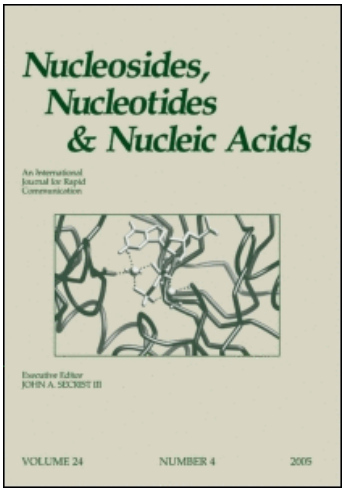

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Construction of Genes or Gene Fragments by Use of Two Long Synthetic Oligonucleotides Representing the Coding and Noncoding Strands

Friedrich Hein[a]; Hans W. Jansen[a]; Eugen Uhlmann[a]
[a] HOECHST AG, Frankfurt (Main) 80, Federal Republic of Germany

To cite this Article Hein, Friedrich , Jansen, Hans W. and Uhlmann, Eugen(1988) 'Construction of Genes or Gene Fragments by Use of Two Long Synthetic Oligonucleotides Representing the Coding and Noncoding Strands', Nucleosides, Nucleotides and Nucleic Acids, 7: 4, 497 — 510
To link to this Article: DOI: 10.1080/07328318808075392
URL: http://dx.doi.org/10.1080/07328318808075392

NUCLEOSIDES & NUCLEOTIDES, 7(4), 497-510 (1988)

CONSTRUCTION OF GENES OR GENE FRAGMENTS BY USE OF TWO LONG SYNTHETIC OLIGONUCLEOTIDES REPRESENTING THE CODING AND NONCODING STRANDS

Friedrich Hein, Hans W. Jansen and Eugen Uhlmann*
HOECHST AG, P.O. Box 80 03 20, D-6230 Frankfurt (Main) 80
Federal Republic of Germany

ABSTRACT

The construction of genes as exemplified for a gene encoding salmon calcitonin-gly(33) from only two long synthetic oligonucleotides (coding/noncoding strand) is described. The type of solid support used in oligonucleotide preparation strongly influences synthesis yields and mutation rates in the cloning of the synthetic DNA.

INTRODUCTION

Since the introduction of 2' - deoxyribonucleoside-3'- O - methyl-N,N- dialkylphosphoramidites and silica gel (Fractosil) supports into solid phase synthesis of oligonucleotides by Caruthers [1,2], further improvements were achieved by the use of 2'- deoxyribonucleoside-3'-O-(2-cyanoethyl)-N,N-dialkylphosphoramidites[3] and CPG (controlled pore glass) supports [4,5]. Additonally increased cloning efficiency due to reduced base modification was reported for 2-cyanoethyl-protected phosphoramidites in comparision with methyl-protected phosphoramidites [6].

Although chemical synthesis of long oligonucleotides (>100 nt) by the phosphoramidite method has been reported[7,8] to our knowledge no details of cloning and sequencing studies are yet available. According to the present status of automated DNA synthesis oligonucleotides up to a length of 150 - 180 nucleotide units can be prepared corresponding

to genes for peptides consisting of up to 60 amino acid residues. Structural genes of such a size, e.g. genes encoding 27-desamidosecretin [9], vasoactive intestinal peptide [10], human epidermal growth factor [11], human growth hormone releasing factor [12-14], usually have been assembled by enzymatic joining of ten or more complementary oligonucleotides by virtue of their sticky ends as originally introduced by Khorana [15]. Quite recently the construction of a gene for human val(8)-calcitonin from 16 oligonucleotides has been reported [16].

Here we describe the construction of a gene for salmon calcitonin (type I) with an additional glycine at position 33 by chemical synthesis of two oligonucleotides, a 109-mer and a 117 - mer, and subsequent hybridization. To our knowledge this is the first time that a gene of such a size has been constructed from only two long oligonucleotides[17]. The influence of pore size of the CPG support, the scale of synthesis and the synthesis cycle on the yields of the two oligonucleotides and their quality as evaluated by the mutation rate in the cloning of the genes is discussed. Synthesis under optimized conditions and cloning of gene fragments from even longer oligonucleotides gave very encouraging results. As a new approach this assembly of genes or gene fragments by mere synthesis and hybridization of two DNA strands cuts down the time consuming task of purifying several shorter oligonucleotides and avoids any enzymatic processing used hitherto.

RESULTS AND DISCUSSION

a) Gene design

Calcitonin is a hypocalcemic and hypophosphatemic peptide hormone involved in the regulation of the serum calcium level [18]. Among the commercially available calcitonins (human, porcine, salmon, eel) fish calcitonins are most active. In mammals calcitonin is produced in the C-cells of

```
SphI          1                                          10
          Met Cys Ser Asn Leu Ser Thr Cys Val Leu Gly

5'      C ATG TGC TCT AAC CTG TCG ACT TGC GTT CTT GGT
3' GT ACG TAC ACG AGA TTG GAC AGC TGA ACG CAA GAA CCA

                                          20
  Lys Leu Ser Gln Glu Leu His Lys Leu Gln Thr Tyr Pro

  AAG CTT TCT CAG GAA CTT CAT AAA CTG CAG ACC TAT CCG
  TTC GAA AGA GTC CTT GAA GTA TTT GAC GTC TGG ATA GGC

                              30                      Eco RI
  Arg Thr Asn Thr Gly Ser Gly Thr Pro Gly Stp Stp

  CGC ACT AAT ACC GGC TCT GGT ACC CCT GGT TAA TAG       3'
  GCG TGA TTA TGG CCG AGA CCA TGG GGA CCA ATT ATC TTA A 5'
```

FIG. 1. Amino acid sequence of salmon calcitonin-gly(33) and nucleotide sequence of the synthetic gene.

the thyroid by posttranslational processing of the precursor at the combined cleavage and amidation signal gly-lys-lys-arg [19] resulting in a carboxyterminal prolinamide. Therefore we designed a gene encoding the 32 amino acids of salmon calcitonin plus an additional codon for gly at the 3'-end. Calcitonin-like activity is expected for salmon calcitonin gly (33), which on the other hand can be converted enzymatically into the amidated form [20].

The total gene for salmon calcitonin-gly (33) was designed from only two long oligonucleotides representing the coding (109-mer, oligonucleotide I) and noncoding (117-mer, oligonucleotide II) strands. Furthermore codons for amino acids of highly expressed proteins in E.coli were selected and regions of self-complementarity on the mRNA level were avoided. A met codon preceded by a SphI site was added at the 5'-end of the gene, two translation stop codons followed by an EcoRI site were placed at the 3'-end, and several restriction endonuclease sites (e.g. SalI, HindIII, PstI, HpaII, KpnI, ScrFI) were distributed throughout the gene to facilitate subsequent modifications.

b) Syntheses of the oligonucleotides

A parameter strongly influencing the efficiency of the solid phase synthesis of oligonucleotides is the accessibility of the surface bound starting nucleoside and of the 5'-end of the growing chain. It can be varied by the pore size and the extent of loading. In the case of the CPG supports commonly used so far, which have a pore size of 500 Å and an attachment of DMT-nucleosides in the range of 20-40 µmol/g, a distinct decrease in the coupling yields is observed when the length of the oligonucleotides exceeds 70-80 nucleotide units. This could be partially caused by mutual steric hindrance of the growing chains as well as by hydrophobic interaction between the chains. Thus, supports with larger pore sizes or lower loading have to be taken into account for the synthesis of long oligonucleotides.

The oligonucleotides I and II were synthesized twice by the phosphoramidite method with an automated DNA synthesizer using 5'-O-DMT-2'-deoxyribonucleoside-3'-O-(2-cyanoethyl)-N, N-diisopropylphosphoramidites [3]. In the first synthesis oligonucleotides I and II were prepared with a 500 Å CPG support having an attachment of 25-35 µmol of nucleoside/g and a surface of about 50 m^2/g. An optimized synthesis cycle comprising a tenfold excess of phosphoramidites was used for a scale of 1 µmol nucleoside bound to the support. In order to avoid guanine modification in the O(6)-position [21] a hydrolysis step was inserted between the capping and the oxidation [22]. Main steps of the synthesis cycle are described in TABLE 1.

A second synthesis of the oligonucleotides I and II was performed using a 1000 Å CPG support with a loading of 20-30 µmol of nucleoside/g and a surface of approximately 25 m^2/g. The excess of phosphoramidites was 25-fold on a synthesis scale of 0.2 µmol (TABLE 2.).

In the second synthesis the repetitive yields of coupling as evaluated by DMT cation assay were about 1 % higher than in the first synthesis, the yields of products

TABLE 1. Synthesis cycle used in the first synthesis (1 μmol scale)

Operation	Reagent	Time (s)
Detritylation	3 % CCl_3COOH / CH_2Cl_2 (w/v)	2 x 60
Condensation	0.1 M fully protected 2'-deoxy-ribonucleoside-N,N-diisopropyl-phosphoramidites in dry CH_3CN, 0.5 M tetrazole in dry CH_3CN (the reagents were used in a 1:1 ratio)	120
Capping	Ac_2O / 2,6-lutidine / THF (1:1:8 v/v/v), 6.5 % DMAP / THF (w/v) (the reagents are used in a 1:1 ratio)	30
Hydrolysis	H_2O / 2,6-lutidine / THF (1:10:40 v/v/v)	30
Oxidation	0.1 M I_2 in H_2O / 2,6-lutidine / THF (1:10:40 v/v/v)	30

TABLE 2. Synthesis cycle used in the second synthesis (0.2 μmol scale; same reagents as described in TABLE 1.)

Operation		Time (s)
Detritylation		5 x 10
Condensation:	base plus tetrazole to column	3
	tetrazole to column	2
	base plus tetrazole to column	2
	tetrazole to column	2
	wait	18
Capping		17
Oxidation		30

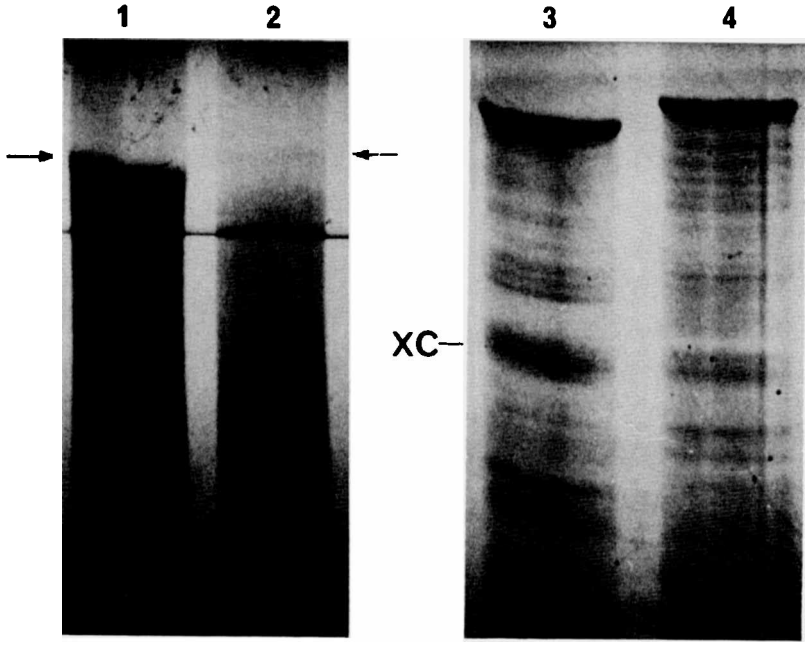

FIG. 2. Preparative 10 % polyacrylamide / 7M urea gels of crude oligonucleotides I and II.
Lane 1: oligomer I, 500 Å support;
lane 2: oligomer II, 500 Å support;
lane 3: oligomer I, 1000 Å support;
lane 4: oligomer II, 1000 Å support;
bands corresponding to oligonucleotides I and II are indicated by arrows; bands were visualized by UV-shadowing; XC = xylene cyanol FF.

purified from 40 A_{260} units of crude material were more than two-fold in the case of oligonucleotide I and nearly nine-fold for oligonucleotide II (TABLE 3.).

These findings correlate well with the shape of the oligonucleotide bands in the preparative polyacrylamide gel electrophoresis (PAGE), which are rather fuzzy in the case of the first synthesis (FIG. 2.). The same is true for the analytical PAGE, where considerably sharper bands were obtained for the oligonucleotides I and II along with a marked reduction of failure sequences in the second synthesis (FIG. 3).

TABLE 3. Yields from the syntheses of oligonucleotide I (109-mer) and oligonucleotide II (117-mer) on CPG supports with pore sizes of 500 and 1000 Å resp.

	1 st synthesis 500 Å support		2 nd synthesis 1000 Å support	
	109-mer	117-mer	109-mer	117-mer
Repetitive yield of detritylation (%) (a)	98.5	98.3	99.4	99.1
Yield of crude product (A_{260} units)	543 (b)	651 (b)	76.3 (c)	77.1 (c)
Yield of purified product (A_{260} units) (d)	1.08	0.34	2.35	2.95

(a) determined by the UV absorption of the DMT cation at 498 nm
(b) synthesis scale 1 µmol
(c) synthesis scale 0.2 µmol
(d) 40 A_{260} units were purified on a 3 mm 10 % polyacrylamidegel

c) Hybridization

The two pairs of the oligonucleotides I and II from the different syntheses were hybridized to yield the salmon calcitonin-gly(33) gene which was analyzed (FIG. 4) and purified by native PAGE. Again, clearly sharper bands were observed for the duplex resulting from the second synthesis.

d) Cloning and sequencing of the calcitonin genes

Both calcitonin genes were ligated into the EcoRI /SphI sites of M13mp18am [23], cloned in E.coli K12 JM101 and sequenced by the Sanger dideoxy method [24]. Only one out of six analyzed clones contained the correct gene sequence when DNA from the first synthesis with the 500 Å CPG support was

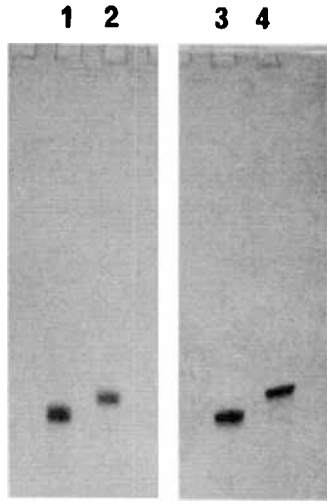

FIG. 3.

Analytical 10 % polyacrylamide/ 7M urea gel of purified oligonucleotides I and II.
Lane 1: oligomer I, 500 Å support;
lane 2: oligomer II, 500 Å support;
lane 3: oligomer I, 1000 Å support;
lane 4: oligomer II, 1000 Å support;
bands were visualized by silver staining.

used. Five point mutations and thirteen insertions were found in total corresponding to a mutation rate of 18/1356 or 1.3 % per nucleotide. With the exception of codons 8 (val) and 21 (thr) where altogether six insertions were centered, the mutations were distributed rather uniformly throughout the gene. Eight clones containing the gene from the second synthesis with the 1000 Å CPG support were sequenced. Only one had a deletion (T) in codon 16 (leu) of the gene sequence. Thus the mutation rate was reduced to 1/1808 or 0.055 % per nucleotide. The reason for the extent and preponderance of insertion mutations in the gene from the first synthesis is not quite clear. One could speculate that the longer condensation times in the first synthesis (120 s) compared to the second synthesis (about 30 s) might

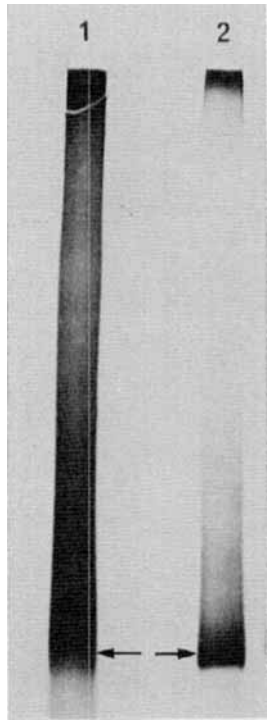

FIG. 4.

Analytical native PAGE of crude salmon calcitonin-gly(33) genes.
Lane 1: gene from the first synthesis, 500 Å support;
lane 2: gene from the second synthesis, 1000 Å support;
arrows indicate the products resulting from the hybridization of oligonucleotides I and II ; bands were visualized by silver staining.

allow partial double coupling reactions due to the minute instability of the DMT group of the purine nucleotide phosphoramidites under these conditions. Recently a frequency of variants of 0.3 % per nucleotide was reported for 63- to 87-mer oligonucleotides prepared with methyl-protected phosphoramidites [25].

TABLE 4. Cloning efficiency of oligonucleotides synthesized on 1000 Å CPG support; for synthesis cycle see TABLE 2.

Length of oligonucleotides in the DNA duplex	Number of sequenced clones	Number of correct clones
109 / 117	8	7
122 / 122	4	2
124 / 124	4	3
125 / 133	10	10
135 / 135	2	1
137 / 145	2	2
140 / 140	2	1
152 / 152	4	2

e) Conclusions

The results of our second synthesis clearly demonstrate the advantage of the 1000 Å CPG support compared to the 500 Å material under the conditions applied. We have used the 0.2 µmol scale cycle (TABLE 2.) also with the 500 Å support, but could not synthesize oligonucleotides longer than approximately 90 nucleotides with a performance comparable to the 1000 Å support. On the other hand, we successfully applied the 1000 Å support to the preparation of several longer oligonucleotides (up to 152-mers) which after hybridization were subcloned and then combined to a larger gene of about 500 base pairs. TABLE 4 gives a survey of our results in the synthesis of genes and gene fragments by use of long oligonucleotides regarding their efficiency in cloning. As can be seen, in every case a correct insert DNA could be found by sequencing a limited number of clones. In praxi, sequencing of four to eight clones at a time can be performed with ease, rendering this long oligonucleotide approach an extremely fast method valuable in the construction of genes or gene fragments.

EXPERIMENTAL

a) Materials and enzymes

5'-O-,N-protected deoxyribonucleoside-3'-O-(2 -cyanoethyl)-N,N- diisopropylphosphoramidites (purity judged by ^{31}P-NMR > 98 %) and dry acetonitrile (water content determined by Karl Fischer titration < 50 ppm) were obtained from Merck, FRG. CPG supports were from Applied Biosystems Inc. Tetrazole and 4-N,N-dimethylaminopyridin (DMAP) were purified by sublimation, tetrazole was dried at 60°C in vacuum for several hours prior to use. Tetrahydrofurane was dried by refluxing with calcium hydride. 2,6-Lutidine was purified by vacuum rectification. E.coli K12 JM101, a ^{35}S DNA sequencing pack and T4 polynucleotide kinase were from New England Biolabs. T4 DNA ligase and restriction enzymes (SphI, EcoRI) were purchased from Boehringer-Mannheim, [α-^{35}S]dATP (650 Ci/mmol) from Amersham.

b) Preparation and purification of oligonucleotides

Oligonucleotides were synthesized with an Applied Biosystems 380 B DNA synthesizer using 2-cyanoethyl-N,N-diisopropyl- phosphoramidites with synthesis cycles shown in TABLES 1. and 2. The oligonucleotides were cleaved from the support automatically and deprotected with concentrated ammonia overnight at 50°C. 40 A_{260} units of the crude oligonucleotide mixtures were loaded onto preparative 10 % polyacrylamide / 7M urea (2 % methylenbisacrylamide) gels (0.3 x 20 x 38 cm, three 4 cm wide slots per gel). The bands were visualized by UV shadowing, gel slices were cut out and crushed with a Dounce homogenizer. DNA was eluted from the gel with 16 ml of 0.2 M triethylammoniumbicarbonate (TEAB) buffer, pH 7.5 at room temperature for 1-2 hours, concentrated and desalted on Sephadex G-50 columns using 20mM TEAB as an eluent. The purity of the oligonucleotides

was checked by applying 0.015 A_{260} units to a 1 mm thick 10 % polyacrylamide/7M urea (2 % methylenbisacrylamide) gel for electrophoresis. The bands were visualized by silver staining [26].

c) 5'-Phosphorylation and hybridization of oligonucleotides I and II

0.5 nmol of oligonucleotides I and II (0.567 and 0.659 A_{260} units resp.) were dissolved in 7.5 µl of Hepes buffer pH 7.6 (20 mM 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT)) at 95°C and cooled on ice. 4.5 µl of 2.5 mM ATP (pH 7.0) and 1 µl of 0.1 M DTT were added and the mixtures were equilibrated at 37°C. After addition of 2 µl (4U) of T4 polynucleotide kinase the mixtures were incubated for 4 hours at 37°C. The kinase was subsequently inactivated by heating at 95°C for 2 min. The reaction mixtures were combined using a rotary Speed Vac and redissolved in 200 µl of 50 mM Tris-HCl pH 7.6 (25 mM $MgCl_2$, 25 mM DTT, 62.5 mM KCl). Oligonucleotides I and II were then hybridized by heating at 95°C for 5 min and cooling to room temperature within 2 hours. The resulting duplex was purified by preparative native PAGE. The concentrated hybridization mixture was loaded into 4 slots in a 10 % native polyacrylamide gel (2 % methylenbisacrylamide, 0.2 x 13.5 x 25 cm) which was electrophoresed for 4000 Vh . After staining with ethidium bromide gel slices containing the gene were excised, crushed and extracted twice with 16 ml of TEAB each for 2 hours at room temperature. The yields of the genes, finally desalted on a Sephadex G-50 column were about 0.25 A_{260} units.

d) DNA sequence analysis

M13mp18am RF (1 µg), digested with SphI and EcoRI, was purified on a 0.8 % agarose gel, isolated with an ion exchange column (ELUTIP-d, Schleicher & Schuell) according

to the protocol of the supplier, and ethanol precipitated. 0.005 A_{260} units of the salmon calcitonin-gly(33) gene and the vector fragment were ligated at 16°C overnight. Transformation of E.coli K12 JM101 , preparation of single stranded DNA from the phages [27] and the dideoxy sequencing reactions [24] as modified for S-labelling [28] were performed as described.

ACKNOWLEDGEMENTS

The authors would like to thank S. Hein and L. Hornung for skillful technical assistance and P.Eckes and W. Wetekam for contributing sequencing data. We also are grateful to B. Baumann for her help in preparing the manuscript.

REFERENCES

1. Matteucci, M.D.; Caruthers, M.H. J. Amer. Chem. Soc. 1981, 103, 3185
2. Beaucage, S.L.; Caruthers, M.H. Tetrahedron Lett. 1981, 22, 1859
3. Sinha, N.D.; Biernat, J.; McManus, J.; Köster, H. Nucleic Acids Res. 1984, 12, 4539
4. Adams, S.P.; Kavka, K.S.; Wykes, E.J.; Holder, S.B.; Galluppi, G.R. J. Amer. Chem. Soc. 1983, 105, 661
5. Köster, H.; Stumpe, A.; Wolter, A. Tetrahedron Lett. 1983, 24, 747
6. Urdea, M.S.; Ku, L.; Horn, T.; Gee, Y.G.; Warner B.D. Nucleic Acids Res., Symp. Ser. No. 16, 1985, 257
7. Warner, B.D.; Warner, M.E.; Karns, G.A.; Ku, L.; Brown-Shimer, S.; Urdea, M.S. DNA 1984, 5, 401
8. Efcavitch, J.W. et al., 2nd International Symposium On Phosphorus Chemistry Directed Towards Biology, Lodz (Poland), September 8-13, 1986, Poster
9. Miyoshi, K.; Hasegawa, A.; Tomiyama, M.; Miyake, T. Nucleic Acids Res., Symp. Ser. No 10, 1981, 197
10. Theriault, N.Y.; Tomich, C.C; Wierenga, W. Nucleosides Nucleotides 1986, 5, 15
11. Urdea, M.S.; Merrywaether, J.P.; Mullenbach, G.T.; Coit, P.; Heberlein, U.; Valenzuela, P.; Barr, P.J. Proc. Natl. Acad. Sci. USA 1983, 80, 7461
12. Cravador, A.; Jacobs, P.; Van Elsen, A.; Lacroix, C.; Colau, B.; Van Alphen, P.; Herzog, A.; Bollen, A. Biochimie 1985, 67, 829

13. Kempe, T.; Chow, F.; Peterson, S.M.; Baker, P.; Hays, W. Opperman, G.; L'Italien, J.J.; Long, G.; Paulson, B. Bio/Technology 1986, 4, 565
14. Engels, J.; Glauder, J.; Müllner, H.; Uhlmann, E.; Wetekam, W.; Hashimoto-Gotoh, T.; Scheikl-Lenz, B. Nucleosides Nucleotides in press
15. Khorana, H.G. Science 1979, 203, 614
16. Ivanov, I.; Gigova, L.; Jay, E. FEBS Lett. 1987, 210, 56
17. Part of this work was presented at the VIIth International Roundtable Nucleosides, Nucleotides and Their Biological Applications, Konstanz (FRG), Sept.29 - Oct.3, 1986
18. Capp, D.H.; Cameron, E.C.; Cheney, B.A.; Davidson, A.G.F.; Henze, K.G. Endocrinology 1962, 70, 638
19. Birnbaum, R.S.; Mahoney, W.; Roos, B.A. J. Biol. Chem. 1983, 258, 5463
20. Bradbury, A.F.; Finnie, M.D.A.; Smyth, D.G. Nature 1982, 298, 686
21. Pon, R.T.; Dahma, M.J.; Ogilvie, K.K. Nucleic Acids Res. 1985, 13, 6447
22. Caruthers, M.H., personal communication
23. Patschinsky, T.; Jansen, H.W.; Blöcker, H.; Frank, R.; Bister, K. J. Virology 1986, 59, 341
24. Sanger, F.; Nicklen, S.; Coulson, A.R. Proc. Natl. Acad. Sci. U.S.A. 1977, 74, 5463
25. McClain, W.H.; Foss, K.; Mittelstadt, K.L.; Schneider, J. Nucleic Acids Res. 1986, 14, 6770
26. Goldman, D.; Merril, C.R. Electrophoresis 1982, 8, 24
27. Zoller, M.J.; Smith, M. DNA 1984, 3, 479
28. Biggin, M.D.; Gibson, T.J.; Hong, G.F. Proc. Natl. Acad. Sci. USA 1983, 80, 3963

Received September 1, 1987.