Nucleotide sequence of bacteriophage fd DNA

E. Beck*, R. Sommer, E.A. Auerswald, Ch. Kurz, B. Zink, G. Osterburg, and H. Schaller
Microbiology, Univ. Heidelberg, Im Neuenheimer Feld 230, Inst. Documentation, Information and
Statistic, German Cancer Research Cent., Im Neuenheimer Feld 280, 6900 Heidelberg, GFR and
Institute for Chemical Research, Kyoto University, Uji, Kyoto-Fu 611, Japan

Received 6 November 1978

ABSTRACT

The sequence of the 6408 nucleotides of bacteriophage fd DNA has been determined. This allows to deduce the exact organisation of the filamentous phage genome and provides easy access to DNA segments of known structure and function.

INTRODUCTION

Small DNA viruses depend during their life cycle largely on host functions and are therefore preferred model systems for the analysis of the organisation, expression and replication of the more complex host genomes. To analyse viral genomes at the nucleotide level has become technically possible with the development of new rapid DNA sequencing techniques. Complete nucleotide sequences have been reported so far for coliphage λ174 and Simian Virus SV40. Here we report the sequence of bacteriophage fd DNA, strain 478 (Heidelberg).

Phage fd along with f1 and M13 belongs to a group of closely related filamentous, male-specific coliphages (for reviews see ref. 9, 10). Its genome is a single-stranded circular DNA of about 6000 nucleotides which is converted to a double-stranded form in the infected cell. Eight genes have been ordered by combined genetic and biochemical analysis within the phage genome. Its detailed organisation remained, however, relatively uncertain due to the lack of protein data for most gene products. Furthermore, analysis on the nucleotide level had concentrated mainly on DNA segments with regulatory functions.
We have previously reported a preliminary nucleotide sequence of fd DNA (11, and personal communications). The aim of this publication is the rapid communication of the final sequence. A more detailed account containing the experimental evidence will be published elsewhere.

RESULTS AND DISCUSSION

Restriction nucleases and cleavage maps. The enzymes used, their recognition sequences and the position of cleavage sites confirmed or newly established during this work are presented in Fig. 1. All cleavage sites shown have also been identified by DNA sequencing the ends of the respective restriction fragments. With one exception, all parts of double-stranded fd can be fragmented by digestion with several of these enzymes into pieces of less than 200 base-pairs.

DNA sequencing. The chemical method of Maxam and Gilbert2 was used which allowed us to read sequences up to 150 (occasionally up to 220) nucleotides. Sequences obtained were stored and processed in a computer (G. Osterburg and R. Sommer, to be published) to yield the composite sequence of 6408 nucleotides presented in Fig. 2. About 75 % of this sequence was determined from both DNA strands in fd 478. Almost all of the missing 25 % have also been sequenced in the second strand, but in the closely related phage f1. Further information was obtained for about 1000 nucleotides by RNA sequencing12 and for about 600 nucleotides by the plus/minus method of Sanger and Coulson1. About 10 % of the fd sequence were also established as recognition sequences for restriction nucleases at known cleavage sites (Fig. 1 and unpublished results).

Nucleotide sequence. According to Fig. 2 fd DNA is composed of 6408 nucleotides (1578A, 2210T, 1325G, 1295C) corresponding to a molecular weight of 2.12 x 10^6 daltons) (sodium salt). The sequence differs from that reported earlier11 mainly by an insert of 18 nucleotides in the
Fig. 1: Fragment maps of restriction nucleases used in the sequence analysis of fd DNA, strain 478. The known maps for Hpal, HpaI, HaeII (HmaII), HaeIII, AluI and HphI were confirmed and refined. Maps for HhaI, HaeIII, TagI, BamH I, Sau3A, (BglII, MboI), EcoRI, MboII, and HphI were newly established (E.A. Auerswald et al., M. Takanami et al., both unpublished). The first nucleotide of the recognition sites for the various restriction nucleases are listed below. An additional Hinf site has been detected in fragment HinfC (position 1858) in the DNA from fd ATCC (M. Takanami, unpublished). The circular phage DNA is opened at the unique HindII (HpaI) cleavage site. The map includes the positions and the orientation of the phage genes. IG is the intergenic space.

AluI: AGCT
63 229 934 1488 1517 2963 3277 3613 4097 5427 5831
BamHI: GGATCC
2220 5645
BglII: GATC
1382 1714 2221 5646
EcoRI: CCTGG
1014 1966
HaeII: RCGCG
2710 4743 5560 5568
HhaI: GCAC
44 873 1011 1085 1177 1470 2196 2467 2711 3040 3096 3599
Hinfl: ACAC
4313 4642 4744 4886 4996 5491 5504 5513 5535 5561 5569
Hpal: GGCC
1396 2245 2554 5082 5240 5346 5415 5726 5829 6181
HpaI: GATC
526 2164 2479 3238
HphI: GATC
4084 5159
MboII: GATC
3913 4073 4272 4938 5256 5588
TaqI: TCGA
336 908 1127 1508 1949 2528 2815 4834 5684 6041
repetitive sequence around position 2380. Except for a G • A transition in position 1859 the identical sequence was obtained in 2000 nucleotides from another fd strain (ATCC).

The nucleotide sequence of the related phage f1 has been determined to about 90 % (E. Beck, unpublished). It differs from the fd sequence by deletion of a single nucleotide (position 3195) and by about 160 base changes. Except for seven, these are all silent mutations which do not alter the amino acid sequence of the fd gene products.

**Genome organisation.** By analysing the fd DNA sequence for continuous translational reading frames - combined with the information obtained from the sequence of amber mutations in f1 and M13 (E. Beck, unpublished; J. Schoenmakers, personal communication) and from the silent base changes in f1 - allows to deduce the exact sizes and positions of the eight known gene products and of known regulatory signals. The DNA sequence predicts the amino acid sequences of known and unknown gene products, and the existence of a new gene (gene IX) in the intergenic space between genes VII and VIII.

According to our analysis (Fig. 2) the overall organisation of the filamentous phage genome differs markedly from that of icosahedral single-stranded DNA phages, like \( \phi \times 174 \): Although genes are generally closely spaced there is only one single short overlap of genes in different reading frames (at the junction of genes I and IV). In addition there is an intergenic region (IG) of 508 nucleotides which harbours the origins of DNA replication\(^1\),\(^3\),\(^4\). Recent experiments show that this space can be further expanded by insertion of foreign DNA\(^1\),\(^6\).

**Applications.** fd DNA is accessible in high yields in both its single-stranded and double-stranded form\(^1\),\(^0\). The knowledge of its nucleotide sequence and of the map positions of a great number of restriction sites provides therefore easy access to well defined DNA molecules which can be used in different investigations on DNA structure...
Fig. 2: Nucleotide sequence of bacteriophage fd. The viral DNA single-strand is shown in 5' → 3' polarity. The circular DNA has been opened at the position of the origin of viral replication. Numbering of nucleotides starts at the unique HindII (HpaII) cleavage site. Genes are boxed, recognition sites for the restriction nucleases shown in Fig. 1 are overlined. The sequence is available on request on magnetic tape.
and function. For example they have been used as size markers in their intact or restricted form, for the search for recognition sequences of restriction nuclease, in the site-specific modification of the fd genome for use as a cloning vehicle, for the isolation and the cloning of regulatory signals from fd DNA, for the analysis of integration and loss of transposon Tn-5 (16,14, E.A. Auerswald, to be published), and for the correlation of thermal denaturation profiles of DNA molecules with their nucleotide sequence.

ACKNOWLEDGEMENT

This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

*Present address: The University of British Columbia, Faculty of Medicine, Department of Biochemistry, 2075 Westbrook Place, Vancouver, B.C. Canada V6T 1W5.

REFERENCES

   eds. Fraenkel-Conrat, H. and Wagner, R.R., Plenum
   Publ. Corp. N.Y. 10011 USA
   Single Stranded DNA Phages, eds. Denhardt, D.T.,
   Dressler, D., and Ray, D.S., Cold Spring Harbor Lab.,
   Cold Spring Harbor, N.Y. 11724 USA
12 Sugimoto, K., Sugisaki, H., Okamoto, T., and Takanami, M.
   results
   Sci. USA 73, 2341 - 2345
   Vol. XLIII, Cold Spring Harbor Lab., Cold Spring Harbor,
   N.Y. 11724, USA
15 Meyer, T., Geider, K., Kurz, Ch., and Schaller, H.
   (submitted for publication)
16 Herrmann, R., Neugebauer, K., Zentgraf, H., and Schaller,
   H. (1978) in Single Stranded DNA Phages, eds. Denhardt,
   D.T., Dressler, D., and Ray, D.S., Cold Spring Harbor
   Lab., Cold Spring Harbor, N.Y. 11724 USA
   (in press)
18 Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1978)
   Nucleic Acids Res. 5, 2547 - 2556