

site, and then covers a whole turn of the DNA helix on each end of the control region. The center of the internal control region (roughly positions 56–77) is covered on only one side of the DNA helix by TFI<sub>II</sub>A, while at the ends of the control region (roughly positions 44–55 and 78–88) TFI<sub>II</sub>A surrounds the helix.

### [34] Mutagenesis with Degenerate Oligonucleotides: An Efficient Method for Saturating a Defined DNA Region with Base Pair Substitutions

By DAVID E. HILL, ARNOLD R OLIPHANT, and KEVIN STRUHL

#### Introduction

Since the advent of recombinant DNA technology, structure–function relationships of genes and genetic elements have been studied primarily by mutating cloned DNA segments and assessing the phenotypic consequences upon introduction into living cells. Although genetic elements can be localized on cloned DNA segments with deletion and insertion mutations, a more detailed description requires single base pair changes. In principle, DNA sequence requirements of a particular genetic element can be determined by making all possible point mutations within the region of interest and then analyzing the phenotypic effects. Such an approach avoids the bias introduced by genetic selections and hence makes it possible to obtain mutations that confer wild-type phenotypes.

Numerous chemical and enzymatic methods for generating point mutations within defined regions of DNA have been described. However, these mutagenesis procedures usually have drawbacks that make them more useful at the earlier stages of an investigation when the region of interest is less defined. Typical problems are (1) severe restrictions in the kinds of mutations that are produced, (2) low frequency of mutagenesis, thus making it necessary to use genetic or physical selections to isolate mutations, and (3) technical difficulties in performing the procedures. As an alternative to these region-specific methods, synthetic oligonucleotides of defined sequence have been used extensively for site-directed mutagenesis of DNA. However, such procedures are extremely expensive as well as time consuming. For example, as there are 30 possible single base pair changes of 10-bp region, saturation would require 30 (or 60) oligonucleotide syntheses, each of which would have to be processed separately.

Here, we describe an efficient method to create numerous point mutations within a given region by using the products of a single oligonucleotide synthesis.

### Principles of the Method

The procedure utilizes synthetic oligonucleotides that are mutagenized by including low concentrations of the three non-wild-type nucleotide precursors at each step of the synthesis.<sup>1-5</sup> The product of such a DNA synthesis is a degenerate oligonucleotide, i.e., a complex mixture of related molecules, each of which has a defined probability of being altered from the wild-type sequence. The frequencies and types of single, double, and higher order mutations can be set simply by choosing appropriate amounts of non-wild-type precursors at each step of nucleotide addition.

For the cloning procedure shown in Fig. 1, degenerate oligonucleotides are converted to the double-stranded form by mutually primed synthesis.<sup>4,6</sup> In most cases, oligonucleotides are synthesized such that their heterogeneous regions are bounded at their 5' and 3' ends by sequences recognized by restriction endonucleases. Because the 3' ends are palindromic, two oligonucleotide molecules can hybridize such that they will serve as mutual primers for extension with DNA polymerase I. The product of this mutually primed synthesis is a double-stranded molecule containing two oligonucleotide units that are separated by the original 3' restriction site and are flanked by the original 5' restriction sites. The double-stranded molecules are cleaved with restriction endonucleases that recognize the 5' and 3' ends to generate homoduplex versions of the original oligonucleotides. After ligation to an appropriate vector, the resulting products are introduced into *Escherichia coli*. Thus, each transformant represents the cloning of a single oligonucleotide from the original collection. DNA preparations from these transformants are then subjected to nucleotide sequence analysis in order to determine the nature of the mutation(s). In principle, mutations should occur at the frequency that was programmed into the DNA synthesis, and they should be located randomly throughout the region of interest.

<sup>1</sup> M. D. Matteucci and H. L. Heyneker, *Nucleic Acids Res.* **11**, 3113 (1983).

<sup>2</sup> J. B. McNeil and M. Smith, *Mol. Cell. Biol.* **5**, 3545 (1982).

<sup>3</sup> C. A. Hutchison, S. K. Nordeen, K. Vogt, and M. H. Edgell, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 710 (1986).

<sup>4</sup> A. R. Oliphant, D. E. Hill, A. L. Nussbaum, and K. Struhl, *Gene* **44**, 177 (1986).

<sup>5</sup> K. M. Derbyshire, J. J. Salvo, and N. D. F. Grindley, *Gene* **46**, 145 (1986).

<sup>6</sup> A. R. Oliphant and K. Struhl, this volume [35].

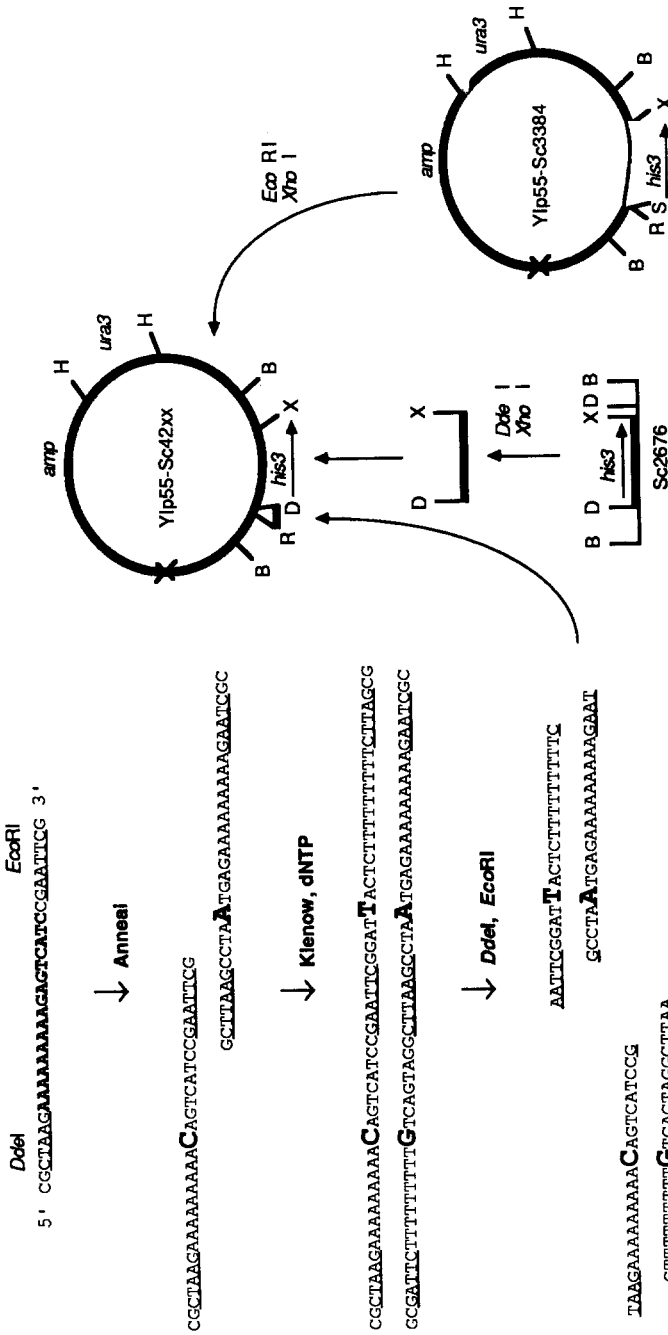


FIG. 1. Saturation mutagenesis using degenerate oligonucleotides. The top line of the left part of the figure indicates the *EcoRI*-*DdeI*-activated oligonucleotide containing 17 central bases (boldface letters) that were synthesized with 90% of the base indicated and 3.3% of the other three bases. Double-stranded, degenerate oligonucleotides suitable for cloning were produced as described in the text. The large, boldface residues indicate base pair substitution mutations of the central region. The *EcoRI*-*DdeI* oligonucleotide mixture was combined with the 9-kb *EcoRI*-*XhoI* fragment of Yip55-Sc3384 and the 0.9-kb *DdeI*-*XhoI* fragment of Sc2676 (the relevant parts of these molecules are indicated in boldface) to produce the desired molecules (indicated as Yip55-Sc42xx). The locations of the *amp*, *ura3*, and *his3* genes as well as restriction endonuclease sites for *Bam*HI (B), *Eco*RI (R), *Sac*I (S), *Dde*I (D), *Xho*I (X), and *Hind*III (H) are indicated. The positions of *Dde*I sites located outside of the 1.8-kb *his3 Bam*HI fragment are not shown. The large X in the circle represents the mutated *Eco*RI site at the junction between vector and *his3* chromosomal sequences. The *Eco*RI-*Sac*I oligonucleotides, as described in the section concerning saturation mutagenesis, were cloned directly between the *Eco*RI and *Sac*I sites of Yip55-Sc3384.

## Methods

### *Oligonucleotide Synthesis*

Oligomers were synthesized on an Applied Biosystems DNA synthesizer (Model 380A) using the phosphite triester method.<sup>7,8</sup> After detachment and removal of all but the 5'-dimethoxytriphenylmethyl protecting groups, the oligomers were separated from shorter congeners by HPLC chromatography on a Waters C-8 column using a 40-min linear gradient of 0.1 M triethylammonium bicarbonate (pH 7.1) and from 0 to 25% acetonitrile. The peak containing the trityl chromogen (emerging near the top of the gradient) was desalted by flash evaporation *in vacuo* at temperatures below 30° and completely deprotected by treatment with 80% aqueous acetic acid at room temperature for 20 min, followed by flash evaporation.

The general procedure for synthesizing a degenerate oligonucleotide proceeds as follows. At positions where mutations are not desired, such as those composing the recognition sequences for restriction endonuclease cleavage, nucleotide addition is performed by standard procedures using a single nucleoside phosphoramidite precursor. In contrast, defined mixtures of nucleoside phosphoramidites are used at positions where mutations are desired. In general, the frequency of addition of particular nucleotides is determined simply by the relative molarities of the precursors, although sometimes G residues are added relatively poorly. This problem may be due to instability of the G precursor,<sup>9</sup> and it can probably be avoided by using freshly prepared solutions. Although certain DNA synthesizers can combine solutions of pure precursors, the most accurate and reproducible way to achieve a desired mixture is to combine appropriate amounts of solid nucleoside phosphoramidites prior to solubilization in anhydrous acetonitrile.

The design of the oligonucleotide depends on the nature of the experiment. In many cases, such as diagrammed in Fig. 1, the object is to alter a region of DNA at a defined mutation rate. This is accomplished by using four mixtures, each composed of one major nucleotide precursor (corresponding to the wild-type sequence) and equal amounts of the three remaining precursors (representing each of the possible base pair changes). At each position to be mutagenized, the mixture containing the wild-type precursor as the major component is used instead of the solution containing the pure precursor. The use of premade mixtures as described above

<sup>7</sup> M. D. Matteucci and M. H. Caruthers, *J. Am. Chem. Soc.* **103**, 3185 (1981).

<sup>8</sup> S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.* **22**, 1859 (1981).

<sup>9</sup> G. Zon, K. A. Gallo, C. J. Samson, K.-L. Shao, M. F. Summers, and R. A. Byrd, *Nucleic Acids Res.* **13**, 8181 (1985).

ensures that the selected mutation rate is maintained at each step of the DNA synthesis. In other experiments where the purpose is to mutate a particular base to all possible alternatives, an equimolar mixture of the three non-wild-type precursors is used at the relevant addition step. More specialized cases can be accommodated simply by choosing appropriate mixtures for the appropriate positions of the oligonucleotide. The expected results from any particular experiment can be calculated according to the laws of probability.

*Conversion of Oligonucleotides to Double-Stranded DNA by Mutually Primed Synthesis<sup>4,6</sup>*

A degenerate oligonucleotide (approximately 1  $\mu\text{g}$ ) is diluted into 10  $\mu\text{l}$  of 3 $\times$  buffer [30 mM Tris (pH 7.5), 150 mM NaCl, 30 mM MgCl<sub>2</sub>, 15 mM dithiothreitol, and 0.1 mg/ml gelatin], hybridized for at least 1 hr at 37°, and then allowed to cool slowly to room temperature. Deoxynucleoside triphosphates (to a final concentration of 250  $\mu\text{M}$  for each of the four) and [<sup>32</sup>P]dATP (10  $\mu\text{Ci}$ ) are then added, and the reaction mixture is diluted to a final volume of 30  $\mu\text{l}$ . Then 5 units of the Klenow fragment of *E. coli* DNA polymerase I are added, and the reaction mixture is incubated at 37° for at least 1 hr. The products of the mutually primed synthesis reaction are extracted with phenol and precipitated with ethanol. A small portion of the resuspended reaction products is then analyzed by electrophoresis in acrylamide gels containing 7 M urea.<sup>10</sup> The desired product, which is visualized by autoradiography, is a homoduplex molecule of length  $2A + 2N + B$  (where  $A$  is the length of the 5'-flanking sequences,  $N$  is the length of the heterogeneous central region, and  $B$  is the length of the 3'-flanking sequences). For calibrating the size of the product, the best markers are oligonucleotides of defined length that have been labeled at their 5' ends with T4 polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP.

In order to clone the degenerate oligonucleotides, the double-stranded molecules produced by mutually primed synthesis are cleaved with appropriate restriction endonucleases and then ligated into vector molecules by standard means. The cleavage reactions are monitored by electrophoresis in denaturing acrylamide gels as described above. At some stage prior to the ligation reaction, it is useful to remove the excess, unreacted, single-stranded oligonucleotides as these reduce the ligation efficiency. In most cases, the initial product is cleaved to completion with the restriction endonuclease recognizing the outside sites (originally the 5' site), extracted with phenol, and concentrated by ethanol precipitation. After

<sup>10</sup> F. Sanger, A. R. Coulson, B. G. Barrell, A. J. Smith, and B. A. Roe, *J. Mol. Biol.* **143**, 161 (1980).

electrophoresis in a native acrylamide gel (6–12% depending on the size of the product), the desired double-stranded molecule containing two oligonucleotide units is eluted in 0.5 M ammonium acetate and 1 mM EDTA for 4–24 hours at 37° and then concentrated by ethanol precipitation. The purified DNA is cleaved with the restriction endonuclease recognizing the central site (the original 3' site) to produce the final product, a double-stranded version of the oligonucleotide mixture with 5' and 3' ends suitable for ligation into standard vector molecules. In our experience, it is better to perform the gel purification step prior to cleavage at the central restriction site; the reasons for this are unknown.

Standard ligation reactions using T4 DNA ligase are carried out at 15° in 20- $\mu$ l reactions containing 50 mM Tris (pH 7.5), 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, and 500  $\mu$ M ATP. As the yield of the final oligonucleotide product is somewhat variable, the amount to be added to a given amount of vector is determined empirically in order to optimize the ligation reaction. Typically, ligation reactions containing varying amounts of oligonucleotide are processed in parallel.

## Results

### *Mutagenesis of the his3 Regulatory Region Using a Single Degenerate Oligonucleotide*

Extensive deletion analysis of the yeast *his3* promoter defines and localizes a positive regulatory site that is critical for the induction of *his3* transcription in response to conditions of amino acid starvation.<sup>11,12</sup> The regulatory region maps between nucleotides –84 and –102 with respect to the transcriptional initiation site and it contains the sequence TGACTC, which is present in front of and implicated in the expression of coregulated genes.

In order to generate a large number of base pair substitution mutations throughout a 17-bp region containing the essential regulatory sequence, we synthesized an oligonucleotide in which the *his3* region was bounded, respectively, at its 5' and 3' ends by the restriction endonuclease cleavage sites for *DdeI* and *EcoRI*. A 10% mutation rate was achieved by using four mixtures, each composed of a wild-type precursor (90%) and equal amounts of the three mutant precursors (10% total). The DNA synthesizer was programmed to use the appropriate 90/10 mix at each position of the 17-base region and the appropriate pure precursors at the remaining

<sup>11</sup> K. Struhl, *Nature (London)* **300**, 284 (1982).

<sup>12</sup> K. Struhl, W. Chen, D. E. Hill, I. A. Hope, and M. A. Oettinger, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 489 (1985).

positions. The number of mutations per oligonucleotide should be defined by a binomial distribution centered around 1.7, the average value; thus, single and double mutations should predominate. The oligonucleotide mixture was converted into double-stranded DNA by mutually primed synthesis, and then cleaved with *EcoRI* and *DdeI* (Fig. 2).

Because of the availability of reliable methods for determining the DNA sequence of double-stranded molecules,<sup>13</sup> we cloned the oligonucleotide mixture such that the resulting molecules could be tested directly for the phenotypes that they conferred *in vivo* (Fig. 1). A ligation reaction containing the *EcoRI*-*DdeI* oligonucleotide mixture, the 9-kb *EcoRI*-*XhoI* vector fragment of YIp55-Sc3384,<sup>12</sup> and the 0.9-kb *DdeI*-*XhoI* fragment of Sc2676<sup>14</sup> was performed, and the ligation products were introduced into *E. coli* by selecting for transformants resistant to ampicillin. Because ligations involving three DNA fragments occur at relatively low frequency, we used colony filter hybridization<sup>15</sup> to identify transformants containing oligonucleotide inserts. Using 5' end <sup>32</sup>P-labeled oligonucleotide as a probe, 41 out of 97 transformants appeared to contain inserted oligonucleotides. The large number of non-oligonucleotide-containing transformants was probably due to incompletely cleaved vector DNA or to incorrect ligation. DNAs were prepared from the 41 putative, oligonucleotide-containing transformants, and the DNA sequence of the *his3* regulatory region was determined by the chain termination method using a *his3*-specific primer corresponding to positions +26 to +42 of the antisense strand.<sup>14</sup>

All 41 plasmids contained oligonucleotide inserts, as expected from the colony filter hybridization. From the entire collection, 23 unique mutants were obtained, and, as expected, single and double mutants comprise the largest classes (Fig. 3). Six of the mutant sequences were duplicates, a frequency much higher than expected. These probably arose during the transformation procedure as a result of incubating the *E. coli* cells for 90 min prior to plating on ampicillin-containing medium. Finally, 12 wild-type sequences were obtained, most of which probably represent independent cloning events. The average number of mutations per oligonucleotide was 1.2, a value in fair agreement with the theoretical prediction of 1.7. The pattern of mutations in terms of the positions of alterations was in good accord with expectations. Thus one degenerate oligonucleotide can be used for creating a large number and variety of base pair substitution mutations within a small, defined region.

<sup>13</sup> E. Y. Chen and P. H. Seeburg, *DNA* 4, 165 (1985).

<sup>14</sup> K. Struhl, *Nucleic Acids Res.* 13, 8587 (1985).

<sup>15</sup> M. Grunstein and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* 72, 3961 (1975).

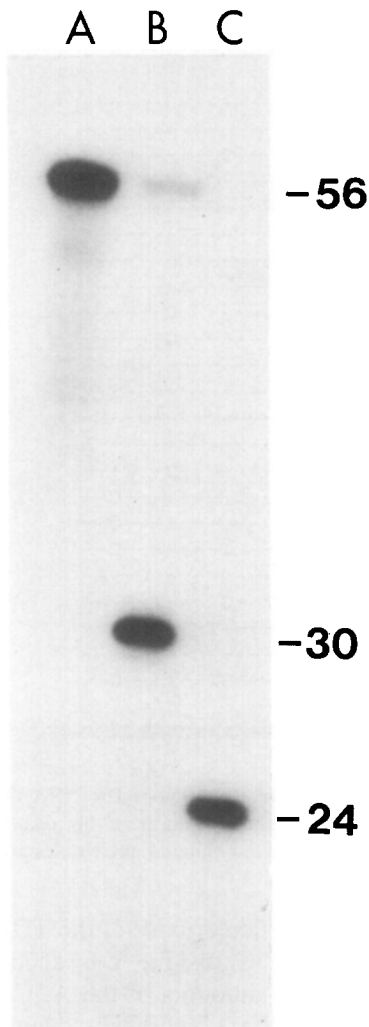


FIG. 2. Conversion of the degenerate oligonucleotide into clonable DNA. The DNAs were electrophoretically separated in a 10% acrylamide gel containing 7 M urea.<sup>10</sup> Lane A corresponds to the initial product of mutually primed synthesis (see Fig. 1). This product was subsequently cleaved with *Eco*RI (lane B) and *Eco*RI + *Dde*I (lane C). The lengths (in nucleotides) of these products are shown at the right side of the autoradiogram.

#### *Saturation Mutagenesis of the TGACTC Core*

As essentially all yeast genes subject to general control contain the sequence TGACTC, we decided to saturate this 6-base sequence with single base changes. Six oligonucleotides were synthesized, each of



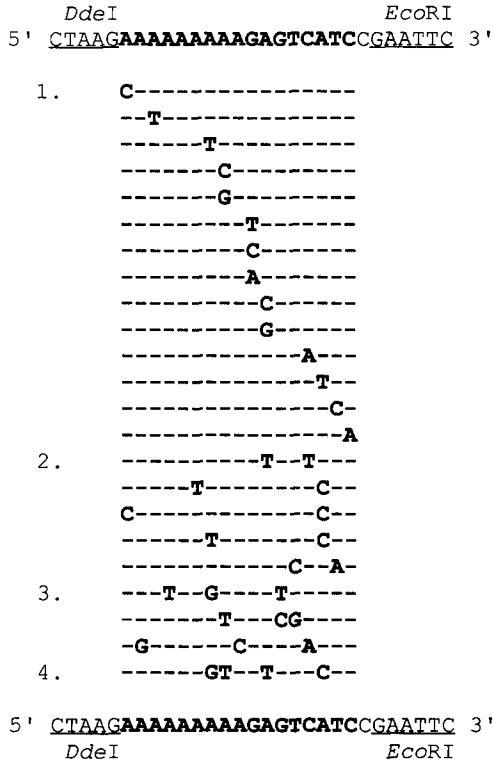


FIG. 3. Mutations of the *his3* regulatory region. The DNA sequences of 23 derivatives containing from one to four substitutions within the 17-bp region that was synthesized at a 10% mutation level (see Fig. 1); only the mutated nucleotides are indicated.

which was mutated at one particular base of the TGACTC sequence. For example, mutagenesis of the G residue was accomplished by using a mixture containing equimolar amounts of the A, C, and T precursors at the appropriate step of the DNA synthesis. Thus, the oligonucleotide product is actually a mixture of three mutant sequences that could be resolved into the individual components by molecular cloning. The degenerate oligonucleotides were cloned as *EcoRI*-*SacI* fragments between the *EcoRI* and *ScaI* sites of YIp55-Sc3384 (Fig. 1). Unlike the first example, these ligation reactions were more efficient because they contained only two DNA segments with heterologous ends. For this reason, it was unnecessary to screen the transformants by colony filter hybridization prior to DNA sequence analysis. From 37 transformants containing an oligonucleotide insertion (out of 52 that were subject to DNA sequence analysis), 15 out of the 18 possible base pair substitution mutations were obtained.

We did not obtain three mutants containing G residues in place of the wild-type residue, possibly for reasons discussed in the methods section.<sup>9</sup>

Thus, it is possible to saturate a short region with all possible single base pair substitutions by using a set of degenerate oligonucleotides that mutagenize one residue at a time. In comparison to more conventional oligonucleotide-directed mutagenesis procedures, this method produces three mutants for the price of one DNA synthesis. Obviously this method becomes more cumbersome as the region of interest becomes larger; in such situations, the method described in the previous section is more practical.

### Comments

Several technical points of the procedures are worth noting. First, the 5' end does not have to be cleavable by a restriction endonuclease because mutually primed synthesis produces blunt ends that are suitable for cloning. Second, in situations involving enzymatic cleavage at the 5' end, it is advantageous to minimize the length of the palindrome in order to disfavor hybridization of the 5' ends that might block complete extension. In addition, it is useful to include from one to three extra nucleotides beyond the endonuclease recognition sequences at the 5' end in order to facilitate cleavage by the enzyme. Third, although the palindrome at the 3' end, which is required for mutually primed synthesis, can be as short as 6 bases, the reaction works more efficiently when the region is 8 bases in length. However, in situations using an 8-base palindrome and a restriction endonuclease with a 6-base recognition sequence, each oligonucleotide unit will contain an extra base between the restriction site and the degenerate central region. Fourth, as a palindromic restriction site at the 3' end of the oligonucleotide is the only requirement for mutually primed synthesis, many different sequences are available. Moreover, as there are almost no limitations on the 5'-end sequences, degenerate oligonucleotides can be cloned into an extremely wide variety of double-stranded DNA molecules. This makes it possible to insert the degenerate oligonucleotides into vectors that can be used directly to examine the phenotypes of the mutant sequences. Fifth, the size of region that one can mutagenize is limited by the length of the oligonucleotide that is synthesized. In other experiments, we have synthesized a degenerate oligonucleotide whose heterogeneous region is 55 bases long.

The mutagenesis procedures described here have several advantages for determining the relationship of structure and function of genetic elements. The cloning efficiency is high because the oligonucleotides are converted to double-stranded restriction fragments, thus making it possi-

ble to obtain a large number of mutations. Moreover, in many cases the vast majority of transformants contain an inserted oligonucleotide, thus eliminating the need for a hybridization screen prior to DNA sequence analysis. Unlike other methods that produce mismatches between mutant oligonucleotides and the wild-type sequence, the oligonucleotides described here are cloned as homoduplex molecules. This avoids biases due to differential stability and preferential repair of heteroduplexes, and to screening procedures that depend on mismatch hybridization to distinguish mutants from nonmutants. Most importantly, essentially all possible mutations can be obtained without regard to their phenotypes *in vivo*. Thus, it is possible to determine directly which nucleotides are critical for a particular genetic function and which ones are unimportant.

### [35] The Use of Random-Sequence Oligonucleotides for Determining Consensus Sequences

By ARNOLD R OLIPHANT and KEVIN STRUHL

#### Introduction

In studying the DNA sequences of various genes and organisms it has become evident that similarity in function is associated with similarity in structure. However, as genetic elements conferring similar functions do not generally have identical DNA sequences, their nucleotide requirements are described as a consensus of related sequences. A common and useful means of describing such a consensus is to construct a matrix listing the number of occurrences of all four nucleotides at each position in the consensus. The reasons for determining a consensus sequence are to increase knowledge about the function of interest and to accurately predict biological meaning and functional activity from newly acquired sequence data.

Consensus sequences are often proposed on the basis of comparing a large number of natural DNA sequences that are believed to encode a particular genetic function. Regions of DNA proposed to contain a genetic element are examined for similarities that would not be expected to occur on a random basis. However, it is difficult to show statistical significance unless the sample size is very large, the elements are localized to small regions of DNA, or the proposed sequence occurs very infrequently on a random basis. More importantly, even when statistically significant