[Chem. Pharm. Bull. 34(11)4724—4730(1986)]

Deoxyribonucleic Acids and Related Compounds. XXII.[1] Synthesis of Genes for Human Nerve Growth Factor and Its Fused Protein

SHIGENORI IWAI,[a] AKIHIRO IMURA,[a] YASUHIKO INOUE,[a] HIDEO INOUE,[a] KAZUNOBU MIURA,[a] TOMOKO TOKUNAGA,[b] MORIO IKEHARA[b] and EIKO OHTSUKA*,[a]

Faculty of Pharmaceutical Sciences, Hokkaido University,[a] Kita-12, Nishi-6, Kita-ku, Sapporo 060, Japan and Faculty of Pharmaceutical Sciences, Osaka University,[b] 1–6 Yamadaoka, Suita 565, Japan

(Received May 19, 1986)

Gene I and II for human nerve growth factor β-subunit (hNGF) (118 amino acids) were synthesized by enzymatic joining of 45 oligodeoxynucleotides with chain lengths of *ca.* 17, which were obtained by the phosphotriester solid-phase synthesis. Gene I contained the methionine codon downstream of a restriction site, *Cla* I, for direct expression under the control of *E. coli trp* promoter, inserted in a plasmid pGH-L9 (M. Ikehara *et al.*, *Proc. Natl. Acad. Sci.*, **81**, 5956 (1984)). Gene II coded for a fused protein containing two-thirds of human growth hormone as well as the sequence Ile–Glu–Gly–Arg, in front of the amino acid sequence of hNGF. Both genes were amplified in *E. coli* by ligation to restriction sites for *Cla* I and *Sal* I of pGH-L9 followed by transformation of *E. coli*. The molecular weight of the expressed products was estimated by electrophoresis on acrylamide gel. The structure of the gene was characterized by restriction analysis and sequencing.

Keywords——phosphotriester method; peptide gene; HPLC; DNA sequence; gene design

Mouse nerve growth factor (mNGF) is known to be responsible for the development and maintenance of neurones.[2] Recently, the levels of mNGF and its messenger ribonucleic acid (mRNA) in the central nervous system were determined.[3] mNGF from male submaxillary gland consists of three subunits, forming a 7 S complex, and two identical β-subunits with 118 amino acids are responsible for its activity.[4] The amino acid sequence of the β-subunit determined some years ago,[5] and recently the nucleotide sequence of the gene for mNGF was determined.[6,7] Structures for a prepro-β-NGF of 27 K and a pro-β-NGF of 25K have been predicted from it. By using the mNGF complementary deoxyribonucleic acid (cDNA) clone, the human β-NGF (hNGF) gene has been isolated and sequenced.[7] Since no proteins with hNGF activity have yet been isolated, in spite of the physiological evidence, it is of importance to prepare a protein which has the amino acid sequence predicted from the hNGF gene. In the present paper we report the design and synthesis of genes which code hNGF with 118 amino acids. Characterization by sequence analysis and expression of the ligated gene in plasmids are also described.

Synthesis of Gene Fragments for hNGF

A gene for hNGF was designed according to the known sequence of 118 amino acids[7] using codons which are used in highly expressed genes in *E. coli*[8] in order to facilitate biosynthesis of hNGF in *E. coli*. Two genes which contained an altered amino terminus were synthesized (Fig. 1). For direct expression, the initiation signal ATG was incorporated in the amino terminus (gene I). For production of a fused protein with two-thirds of the human

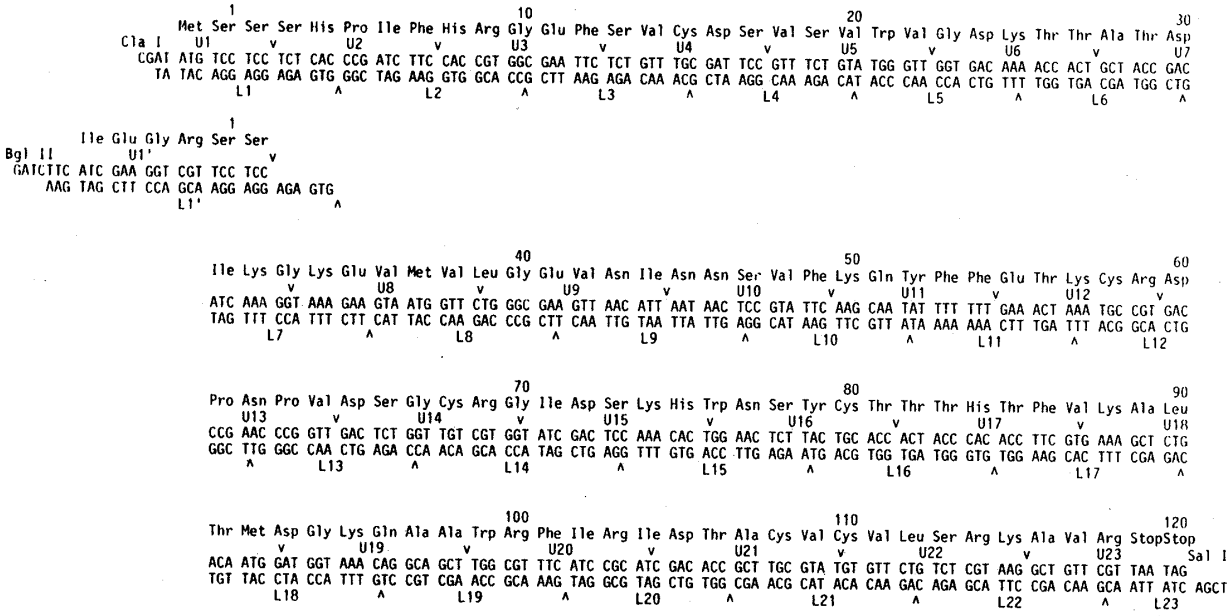


Fig. 1. Amino Acid Sequence and Synthetic Genes for hNGF

Synthetic fragments are numbered.

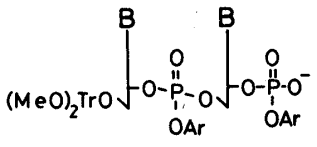
TABLE I. Codon Usage in Synthetic (and Natural) hNGF Gene

First position (5′ end)	Second position								Third position (3′ end)
	U		C		A		G		
U	Phe	2 (4)	Ser	6 (0)	Tyr	1 (1)	Cys	2 (4)	U
	Phe	5 (3)	Ser	5 (1)	Tyr	1 (1)	Cys	4 (2)	C
	Leu	0 (0)	Ser	0 (4)	Term	1 (0)	Term	0 (0)	A
	Leu	0 (1)	Ser	0 (1)	Term	1 (0)	Trp	3 (3)	G
C	Leu	0 (0)	Pro	0 (0)	His	0 (1)	Arg	6 (0)	U
	Leu	0 (1)	Pro	0 (2)	His	4 (3)	Arg	1 (0)	C
	Leu	0 (0)	Pro	0 (1)	Gln	1 (0)	Arg	0 (0)	A
	Leu	3 (1)	Pro	3 (0)	Gln	1 (2)	Arg	0 (4)	G
A	Ile	1 (2)	Thr	3 (1)	Asn	1 (1)	Ser	0 (2)	U
	Ile	5 (3)	Thr	6 (6)	Asn	4 (4)	Ser	0 (3)	C
	Ile	0 (1)	Thr	1 (1)	Lys	7 (1)	Arg	0 (1)	A
	Met	3 (2)	Thr	0 (2)	Lys	2 (8)	Arg	0 (2)	G
G	Val	8 (2)	Ala	5 (2)	Asp	2 (3)	Gly	5 (0)	U
	Val	0 (2)	Ala	0 (3)	Asp	6 (5)	Gly	2 (4)	C
	Val	4 (1)	Ala	1 (0)	Glu	4 (1)	Gly	0 (1)	A
	Val	1 (8)	Ala	0 (1)	Glu	0 (3)	Gly	0 (2)	G

growth hormone (hGH), a dodecanucleotide which codes for the recognition sequence of a restriction endopeptidase, blood coagulation factor Xa (Ile–Glu–Gly–Arg)[9] was inserted upstream of the amino terminus (gene II). Codon usage in the present synthetic gene is compared with that in the human gene[7] in Table I. Gene I was designed for insertion into the *Cla* I–*Sal* I sites of an expression vector which has been used for expression of a synthetic hGH gene.[10] Gene II contained *Bgl* II–*Sal* I sites which could be used to join it to the plasmid at two-thirds along the hGH gene. Internal restriction sites were inserted by searching possible amino acid sequences. Recognition sequences for *Eco* RI and *Hpa* I are shown at Glu (11)–Phe (12) and Val (42)–Asn (43) in Fig. 1. These genes were divided into 45 oligonucleotides with

TABLE II. Oligonucleotides and Synthetic Units

U1	CGATATGTCCTCC	L1	GTGAGAGGAGGACATAT
U1'	GATCTTCATCGAAGGTCGTTCCTCC	L1'	GTGAGAGGAGGAACGACCTTCGATGAA
U2	TCTCACCCGATCTTC	L2	CCACGGTGGAAGATCGG
U3	CACCGTGGCGAATTC	L3	CAAACAGAGAATTCG
U4	TCTGTTTGCGATTCC	L4	ACAGAAACGGAATCG
U5	GTTTCTGTATGGGTT	L5	TTGTCACCAACCCAT
U6	GGTGACAAAACCACT	L6	TCGGTAGCAGTGGTT
U7	GCTACCGACATCAAAGG	L7	TTCTTTACCTTTGATG
U8	TAAAGAAGTAATGGTTC	L8	TCGCCCAGAACCATTAC
U9	TGGGCGAAGTTAACATT	L9	AGTTATTAATGTTAACT
U10	AATAACTCCGTATTC	L10	ATTGCTTGAATACGG
U11	AAGCAATATTTTTTT	L11	TAGTTTCAAAAAAAT
U12	GAAACTAAATGCCGT	L12	TCGGGTCACGGCATT
U13	GACCCGAACCCGGTT	L13	CAGAGTCAACCGGGT
U14	GACTCTGGTTGTCGTGG	L14	GAGTCGATACCACGACAAC
U15	TATCGACTCCAAACACT	L15	AAGAGTTCCAGTGTTTG
U16	GGAACTCTTACTGCACC	L16	GTGGGTAGTGGTGCAGT
U17	ACTACCCACACCTTCGT	L17	AGAGCTTTCACGAAGGT
U18	GAAAGCTCTGACAATGG	L18	GTTTACCATCCATTGTC
U19	ATGGTAAACAGGCAGCT	L19	GAAACGCCAAGCTGCCT
U20	TGGCGTTTCATCCGCAT	L20	GGTGTCGATGCGGAT
U21	CGACACCGCTTGCGTAT	L21	AGAACACATACGCAAGC
U22	GTGTTCTGTCTCGTAAG	L22	GAACAGCCTTACGAGAC
U23	GCTGTTCGTTAATAG	L23	TCGACTATTAAC



B = thymin-1-yl,
N-benzoylcytosin-1-yl,
N-benzoyladenin-9-yl,
N-isobutyrylguanine-9-yl.
Ar = *o*-chlorophenyl.
$(MeO)_2Tr$ = 4,4′-dimethoxytrityl.

Chart 1

chain lengths of 17, which were synthesized by the phosphotriester method on 1% cross-linked polystyrene.[11] Table II shows synthetic fragments used for the synthesis. U1′ and L1′ which are located at the N-terminus of gene II, were synthesized by the amidite method.[12] Protected dinucleotides[13] (Chart 1) were purified by reversed-phase chromatography and used as condensing units. They were examined by reversed phase thin layer chromatography (TLC) before condensation. 1-(Mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole[14](MSNT) was employed as the condensing reagent at 40 °C for 20 min.[13] Oligonucleotides were purified by chromatography on alkylated silica gel before removal of the 5′-dimethoxytrityl group. The completely deprotected product was further purified by high-pressure liquid chromatography

(HPLC) on C-18 silica gel and analyzed by anion-exchange HPLC. The sequence and purity of the oligonucleotides were confirmed by mobility shift analysis.[15]

Ligation and Insertion of the Genes into Plasmids

The synthetic oligonucleotides were ligated as shown in Fig. 2 using T4 DNA ligase. The 5′-hydroxyl groups of oligonucleotides were phosphorylated with polynucleotide kinase and adenosine triphosphate (ATP), except for the 5′-terminal fragments (U1, U1′ and L23). Gene I for direct expression of hNGF was constructed using subfragment I which was prepared by

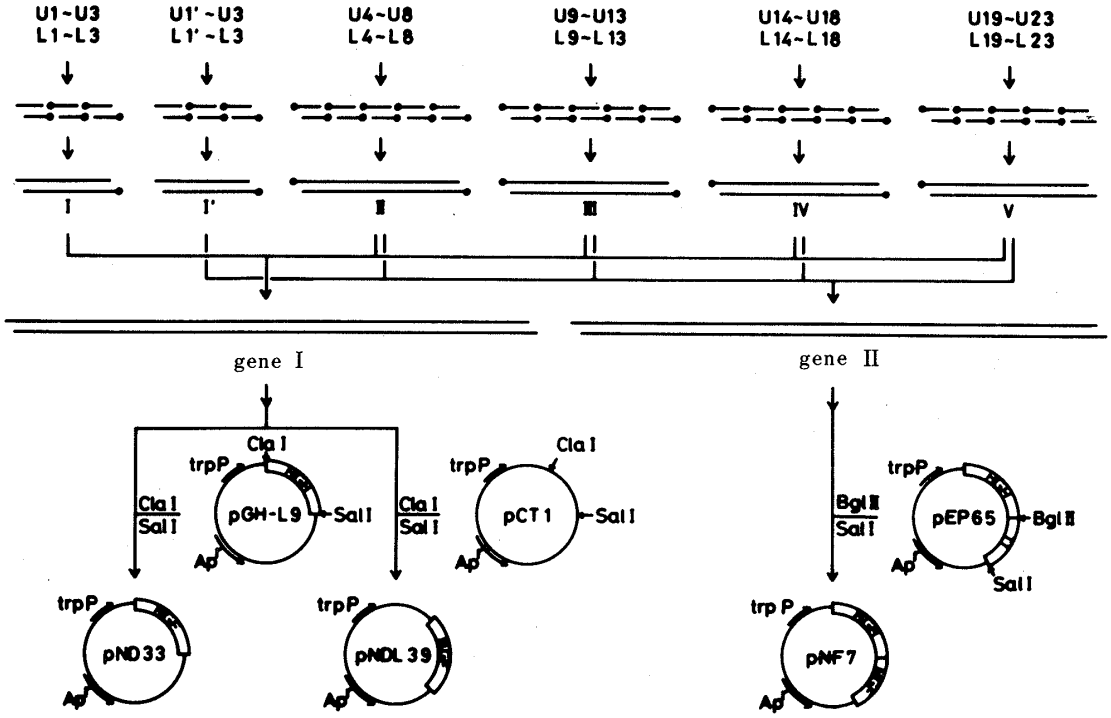


Fig. 2. Construction of Plasmids Containing the hNGF Gene

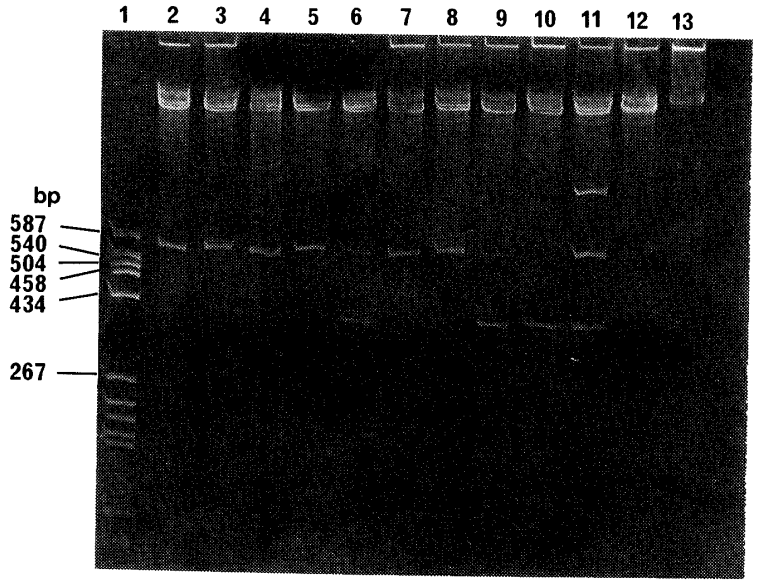


Fig. 3. Restriction Analysis of Plasmids Containing the hNGF Gene

pND25 (lane 2) contained the hGH gene and pND33 (lane 10) contained an hNGF gene. Lane 1, *Hae* III digests of pBR322; lane 2—13, *Cla* I and *Sal* I digests of pND25—36.

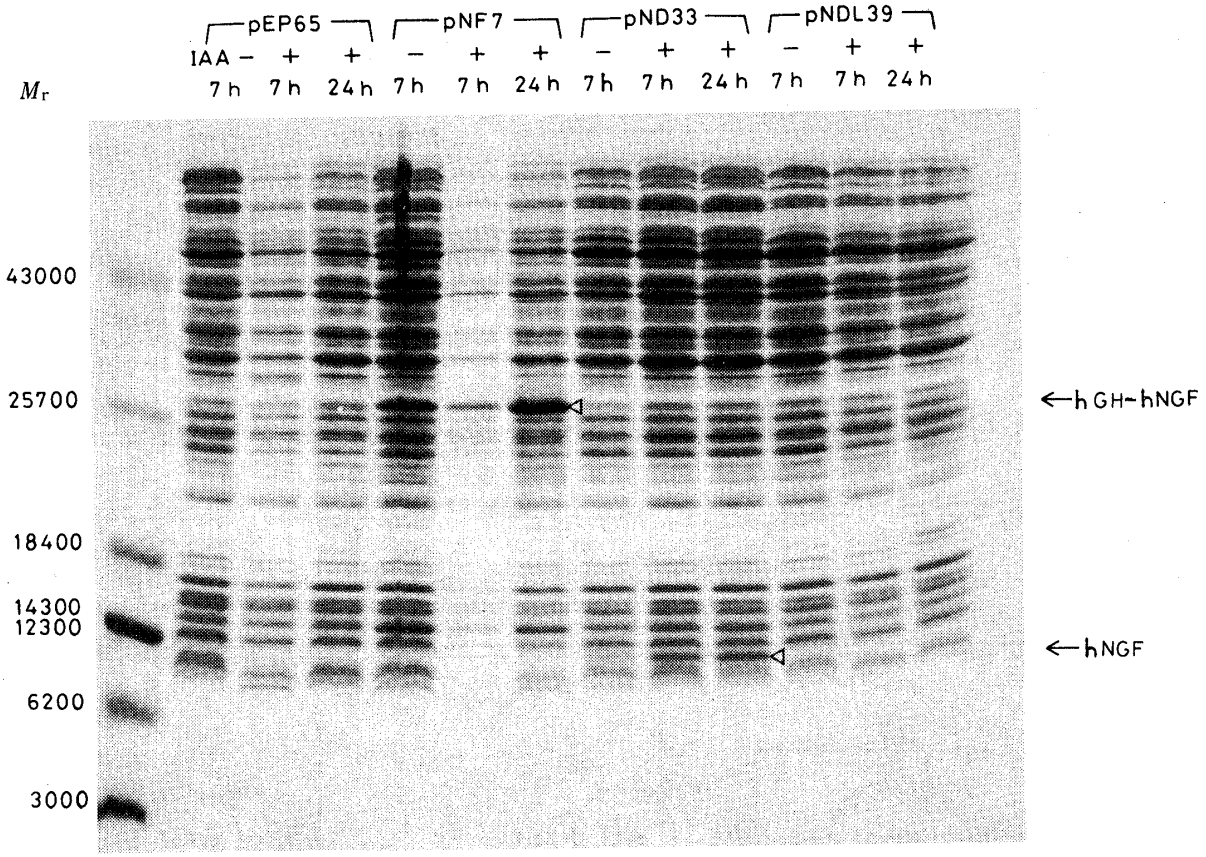


Fig. 4. Detection of Induced Proteins on 15% PAGE

ligation of U1 and L1 containing the methionine codon downstream of the *Cla* I site, as shown in Fig. 2. For the synthesis of gene II coding fused hNGF, subfragment I′ was prepared by using U1′ and L1′, which contain the *Bgl* II site and the coding sequence for the recognition sequence Ile–Glu–Gly–Arg for blood coagulation factor Xa. Gene I (365 bp) was inserted into two expression vectors having the *trp* promoter of *E. coli* without the leader sequence (hGH-L9)[10] or with the attenuator (pCT1)[10] at the *Cla* I–*Sal* I site to yield pND33 and pNDL39, respectively. Gene II (375 bp) was joined at the *Bgl* II–*Sal* I site of pEP65 which contained a gene for RNase T_1 at this site. The resulting plasmid pNF7 had a coding region for 262 amino acids.

Identification of the Synthetic Gene

E. coli HB 101 were transformed using plasmids containing gene I or gene II. Transformants were screened by the rapid boiling method,[16] and plasmids pND33, pNDL39, and pNF7 contained *Cla* I–*Sal* I fragments of proper length (365 bp, 365 bp, and 794 bp respectively) (Fig. 3). Each plasmid was isolated and denatured with 0.1 N NaOH to determine the sequence of the gene by Sanger's dideoxy method[17] using synthetic gene fragments as primers. U1, U8, U15, L10, L17, and L23 were used for gene I and U1′ instead of U1 was used for gene II, and it was found that both genes had the correct sequences, as expected. The transformants were also examined for expression of the hNGF gene (Fig.4). Although no induced protein was detected in *E. coli* carrying pNDL39, the transformant harboring pND33 produced a protein with molecular weight of *ca.* 13000 which was considered to be hNGF (M_r, 13398). *E. coli* carrying pNF7 produced more efficiently a protein with a molecular weight of *ca.* 30000, which was regarded as the fused protein (M_r, 29293). The contents of these overproduced proteins were estimated by densitometry and found to be 5 and 17% of

the total protein in *E. coli*, respectively.

Discussion

Synthetic genes hNGF have been designed according to the amino acid sequence derived from the human genomic sequence. The amino acid codons of these synthetic genes were chosen from prefered codons in *E. coli*, unless other codon usage was required to prevent direct repeats which might cause the formation of base pairs between unexpected fragments, or to insert restriction sites. The present approach involves a synthesis of proteins which have not yet been isolated from natural sources. Production of relatively small peptides as a fused protein may be advantageous in obtaining the product from cell extracts because such peptides are usually degraded by proteases in the cell. If cleavage by specific endopeptidases occurs correctly, an extra methionine is not included at the N-terminal position, when the terminus contains other amino acids. hNGF contains serine at the terminus. Unless the initiation methionine is removed *in vivo*, the directly expressed product has the extra amino acid. Fused genes with two-thirds of the hGH and RNase T1 have been efficiently expressed in *E. coli*.[18] It was reported that the structure of the N-terminal gene or mRNA affected the efficiency of expression.[19] Efficiency of expression of fused genes with the N-terminal hGH gene may be expected to be high. Preliminary results on the expression of the present genes, gene I for direct expression and gene II for a fused protein, indicated that the efficiency of expression of gene I was *ca.* one-fourth of that of gene II, although inverted repeat sequences which were reported to reduce the efficiency of gene expression were not found between the promoter and N-terminus of the structure gene (gene I). Isolation and characterization of the gene products are required to investigate the biological properties of hNGF. In particular, characterization of disulfide bonds may be important when a foreign gene is expressed in *E. coli*, whose cytoplasmic proteins mostly contain sulfhydryl groups.

Experimental

General Method——Kieselgel 60 F_{254} plates (Merck) in chloroform–methanol (10 : 1, v/v) and Kieselgel 60 F_{254} silanisiert (Merck) in acetone–20 mM TEAA (7 : 3, v/v) were used for TLC and reversed-phase TLC, respectively. Protected nucleosides and nucleotides were purified by column chromatography using Wakogel C-300 (Wako Pure Chemical Industries) in chloroform–methanol. The purity of dinucleotide units was examined by reversed-phase TLC in acetone–20 mM triethylammoniumacetate (55 : 45, v/v) and a dual-wavelength TLC scanner (Shimadzu CS-900) (λ_s, 254 nm and λ_R, 350 nm). Dinucleotide units whose purity was less than 90% were purified by column chromatography using preparative C18 (Waters) in acetone–0.1% aqueous pyridine (4 : 6—6 : 4, v/v). Protected dinucleotides were synthesized by condensation of 5′-dimethoxytrityl-3′-(*o*-chlorophenyl)phosphoryl-*N*-protected deoxynucleosides and *N*-protected deoxynucleoside 3′-(*o*-chlorophenyl)-β-cyanoethylphosphates using MSNT as the condensing reagent under the conditions described previously.[13] The solid-phase phosphotriester synthesis of heptadecamers was performed manually.[10,13] Larger fragments (U1′ and L1′) were synthesized by the amidite method using a DNA synthesizer (Applied Biosystems model 380 A). Deprotected oligonucleotides were analyzed and purified by HPLC on a Shimadzu LC-3A or Gilson MS-4 system using columns of M&S PACK C18 (M&S Instruments) and TSK gel diethylaminoethyl (DEAE)-2SW (Toyo Soda Manufacturing).[13]

Enzymatic joining of gene fragments, transformation of *E. coli* and induction of gene expression were performed as described.[10] Molecular weight standards for gel electrophoresis of proteins (M_r range 3000—43000 daltons) were purchased from Bethesda Research Laboratories Inc.

Ligation of Gene Fragments and Construction of Plasmids——Oligonucleotides (0.2 A_{260} unit) except for U1 and L23 were phosphorylated by polynucleotide kinase plus ATP as described.[10] Six to ten fragments were treated with T4 DNA ligase in the standard media at 15 °C for 2 h. Subfragments (I—V, in Fig. 3) were isolated by 10% polyacrylamide gel electrophoresis (PAGE) and dissolved in 10 mM Tris–HCl (pH 8.0), 1 mM ethylene diamine tetraacetic acid (EDTA). Genes I and II were prepared by joining subfragments with T4 DNA ligase at 20 °C for 2 h and isolated by 5% PAGE. Gene I (*ca.* 3 μg, 9 pmol) was phosphorylated by polynucleotide kinase (6 unit) and ATP (10 nmol) in 10 μl medium at 37 °C for 1 h. An aliquot of the gene (*ca.* 1 μmol) was joined to the larger fragment of

pGHL9[10] (1.5 μg), which was obtained by *Cla* I plus *Sal* I treatment, with T4 DNA ligase (350 units) in 20 μl of media at 20 °C for 1 h. Plasmids thus obtained were precipitated with ethanol and dissolved in 20 μl of 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA.

Characterization of Synthetic Genes——Gene I inserted into plasmid pND33 and gene II inserted into pNF7 were sequenced by the dideoxy method after cloning in *E. coli*. Plasmids (1.5 μg in 10 μl of 10 mM Tris–HCl–1 mM EDTA) were used to transform *E. coli* HB101 which was treated with $CaCl_2$. The rapid boiling method[16] was slightly modified by introduction of isopropyl alcohol precipitation before RNase A treatment. Restriction analyses of these plasmids were performed using *Cla* I (5 units) and *Sal* I (12 units) for gene I. *Bgl* II (8 units) and *Sal* I (12 units) were used for gene II. Obtained gene fragments were detected by 5% PAGE.

Examination of induced proteins from genes I and II was performed by 15% PAGE in the presence of 0.1% sodium dodecyl sulfate (SDS). *E. coli* harboring plasmids was cultured in M9 media (5 ml) overnight. After incubation of an aliquot (0.1 ml) in M9 (5 ml) for 1 h, 3-indolacrylic acid (IAA) (20 μl, 10 mg/1 ml of ethanol) was added. Aliquots of 0.1 ml were taken at intervals (7 and 24 h) and centrifuged at 13000 rpm for 5 min. Precipitates were stored at −80 °C and dissolved in 62.5 mM Tris–HCl (pH 6.8), 2% SDS, 5 mM EDTA, 10% glycerol, 35 mM 2-mercaptoethanol and 0.001 % bromphenol blue for PAGE.

Acknowledgement This work was supported in part by a Grant from the Takeda Science Foundation.

References

1) Part XXI: Y. Ishino, S. Shibahara, E. Ohtsuka and M. Ikehara, *Nucleosides & Nucleotides*, in press.
2) B. A. Yankner and E. M. Shooter, *Annu. Rev. Biochem.*, **51**, 845 (1982).
3) S. Korsching, G. Auburger, R. Heumann, J. Scott and H. Thoehen, *EMBO J.*, **4**, 1389 (1985).
4) L. A. Greene, S. Varon, A. Piltch and E. M. Shooter, *Neurobiology*, **1**, 37 (1971).
5) E. H. Angeletti and R. A. Bradshaw, *Proc. Natl. Acad. Sci. U.S.A.*, **68**, 2417 (1971).
6) J. Scott, M. Selby, M. Urdea, M. Quiroga, G. L. Bell and W. J. Rutter, *Nature* (London), **302**, 538 (1983).
7) A. Ullrich, A. Gray, C. Berman and T. J. Dull, *Nature* (London), **303**, 821 (1983).
8) T. Ikemura, *J. Mol. Biol.*, **151**, 389 (1981).
9) K. Nagai and H. C. Thegersen, *Nature* (London), **309**, 819 (1984).
10) M. Ikehara, E. Ohtsuka, T. Tokunaga, Y. Taniyama, S. Iwai, K. Kitano, S. Miyamoto, T. Ohgi, Y. Sakuragawa, K. Fujiyama, T. Ikari, M. Kobayashi, T. Miyake, S. Shibahara, A. Ohno, T. Ueda, T. Tanaka, H. Baba, T. Miki, A. Sakurai, T. Oishi, O. Chisaka and K. Matsubara, *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 5956 (1984).
11) H. Ito, Y. Ike and K. Itakura, *Nucleic Acids Res.*, **10**, 1755 (1982).
12) S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.*, **22**, 1859 (1981).
13) E. Ohtsuka, S. Iwai, T. Tokunaga and M. Ikehara, *Chem. Pharm. Bull.*, **33**, 3153 (1985).
14) S. S. Jones, B. Rayner, C. B. Reese, A. Ubasawa and M. Ubasawa, *Tetrahedron*, **36**, 3075 (1980).
15) *a*) F. Sanger, J. E. Donelson, A. R. Coulson, H. Kossel and D. Fischer, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 1209 (1973); *b*) M. Silberklung, A. M. Gillum and U. L. RojBhandary, *Nucleic Acids Res.*, **4**, 4091 (1977).
16) D. S. Holmes and M. Quigley, *Anal. Biochem.*, **114**, 193 (1981).
17) *a*) F. Sanger, S. Nicklen and A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5463 (1977); *b*) J. Messing, *Methods Enzymol.*, **68**, 326 (1979).
18) M. Ikehara, E. Ohtsuka, T. Tokunaga, S. Nishikawa, S. Uesugi, T. Tanaka, Y. Aoyama, S. Kikyotani, K. Fujimoto, K. Yanase, K. Fuchimura and H. Morioka, *Proc. Natl. Acad. Sci. U.S.A.* in press.
19) M. D. Edge, A. R. Greene, G. R. Heathcliffe, V. E. More, N. J. Faukner, R. Camble, N. N. Retter, P. Trueman, W. Schuch, J. Hennam, T. C. Atkinson, C. R. Newton and A. F. Markham, *Nucleic Acids Res.*, **11**, 6419 (1983).