Volume 15 Number 12 1987 Nucleic Acids Research

Solid-phase assembly of cow colostrum trypsin inhibitor gene

Zdeněk Hostomský, Jiří Smrt[1], Luboš Arnold[1], Zdeněk Točík and Václav Pačes

Institute of Molecular Genetics and [1]Institute of Organic Chemistry and Biochemistry, Czechoslovak Academy of Sciences, Flemingovo nám. 2, 166 37 Prague 6, Czechoslovakia

Received April 28, 1987; Accepted May 26, 1987 Accession no. Y00351

ABSTRACT

A gene for cow colostrum trypsin inhibitor (CTI) was constructed from synthetic oligonucleotides using a novel method of solid-phase gene assembly. In the first step an anchor oligonucleotide was covalently bound to the CNBr-activated Sephacryl S-500 support. Next, triads or tetrads of separately annealed oligonucleotides were stepwise hybridized to the immobilized complementary sequence, with washing after each step. In the last step a linearized vector molecule was ligated to the assembled gene. The whole construct was released from the solid support with a restriction enzyme, circularized, and used for transformation, with a high yield of recombinant clones being obtained. The method represents a generally applicable approach to rapid and efficient assembly of extended DNA duplexes.

INTRODUCTION

The cow colostrum trypsin inhibitor (CTI) is a protein of 67 amino acid residues of known sequence (1). It belongs to a family of well-characterized inhibitors of serine proteinases (2). CTI has a compact tertiary structure stabilized by three intrachain disulphide bridges. In order to study the predicted changes in inhibitory specificity due to replacements of particular amino acid residues in the reactive site of the CTI molecule, we decided to prepare the inhibitor via total gene synthesis. This enables to prepare sufficient amounts of not only the wild-type inhibitor polypeptide, but also of its defined variants. To construct the CTI gene from synthetic oligonucleotides a novel method of gene assembly was developed.

MATERIALS AND METHODS

Materials

Sephacryl S-500 was from Pharmacia. All enzymes were from

New England Biolabs or Boehringer Mannheim, except T4 polynucleotide kinase, which was a gift from Dr. Turňa.

Oligonucleotide synthesis

5'-O-dimethoxytrityl-protected deoxyribonucleosides were prepared as described: d-T (3), N-benzoyl-d-C (4), N-dimethylaminoacetamidin-d-A (5) and N-isobutyryl-d-G (6). Internucleotidic linkage was synthesized by a modified methylphosphochloridite method (7) or by the 2-cyanoethyl-N,N-diisopropylphosphoramidite method (8) mediated by methylimidazol trifluoromethanesulphonate (9). Chain assembly of the protected oligonucleotides was performed on an automatic DNA-synthesizer (Synen-1, Czechoslovak Academy of Sciences) using Fractosil-1000 with a long aliphatic chain as the solid phase (0.7 μmol/100 mg). After completion of the synthesis, the methoxy (or 2-cyanoethyl) and N-protecting groups were removed (10); ammonia was evaporated in the presence of 1 M potassium acetate (0.4 ml) to a volume of 1 ml. Ethanol (8 ml) was added and the mixture was kept at 0°C for 20 h. The pellet after centrifugation was dissolved in water (2.5 ml) at 80°C, centrifuged again and the supernatant (60-100 A_{260} units) was lyophilized, affording 2-4 mg of crude product. The oligonucleotides were purified by preparative electrophoresis in 16% polyacrylamide gel under denaturing conditions and phosphorylated using polynucleotide kinase and ATP.

Covalent coupling of anchor oligonucleotide to solid support

A synthetic 29-mer was coupled to Sephacryl S-500 via cyanogen bromide activation as described (11). The Sephacryl S-500 support with the DNA ligand was stored in 10 mM potassium phosphate buffer pH 8 at +4°C for one year without substantial loss of binding capacity.

Gene assembly

Purified and phosphorylated oligonucleotides (except oligonucleotide #1, which was not phosphorylated) were distributed into overlapping groups of three or four. Each triad or tetrad was brought to 90°C in 20 μl of 50 mM Tris-HCl, 10 mM $MgCl_2$, pH 8 (1xLIG) and allowed to cool to 30°C for at least 2 hours. The molarity of neighbouring oligonucleotides in each triad or tetrad increased by a factor of 1.5 in the direction of the growing chain, away from the solid support (cf Fig.1); e.g. there was 1

pmol of oligonucleotide #18, 1.5 pmol of #19 and 2.3 pmol of #17 in 20 µl of triad 1.

20 µl of wet sediment of Sephacryl S-500 with DNA ligand was washed 3 times with 1 ml of 1xLIG and incubated with annealed triad 1 (40 µl total incubation volume) at 37°C under gentle motion of the suspension for 2 hours. Then the supernatant was removed and the sediment was washed twice with 600 µl 1xLIG. In the next step the annealed triad 2 was added and the procedure was repeated. The same was carried out with triads 3, 4, 5 and tetrad 6. The incubation times varied according to convenience from 2 to 16 hours, with longer periods of time at the later elongation steps.

Gene cloning and sequencing

An aliquot (2 µl) of the wet sediment of Sephacryl S-500 with the completed gene duplex was incubated in 15 µl of 1xLIG for 45 min. at 37°C in the presence of ATP and polynucleotide kinase to phosphorylate the free 5'-end of the gene in solid phase. To the sediment, washed with 1xLIG, was added 30 ng of M13mp18 vector DNA (12) linearized with EcoRI and HindIII, ATP (1 mM final concentration), and T4 DNA ligase, and the mixture was incubated in 20 µl final volume of 1xLIG at room temperature for 16 hours. After the ligation the sediment was washed twice with 50 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM $MgCl_2$, and the resulting structure was released by adding 8 units of EcoRI to the 10 µl of the suspension and incubation for 2 hours at 37°C. The supernatant was then incubated at 60°C for 10 min, diluted 4 times with 1xLIG, incubated with T4 DNA ligase in the presence of ATP, and used to transform competent JM109 host bacteria (13).

Alternatively, to analyse the physical characteristics of the completed gene and the efficiency of assembly, the resulting structure after addition of tetrad 6 and the ligation in solid phase was released by EcoRI, precipitated by ethanol, end-labelled using polynucleotide kinase and gamma[^{32}P]ATP, purified by electrophoresis in 5% polyacrylamide gel, and an aliquot of it was ligated with M13mp18 for transformation.

The nucleotide sequences of the cloned products were determined using Sanger's dideoxy chain termination method (14).

RESULTS AND DISCUSSION

Preparation of the solid support

The anchor oligonucleotide A was covalently coupled to the CNBr-activated Sephacryl S-500. The sequence of the first 20 nucleotides (from the 3'-end) of this oligonucleotide are complementary to the 3'-portion of exon II of the Saccharomyces cerevisiae actin gene. This support was at first used successfully for the immobilization of the intron-containing RNA transcript. It was demonstrated that the RNA immobilized in this way was recognized as substrate by the yeast splicing machinery ("solid phase splicing") (15). This system was then used for purification of the assembled splicing complex and identification of snRNAs present in the spliceosome (15).

The binding capacity of the support was measured using a radioactively labelled complementary oligonucleotide (#18) of known molarity and was determined to be approx. 2 pmol/10 µl of wet sediment.

Solid-phase assembly of the gene

Annealed oligonucleotide triads or tetrads were chosen as building blocks for the stepwise assembly of the double-stranded DNA via hybridization to the immobilized complementary oligonucleotide. It has been observed also by others (16) that while four adjacent oligonucleotides can be joined in very high yields (more than 90%) in the T4-ligase-catalysed reaction, the yields of full-length products decrease substantially in more complex oligonucleotidemixtures.

In most projects of synthetic gene assemblies the length of the protruding cohesive ends has been in the range of 4 to 7 nucleotides. Contrary to this, we used 15 nucleotide overhanging ends (Fig.1). A potential disadvantage of the presence of single-stranded portions of this length is that, owing to a hidden self-complementarity, unnoticed at the design stage, they could form secondary structures preventing the efficient joining to the other oligonucleotide subassemblies.

As yet we have not encountered any difficulties of this kind and could take advantage of the following observations: It was not necessary to ligate the annealed oligonucleotide triads or tetrads. They, as well as the whole growing chain, were held

together solely by the intermolecular hydrogen bonds. The resulting structure was stabilized by ligation en bloc atthe last elongation step. Residual nicks, which were observed in the final gene structure after electrophoresis under denaturing conditions (data not shown) as a result of incomplete phosphorylation and/or ligation, did not prevent the formation of recombinant molecules and were probably removed in the bacterial cell by mechanisms of DNA repair.

Release of the completed gene

In order to release the assembled gene by digestion with a restriction endonuclease, a spacer region containing multiple restriction sites (polylinker) was designed between the gene sequence and the solid support (Fig.1). In a model system, the assembled triads 1 and 2 with terminally labelled oligonucleotide #16 were incubated with restriction endonuclease BamHI, EcoRI, NcoI, NdeI or ApaI. It was shown that the assembled DNA duplex could be released by any of the enzymes except ApaI. It follows from this experiment that the minimal distance of the restriction site for the enzymatic release of DNA duplex is 7 bp, measured from the 3'-end of the anchor oligonucleotide covalently bound to the Sephacryl S-500 support.

Cloning and sequencing of the synthetic gene

Addition of the double-digested M13mp18 vector in the presence of T4-DNA ligase to the assembled gene may be formally considered the last elongation step. After thorough washing, only those vector molecules remained in the system which were covalently attached to the assembled gene. The solid-phase vector-insert ligation ("solid-phase cloning") turned out to be both less laborious and more efficient than the standard ligation of the isolated gel-purified assembled DNA duplex with vector DNA. In the former case, after the completed gene with the attached vector molecule is released from the solid support, the subsequent ligation is a simple circularization, i.e. a monomolecular reaction that gives at least 10 times more recombinants than the standard bimolecular vector-insert ligation.

Three independent clones were sequenced and the expected sequence of CTI gene was confirmed in all three cases.

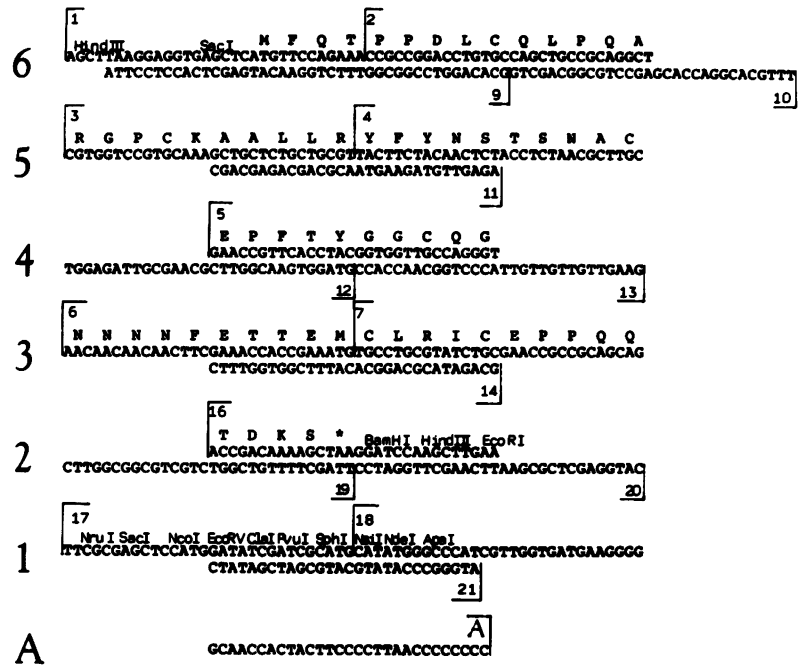

Figure 1. Design of the CTI gene. For stepwise assembly of the gene on solid support the oligonucleotides were annealed as triads or tetrads that are identified by large numbers to the left. Pointer with a small number identifies the 5'-end of each oligonucleotide. In order to determine the sufficient distance from the solid support for an efficient release of the completed gene via restriction endonuclease digestion, multiple restriction sites were introduced downstream from the coding region.
A - anchor oligonucleotide, which was covalently bound to Sephacryl S-500 support.

Advantages of the "solid-phase approach"

The most laborious part of the method described - CNBr-coupling of an anchor oligonucleotide to Sephacryl S-500-particles - need not, in principle, be performed more than once. The solid support can then be used for the assembly of variant forms of one and the same gene, for different genes, or for any other purposes where only one end of a temporarily immobilized DNA molecule should be involved in the reactions. The other end of the DNA duplex can be reversibly protected by ligation to the spacer polylinker, which is cleaved by an appropriate restriction enzyme.

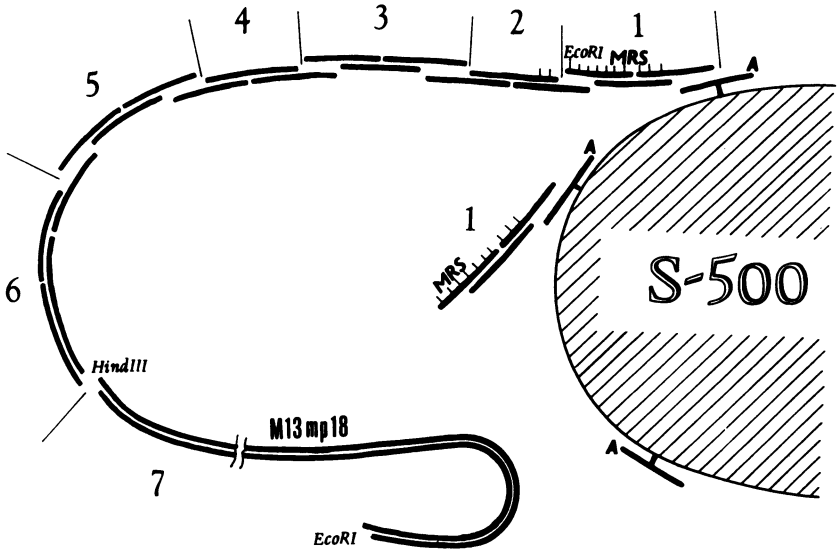

Figure 2. Schematic representation of the solid-phase gene assembly approach. S-500 - Sephacryl S-500 particle. A - anchor oligonucleotide. MRS - multiple restriction sites. Numbers identify annealed triads or tetrads from Fig.1 as well as chain elongation steps. Joining of the linearized M13mp18 vector molecule to the assembled gene is considered to be the 7th elongation step. An example of prematurely terminated chain is also shown.

It was demonstrated that many enzymes or enzyme systems, including T4-DNA ligase, polynucleotide kinase, restriction endonucleases and splicing extracts, are active in the presence of Sephacryl S-500 with bound nucleic acids, to a degree which is comparable with that of homogeneous solutions. It could be extrapolated that most reactions involving nucleic acids and enzymes could be performed in Sephacryl suspensions. This system enables rapid and efficient change of buffers and removal of unbound reaction products by washing.

The gene assembly itself, starting from the ready support and purified oligonucleotides, is quite straightforward, with no need of any electrophoretic separations up to the stage of harvesting the recombinant clones. The principle of the method is schematically summarized in Fig.2.

REFERENCES

1. Čechová, D., Jonáková, V. and Šorm, F. (1971) Collect. Czech. Chem. Commun. **36**, 3342-3357.
2. Laskowski, M. and Kato, I. (1980) Ann. Rev. Biochem. **49**, 593-626.
3. Schaller, H., Weimann, G., Lerch, B. and Khorana, H.G. (1963) J. Am. Chem. Soc. **85**, 3821-3827.
4. Ti, G.S., Gaffney, B.L. and Jones, R.A. (1982) J. Am. Chem. Soc. **104**, 1316-1319.
5. Caruthers, M.H., McBride, L.J., Bracco, L.P. and Dubendorff, J.W. (1985) Nucleosides and Nucleotides **4**, 95-105.
6. McGee, D.P.C., Martin, J.C. and Webb, A.S. (1983) Synthesis, 540-541.
7. Smrt, J., Večerková, H., Forstová, J., Hostomský, Z. and Rosenthal, A. (1985) Collect. Czech. Chem. Commun. **50**, 2201-2209.
8. Sinha, N.D., Biernat, J., McManus, J. and Köster, H. (1984) Nucleic Acids Res. **12**, 4539-4557.
9. Arnold, L., Točík, Z., Bradková, E., Hostomský, Z. and Smrt, J. (manuscript in preparation).
10. Caruthers, M.H. (1982) In Gassen, H.G. and Lang, A. (eds), Chemical and Enzymatic Synthesis of Gene Fragments, Verlag Chemie, Einheim, pp.71-79.
11. Bünemann, H. and Westhoff, P. (1983) Methods Enzymol. **100**, 400-407.
12. Norrander, J., Kempe, T. and Messing, J. (1983) Gene **26**, 101-106.
13. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103-119.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5468.
15. Hostomský, Z., Ruby, S. and Abelson, J. (submitted for publication).
16. Ferretti, L., Karnik, S.S., Khorana, H.G., Nassal, M. and Oprian, D.D. (1986) Proc. Natl. Acad. Sci. USA **83**, 599-603.