SEQUENCING STRATEGY FOR A PROTEIN-LINKED GENOME: SPONTANEOUS REVERSIONS OF OCHRE TRIPLET IN THE PHAGE φ29 GENE 17

Vladimir FUCIK[a], Vladimir CHALOUPECKY[b], Jaroslav BERAN[a] and Jiri JONAK[a]

[a] *Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 166 37 Prague 6, Czech Republic; e-mail: jjon@img.cas.cz*

[b] *Institute of Medical Microbiology, 1st Faculty of Medicine, Charles University, 128 06 Prague 2, Czech Republic*

Received February 14, 1997
Accepted April 8, 1997

A direct method using T7 gene 6 exonuclease to prepare single-stranded templates was adapted for the sequencing of the phage φ29 genome carrying a covalently linked terminal protein at the 5′ ends of both DNA chains. The terminal protein was found to block completely the action of the exonuclease. The treatment of the phage DNA by proteinase K did not restore the accessibility of the 5′ ends to the exonuclease to the full extent but only partially. The digestion of proteinase K-treated terminal fragments of the genome by 5′ exonuclease could proceed now from both 5′ ends, however, at different rates. The delay of exonucleolytic digestion apparently due to residual amino acid(s) was calculated to be about 800 nucleotides. The relative rates of exonucleolytic splitting of both chains were of decisive significance for the choice of suitable primers for the sequencing of desired regions. In all Sus^+ revertants sequenced by the described method the TAA(ochre) codon in the gene 17 was found changed either to the original CAA(Gln) or to TAT(Tyr) with the same frequencies.

Key words: Bacteriophage φ29; Spontaneous mutations; Terminal protein; Proteinase K; T7 gene 6 exonuclease.

Most sequencing of linear double-stranded genomes is now performed after cloning in single-stranded vectors. In experiments in which only the nucleotide variations in a short region of otherwise known sequence are of interest, the cloning can be considered as time-consuming. To avoid the cloning step when sequencing the mutants of fully sequenced phage φ29 (ref.[1]), we tried first to prepare single-stranded templates by alkaline or heating denaturation. Because of unsatisfactory results we turned to using T7 gene 6 exonuclease to prepare single-stranded molecules as sequencing templates. This exonuclease degrades both strands on double-stranded substrates in the 5′-3′ direction. When a DNA molecule is synchronously degraded from both 5′ ends the reaction stops in the middle of the molecule and two single-stranded DNA half-molecules are produced[2,3]. Although the exonuclease requires double-stranded substrates, it can start digesting fragments containing either 5′ or 3′ "overhangs" (*e.g.* fragments after DNA digestion with restriction endonucleases).

We found that this strategy of the preparation of single-stranded templates suitable for the sequencing reaction was also applicable to genomes with protein molecules covalently linked to the 5′ ends of their genomic DNA. To this category of organisms belong not only certain bacteriophages like ϕ29 but also adenoviruses and some linear plasmids or virus-like particles[4].

In the phage ϕ29 the terminal protein contains 266 amino acids and is encoded by the phage gene No. 3. One of the serine residues is covalently attached to the first deoxyribonucleotide which is dAMP on both 5′ ends of the genome. It is assumed that the protein cannot be fully removed by a proteinase treatment of the DNA. This was documented by the polynucleotide kinase failure to phosphorylate the 5′ ends of the DNA unless the last amino acid or a short peptide were removed by piperidine treatment[5].

As a biological object for the development of the method we used the spontaneous revertants of ϕ29 *sus17(112)* phage mutant. Gene No. 17 is one of the 18 genes identified in the genome of ϕ29. It was mapped near the right end of the genome and belongs to the category of early genes playing certain role in DNA replication[6]. Sometimes, the gene 17 is described as dispensable[7]. The comparison of the nucleotide sequences of the gene 17 even in the most closely related phages (ϕ15, PZA, ϕ29) demonstrated an extreme variability especially in the central and 3′ end proximal region[8]. The suppressor-sensitive phenotype of the ϕ29 *sus17(112)* mutant is caused by a stop codon in the more conservative sequence close to the beginning of the gene-coding region. The stop codon (TAA) was identified as the fifth triplet in the gene sequence[9]. The preparation of single-stranded templates by T7 exonuclease digestion enabled a rapid screening of base substitutions in ϕ29 *sus17*$^+$ revertants.

EXPERIMENTAL

Bacteria

Strains of *Bacillus subtilis* 168 were constructed in our laboratory by marker congression as sporulation-deficient mutants with the following genotypes:

9/3: *leuA8* r^-_{168} m^-_{168} *spo0A3*

13/7: leuA8 r^-_{168} m^-_{168} (*purB6 metB5 thr-5*)$^+$ *sup*$^{+3}$ *spo0A3*

The bacteriophages used in this study were kindly provided by M. Salas (Madrid).

Selection of Spontaneous Revertants

Diluted suspensions of ϕ29 *sus17(112)* were plated on 13/7 host bacteria. Single plaques were isolated and suspended in 1 ml aliquots of dilution buffer TMS (50 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$ and 200 mM NaCl). These suspensions were plated with bacteria 9/3 lacking a suppressor gene. From each plate where revertants appeared, only one plaque was picked up and further propagated for DNA isolation. The sequenced revertants could not therefore originate from the same clone but represented independent mutational events.

Isolation of Phage DNA

A) *Proteinase K-treated DNA*. Phages purified in CsCl gradient in a vertical rotor VTi80 were dialyzed against TMS buffer and disrupted by the addition of sodium dodecylsulfate, proteinase K and EDTA for several hours[10]. The viscous material was further treated with phenol, phenol–chloroform and chloroform, and after the addition of sodium acetate (0.3 M final concentration), the solution was precipitated with the same volume of isopropyl alcohol.

B) *Terminal protein p3-DNA complex*. Purified phages were dialyzed against TE (10 mM Tris-HCl pH 8.0 with 1 mM EDTA) and lysed with 4 M guanidine . HCl according to Penalva and Salas[11]. The lysates were purified in CsCl gradient (5 ml tubes in a Beckman SW55 rotor). The tubes contained two layers of CsCl solution (52% and 43%, 2.2 ml each) and the samples were layered on the top. One tube served as a control, on which only a small amount of the lysate together with ethidium bromide and a necessary amount of the dilution buffer was loaded. After the run the position of the lower UV-fluorescent band was marked on the control tube and, from the corresponding positions, samples (0.5–0.7 ml) were withdrawn with a syringe through the punctured wall. Samples were dialyzed against diluted TE and concentrated in Eppendorf tubes by a flow of nitrogen.

Sequencing reaction was performed with the Sequenase Version 2.0 sequencing kit (USB) according to the producers' protocol.

Templates for sequencing were prepared by the T7 gene 6 exonuclease digestion as recommended by the producer (USB). For one sequencing reaction, 0.5 pmol of the phage DNA was used. Primers were designated in accordance with the genetic map[12] of the fully sequenced phage[1]; those oriented leftward or rightward carry the letter L or R, respectively, in addition to a number. Their nucleotide sequences and positions were the following:

L3 (20-mer): 5′-GTAACTGTTGAAGAGTTAGG-3′ (18892 – 18872)
L4 (19-mer): 5′-GGTAATAAGACAACCAATC-3′ (19119 – 19100)
R5 (19-mer): 5′-GTTTGTGTTGATGATGTCG-3′ (19030 – 19049)

RESULTS AND DISCUSSION

In accord with our aim to develop a simple method of sequencing a number of ϕ29 *sus17*$^{+}$*(112)* spontaneous revertants, we first designed two opposite-direction primers L4 and R5 (Fig. 1) situated on the right and left side, respectively, of the nonsense triplet (TAA) in the mutant allele of the gene 17.

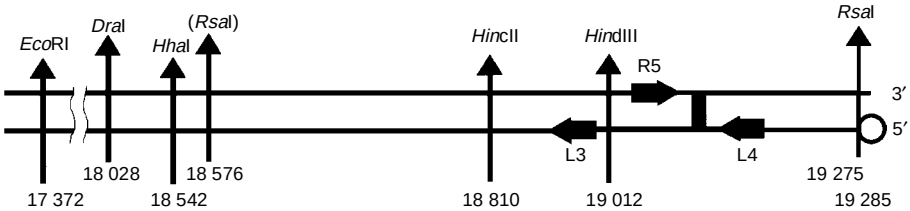

FIG. 1
Right-hand terminal region of ϕ29 double-stranded DNA. Numbers indicate the positions of nucleotides in the phage DNA. ↑ Restriction sites, (*Rsa*I) restriction site is absent in ϕ29 *sus 8.5 sus 14*; →, ← orientation of primers; ▮ nonsense codon in the mutated gene 17; ❍ terminal protein

Phage DNA was prepared in two ways, as a p3-DNA with intact protein molecules on both 5′ ends or as a proteinase K-treated DNA.

Table I shows that different approaches, as far as the usage of restrictases or the form of the DNA is concerned, gave positive results. Sequences encountered in the place of the original TAA nonsense triplet are shown in Tables II and III.

Originally, we supposed that both the intact protein and the residual amino acid(s) on proteinase K-treated DNA would protect the 5′ ends of DNA molecules against the exonuclease digestion. We therefore routinely treated the DNA preparations with restriction enzymes before the exonuclease digestion. As the gene No. 17 is situated close to the right end of the genome, most restriction fragments carrying the gene 17 were terminal, having at one end the terminal protein or a residual peptide, while the other end was blunt or with "overhangs" according to the restrictase used (Fig. 1).

TABLE I
Sequencing of the p3-DNA complex and the proteinase-treated DNA digested by different restrictases followed by 5′ exonuclease

Restrictase	Length of terminal fragment (bp)	Proteinase treated DNA			p3-DNA	
		L3	L4	R5	L4	R5
*Bcl*I	13 583		+			+
*Xba*I	10 357		+			+
*Bst*EII	6 820		+			+
*Nco*I	4 259	+	+	–	–	+
*Pvu*I	2 990	+				
*Hpa*II	2 869		+			
*Hpa*I	2 549	+	+			
*Eco*RI	1 913	+	+	–	–	+
*Dra*I	1 257		–			
*Hha*I	743		–	+	–	+
*Hinc*II	475		–	+	–	+
*Hind*III	273		–	+	–	+
*Hha*I + *Rsa*I			+	–	+	–
*Hinc*II + *Rsa*I			+	–	+	–
*Hind*III + *Rsa*I			+	–	+	–
*Rsa*I			+	–	+	–

Symbols: + primer elongated, – primer not elongated.

When the p3-DNA was treated with a restrictase having the last recognition site upstream from the nonsense triplet, only one 5′ end of the terminal fragment was accessible to the exonuclease and, as a result, only the R5 primer (Fig. 1) complementary to the protein-carrying strand was applicable to sequencing. The opposite template for the other primer L4 was evidently digested away. The length of the fragment played no role (see Table I).

A different situation was observed when DNA preparations treated with proteinase K were used for exonuclease digestion and sequencing assay. In that case the experimental results were dependent on the length of the prepared restriction fragments (Table I). Single-stranded templates for R5 priming could be prepared from the restriction digests with shorter right-hand terminal fragments (*Hin*dIII, *Hinc*II, *Hha*I), while with the longer terminal fragments produced by *Eco*RI or *Nco*I, the primer R5 became ineffective and the sequencing of the site of our interest could be achieved with the opposite-direction primer L4 complementary to the other strand.

Our experimental results obtained with terminal fragments from proteinase-treated DNAs indicate that the rate of exonuclease digestion was nonsymmetrical and significantly slowed down (but not completely blocked) by the residual amino acid(s) at one of the 5′ ends. This was confirmed by the finding that the removal of the last ten base

TABLE II
Possible one-base substitutions in TAA triplet. Codons encountered among the ϕ29 *sus17(112)*$^+$ spontaneous revertants are underlined

TAA(ochre)		
CAA(Gln)	TGA(opal)	TAG(amber)
AAA(Lys)	TCA(Ser)	TAC(Tyr)
GAA(Glu)	TTA(Leu)	TAT(Tyr)

TABLE III
Frequencies of different codons found in place of TAA in the spontaneous Sus$^+$ revertants of ϕ29 *sus17(112)* phage

Type of revertant	Number
CAA	11
TAT	10
Total	21

pairs from the right-hand end of the genome (including the inhibiting residual amino acid) by *Rsa*I made possible the usage of L4 primer for the sequencing reaction also with *Hin*cII or *Hha*I digests. The switch of applicability of the primers after the double digestion could not be achieved in the combination of *Hin*dIII with *Rsa*I because both primers are situated in the left half of the *Hin*dIII terminal fragment.

For the experiments with *Hha*I and *Rsa*I double digestion, we used the mutant strain ϕ29 *sus8,5 sus17* lacking the *Rsa*I restriction site at the position 18 576 bp (Fig. 1), otherwise the *Hha*I–*Rsa*I double digestion would be identical to *Rsa*I digestion alone within the fragment of our interest.

Taking into consideration the distances of L3 and L4 primers from the right-hand end of the genome and the results of sequencing of the *Dra*I and *Eco*RI restriction digests, which mark the boundaries within which the switch for opposite-direction primer was observed, an estimate could be made that the delay of exonucleolytic digestion caused by the presence of residual amino acid(s) equals approximately 800 nucleotides.

In the course of our experiments aimed at working out the method of preparation of sequencing-templates we used DNAs from different revertants (back mutants) of ϕ29 *sus17* bacteriophage. Tables II and III summarize the results showing that the TAA(ochre) codon in the gene 17 changed to either CAA(Gln) or TAT(Tyr) with the same frequency.

Double-strand specific exonucleases proved to be suitable tools to generate single-stranded templates for direct sequencing of genomes with unprotected 5′ termini[13,14]. Exceptions were described when a terminus of a DNA molecule was, for some reasons, not acceptable to the enzyme. Exonuclease III of *Escherichia coli* rejects, for instance, molecules with 3 overhangs and this property could be used in combination with the nuclease S1 for shortening the molecules having one single-stranded 3′ terminus[15].

Our results demonstrated that the double-stranded DNA of ϕ29 with protein molecules covalently bound to its 5′ ends was resistant against the digestion with T7 gene 6 exonuclease. Various restriction enzymes were, therefore, used and from the right-hand terminal fragments of ϕ29 DNA, the whole nonprotected strand was digested irrespective of its length, so that the strand with the protein could serve as a template. Another possibility how to sequence a region close to the end of the genome was to cut off the terminal part with a suitable restriction enzyme (in our case it was *Rsa*I, see Fig. 1) and to sequence the complementary strand using a reverse sequencing primer.

It was described that T7 gene 6 exonuclease, in contrast to lambda exonuclease, could also attack the non-phosphorylated 5′ terminus of a DNA polynucleotide chain and the first residue liberated by hydrolysis was dinucleoside monophosphate[3]. Our results extend this picture of the lack of strict preference of the exonuclease for a free 5′ phosphorylated terminus. They revealed only partial inhibition of the nucleolytic activity by amino acid residue(s) covalently attached at the 5′ terminus of the DNA treated with proteinase K. Whereas the intact 5′-terminal protein proved to be com-

pletely inhibitory for the exonuclease action, the amino acid residues only delayed its onset.

REFERENCES

1. Vlcek C., Paces V.: Gene *46*, 215 (1986).
2. Kerr C., Sadowski P. D.: J. Biol. Chem. *247*, 305 (1972).
3. Kerr C., Sadowski P. D.: J. Biol. Chem. *247*, 311 (1972).
4. Salas M.: Annu. Rev. Biochem. *60*, 39 (1991).
5. Escarmis C., Salas M.: Proc. Natl. Acad. Sci. U.S.A. *78*, 1446 (1981).
6. Carrascosa J. L., Camacho A., Moreno F., Jimenez F., Mellado R. P., Vinuela E., Salas M.: Eur. J. Biochem. *66*, 229 (1976).
7. Escarmis C., Guirao D., Salas M.: Virology *169*, 152 (1989).
8. Benes V., Arnold L., Smrt J., Paces V.: Gene *75*, 341 (1989).
9. Smarda J.: *Ph.D. Thesis*. Czechoslovak Academy of Sciences, Prague 1989.
10. Inciarte M. R., Lazaro J. M., Salas M., Vinuela E.: Virology *74*, 314 (1976).
11. Penalva A., Salas M.: Proc. Natl. Acad. Sci. U.S.A. *79*, 5522 (1982).
12. Mellado R. P., Moreno F., Vinuela E., Salas M., Reilly B. E., Anderson D. L.: J. Virol. *19*, 495 (1976).
13. Malik A. N. in: *Methods in Molecular Biology* (H. G. Griffin and A. M. Griffin, Eds), Vol. 23, p. 141. Humana Press, Totowa, N. J. 1993.
14. Ruan C. C., Fuller C. W.: *U. S. Biochemical, Sequencing Support Service*, Vol. 18, No. 1, p. 1. United States Biochemical, Cleveland, Ohio 1991.
15. Hennikoff S.: Gene *27*, 351 (1984).