Volume 9 Number 21 1981 Nucleic Acids Research

Specific deletion of DNA sequences between preselected bases

Nikos Panayotatos and Kim Truong

Biogen S.A., 3, route de Troinex, 1227 Carouge/Geneva, Switzerland

Received 18 September 1981

ABSTRACT

Blunt-end ligation of a "filled-in" HindIII, Sal I, Ava I or Bcl I restriction site with a DNA fragment having A, G, C, or T as the terminal 3' nucleotide regenerates the corresponding restriction site. A combination of this property with the action of BAL 31 nuclease which progressively removes base-pairs from the ends of linear DNA, can generate deletions extending to desired pre-selected nucleotides, and introduces unique restriction sites at those positions. Similarly other restriction sites can be used to select for the deletion of sequences between specific di-, tri-, tetra- and penta-nucleotides. Using this method, 10 base pairs were deleted from the end of a restriction fragment carrying the late promoter for bacteriophage T7 gene 1.1, to create a molecule with a unique restriction site at the initiation codon for translation.

INTRODUCTION

An important aspect of recombinant DNA techniques is the capacity to introduce specific deletions in a genome. Selective deletion of DNA sequences can be used to locate a regulatory site, to delineate the boundaries of a gene, and to generate new mutant strains. This approach has found extensive application in the construction of "perfect" eukaryotic genes which can produce, in bacteria, a protein having no extraneous amino acids, except for the initiating methionine.

Methods described so far, take advantage of the single-stranded nature of "looped-out" structures (1) or D-loops (2) in order to generate somehow specific deletions in double-stranded circular DNA molecules. An alternative approach makes use of synthetic linkers (3, 4) and, either the double-stranded exonucleolytic activity of BAL 31 nuclease (5, 6) or the combined action of an exonuclease and SI nuclease (7). One advantage of

the latter approach is the generation of deletions, with the simultaneous introduction of restriction sites close to the desired positions. This method, however, has two serious drawbacks: (a) selection of the desired deletions among the background of unwanted transformants must be done by DNA sequencing which is laborious, and b) synthetic linkers are used which introduce extraneous nucleotides.

Here, we describe an approach which circumvents these problems. Selection of the desired deletion mutants does not involve DNA sequencing and unique restriction sites are introduced without the use of synthetic linkers. Most importantly, deletions terminating at specific base-pairs can be selected.

MATERIALS AND METHODS

All restriction enzymes, polynucleotide kinase and T4 DNA ligase were purchased from New England Biolabs. Conditions for these enzymatic reactions have been described (8, 9). Nuclease BAL 31 (5) was purchased from Bethesda Research Labs and assayed in 20mM Tris-HCl pH 8.0, 12mM $MgCl_2$, 12mM $CaCl_2$, 600mM NaCl, and 1mM EDTA. Plasmid DNA (20µg in 400µl buffer) was pre-incubated at 30°C for 5 min. 6.4 units (1µl) of BAL 31 nuclease was added, aliquots were removed at the desired times and quenched with buffer-equilibrated phenol. The DNA was prepared for subsequent reactions by ether extraction, followed by ethanol precipitation. Agarose and polyacrylamide gel electrophoresis were performed as described previously (8). "Fill-in" reactions with DNA polymerase I-large fragment (Boehringer) were carried out in 20mM Tris-HCl pH 7.6, 10mM $MgCl_2$, 0.5mM EDTA, 0.25mM dithiothreitol and 60µM each of the four deoxynucleotide triphosphates (Sigma) for 30 min at 37°C. *E. coli* C600SF8 was transformed as described previously (8). Selection plates contained 25mg/liter ampicillin.

RESULTS

Construction of Deletions The region between 14.7 and 15.0% of bacteriophage T7 DNA carries the late promoter, the ribosome binding site and the DNA sequence for the first three amino acids of T7 gene 1.1 (ref. 9, 10, 11 and Fig. 4). In order to exploit

these regulatory signals for the expression of eukaryotic genes, it would be useful to create deletions in which the DNA coding for all but the initiating (methionine) amino acids was removed. At the same time, a restriction site at the end of the deletion was, also, desired, so that new genes could subsequently be introduced. For this purpose, a restriction fragment was isolated from T7 DNA (8) and inserted between the EcoRI-BamHI sites of pBR322 in a manner that regenerated these sites (see legend to Fig. 1). The resulting plasmid (pNKB107) was "linearized" with BamHI and treated with BAL 31 nuclease. In order to determine the optimal reaction time for the exonucleolytic removal of 10 bp, aliquots of the BAL 31 reaction mixture were removed at various times, cleaved into smaller fragments with HinfI and analyzed on 5% polyacrylamide gels. The two HinfI sites which span the BamHI site in pNKB107 are located 107 bp to the left (inside the T7 insertion) and 256 bp to the right. The course of the reaction was monitored by the gradual decrease in the sizes of these two fragments. The results (Fig. 2) show that, under the conditions used, the size of the 256 bp fragment gradually decreases to an approximate mean size of 230 bp over a period of 6 min. The size of the 107 bp fragment also decreases at a similar rate. Accordingly, a reaction time of 3 min. was chosen for the exonucleolytic removal of 10 bp. The results of Fig. 2, also, show that the shortened fragments appear on the gel as rather sharp bands indicating a narrow (5-10 bp) distribution for the majority of the products. This observation is consistent with the known properties of BAL 31 (5).

Following incubation with BAL 31, pNKB107 DNA was restricted with Sal I, at the single site located 275 bp away from the BamHI site, i.e. beyond the range of BAL 31. The product molecules were, thus, flanked by a Sal I site at one end and the various base pairs up to which BAL 31 had reached, at the other (Fig. 1). Filling-in of the Sal I site followed by ligation regenerated circular molecules which were transformed into E. coli C600SF8, and selected for ampicillin resistance.

Deletion Selection. Of the ampicillin resistant transformants obtained, only plasmids into which a terminal G·C pair had been ligated with the filled-in Sal I site at the other end

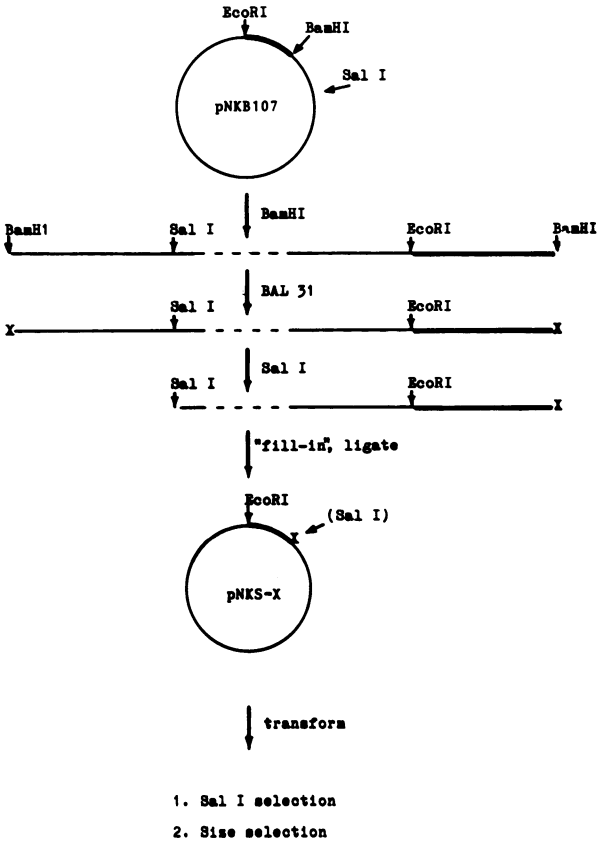

Fig. 1 - Outline of the method. pNKB107 carries a 122 bp fragment from the 14.7-15.0% region of bacteriophage T7 (9) inserted between the EcoRI and BamHI sites of pBR322. Originally, this fragment was isolated as an Hpa II-Tac I fragment (9), and converted into an EcoRI-EcoRI fragment by filling-in and blunt-end ligation into the EcoRI site of pVH51 (16). By coincidence, the "downstream" EcoRI site also partially overlaps with an Asu II site. An 122 bp EcoRI-Asu II fragment was then inserted between the EcoRI-BamHI sites of pBR322 (A filled-in Asu II site blunt end ligated to a filled-in BamHI site regenerates the latter site). The resulting plasmid pNKB107 has unique EcoRI and BamHI sites flanking the insertion. (X) denotes that a range of termini are created by BAL 31. The brackets around the Sal I site in the pNKS series emphasize the fact that only a small fraction of these molecules carries that site.

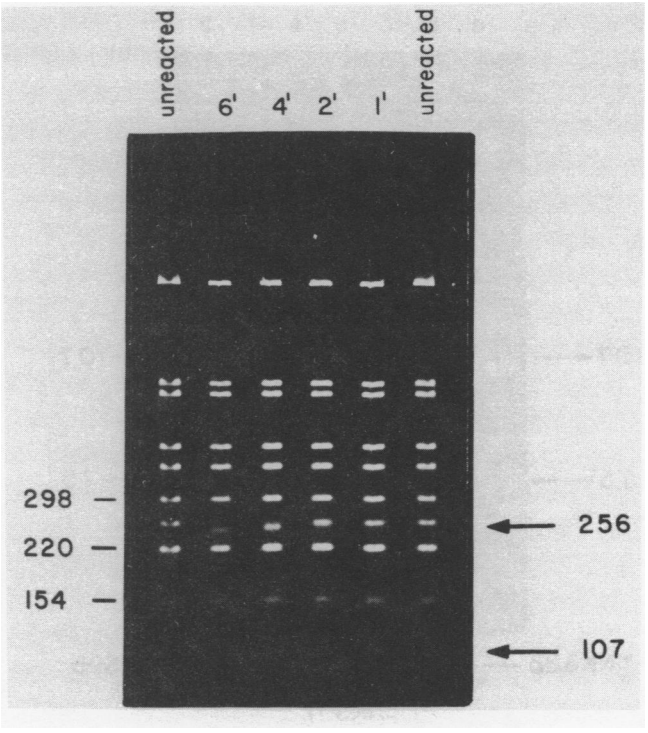

Fig. 2 - Progress of BAL 31 exonucleolytic activity on pNKB107 DNA linearized with BamH1. The DNA from each reaction time with BAL 31 was subsequently incubated with HinfI and analysed by electrophoresis on 5% polyacrylamide gels. The 256 bp and 107 bp HinfI-BamH1 fragments which flank the BamH1 site are indicated by arrows. The average decrease in size on BAL 31 digestion was estimated from this gel to be 2-3 bp per minute.

of the molecule would have regenerated the site. Plasmids were prepared from single bacterial colonies using a rapid method (15), incubated with Sal I and analysed by agarose gel electrophoresis. Eleven out of the 80 colonies tested contained plasmids that were cleaved by Sal I (data not shown).

In order to determine the size of the deletions introduced into these plasmids, the DNA was restricted with Sal I and HinfI, and analysed on 8% polyacrylamide gels. Typical results of such digestions are shown in Fig. 3. Overall, of the 11 candidates examined, 7 produced 107 bp long HinfI-Sal I fragments which apparently arose from molecules having the single-stranded

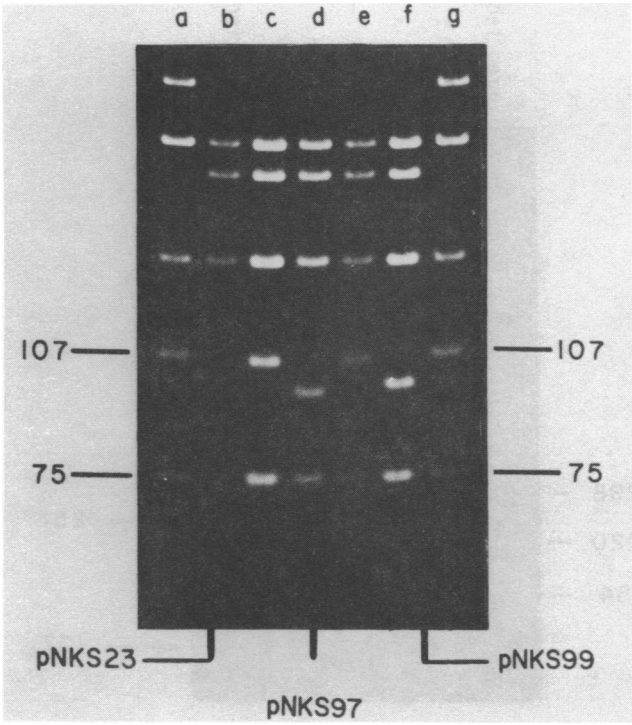

Fig. 3 - Size selection of desired deletions. Plasmids which were found to contain a Sal I site were digested with Sal I and HinfI and electrophoresed on 32 cm long 8% polyacrylamide gels. pNKB107 (lanes a, and g) digested with BamH1 and HinfI served as the intact size marker fragment. The sizes of the new HinfI-Sal I fragments are 23 bp, lane b, 107 bp, lane c, 97 bp, lane d, 107 bp, lane e, and 99 bp, lane f. For optimal resolution, only the small fragments region of the gel is shown.

protruding ends of the BamHI site removed by the BAL 31 nuclease. Three other candidates produced fragments which were 23, 97, and 99 bp in length, within an estimated error of ± 1-2 bp. The extent of the deletion in the eleventh molecule could not be determined.

Since Sal I sites could be introduced only at the G·C pairs located within the range of the BAL 31 nucleolytic activity, the sizes of the expected deletions could easily be calculated from the known DNA sequence of the HinfI-Sal I fragment (Fig. 4). The results of Fig. 2 indicated that BAL 31 had progressed

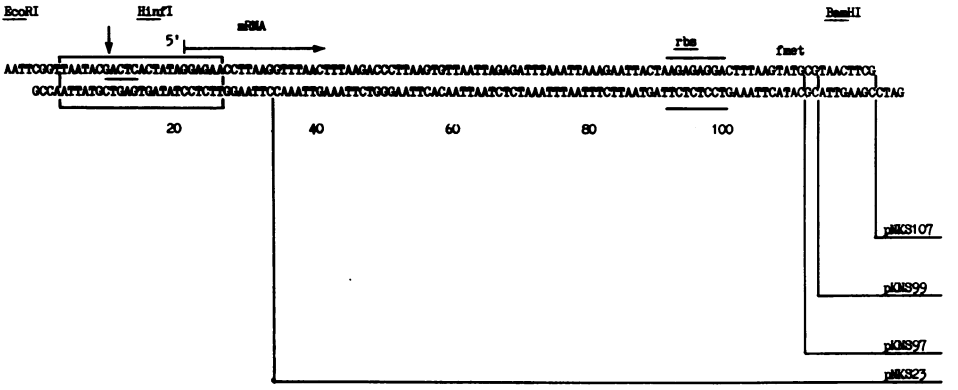

Fig. 4 - The DNA sequence of the deletions obtained. pNKS97 is a molecule in which the sequence of 3 amino acids of the T7 gene 1.1 has been deleted. The bars indicate the position where the deletions end, and the Sal I sites are introduced.

approximately 10 bp from the BamHI site, with the large majority of the deletions being within a range of ± 10 bp from the mean. This 20 bp region (Fig. 4) contains four G·C pairs at positions 105, 109, 111, and 118 which would give rise to HinfI-Sal I fragments 93, 97, 99, and 107 bp in length, respectively. Fragments 97, 99, and 107 ± 1-2 bp were, indeed, identified among the deletions. An additional deletion having a fragment of 23 bp was, also, generated, apparently as the result of more extensive degradation by BAL 31.

The assignment of the Sal I sites to the G·C pairs of the HinfI-Sal I sequence was based on the sizes of the corresponding fragments determined from the 8% polyacrylamide gels. In order to test the accuracy of these assignments, the DNA of the deletions generating the 23, 97, 99, and 107 bp fragments was isolated in larger scale and sequenced. The DNA sequences verified the size assignment (data not shown). This result demonstrated that the correct positions of the deletions can be identified, without the need of DNA sequencing as a screening procedure. Because of the selective introduction of a Sal I site, only at G·C bp, a characteristic pattern of the expected fragments can be visualized. Based on this pattern, accurate band sizes can be determined for individual base pairs separated

by as little as two bases on the DNA sequence.

DISCUSSION

To exploit the transcription and translation initiation signals of bacteriophage T7 gene 1.1 for the expression of eukaryotic genes, we desired a specific restriction site at the G.C base pair of the initiating ATG codon, with the simultaneous deletion of 10 base pairs of unwanted DNA sequence. Using the approach outlined in Fig. 1, pNKS97 which carries a unique Sal I site at the desired position was constructed. A two-step approach was employed for the selection of the desired deletion. First, DNA from ampicillin resistant transformants were screened for the presence of Sal I sites by agarose gel electrophoresis. Subsequently, candidates carrying Sal I sites were screened for the size of the deletion on polyacrylamide gels. The use of polyacrylamide gels for the simultaneous screening of the desired deletions according to both size and the presence of Sal I sites can also be employed.

The identification of deletions extending up to specific base pairs can be based solely on the results of the polyacrylamide gels. Exact size assignment is possible because the method described introduces Sal I sites exclusively at G·C base pairs. In theory, new HinfI-Sal I restriction fragments can be generated from all possible G·C base pairs located within the range of the BAL 31 treatment. The relative lengths of these fragments constitute a characteristic, size-diagnostic pattern. Comparison of this pattern with the relative sizes of the new HinfI-Sal I fragments observed on the polyacrylamide gel can accurately identify the exact extent of each deletion. The evidence presented here, demonstrates the feasibility of this approach.

Selection of other specific nucleotides. Selection of deletions extending up to particular base pairs can be extended to all possible combinations of terminal base pairs. Any restriction fragment inserted at the Eco RI site of pBR322, linearized, and treated with BAL 31 could be screened for deletions ending at A·T, G· C, C·G, and T·A base pairs using blunt-end ligation to a filled in Hind III, Bam HI or Sal I, Ava

I, and Bcl I site respectively. These sites exist naturally in pBR322 except for the Bcl I site which can be introduced at the Pvu II or the Bal I sites by blunt-end ligation of a synthetic linker.

Selection for specific di-, tri-, tetra-, and penta-nucleotide sequences is also possible, the only limiting factor being the availability of the corresponding restriction enzymes. Within the sequence of the 122 bp T7 fragment (Fig. 4), contribution of five (TCGAC) base pairs of the incoming "filled-in" Sal I site to any G·C pair regenerates a Sal I site. At the same time, contribution of the three terminal base pairs of the incoming Sal I site blunt-end ligated to the sequence CGA which occurs once at position 12 (Fig. 4) would regenerate a Pvu I site (CGATCG). Similarly, an Eco RI site would, also, be generated from the blunt-end ligation of the filled-in Sal I site to the sequence GAAT which occurs at position 85 (Fig. 4). Overall, contribution of up to five terminal base pairs from the incoming Sal I site, blunt-end ligated to the appropriate sequence at the other end of the molecule, can regenerate twelve different 6 bp restriction sites. If less than a six base specificity is required deletions extending between virtually any set of defined di-, tri-, tetra-, and penta-nucleotides should be easily obtained by this method.

REFERENCES

1. Wallace, R.B. et al. (1980). Science, 209, 1396-1400.
2. A. Green, C. and Tibbetts, C. (1980). Proc. Natl. Acad. Sci. USA 77, 2455-2459.
3. Heffron, F., So, M. and McCarthy, B.J. (1978). Proc. Natl. Acad. Sci. USA 75, 6012-6016.
4. Ullrich, A. et al (1977). Science 196, 1313.
5. Gray. J.B. et al. (1975) Nucleic Acids Res. 2, 1459-1472.
6. Talmadge, K., Stahl, S. and Gilbert, W. (1980). Proc. Natl. Acad. Sci. USA 77, 3369-3373.
7. Roberts, T. and Lauer, G. (1979). Methods Enzymol. 68, 473-482.
8. Panayotatos, N., and Wells, R.D. (1979). J. Biol. Chem. 254, 5555-5561.
9. Panayotatos, N., and Wells, R.D. (1979). J. Mol. Biol. 135, 91-109.
10. Panayotatos, N., and Wells, R.D. (1979). Nature 280, 35-39.
11. 1. Dunn, J.J, and Studier, F.W. (1981). J. Molec. Biol. 148, 303-330.
12. Chapman, K. (1981). Fed. Proc. 40, 1577.

13. Shine, J., and Dalgarno, L. (1974). Proc. Natl. Acad. Sci. USA 71, 1342-1346.
14. Sutcliffe, J.G. (1978). Cold Spring Harb. Symp. quant. Biol. 42, 77-90.
15. Klein, R.D., Selsing, E., and Wells, R.D. (1980) Plasmid 3, 88-91.