

Chemical DNA Synthesis and Recombinant DNA Studies

Keiichi Itakura and Arthur D. Riggs

Twenty years ago, the design and construction of totally synthetic genes was just a dream, something that most biochemists and geneticists did not expect to happen in their lifetimes. Fortunately, nucleic acid chemistry was considered interesting in its own right, and a few organic chemists, foremost among them H. G. Khorana (1), made steady progress on oligonucleotide synthesis throughout the 1960's and early 1970's.

The work of assembling nucleotides was slow and tedious, but when the natural nucleotide sequence of a small transfer RNA (tRNA) became known in 1965 (2), it was possible to begin the step-by-step assembly of nucleotides to form this

Thus cloning is of great importance for synthetic DNA chemistry, and, conversely, synthetic DNA has proved to be generally useful for recombinant DNA work. Table 1 outlines most of the proven applications of chemical DNA synthesis to recombinant DNA work. In the following sections, we review most of these applications.

Methods for Chemical DNA Synthesis

Figure 1 illustrates the three methods for oligodeoxyribonucleotide synthesis that have been useful for recombinant DNA studies.

Summary. Chemically synthesized DNA has been used in many recombinant DNA studies. These uses have included the total synthesis and cloning of functional genes, the cloning and expression of natural genes, and editing and changing genes by directed mutation.

gene [77 base pairs (bp)]. A large team of scientists worked on this for several years, and by 1970 they successfully produced a small amount of the alanine tRNA gene (3). However, critics were quick to point out that with the methods available then, sequence errors would have been missed unless they were present in the majority of molecules. The aggregate probability of having a perfect molecule was very low. Of what value was it for a large team of scientists to work for years to make a few micrograms of a gene that might not be functional? Fortunately, the critics did not prevail, probably because of the basic elegance of this intellectual exercise.

In the mid-1970's, chemical DNA synthesis suddenly became of great practical value because of the unexpected development of recombinant DNA techniques. Molecular cloning requires only microgram quantities of DNA, and once a desired clone is obtained, a continuous supply of the DNA fragment is ensured.

Diester method. The diester method was the first to be developed to a usable state, primarily by Khorana and co-workers (1). The basic step is the joining of two suitably protected deoxynucleotides to form a dideoxynucleotide containing a phosphodiester bond (Fig. 1). The diester method is well established and has been used to synthesize a number of DNA molecules, including two genes for tRNA's (1).

Triester method. Although perfected more recently (4), this method has advantages in speed and yield. The main difference between the diester and triester methods is the presence in the latter of an extra protecting group on the phosphate atoms of the reactants and products. The phosphate protecting group is usually a chlorophenyl group, which renders the nucleotides and polynucleotide intermediates soluble in organic solvents. Therefore purifications are done in chloroform solutions. Additional features of the method are given in

the legend to Fig. 2. Some recent improvements in the method are (i) the block coupling of trimers and larger oligomers, (ii) the extensive use of high-performance liquid chromatography for the purification of both intermediate and final products, and (iii) solid-phase synthesis.

Polynucleotide phosphorylase method. Smith and his colleagues (5, 6) developed an enzymatic method of DNA synthesis and successfully synthesized many useful oligodeoxynucleotides. Under controlled conditions, polynucleotide phosphorylase adds predominantly a single nucleotide to a short oligodeoxynucleotide. Chromatographic purification allows the desired single adduct to be obtained. At least a trimer is required to start the procedure, and this primer must be obtained by some other method. The polynucleotide phosphorylase method works well and has the advantage that the procedures involved are familiar to most biochemists.

Solid-phase methods. Drawing on the technology developed for the solid-phase synthesis of polypeptides, it has been possible to attach the initial nucleotide to solid support material and proceed with the stepwise addition of nucleotides (7). All mixing and washing steps are simplified, and the procedure becomes amenable to automation. Several research groups are developing automatic DNA synthesizers.

Initial Synthetic Gene-Recombinant DNA Studies

In 1976, two small synthetic genes were cloned: a regulatory gene (the lac operator) (8) and a tyrosine suppressor gene (9). These studies established that chemical DNA synthesis methods were indeed capable of making biologically functional DNA. Both genes were copies of natural gene sequences. The lac operator functioned by binding the lac repressor protein in vivo; the synthetic tyrosine tRNA gene "suppressed" a phage mutation.

The lac operator cloning experiments also demonstrated that restriction enzyme sites can be added to the ends of DNA molecules by blunt-end ligation. Figure 3, which is adapted from Scheller *et al.* (10), illustrates this method, which has proven to be of great value for recombinant DNA studies. Several linkers are now available commercially. Early applications included cloning sea urchin

The authors are staff members at the Division of Biology, City of Hope Research Institute, Duarte, California 91010.

Table 1. Applications of chemical DNA synthesis.

| Key words | Explanation | References |
|-----------------|--|---------------------|
| Linkers | Restriction sites that can be ligated to DNA fragments to facilitate cloning | (8, 10-13, 23) |
| Primers | Specific priming of DNA synthesis; DNA and RNA sequencing; directed mutation; gene editing | (24-30) |
| Gene synthesis | Transfer RNA genes; regulatory genes; hormone genes | (1, 8, 9, 14, 17) |
| Gene isolation | Linkers; specific primers; hybridization probes | (8, 10-13, 23-30) |
| Gene alteration | Shortening, lengthening, or changing natural genes; hormone analogs; directed mutation; editing for efficient expression | (21, 31, 32, 34-38) |

genomic fragments (11) and constructing lambda libraries containing mouse, rabbit, or human DNA (12). The cloning of complementary DNA (cDNA) is discussed later.

Recently, nonsymmetrical linkers were made (13). With these, sticky ends are present immediately after ligation; no treatment with restriction enzyme is required, an advantage for the cloning of large DNA fragments that may contain the same restriction sites as the linker.

Hormones from Synthetic Genes

Somatostatin. The natural nucleotide sequences were copied for both the lac operator and the tRNA genes. Thus synthetic copies of natural genes were made, cloned, and shown to function in

vivo. However, as expected from the nature of the genes, no peptide products were made. The first peptide product from a synthetic gene was somatostatin, a mammalian hormone (14). Somatostatin was the first potentially useful peptide product of recombinant DNA work. As outlined in Fig. 4, DNA fragments for a 57-bp somatostatin gene were synthesized by the triester method, assembled and joined by ligase to form double-stranded DNA, and inserted near the end of the gene for β -galactosidase. In vivo, somatostatin was made as a short tail on the end of the β -galactosidase polypeptide chain. Somatostatin was cleaved from the fused protein product by treatment in vitro with cyanogen bromide, which very specifically and efficiently cleaves polypeptide chains at methionine. Because the gene was made syn-

thetically, it was easy to arrange for a methionine to be at the junction of β -galactosidase and somatostatin.

Since the nucleotide sequence of the natural gene for somatostatin was still not known, investigators designed the gene for somatostatin by using the genetic code and knowing only the desired sequence of amino acids. Thus the somatostatin gene was a totally man-made gene; yet it functioned well, with more than 10^4 molecules of somatostatin being made per bacterial cell. In this case, the fused protein approach was essential. When the somatostatin gene was moved just downstream of the lac operator so that the product would be made without a precursor peptide, no detectable somatostatin was made (14)—presumably since *Escherichia coli* rapidly degrades most small peptides (15).

Human insulin. After the somatostatin project established the feasibility of the synthetic gene approach for bacterial production of hormones, almost identical methods were used for insulin (16, 17). As reported by Crea *et al.* (16), genes for the A and B chains of human insulin were designed using the genetic code and then synthesized by the triester method. To make these insulin genes required the synthesis of 29 oligonucleotides averaging 14 nucleotides in length. These fragments, enough for a total of 181 bp of duplex DNA, were made by four scientists in about 3 months. Today, with new improvements and solid-phase methods, this same synthesis would probably require only 1 month.

The insulin deoxyoligonucleotides were assembled and joined covalently by ligation to form the two insulin genes. Each gene was cloned and attached to the gene for β -galactosidase (as was done for somatostatin). Thus two bacterial strains were constructed: one producing a fused protein carrying the human insulin A chain, and the other a similar fused protein carrying the B chain. The insulin chains were clipped from the β -galactosidase precursor by treatment with cyanogen bromide and then purified. The two insulin chains were joined by the Katsoyannis procedure (18) to yield complete insulin (17). Good yields were obtained (more than 10^5 molecules per cell) (17). The human insulin produced after purifying and joining the chains is fully active by a number of tests, including radioimmunoassay, receptor binding assays, fat cell assays, and the rabbit hypoglycemia assay (17, 19).

Several peptide products have been made in *E. coli* from cloned foreign

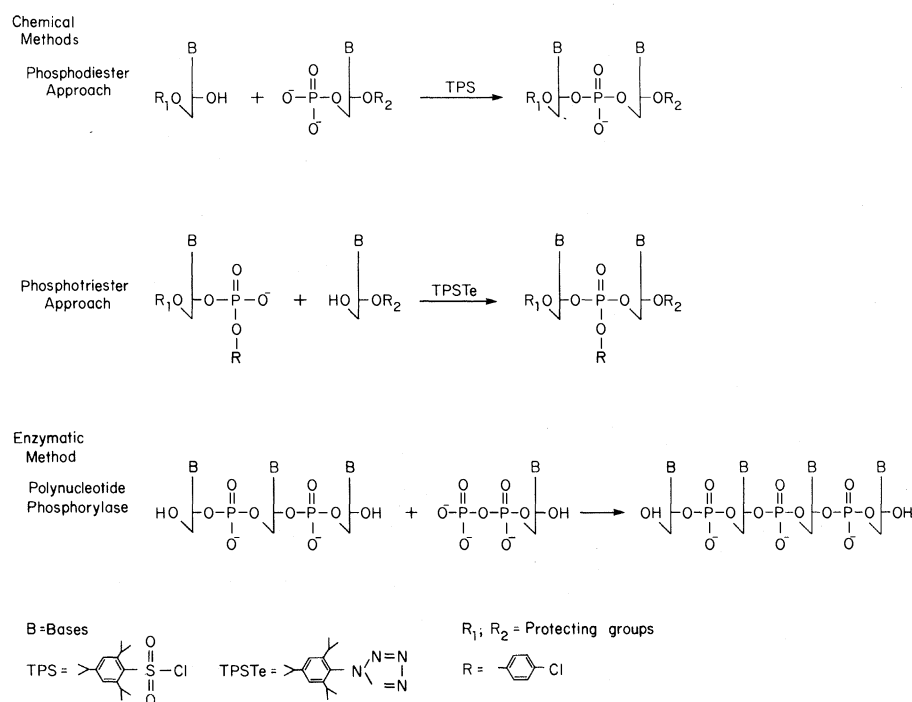


Fig. 1. Three methods for the synthesis of DNA. All of these methods have been used to make DNA for recombinant DNA experiments.

genes. Those whose production involved genes made either totally or partly from chemically synthesized DNA include, to date, somatostatin (14), insulin (17), thymosin (20), and human growth hormone (hGH) (21). The growth hormone project is especially interesting because part of the gene coding for the first 24 amino acids was made chemically, whereas the rest of the gene came from cDNA.

Gene Editing

Although a gene of any size can in principle be made by chemical DNA synthesis, for large genes it is probably more practical to isolate the natural gene and then, when necessary, merely "edit" the gene with chemically synthesized DNA. The efficacy of this approach was demonstrated by Goeddel *et al.* (21) with hGH. As obtained from messenger RNA (mRNA), the hormone cDNA was either too short or too long (pregrowth hormone) for correct expression of biologically active hGH. However, as illustrated schematically in Fig. 5, a Hae III restriction site is located 70 nucleotides from the beginning of the hGH sequence. A DNA fragment was then synthesized chemically which completed the sequence and allowed for easy insertion of the complete gene at the correct position behind the lac promoter for efficient expression (21). Active hGH was made in extremely good yields—more than 10^5 molecules per cell. In this case, the hormone is reasonably stable in *E. coli*; perhaps because of its size (191 amino acids), no precursor peptide is needed.

Very recently, Goeddel and co-workers (22) used a similar approach to produce biologically active interferon in *E. coli*. This approach to gene editing and gene completion seems generally applicable, especially since, if convenient restriction sites are not available in the natural sequence, they could be made by directed mutagenesis.

Gene Isolation

Because of the obvious importance of natural gene isolation, it is fortunate that chemical DNA synthesis can aid this effort. The most common approach for gene isolation from natural sources has been to make cDNA from mRNA and then clone double-stranded cDNA. Linkers, as described above, have frequently been used to aid the cloning process. Ullrich *et al.* (23) were the first to

demonstrate the feasibility of this approach by adding Hind III linkers to the ends of rat insulin cDNA. The method works well, but has the disadvantage that the addition of the linkers often is a low-efficiency step; thus far it has been

practical only when the desired message is abundant. A great many genes of interest, probably the majority, are not expressed at high levels. With proven methodology it is very difficult to clone genes expressed to less than 1 percent of

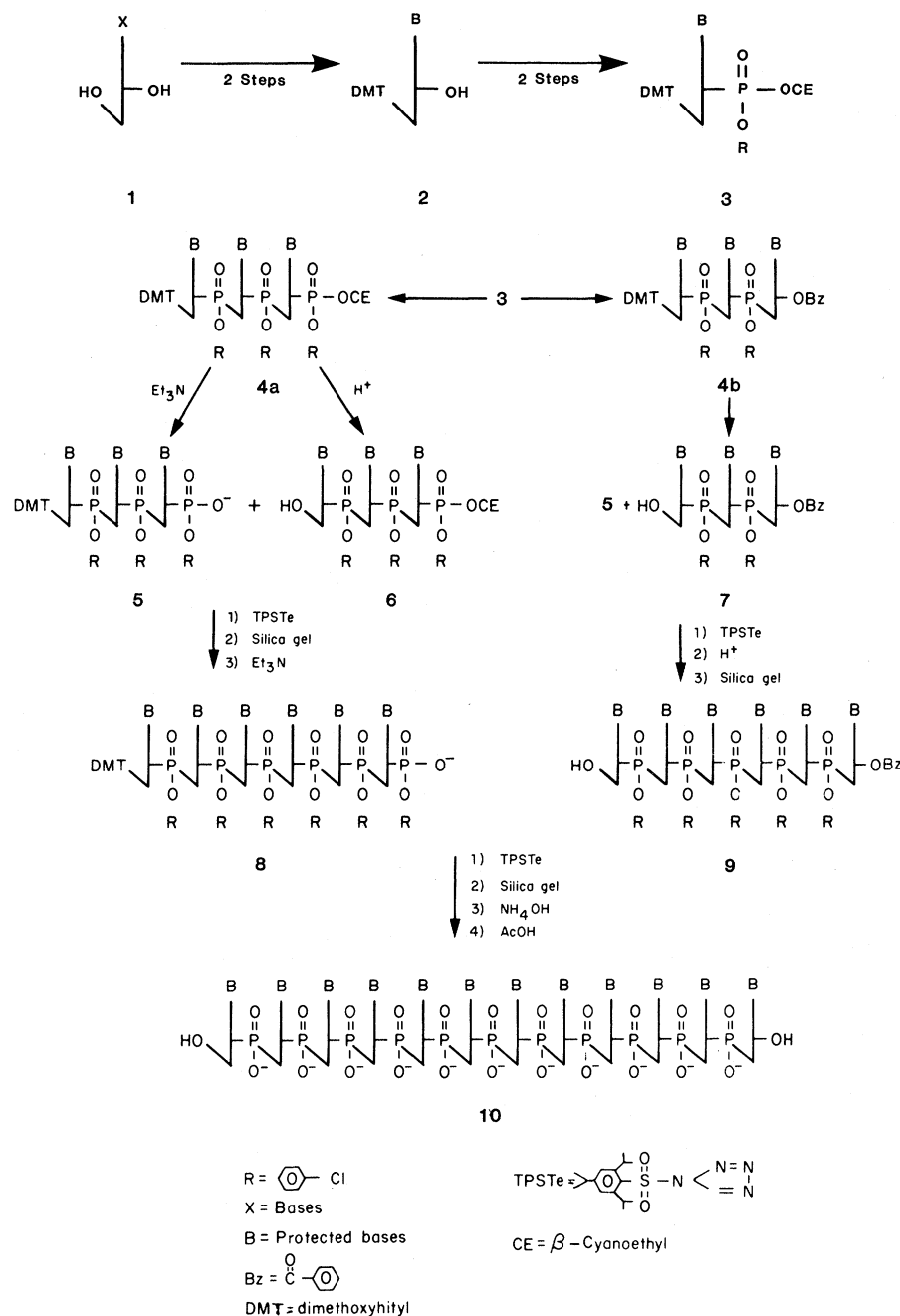
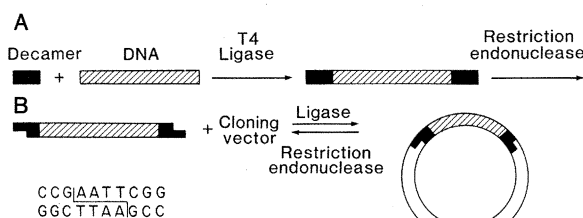


Fig. 2. Chemical synthesis of oligodeoxyribonucleotides by the improved triester method. Starting with nucleosides, it is possible to make a library of fully protected triester trimers, such as 4a and 4b. The type 4b trimer, with the benzoyl 3'-protecting group, will become the 3' end of the oligonucleotide. The type 4a trimer is bifunctional and, depending on the treatment (either mild acid or base), will be either the 5' end component or an internal sequence component. Because of the chlorophenyl protecting groups attached by an ester linkage to the phosphate groups (forming phosphotriesters), the trimers and intermediate oligonucleotides are not soluble in water. Therefore all purifications are carried out in nonaqueous solvents such as chloroform. Trimers can be condensed to yield hexamers (for example, 5 + 6 yields 8), and hexamers can be condensed to yield dodecamers (for example, 8 + 9 yields 10, still in fully protected triester form). The next to last step in a typical synthesis is the removal of all protecting groups by treatment with NH_4OH and acetic acid, generating the desired water-soluble single-stranded DNA fragment. The last step is a careful purification of the DNA fragment by high-pressure liquid chromatography.

Fig. 3. (A) Decameric linker (10-mer) bearing a restriction enzyme site is joined by blunt-end ligation to both ends of the DNA to be cloned. Cohesive ends are produced by treatment with the restriction endonuclease; the DNA can then be incorporated into a vector that was cut with the same restriction endonuclease. (B) Nucleotide sequence of an Eco RI linker.



the total mRNA. Therefore it is fortunate that oligodeoxynucleotides show great promise in providing additional tools to aid gene isolation.

Specific primers and hybridization probes. Reverse transcriptase does not initiate DNA synthesis on single-stranded RNA. It is similar to DNA polymerase in that it requires a primer with a free 3'-hydroxyl to be base-paired to the template strand. Oligodeoxythymidylate (dT) has been used to prime most reverse transcription of mRNA, but of course this is nonselective. When sequence information is known or can be guessed, it is possible to selectively prime cDNA synthesis for the gene of interest. For molecular cloning, specific priming was first used successfully by Villa-Komaroff *et al.* (24) to clone the gene for rat proinsulin. The first two nucleotides preceding the poly(A) (polyadenylate) tail of the message were known to be GC (G, guanine; C, cyto-

sine), so a (dT)₃dGdC primer was used. An enrichment for insulin cDNA was probably obtained. Recently, a specific deoxydecanucleotide primer aided the cloning of full-length cDNA's for rat proinsulin I and II (25). Specific priming gave single-stranded cDNA's that were pure enough to use as hybridization probes for screening colonies by the Grunstein-Hogness colony hybridization procedure (26). Earlier, use of a tridecamer as a hybridization probe allowed the identification of a clone carrying the gene for yeast cytochrome c (27).

The exact sequence of nucleotides in the desired mRNA was known for the studies described above. In most cases, however, such information is unavailable. However, if the sequence of the protein is known, the genetic code can be used to predict possible DNA sequences. Because of the degeneracy of the genetic code, there will usually be several possible sequences. However,

with the recent improvements in DNA synthesis rates, all these sequences can be made. As an application of this approach, Noyes *et al.* (28) used a dodecamer to specifically prime gastrin cDNA synthesis. Even with total mRNA from the pyloric gland area, only two prominent cDNA bands were seen by gel electrophoresis analysis. Maxam-Gilbert sequencing established that one of the bands was indeed derived from gastrin mRNA. Two pentadecamers based on the sequence of human interferon have also been used to specifically prime reverse transcriptase (29). Priming was so specific that sequencing of the interferon message was possible. Other studies in which mixed oligodeoxynucleotides are used as primers and hybridization probes are under way in several laboratories. Wallace *et al.* (30) have shown that short oligonucleotides (a heptadecamer, for example) have an advantage over longer DNA fragments as hybridization probes in that single-base-pair mismatches have a large effect on the stability, thus allowing one to distinguish mutant from wild-type.

Gene Alteration

Chemical DNA synthesis allows essentially complete flexibility with regard to gene manipulation and construction. In most cases, the natural sequence will

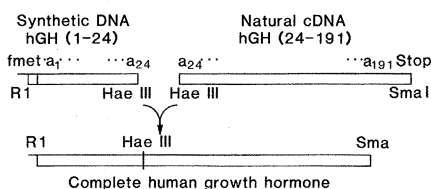
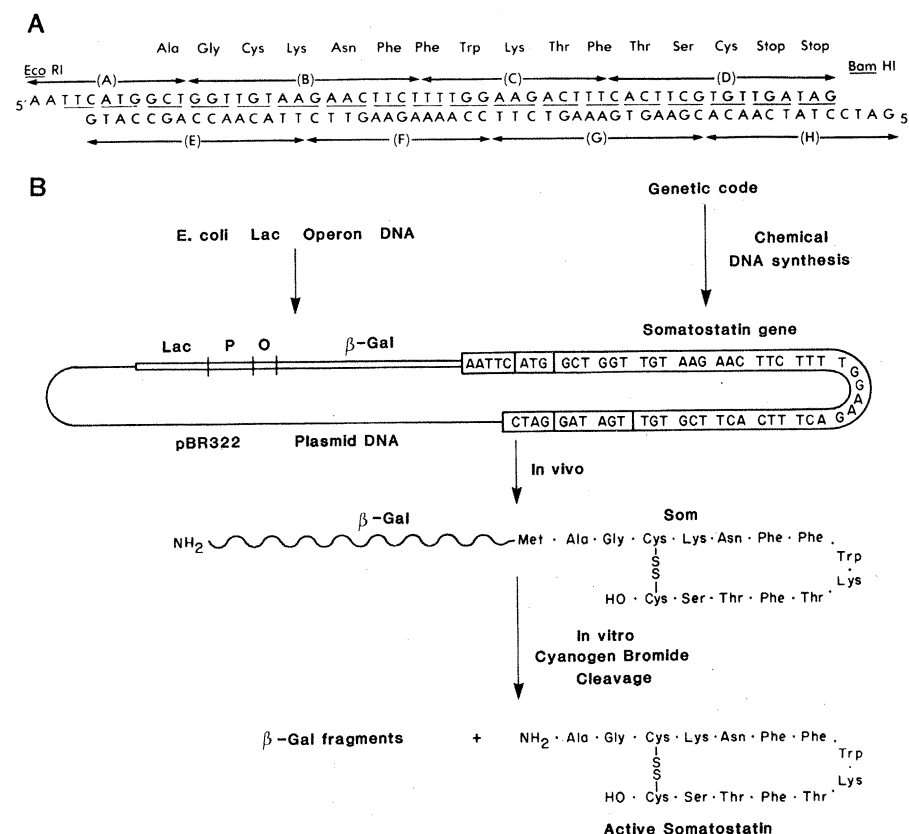


Fig. 4 (left). (A) Somatostatin gene. Eight oligodeoxynucleotides (fragments A to H) were synthesized by the modified triester method (4). The codons are underlined and their corresponding amino acids are given. The eight fragments were designed to have at least five nucleotide complementary overlaps to ensure correct joining by T4 DNA ligase to give duplex DNA 60 nucleotides long. (B) Schematic overview of the somatostatin project. The gene for somatostatin was fused to the *E. coli* β -galactosidase gene on the plasmid pBR322. After transformation into *E. coli*, the plasmid directs the synthesis of a chimeric protein that has somatostatin on a short tail on the end of β -galactosidase. Somatostatin is clipped off by treatment with cyanogen bromide, which cleaves specifically at methionine residues. Fig. 5 (right). Combination of chemically synthesized DNA with an incomplete human growth hormone (hGH) gene to produce a complete gene. The natural gene sequence for amino acids 24 to 191 (the COOH terminus) was obtained by the molecular cloning of pituitary cDNA. This natural gene sequence was then completed by the addition of chemically synthesized DNA (21).

be just the starting point. An excellent example of this is provided by the elegant studies on the lac operator by Caruthers and his colleagues (31). Many changes have been made in the lac operator by combined chemical and enzymatic synthesis. These operator "mutants," most of which are different from operator mutants obtained by classical genetics, have led to a greater understanding of specific protein-DNA interactions (31, 32). Studies are under way in several laboratories to make altered promoters.

Hormone and enzyme analogs. Organic chemists and peptide chemists have long been making analogs, often with good results, but analogs of the larger peptide hormones have been difficult to make. However, these hormones now can be changed at will. With directed mutation and other editing techniques, specifically planned changes can be made even in large hormones and enzymes, so we expect many hormone analog experiments to be done in the near future by recombinant DNA techniques. In the long range, "natural" hormones and enzymes may not be most suitable for many applications.

Gene editing. We use the term editing in the same sense as a film maker would. Most natural genes will probably need to be cut, spliced, or "mutated" for optimum expression in *E. coli*. Unwanted leader sequences, or, in some cases, internal sequences will need to be removed. If the natural sequence near the ribosome binding site and translational start signal is such that inappropriate mRNA secondary structure occurs, the early portion of the gene can be reworked to change the mRNA secondary structure.

Directed Mutation

A powerful tool for gene alteration is provided by directed mutation. There are several methods for directing mutations to specific sites (33), but the most precise, flexible methods use synthetic oligodeoxynucleotides. Hutchinson *et al.* (34) and Razin *et al.* (35) first demonstrated, using single-stranded ϕ X174 DNA, that AT (A, adenine; T, thymine) to GC transitions could be directed efficiently by oligodeoxyribonucleotides. The process, which depends on DNA synthesis by DNA polymerase specifically primed by synthetic DNA, is illustrated in Fig. 6. All types of single-base changes have been made, both transition and transversions (36). In vitro methods for selection of the desired "mutant" DNA were developed by Gillam and Smith (37). Wallace *et al.* (38) recently extended the method for use on double-stranded DNA (pBR322) and showed that specific deletions can be made. In this case, an intron was removed from a yeast suppressor tRNA precursor gene that had been cloned in pBR322. It is now clear that almost any well-defined, simple change can be made by directed mutation techniques.

Conclusion

Fortunately, the fields of chemical DNA synthesis and recombinant DNA came of age at approximately the same time. From this short review, we hope it is apparent that there already has been a synergistic interaction between these fields and probably the best is yet to come.

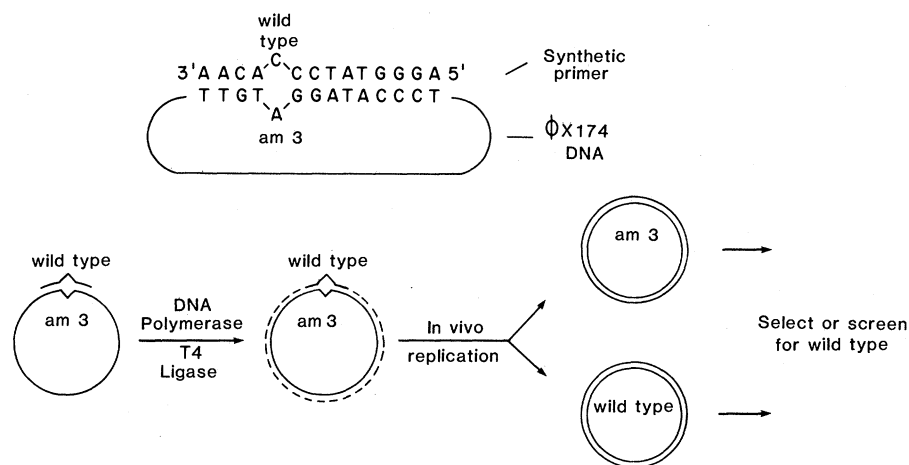


Fig. 6. Correction of a mutant by directed mutation. A chemically synthesized oligodeoxynucleotide with a wild-type sequence is hybridized (with a single-base-pair mismatch) to a ϕ X174 mutant, amber 3 (*am 3*). The oligonucleotide serves as a primer for DNA synthesis in vitro. Transfection and replication in vivo yield a sizable percentage of "corrected" progeny.

References

1. H. G. Khorana, *Science* **203**, 614 (1979).
2. R. W. Holley, J. Appgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *ibid.* **147**, 1462 (1965).
3. K. L. Agarwal *et al.*, *Nature (London)* **227**, 27 (1970); H. G. Khorana *et al.*, *J. Mol. Biol.* **72**, 209 (1972).
4. K. Itakura, N. Katagiri, S. A. Narang, C. P. Bahl, K. J. Mariani, R. Su, *J. Biol. Chem.* **250**, 4592 (1975).
5. S. Gillam, P. Jahnke, M. Smith, *ibid.* **253**, 2532 (1978).
6. S. Gillam, P. Jahnke, C. Astell, S. Phillips, C. A. Hutchinson, M. Smith, *Nucleic Acids Res.* **6**, 2973 (1979).
7. *Nucleic Acids Research*, special edition No. 7, in press.
8. H. L. Heyneker *et al.*, *Nature (London)* **263**, 748 (1976); C. P. Bahl, K. J. Mariani, R. Wu, J. Stawinsky, S. A. Narang, *Gene* **1**, 81 (1976).
9. M. J. Ryan *et al.*, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **36**, 732 (Abstr.) (1977).
10. R. H. Scheller, R. E. Dickerson, H. W. Boyer, A. D. Riggs, K. Itakura, *Science* **196**, 177 (1977).
11. R. H. Scheller, T. L. Thomas, A. S. Lee, W. H. Klein, W. D. Niles, R. J. Britten, E. H. Davidson, *ibid.*, p. 197.
12. T. Maniatis, R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, A. Efstratiadis, *Cell* **15**, 687 (1978).
13. C. P. Bahl, R. Wu, R. Brousseau, A. K. Sood, H. K. Hsiung, S. A. Narang, *Biochem. Biophys. Res. Commun.* **81**, 695 (1978).
14. K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, H. W. Boyer, *Science* **198**, 1056 (1977).
15. A. L. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 422 (1972); S. Lin and I. Zabin, *J. Biol. Chem.* **247**, 2204 (1972); K. Murakami, R. Voellmy, A. L. Goldberg, *ibid.* **254**, 8194 (1979).
16. R. Crea, A. Kraszewski, T. Hirose, K. Itakura, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5765 (1978).
17. D. V. Goeddel *et al.*, *ibid.* **76**, 106 (1979).
18. P. G. Katsoyannis, A. C. Trekatellis, S. Johnson, C. Zalut, G. Schwartz, *Biochemistry* **6**, 2642 (1967).
19. K. Kikuchi, personal communication; R. Chance, personal communication.
20. R. Wetzel, H. L. Heyneker, D. V. Goeddel, P. Jhurani, J. Shapiro, R. Crea, in preparation.
21. D. V. Goeddel *et al.*, *Nature (London)* **281**, 544 (1979).
22. D. V. Goeddel, H. M. Shepard, E. Yelverton, D. W. Leung, A. Sloma, S. Peska, R. Crea, *Nucleic Acids Res.*, in press; D. V. Goeddel *et al.*, in preparation.
23. A. Ullrich, J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. J. Rutter, H. M. Goodman, *Science* **196**, 1313 (1977).
24. L. Villa-Komaroff, A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3727 (1978).
25. S. J. Chan, B. E. Noyes, K. L. Agarwal, D. F. Steiner, *ibid.* **76**, 5036 (1979).
26. M. Grunstein and D. S. Hogness, *ibid.* **72**, 3961 (1975).
27. D. L. Montgomery, B. D. Hall, S. Gillam, M. Smith, *Cell* **14**, 673 (1978).
28. B. A. Noyes, M. Mevarech, R. Stein, K. L. Agarwal, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1770 (1979); M. Mevarech, B. E. Noyes, K. L. Agarwal, *J. Biol. Chem.* **254**, 7472 (1979).
29. M. Houghton *et al.*, *Nucleic Acids Res.* **8**, 1913 (1980).
30. R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, K. Itakura, *ibid.* **6**, 3543 (1979).
31. E. F. Fisher and M. H. Caruthers, *ibid.* **7**, 401 (1979).
32. D. V. Goeddel, D. G. Yansura, M. H. Caruthers, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3578 (1978).
33. D. Shortle and D. Nathans, *ibid.*, p. 2170; F. Heffron, M. So, B. J. McCarthy, *ibid.*, p. 6012.
34. C. A. Hutchinson, S. Phillips, M. H. Edgell, *J. Biol. Chem.* **254**, 6551 (1978).
35. A. Razin, T. Hirose, K. Itakura, A. D. Riggs, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4268 (1978).
36. S. Gillam, P. Jahnke, C. Astell, S. Phillips, C. A. Hutchinson, M. Smith, *Nucleic Acids Res.* **6**, 2973 (1979); S. Gillam and M. Smith, *Gene* **8**, 81 (1979).
37. S. Gillam and M. Smith, *Gene* **8**, 99 (1979).
38. R. B. Wallace, P. F. Johnson, S. Tanaka, M. Schödl, K. Itakura, J. Abelson, *Science* **209**, 1396 (1980).

21 July 1980