Chemical synthesis and cloning of a gene for human β -urogastrone

John Smith, E.Cook, Ian Fotheringham, Sally Pheby, Roger Derbyshire, Michael A.W.Eaton¹, Mike Doel¹, David M.J.Lilley², John F.Pardon, Thakor Patel¹, Hilary Lewis³ and Leslie D.Bell

Searle Research and Development, Lane End Road, High Wycombe, Bucks HP12 4HL, UK

Received 14 June 1982; Accepted 6 July 1982

ABSTRACT

A DNA duplex coding for the 53 amino acids of human β -urogastrone has been synthesised. Computer assisted design of the gene included restriction endonuclease sites for plasmid insertion, a termination codon and two triplets coding for lysine at the 5'-end of the structural gene.

The synthesis involved preparation of 23 oligodeoxyribonucleotides by phosphotriester procedures coupled to rapid HPLC techniques. The gene was constructed in two halves by enzymatic ligation of the oligonucleotides and cloned into a specially constructed chimeric plasmid vector. *Escherichia coli* K12 MRC8 was transformed by the plasmid and clones containing the full gene sequence were isolated and characterised.

INTRODUCTION

Urogastrone is a polypeptide hormone found predominantly in the duodenum and in the salivary glands of normal humans (1). It is a potent inhibitor of gastric acid secretion (2) and promotes epithelial cell proliferation both in vitro (3) and in vivo.

The anti-secretory activity of urogastrone was first observed over 40 years ago (4) but it is only in the last six years that two forms of urogastrone, with similar biological activities, have been isolated in a sufficiently pure form for detailed structural studies (5). β -Urogastrone contains a single polypeptide chain of 53 amino acids, while γ -urogastrone contains the same amino acid sequence 1-52 but lacks the carboxyl terminal arginine of the β -form (6). Similar biological activities have been found for mouse epidermal growth factor (EGF) and determination of the amino sequences has shown a high degree of homology between the two molecules (7).

The antisecretory and cell-proliferative activities of β -urogastrone may prove useful in the clinical treatment of

gastric and duodenal ulcers by reducing the production of acid in the stomach and promoting the healing of ulcerated tissue. Until now, β -urogastrone has been isolated from human urine and purified by conventional techniques in yields of only about 1mg per 1000 litres (5) and consequently this approach is impractical for the production of sufficient urogastrone to allow significant clinical studies. We have therefore investigated the application of recombinant DNA techniques to the production of human β -urogastrone.

The cloning of urogastrone cDNA made from natural mRNA faced the problems of the poor availability of fresh human duodenum and the high level of nucleases present in such tissue. This led us to consider the possibility of assembling and cloning a synthetic gene since recent improvements in the synthesis of oligodeoxynucleotides by the phosphotriester procedure (8) have allowed us to rapidly produce oligomers of single stranded DNA.

Recently, this strategy has been applied to the synthesis of genes encoding a precursor of aspartame (9), human somatostatin (19), human insulin A and B chains (11) and human leukocyte interferon (12). We now report the synthesis and cloning of a gene coding for human β -urogastrone.

MATERIALS

Analytical grade reagents were used for the chemical synthesis of the oligodeoxynucleotide blocks. 4,4'-Dimethoxytrityl,N-acyl protected nucleosides were obtained from Lancaster Synthesis and further purified by silica gel chromatography. Fully protected nucleotide triesters were prepared essentially by the method of Gough *et al* (13). 2,4,6-Trisopropylbenzenesulphonyl tetrazole (TPS-tet) was prepared as described (14). HPLC purification was performed on a Dupont 830 Liquid Chromatograph. ODS modified silica (for the preparative purification of intermediates) was prepared as described (15). T4 polynucleotide kinase from New England Biolabs and T4 DNA ligase from Bethesda Research Laboratories Inc., were routinely tested for the presence of contaminating nucleases. Restriction enzymes were obtained from New England Biolabs and were used as

directed by the manufacturer. DNA polymerase I was purchased from Boehringer Mannheim.

METHODS

Design of the Gene

The molecule was designed to be expressed initially as a fusion product with a portion of the E.coli Tryptophan E gene, in a previously constructed expression vehicle (16). Linkage of the synthetic gene to the Trp E gene was via a HindIII restriction enzyme site (Figure 1). A Lys-Lys sequence was incorporated immediately preceding the first β -urogastrone amino acid, Asn¹, to allow cleavage by the action of trypsin. Since β -urogastrone is known to be stable to the action of trypsin (6) the mature peptide can be produced from the fused product. An alternative method, that of chemical cleavage adjacent to methionine residues by the action of cyanogen bromide (11) cannot be used here due to the presence of an internal labile methionine residue (6).

A favoured nucleotide sequence for the synthetic urogastrone gene was selected from the many alternatives allowed by the redundancy in the genetic code by choosing the preferred codons in E.coli (17). A single TAA termination codon was used following the coding sequence. The 'Trp E' HindIII and terminal BamHI restriction endonuclease sites allowed insertion of the gene in the desired orientation into a non-expressive pBR322 derived plasmid. To allow the gene to be cloned in two halves and to be specifically dissected, a number of other restriction sites were strategically placed along the synthetic gene sequence. The site of greatest importance is the EcoRI restriction endonuclease site positioned at Asn¹ which allows for expression of mature urogastrone in an appropriate expression vector (unpublished results).

The resultant double stranded molecule was scanned by a computer programme designed to detect repeated and self-complementary sequences. Where possible, these were removed within the constraints of the genetic code to maximise correct ligation of the oligonucleotides. The two strands were divided into 23 oligodeoxynucleotides of between 12 and 20 nucleotides in length, for convenience of chemical synthesis, purification and ligation,

FIGURE 1. The synthetic β -urogastrone gene showing the nucleotide and amino acid sequence, the individual oligonucleotides and the 6 base pair restriction enzyme sites.

ensuring in each case an overlap of at least four nucleotides (Figure 1).

Oligonucleotides 1 to 11, up to the central XbaI restriction endonuclease site are referred to as the 'left half' of the gene and oligonucleotides 12 to 23 up to the BamHI restriction endonuclease site are referred to as the 'right half' of the gene.

Chemical Synthesis Of Oligodeoxynucleotides

Oligodeoxynucleotides 1 to 23 (Figure 1) were synthesised by the phosphotriester procedure coupled to rapid HPLC separation of the intermediates.

Starting materials for the syntheses were the 5'-O-(di-p-methoxytrityl)-3'-p-chlorophenyl β -cyanoethyl phosphate derivatives of N⁶-benzoyladenine, N⁴-benzoylcytosine, N²-isobutyryl-guanosine and thymidine. The 3'-hydroxyl end of each oligonucleotide was protected with a benzoyl group. To proceed to condensation reactions 5'-dimethoxytrityl protecting

groups were removed with 2% benzene sulphonic acid in chloroform/methanol (7:3 v/v) for 5 minutes at 0°C and cyanoethyl protecting groups were removed by treatment with pyridine/triethylamine (1:1 v/v) for 2 hours at 40°C.

Typically 80mg of the 5'-hydroxyl component was condensed with a 1.5 times molar excess of the 3'-phosphodiester component in anhydrous pyridine in the presence of a 3 times molar excess of 2,4,6-tri-isopropylbenzene sulphonyltetrazolide. After 1 hour or when the 5'-hydroxyl component was used up (as judged by silica tlc eluted with 10% methanol in chloroform) the reactions were quenched by the addition of 5% (w/v) aqueous sodium bicarbonate and extracted with chloroform. The chloroform extracts were dried with sodium sulphate and chromatographed on 15-25µm ODS silica using a solvent gradient from chloroform/methanol/water (2:6:3 v/v) to chloroform/methanol/water (2:6:0.5 v/v). Complete separation was obtained in under two hours and the fractions containing product were extracted into chloroform, dried and evaporated under vacuum to give an off white solid. Yields typically varied between 50% and 80%. This process was repeated as necessary until at the completion of the synthesis all the protecting groups were removed by sequential treatment with 0.5M N¹,N¹,N²,N²-tetramethyl-guanidinium 4-nitrobenzaldoximate in dioxan/water (1:1 v/v) at room temperature for 60 hours, concentrated ammonia at 70°C for 4 hours and 80% aqueous acetic acid at room temperature for 30 minutes. After evaporation, an aqueous solution of the oligonucleotide was extracted with diethyl ether and filtered. The product was purified by chromatography on a Partisil 10 SAX HPLC column, using a solvent gradient from 1M to 4M triethylammonium acetate pH4.9 at 50°C.

Sequence analysis was carried out using the mobility shift method of Wu *et al* (18).

Construction of the β -urogastrone gene

Aliquots of the chemically synthesised oligonucleotides 1-23 (up to 7µg of each) were phosphorylated with 4 units of T4 induced polynucleotide kinase in 20µl of a solution containing 50mM Tris-HCl (pH 7.8), 10mM magnesium chloride, 0.25mM [γ -³²P] ATP (14Ci/mmol) and 10mM mercaptoethanol for 15 minutes at 37°C.

The enzyme was inactivated by incubation at 100°C for 5 minutes.

Ligation of the phosphorylated oligonucleotides was carried out as shown in Figure 2 to give firstly fragments A-G and ultimately the two halves of the gene for cloning.

Reactions contained an equimolar quantity of phosphorylated oligonucleotides (300 p mole) except for oligonucleotides 2,3,7,8 and 10 which were used in a limiting quantity (0.5:1) for number 2 and 0.75:1 for the others). Ligations were carried out with 6 units of T4 DNA ligase in a solution containing 50mM Tris-HCl pH 7.8, 10mM magnesium chloride, 1mM APT, 20mM dithiothreitol (standard conditions) for 3 to 16 hours at 25°C. The ligated products of oligonucleotides 12-23 were identified by autoradiography following electrophoretic separation on a 20% polyacrylamide gel run in 90mM Tris-HCl pH 8.3, 90mM boric acid and 2.5mM EDTA (Figure 3).

The correctly assembled products were recovered from gel slices by electroelution onto DE52 DEAE cellulose which was then washed with a solution of 100mM ammonium acetate, 2mM magnesium acetate, 0.02% w/v SDS and 20mM EDTA and the DNA eluted by 2ml

FIGURE 2. Ligation strategy for the synthetic β -urogastrone gene.

FIGURE 3. Autoradiograph of a 20% polyacrylamide gel showing the assembled fragments D,E,F,G. Lane D shows 12+13 dimerised about the XbaI site. Lane E shows 14+15+17. Lane F shows 16+18+19. Lane G shows 20+22+23 dimerised about the BamHI site.

of 1.1M sodium chloride in the same buffer. The solution was diluted 3 times with water and the DNA was recovered by ethanol precipitation. The recovery was estimated by counting a small aliquot on a GFC filter disc.

Ligated 10b.p long HindIII linkers were run in parallel on the gels to provide accurate size markets in the range 10 to

80 base pairs.

Construction Of The Chimeric Cloning Vector pLFI

As shown in figure 4 plasmids pUB110 (5µg) and pAT153 (4µg) were each digested with BamHI and EcoRI restriction endonucleases. Reactions were terminated by heating at 65°C for 3 minutes and the fragments separated by electrophoresis on a 1% agarose gel. The 3280 b.p large fragment of pAT153 and the approximately 840 b.p small fragment of pUB110 were detected with ethidium bromide and recovered by electroelution as described above. The two fragments were ligated with 10 units of T4 DNA ligase for 18 hours at 15°C. After ethanol precipitation the ligated DNA was dissolved in 100µl of a solution containing 10mM Tris-HCl pH7.5, 10mM magnesium chloride and 10mM calcium chloride, and transformed into E.coli K12, strain MRC8 (F⁻,dap 103, hsd R, met B1, glm 533, upp 1, dap 101, sup E, rec A) (19). The bacteria were plated on L-agar plates and selected for resistance to ampicillin (100µg/ml). Transformants were analysed by EcoRI, BamHI and XbaI restriction endonuclease cleavage of plasmid DNA (20), and one full length clone designated pLFI was selected for use as the cloning vector.

Cloning Of The Synthetic Gene

The two halves of the gene were cloned independently as shown in Figure 5.

pLFI DNA (5µg) was digested with XbaI and BamHI and the two fragments separated by electrophoresis on a 1% agarose gel. The larger fragment was recovered as described above and ligated at 15°C for 18 hours to a 10 fold excess of the assembled 'right half' of the gene, fragment P (Figure 1). The latter was generated by XbaI and BamHI double digestion of concatameric fragments N.

Aliquots were used to transform E.coli K12 MRC8 as described above and transformants resistant to (100µg/ml) ampicillin isolated and characterised.

DNA from 4 clones was digested with XbaI and BamHI and compared with the digestion of pLFI on a gel with appropriate size markers (Figure 6). The expected 88 b.p urogastrone fragment was generated. Similarly digestion with HaeII restriction endonuclease yielded the predicted 82 b.p fragment.

FIGURE 4. Construction of hybrid cloning vector.

The 2.5 Kb HaeII fragment of pLFI containing the pUB110 fragment was cleaved to 2.17 Kb. One of these 4 clones designated pURI was shown by sequence analysis (21) to contain the correct insert.

pURI DNA (2μg) was digested with 10 units of EcoRI for 3 hours at 37°C. The reaction was terminated by heating at 65°C for 5 minutes, phenol/chloroform (1:1 v/v) extracted and the DNA recovered by ethanol precipitation. The linearised DNA was treated with 5 units of DNA polymerase I in a solution containing 0.25mM of each of the four deoxynucleoside triphosphates, 50mM Tris-HCl pH 7.5, 10mM MgCl₂ and 1mM β-mercaptoethanol for 20 minutes at 4°C. The reaction was terminated as above and the recovered DNA ligated to a 10 fold excess of HindIII linker using

FIGURE 5. Cloning of synthetic β -urogastrone gene.

6 units of T4 DNA ligase for 18 hours at 15°C.

Following termination of the reaction, the recovered DNA was digested with 5 units of HindIII for 1 hour and the DNA electrophoresed on a 1% Agarose gel. The linearised plasmid DNA band was recovered as described above, self ligated using 6 units of T4 DNA ligase for 18 hours at 15°C, and aliquots used to transform MRC8 as above. A transformant containing a single HindIII site inserted at the EcoRI site of the plasmid was isolated, designated pUR2 and used for cloning of the left half of the urogastrone gene.

pUR2 DNA (5 μ g) was digested to completion with HindIII and XbaI and the larger fragment purified by electrophoresis as above. This fragment was ligated to a 2 fold molar excess of the left hand half of the synthetic gene with 6 units of T4 DNA

FIGURE 6. 5% Polyacrylamide gel showing the digestion of pURI DNA and presence of the cloned right half of the β -urogastrone gene. Lanes A-E show $XbaI + BamHI$ cleavage of pLFI, clone RH1, clone RH2 (pURI), clone RH3 and clone RH4 respectively. Lane F shows OX174 RF DNA digested with HaeIII as a marker. Lanes G to K show HaeII cleavage of clone RH1, clone RH2 (pURI), clone RH3, clone RH4 and pLFI respectively.

ligase for 18 hours at 15°C. The reaction was terminated by heating at 65°C for 3 minutes and aliquots used to transform *E.coli* K12 MRC8, selecting for ampicillin resistance (100 μ g/ml).

One clone designated pUR3 was characterised by restriction analysis (Figure 7) and complete nucleotide sequencing of the urogastrone gene.

FIGURE 7 5% Polyacrylamide gel showing digestion of pUR3 and the presence of the complete synthetic β -urogastrone gene. Lanes A and G show OX174 RF DNA cleavage with HaeIII. Lane B shows HindIII + BamHI cleavage of pUR3 giving the complete 174 b.p urogastrone gene. Lane C shows HinF1 cleavage of pUR3 liberating expected fragments of 395 and 1040 b.p as compared to the 1631 b.p fragment of pAT153 in lane D. The expected 398 b.p fragment from digestion of pUR3 with AccI can be seen in Lane E, which is not present in an AccI digestion of pAT153, track F.

RESULTS AND DISCUSSION

Developments in the chemical synthesis of oligodeoxyribonucleotides by the triester procedure (8) now ensure that this method of gene production, in the case of relatively short genes

coding for a protein whose amino acid sequence is known, is a more viable alternative to the more laborious techniques required to isolate genetic material from natural sources. The construction of the 174 base pair gene coding for human β -urogastrone was undertaken in two halves so that one half could be cloned, amplified and analysed whilst synthesis of the other half was in progress.

The synthetic urogastrone gene was designed using, where possible, preferred *E.coli* codons (17) and incorporating specific restriction sites both terminally and internally. This allows two advantages over the natural gene for expression in *E.coli*. Firstly, it should be possible to circumvent translational limitations which might otherwise be imposed on the bacterial cell by the presence in the eukaryotic sequence of codons which are rarely used in prokaryotes. Secondly, insertion of the urogastrone gene into cloning and expression vectors both for production of hybrid polypeptides and mature urogastrone and subsequent sequence analysis of recombinants is greatly facilitated by careful positioning of unique restriction sites. The need to avoid repeated and palindromic sequences which may cause considerable problems during assembly of the gene was taken into account in the choice of codons and strategic restriction sites.

The use of 25 μ m ODS silica particles for the HPLC separation of chemical intermediates allowed considerable savings in time previously spent purifying these intermediates. Also, recent developments in solid phase methodologies (22) and the use of mesitylenesulphonyl-3-nitro-1,2,4-triazole (23) have allowed us to make further time savings whilst maintaining high yields at each condensation stage.

The assembly of each half of the gene was carried out by slightly different strategies and cloned using a hybrid pAT153/pUB110 cloning vehicle. The assembly of oligomers as dimers about the terminal restriction sites helped in identification of correct assembly intermediates. Purification by extraction of individual bands from polyacrylamide gels allowed removal of unreacted, degraded and wrongly ligated material. The right half of the gene was assembled using this strategy, whereas the left half was assembled without purification and the correct

product selected at the cloning stage. In order to facilitate the two stage cloning of the synthetic gene, a vector having unique HindIII, BamHI sites and a central XbaI site was required. The Staphylococcus aureus plasmid pUB110, which cannot be propagated in E.coli (24) contains a suitable XbaI site roughly centrally placed between sites for EcoRI and BamHI. A hybrid plasmid, designated pLFI was constructed by cleaving out this approximately 840 b.p. fragment from pAT153. The right half of the urogastrone gene was inserted into pLF1 and characterised by digestion with particular restriction endonucleases each of which gave the desired bands upon cleavage (Figure 6). The insert in one clone, designated pURI, was sequenced and shown to be correctly assembled. A HindIII cleavage site was then introduced at the EcoRI site (with a linker which allowed retention of the EcoRI site) and following ligation to the assembled synthetic left half of the gene the full gene clone, pUR3 was fully characterised by restriction endonuclease digestion (Figure 7) and nucleotide sequencing of the insert.

To determine whether the gene could be expressed in E.coli as a biologically active polypeptide, pUR3 was cleaved with HindIII and BamHI and the fragment containing the urogastrone gene inserted in a specially constructed expression plasmid based on the E.coli Trp promotor (16) cloned in a derivative of pAT153 (25).

The nucleotide sequence expected around the HindIII site of the HindIII - BamHI fragment predicted that a hybrid polypeptide could be expressed consisting (in order from the N-terminus) of 7 amino acids from TrpE, 7 linking amino acids and 53 urogastrone amino acids. Detailed biological studies (to be published elsewhere) have shown that a polypeptide can be expressed at high level in E.coli (~2-3mg/litre) which has the same biological properties as authentic human urogastrone by the following criteria:

- (i) The cloned polypeptide in crude E.coli extracts (but not in control extracts) successfully competes with authentic ^{125}I -labelled purified human β -urogastrone in a cell surface receptor binding assay (26).
- (ii) The E.coli extract stimulates [^3H] thymidine uptake by

human foreskin fibroblasts using a modification of the method of Hollenberg and Gregory (26).

- iii) The cloned material exhibits anti-secretory properties in an isolated rabbit gastric gland preparation stimulated by histamine (27).

ACKNOWLEDGEMENTS

We thank Dr. B.M. Richards and Dr. P. Fellner for many helpful and stimulating discussions, and Dr. A. Porter for his critical reading of the manuscript.

¹Present address: Celltech, 250, Bath Road, Slough, Bucks, UK

²Present address: Biochemistry Department, Medical Sciences Institute, University of Dundee, Dundee DD1 4HN, UK

³Present address: Department of Microbiology, University of Sheffield, Sheffield, UK

REFERENCES

1. Heitz, P.U., Kasper, M., Van Neoden, S. Pola, J.M., Gregory, H. and Pearse, A.G.E. (1978) *Gut* 19, 408-413.
2. Elder, J.B., Gauguli, P.C., Gillespie, I.E., Delamore, I. and Gregory, H. (1975) *Lancet*, 424-427.
3. Elder, J.B., Gauguli, P.C., Gillespie, I.E., Gerring, E.L. and Gregory, H. (1975) *Gut* 16, 887-893.
4. Gray, J.S., Wieczorowski, E. and Ivy, A.C. (1939) *Science* 89, 489-490.
5. Gregory, H. and Willshire, I.R. (1975) *Hoppe-Seylers Z. Physiol. Chem.* 356, 1765-1774.
6. Gregory, H. and Preston, B.M. (1977) *Int. J. Peptide Protein Res.* 9, 107-118.
7. Morley, D., Hollenberg, M.D. and Gregory, H. (1980) *Mol. Pharmacol.* 17, 314-320.
8. Hsuing, H.M., Brousseau, R., Michniewicz, J. and Narang, S.A. (1979) *Nucleic Acids Res.* 6, 1371-1385.
9. Doel, M.T., Eaton, M., Cook, E.A., Lewis, H., Patel, T. and Carey, N.H. (1980) *Nucleic Acids Res.* 8, 4575-4591.
10. Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heymeker, H.L., Bolivar, F. and Boyer, H.W. (1977) *Science* 198, 1056-1063.
11. Crea, R., Kraszewski, A., Hirose, T. and Itakura, K. (1978) *Proc. Natl. Acad. Sci. USA* 75, 5765-5769.
12. 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* 292, 756-762.
13. Gough, G.R., Singleton, C.K., Weith, H.L. and Gilham, P.T. (1979) *Nucleic Acids Res.* 6, 1557-1570.
14. Stawinski, Y., Hozurni, T., Narang, S.A., Bahl, C.P. and Wu, R. (1977) *Nucleic Acids Res.* 4, 353-372.
15. Little, C.J., Dale, A.D., Whatley, J.A. and Evans, M.B.

- (1979) J. Chromat. 171, 431-438.
16. Tacon, W., Carey, N. and Emtage, S. (1980) Molec. Gen. Genet. 177, 427-438.
 17. Grantham, R., Gautier, C. and Guoy, M. (1980) Nucleic Acids Res. 8, 1893-1912.
 18. Chen-Pei, D., Jay, E., Bahl, C.P. and Wu, R. (1976) Anal. Biochem. 74, 73-93.
 19. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. 69, 2110-2114.
 20. Katz, L., Kingsbury, D.K., Helinski, D.R. (1973) J. Bacteriology 114, 577-591.
 21. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
 22. Miyoshi, K., Arentzen, R., Huang, T. and Itakura, K. (1980) Nucleic Acids Res. 8, 5507-5517.
 23. Reese, C.B., Titmas, R.C. and Yau, L. (1978) Tetrahedron Lett. 2727-2730.
 24. Keggins, K.M., Lovett, P.S. and Duvall, E.J. (1978) Proc. Natl. Acad. Sci. USA 75, 1423-1427.
 25. Brown, K.D., Bennet, G.N., Lee, F., Schweingruker, M.E. and Yanofsky, C. (1978) J. Mol. Biol. 121, 153-177.
 26. Hollenberg, M.D. and Gregory, H. (1980) Mol. Pharmacol. 17, 314-320.
 27. Berglindh, T. and Obrink, K.J. (1976) Acta. Physiol. Scand. 96, 150-159.