An Alternate Method for Synthesis of Double-stranded DNA Segments* 

(Received for publication, March 29, 1982, and in revised form, June 2, 1982)

John J. Rossi, Ryszard Kierzek†, Ting Huang, Peter A. Walker, and Keiichi Ikura§

From the Department of Molecular Genetics, City of
Hope Research Institute, Duarte, California 91010

Recent progress in the chemical synthesis of DNA has now made it possible to rapidly synthesize singlestranded DNAs over 40 bases in length. We have taken advantage of these longer DNAs in assembling and cloning a 132-base pair gene segment coding for amino acids 126 through the stop codon of human leucocyte interferon α2. The method used involves DNA polymerase I-mediated repair synthesis of synthetic oligonucleotide substrates having short stretches of complementary sequence at their 3' termini. In the presence of DNA polymerase I and the four deoxyribonucleoside triphosphates, those primer-templates are converted to full length double-stranded DNAs. The economy in chemical synthesis using this approach is substantial with a greater than 40% reduction in the amount of chemical synthesis required as compared with the conventional approach. We describe in detail this methodology for the biochemical assembly of long gene segments from synthetic oligodeoxyribonucleotides.

Within recent years, a great deal of interest has been focused on the cloning of DNA sequences coding for biologically and medically important proteins. Recent advances in the chemical synthesis of oligodeoxyribonucleotides have now made it possible to chemically synthesize DNA sequences coding for proteins composed of well over 100 amino acids (1). The advantages of chemically synthesizing such sequences, as compared with cloning the cDNA or genomic sequences, reside in the greater potential for engineering certain desired features medically important proteins. Recent advances in the chemical synthesis of DNA have now made it possible to rapidly synthesize singlestranded DNAs over 40 bases in length. We have taken advantage of these longer DNAs in assembling and cloning a 132-base pair gene segment coding for amino acids 126 through the stop codon of human leucocyte interferon α2. The method used involves DNA polymerase I-mediated repair synthesis of synthetic oligonucleotide substrates having short stretches of complementary sequence at their 3' termini. In the presence of DNA polymerase I and the four deoxyribonucleoside triphosphates, those primer-templates are converted to full length double-stranded DNAs. The economy in chemical synthesis using this approach is substantial with a greater than 40% reduction in the amount of chemical synthesis required as compared with the conventional approach. We describe in detail this methodology for the biochemical assembly of long gene segments from synthetic oligodeoxyribonucleotides.

Within recent years, a great deal of interest has been focused on the cloning of DNA sequences coding for biologically and medically important proteins. Recent advances in the chemical synthesis of oligodeoxyribonucleotides have now made it possible to chemically synthesize DNA sequences coding for proteins composed of well over 100 amino acids (1). The advantages of chemically synthesizing such sequences, as compared with cloning the cDNA or genomic sequences, reside in the greater potential for engineering certain desired features medically important proteins.

Recent advances in the chemical synthesis of DNA have now made it possible to rapidly synthesize singlestranded DNAs over 40 bases in length. We have taken advantage of these longer DNAs in assembling and cloning a 132-base pair gene segment coding for amino acids 126 through the stop codon of human leucocyte interferon α2. The method used involves DNA polymerase I-mediated repair synthesis of synthetic oligonucleotide substrates having short stretches of complementary sequence at their 3' termini. In the presence of DNA polymerase I and the four deoxyribonucleoside triphosphates, those primer-templates are converted to full length double-stranded DNAs. The economy in chemical synthesis using this approach is substantial with a greater than 40% reduction in the amount of chemical synthesis required as compared with the conventional approach. We describe in detail this methodology for the biochemical assembly of long gene segments from synthetic oligodeoxyribonucleotides.

Within recent years, a great deal of interest has been focused on the cloning of DNA sequences coding for biologically and medically important proteins. Recent advances in the chemical synthesis of oligodeoxyribonucleotides have now made it possible to chemically synthesize DNA sequences coding for proteins composed of well over 100 amino acids (1). The advantages of chemically synthesizing such sequences, as compared with cloning the cDNA or genomic sequences, reside in the greater potential for engineering certain desired features medically important proteins. Recent advances in the chemical synthesis of DNA have now made it possible to rapidly synthesize singlestranded DNAs over 40 bases in length. We have taken advantage of these longer DNAs in assembling and cloning a 132-base pair gene segment coding for amino acids 126 through the stop codon of human leucocyte interferon α2. The method used involves DNA polymerase I-mediated repair synthesis of synthetic oligonucleotide substrates having short stretches of complementary sequence at their 3' termini. In the presence of DNA polymerase I and the four deoxyribonucleoside triphosphates, those primer-templates are converted to full length double-stranded DNAs. The economy in chemical synthesis using this approach is substantial with a greater than 40% reduction in the amount of chemical synthesis required as compared with the conventional approach. We describe in detail this methodology for the biochemical assembly of long gene segments from synthetic oligodeoxyribonucleotides.

The total chemical synthesis for a final product of 144-b.p. nucleotides involved only four synthetic fragments of 43, 42, 39, and 39 bases long (or a total of 163 bases). This represents a 44% reduction in the amount of chemical synthesis required by the conventional approaches (1, 6). The IFα gene segment has been cloned into an E. coli plasmid vector using a three-part ligation, and the correct DNA sequence of the synthetic gene segment has been verified subsequent to its cloning.

We report in detail here the biochemical methodology used to construct and clone the synthetic IFα gene segment. This approach to gene synthesis should prove generally useful for the synthesis and cloning of DNA encoding regulatory and protein sequences. The chemical synthesis of the starting oligodeoxyribonucleotides by the solid phase method will be reported elsewhere.

EXPERIMENTAL PROCEDURES

DNA Synthesis and Purification—Oligonucleotides were chemically synthesized by the phosphotriester method using a solid support. Details of the chemical synthesis will be reported elsewhere (7). Individual oligodeoxyribonucleotides were purified by gel electrophoresis using 8% polyacrylamide gels containing 7 M urea. The DNA fragments were visualized in the gel by ultraviolet shadowing using fluorescent thin layer chromatographic plates and a short wavelength UV lamp. DNAs of the desired chain length were excised from the gel, electroeluted, and further purified by benzoylated naphthoylated DEAE-cellulose (Serva) column chromatography (8).

Enzymes and Chemicals—Restriction endonucleases Eco RI and Pst I, Polynucleotide kinase, T4 DNA ligase, and bacterial alkaline phosphatase were purchased from Bethesda Research Laboratories. The large fragment of E. coli DNA polymerase I (Klenow) was purchased from Boehringer Mannheim. Acrylamide and bisacrylamide were from Bio-Rad. The [γ-32P]ATP (60000-10000 Ci/mmol) was synthesized according to published procedures (9) using commercially available 32P (ICN or New England Nuclear); [α-32P]CTP was purchased from New England Nuclear.

 Gel and Enzyme Buffers—Restriction endonuclease digestions and DNA polymerase I reactions were all performed in Easo/R buffer (10 mM Tris-HCl (pH 7.9), 6.6 mM MgCl2, 6 mM β-mercaptoethanol, 60 mM NaC1). Gel electrophoresis was carried out using a one-half concentrate of Tris-borate-EDTA buffer (10). Bacterial alkaline phosphatase treatment, ligations, and kinases were carried out under

---

*The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

†Present address, Institute of Organic Chemistry, Polish Academy of Sciences, Department of Stereochromene of Natural Products, Poznan, Poland.

§Recipient of United States Public Health Service Grant GM26408 and a member of the Cancer Research Center (Grant CA16434) at the City of Hope Research Institute.

The abbreviations used are: IFα, human leucocyte interferon α2; b.p., base pair.
conditions recommended by the supplier.

DNA Sequencing—DNA sequences were determined using the method of Maxam and Gilbert (11) on restriction fragments labeled at their 3' termini with [\alpha-^{32}P]dCTP and DNA polymerase I (Klenow). Labeling was carried out in *Endo* R buffer at room temperature for 5 min.

**Bacterial Transformations**—E. coli strain MC1061 (12) was used as the transformation host, employing the procedure of Kushner (13). Antibiotics, ampicillin or tetracycline, were used in the growth media.

Labeling was carried out in the presence of [\alpha-^{32}P]dCTP and

The polymerization products were cleaved with either Eco RI for A and B or *Pst* I for C and D. The desired products of these cleavages were purified by gel electrophoresis. The appropriate fragments were eluted from the gel and ligated with the Eco RI-*Pst* I-cleaved plasmid pXJ01 as depicted. See text for additional details concerning screening and selection of colonies containing the cloned IFo2 gene segment.

The complete nucleotide sequence of the synthetic, polymerized products is depicted below. The heavy underlining depicts the original, synthetic oligonucleotides A, B, C, and D. The underlinings overlap at the regions of 3' complementary sequences.

The following amino acid sequences were determined using the amino acid sequence for IFo2 amino acid residues 14-17. We have utilized the amino acid coding sequences deduced for IFo2 (17) to chemically synthesize four oligodeoxyribonucleotides which, when assembled, comprise the coding sequence for IFo2 amino acid residues 126 through 165. In addition, restriction endonuclease recognition sites for Eco RI and *Pst* I were introduced into the synthetic design to facilitate cloning of the IFo2 segment into a bacterial plasmid vector. As stated under "Introduction," the codon usage takes advantage of the most abundant *E. coli* tRNA species (5) and, therefore, differs considerably from that found in the corresponding natural IFo2 (Table I).

Four oligodeoxyribonucleotide fragments, 43, 42, 39, and 39 bases long, were chemically synthesized by the phosphotriester method. Each pair of fragments was designed to have a short region of complementary sequence at its 3' termini (Fig. 1). For the A and B fragments, this stretch is 10 b.p., while for the C and *D* fragments there is 9 b.p. of complementary sequence. When the pairs of fragments are annealed in the presence of deoxyribonucleoside triphosphates, they become a substrate for DNA polymerase I (Klenow)-mediated polymerization to give a full length duplex product (Figs. 1 and 2).

**Synthetic Design**—The complete nucleotide sequences of several different human leukocyte interferon species have been determined from analyses of cDNA and genomic clones (14-17). We have utilized the amino acid coding sequences deduced for IFo2 (17) to chemically synthesize four oligodeoxyribonucleotides which, when assembled, comprise the coding sequence for IFo2 amino acid residues 126 through 165. In addition, restriction endonuclease recognition sites for Eco RI and *Pst* I were introduced into the synthetic design to facilitate cloning of the IFo2 segment into a bacterial plasmid vector. As stated under "Introduction," the codon usage takes advantage of the most abundant *E. coli* tRNA species (5) and, therefore, differs considerably from that found in the corresponding natural IFo2 (Table I).

Four oligodeoxyribonucleotide fragments, 43, 42, 39, and 39 bases long, were chemically synthesized by the phosphotriester method. Each pair of fragments was designed to have a short region of complementary sequence at its 3' termini (Fig. 1). For the A and B fragments, this stretch is 10 b.p., while for the C and D fragments there is 9 b.p. of complementary sequence. When the pairs of fragments are annealed in the presence of deoxyribonucleoside triphosphates, they become a substrate for DNA polymerase I (Klenow)-mediated polymerization to give a full length duplex product (Figs. 1 and 2).

**In Vitro Construction of Double-stranded Gene Fragments**—Prior to polymerization, each synthetic fragment was phosphorylated at its 5' terminus, using polynucleotide kinase and ATP. This was done to facilitate detection and subsequent cleavage of the double-stranded products by the appropriate restriction endonucleases. In a typical experiment, equimolar amounts of 5' terminally phosphorylated fragments A and B or C and D were mixed together, heated in a boiling water bath for 3 min, quickly chilled, and then allowed to form the desired annealing at their 3' termini. The fragments were then incubated with all four deoxyribonucleoside triphosphates and DNA polymerase I. The products of these reactions were electrophoresed in acrylamide to resolve duplexes from the single-stranded oligonucleotides (Fig. 2). From the relative
ments using E. coli DNA polymerase I (Klenow). A, B, C, and D were individually phosphorylated at their 5' termini with polynucleotide kinase and a mixture of \( [\gamma-^{32}P]ATP \) and nonradioactive ATP. The reactions were terminated by extractions with phenol and chloroform and, finally, ethanol-precipitated after the addition of carrier RNA. The precipitates were dried and resuspended in distilled water, and equal molar amounts of fragments A and B or C and D were mixed together in a final volume of 40 \( \mu l \) at individual concentrations of 1.5 mM. The mixtures were heated in a boiling water bath for 3 min, cooled quickly in an ice water bath, and brought to 10 \( \text{mM} \) Tris, pH 7.9, 60 \( \text{mM} \) NaCl, 6 \( \text{mM} \) \( \beta \)-mercaptoethanol, 6.6 \( \text{mM} \) MgCl\(_2\). A mixture of all four deoxyribonucleoside triphosphates was added to a final concentration of 60 \( \mu \text{g/ml} \) for each dNTP. The polymerization reaction was initiated by the addition of the 1.5 units of DNA polymerase I (Klenow) and allowed to proceed for 30 min at room temperature. The polymerization was stopped by two phenol extractions, followed by two chloroform extractions and, finally, an ethanol precipitation. Aliquots of the individual (A, B, C, or D) and polymerized (A + B Pol. or C + D Pol.) product were electrophoresed on an 8% polyacrylamide-7 M urea gel. The single-stranded 43- (A), 39- (B), 42- (C), and 39-base pair (D), as well as the polymerized 72-base pair (A + B Pol.) and 72-base pair (C + D Pol.), fragments are indicated. The background products in the polymerized reactions are due to the 3'-5' exonuclease activity of DNA polymerase I.

The development of more efficient methods for the chemical synthesis of oligodeoxyribonucleotides by the solid phase approach allows us to chemically synthesize DNAs of over 40 bases in length (7). A marked advantage of utilizing long DNAs for cloning is that it greatly simplifies the gene assembly process. Prior to this communication, the only approach for gene assembly was that pioneered by Khorana (6) and his co-workers, template-dependent ligations of short oligonucleotides. However, this approach has some disadvantages. The entire DNA sequence to be synthesized must be carefully screened by computer analyses to search for regions of overlapping homologies (1). These regions could hamper desired placements of restriction endonuclease sites as well as codon usage in certain regions of the gene sequence. The methodology we have utilized to assemble the IF\(_{2}\) gene segment provides a rapid and convenient alternative to constructing double-stranded gene segments. Although homologous complementary sequences other than the target sequences could lead to undesired polymerization products, this is more easily controlled by adding or deleting a few nucleotides in the design of the region of 3' complementary sequence. It should be noted that DNA polymerase has been previously used to complete a double-stranded sequence, but in this case, only one strand served as primer (6).

The fidelity of the overlapping polymerization technique is demonstrated by the nucleotide sequence analyses of the cloned gene segment (Fig. 3) which agree precisely with the designed sequence (Fig. 1). The three-part ligation methodology used to clone the IF\(_{2}\) gene segment also greatly simplifies the gene assembly process (Fig. 1). The frequency of clones containing the desired insert was somewhat low, approximately 4%. This could be due to the lowered probability of properly aligning all three segments of DNA for the desired joinings. Although we used a DNA miniscreen procedure to analyze the phenotypically correct clones, colony hybridiza-
tion with radioactively labeled synthetic DNA probes could greatly facilitate the screening process (21).

One of the most important features in synthesizing a designed gene coding for a protein would be higher expression than the corresponding natural gene in a microbial environment. It has been demonstrated for several organisms that a definite bias in codon usage exists for highly expressed proteins. In E. coli, where many tRNA species have been identified, this codon bias seems to be well correlated with the relative abundance of iso-accepting tRNA species (5). We have incorporated the E. coli preferred codons into the synthetic IFα2 segment (Table I). Obviously, we will have to await the total gene synthesis for IFα2 in order to draw any conclusions as to whether or not codon usage will play a significant role in the expression of this gene.

Acknowledgment—We would like to thank Dr. T. Ikemura for the preprint (Ref. 5).

REFERENCES
1. Edge, M. D., Greene, A. R., Heathcliffe, G. R., Meacock, P. A., Schuch, W., Scanlon, D. B., Atkinson, T. C., Newton, C. R., and Markham, A. F. (1981) Nature (Lond.) 292, 756-762
2. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., and Boyer, H. W. (1977) Science 198, 1086-1093
3. Crea, R., Hirose, T., Kraszewski, A., and Itakura, K. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 5765-5769
4. Goeddel, D. V., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Vansura, D. G., Ross, M. G., Miozzari, G., Crea, R., and Seeburg, P. H. (1979) Nature (Lond.) 281, 544-548
5. Ikemura, T. (1981) J. Mol. Biol. 151, 389-409
6. Khorana, H. (1979) Science 205, 614-625
7. Ito, H., Ike, Y., Ikuta, S., and Itakura, K. (1982) Nucleic Acids Res. 10, 1755-1768
8. Rossi, J., Ross, W., Egan, J., Lipman, D., and Landy, A. (1979) J. Mol. Biol. 128, 21-47
9. Walseth, T. F., and Johnson, R. A. (1979) Biochim. Biophys. Acta 526, 11-31
10. Peacock, A., and Dingham, C. (1968) Biochemistry 7, 668-674
11. Maxam, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564
12. Casadaban, M., and Cohen, S. (1980) J. Mol. Biol. 138, 179-207
13. Kushner, S. (1978) in Genetic Engineering (Boyer, H., and Nicosia, S., eds) pp. 17-23, Elsevier-North Holland, Amsterdam
14. Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A., Yelverton, E., and Gray, P. W. (1981) Nature (Lond.) 290, 26-30
15. Lawn, E. M., Adelman, J., Dull, T., Gross, M., Goeddel, D. V., and Ullrich, A. (1981) Science 212, 1159-1162
16. Nagata, S., Mantei, N., and Weissmann, C. (1980) Nature (Lond.) 287, 401-408
17. Goeddel, D., Yelverton, E., Ullrich, A., Heyneker, H., Miozzari, G., Holmes, W., Seeburg, P., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandliss, R., Sloma, A., Tabor, J., Gross, M., Famielli, P., and Pestka, S. (1980) Nature (Lond.) 287, 411-416
18. Soberon, X., Rossi, J. J., Larson, G. P., and Itakura, K. (1982) Nucleic Acids Res. Symp. Ser., No. 9, in press
19. Soberon, X., Covarrubias, L., and Bolivar, F. (1980) Gene, 1, 305-308
20. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523
21. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., and Itakura, K. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 6615-6617

FIG. 3. DNA sequence of cloned IFα2 gene segment. The DNA sequences presented were determined from 3'-labeled fragments using the DNA sequencing procedure of Maxam and Gilbert (6). For the sequence presented in A, a Hpa II site corresponding to the site at position 2490 of the pBR322 sequence (19) was 3'-labeled with [α-32P] dCTP and DNA polymerase I (Klenow). The plasmid was secondarily cleaved with HindIII and the appropriate fragment was purified from the gel and sequenced. The nucleotide sequence presented in A corresponds to the upper strand of the Fig. 1 sequence beginning with the Pst I site. For clarity, two different lengths of sequencing gel runs are presented. For the B sequence, the HindIII site was 3'-labeled with [α-32P] dCTP and DNA polymerase I (Klenow). The plasmid was secondarily cleaved with Pst I. The sequence presented corresponds to the lower strand of the Fig. 1 sequence beginning with amino acid 128. The correct sequence upstream of this position has also been confirmed (data not presented).