is carried out on 40 μl of XL1-Blue Electro-Competent cells (Stratagene) by electroporation with 20% of the ligation reaction in a 1-mm cuvette using the Gene Pulser/E. coli Pulser system (Bio-Rad). After electroporation, 1 ml of chilled sterile SOC outgrowth medium (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose) is immediately added to the cuvette. The cell suspension is transferred to a 10-ml culture tube and incubated for 1 hr in a 37°C shaker to allow phenotypic expression. Cells are spread on LB/ampicillin plates, which are then incubated for 16 hr at 37°C. Ampicillin-resistant colonies are streak-purified on fresh LB/ampicillin plates, and single colonies are selected for inoculating liquid LB/ampicillin (75 μg/ml) cultures.
DNA minipreps of 3-ml overnight cultures are prepared following the Qiagen-tip 20 protocol (Qiagen). Restriction analysis and subsequent dideoxy sequencing on double-stranded plasmid DNA (Sequenase, United States Biochemical) are carried out. A clone with the correct DNA sequence is selected for assembly of the full-length synthetic gene into an expression vector.

**Cassette Mutagenesis**

Site-directed mutagenesis of the synthetic gene is accomplished by synthesizing a pair of complementary oligonucleotides to form a duplex containing the desired codon alteration and the appropriate cohesive terminal overhangs. After purification and annealing as previously described, the 5' end nonphosphorylated synthetic duplex is ligated into the plasmid/gene DNA fragment linearized with the appropriate restriction endonucleases. Molar ratios of synthetic duplex to linearized vector are 3:1. Procedures carried out after ligation are as presented above. Procedures for other types of cassette mutagenesis, such as combinatorial cassette mutagenesis, can be derived from published sources.

**Expression of Synthetic Genes**

As discussed above, two of the advantages of the use of synthetic genes are that they can be easily transferred among a variety of vectors and that codon usage can be optimized where relevant to achieve maximal levels of expression. Synthetic receptor genes will generally be expressed in mammalian cells in tissue culture where pharmacological and cellular physiological effects can be correlated with structural changes introduced by mutation. In the case of the synthetic gene for bovine rhodopsin, large quantities of the opsin apoprotein can be produced in monkey kidney cells by transfection where transcription is under the control of the human adenovirus major-late promoter (78). The apoprotein in the plasma membrane can be regenerated with the chromophore 11-cis-retinal to form rhodopsin. The recombinant rhodopsin can be solubilized with detergent treatment and purified using an affinity adsorption method (3, 78).

**Conclusions**

Synthetic genes have been expressed in a variety of heterologous expression systems. For example, synthetic genes have been expressed in E. coli (64), in monkey kidney cells in tissue culture (61), in insect Sf9 cells (79), and in yeast (80–82). Perhaps the most extensively studied synthetic genes are those of bacteriorhodopsin (64) and bovine rhodopsin (61). In visual pigment structure–function studies, synthetic receptor genes for the rhodopsin and
for the human blue, green, and red cone pigment genes have been expressed in mammalian cells and purified from cell extracts after reconstitution with 11-cis-retinal chromophore (59). Purified site-directed mutant pigments have been studied by a variety of biochemical (83–86) and biophysical techniques (87–89). These studies have led to a greater understanding of the mechanism of wavelength regulation by visual pigments (84, 90–93) and of the mechanism of rhodopsin–transducin interaction (4, 86). Synthetic genes for the α subunit and γ subunit of the rod cell G protein, transducin, have also been expressed in a variety of systems including E. coli (γ subunit) (94), COS cells (α subunit) (60), and insect Sf9 cells (α subunit) (79).

In conclusion, gene synthesis should be considered when extensive long-term structure–function studies are planned. The initial investment in time and expense is offset by the ease and flexibility of DNA manipulation using a synthetic gene. Improved economical automated DNA synthesis, PCR techniques, and the availability of a large number of quality restriction endonucleases have combined to make gene synthesis possible for nearly all molecular biology laboratories.

Acknowledgment

This chapter is dedicated with respect and admiration to H. Gobind Khorana.

References


III STUDIES OF FUNCTIONAL DOMAINS OF RECEPTOR AND CHANNELS