Volume 4 Number 7 July 1977 Nucleic Acids Research

The nucleotide sequence of two restriction fragments located in the gene AB region of bacteriophage S13

Frank G. Grosveld and John H. Spencer

Department of Biochemistry, McGill University, 3655 Drummond Street, Montreal, Quebec H3G 1Y6, Canada

Received 22 April 1977

ABSTRACT

The nucleotide sequence of a double stranded DNA fragment from the gene AB region of bacteriophage S13 DNA has been determined. The fragment was isolated as two adjacent shorter fragments by cleavage of S13 replicative form (RF) DNA with restriction endonuclease III from *Hemophilus aegyptius*. The strands of the fragments were separated electrophoretically and hydrolyzed with T4 endonuclease IV to yield short oligonucleotides which were then sequenced by partial exonuclease digestion. The complete nucleotide sequence of the restriction fragments was obtained by ordering the inter- and intrastrand overlapping oligonucleotide sequences. The adjacent fragments were 190 nucleotides in length. The sequences included a *Hin*dII site, an *Alu*I site and two sequences which may be possible transcription initiation sequences, one with an adjacent sequence homologous to the canonical promoter site sequence T-A-T-Pu-A-T-Pu. Examination of the three possible reading frames for translation of the sequence revealed only one possible complete translation product. The postulated partial sequence of gene A protein has a highly positively charged arginine-rich area which may have importance in DNA binding.

INTRODUCTION

Bacteriophages S13 and φX174 are almost identical in size, physical and biological characteristics. The two phages are closely related in biological function as demonstrated by the criteria of serological cross-reaction[1] and genetic recombination[2]. Complementation tests have shown that they have eight homologous genes[3] and the genes occur in the same order[4,5]. Comparative studies of the two phage DNA's by pyrimidine cluster[6] and restriction endonuclease[7,8,9] analyses indicate extensive sequence homology. Most of the longer pyrimidine oligonucleotides in the two DNA's have identical sequences[6], and the various restriction enzyme maps are very similar[7,8,9]. The gene products of five of the genes of the two phages comigrate on SDS polyacrylamide gels and are assumed to be almost identical in size[10]. Gene D and H proteins from each phage have similar yet different electrophoretic mobilities but the gene B proteins differ markedly and Tessman *et al.*[10] have concluded that the gene B protein

of S13 is considerably smaller than that of ϕX174. In a study of the *in vitro* transcripts of ϕX174 DNA and their hybridization to various restriction fragments, Smith and Sinsheimer[11] have placed the start of *in vitro* RNA transcription of gene B in restriction fragment ϕX174 *Hin*dIII 8.

The gene AB regions of the S13 and ϕX174 genomes thus have a good potential for providing information on the control of transcription and translation in the two phages. This report presents the nucleotide sequence of S13 restriction enzyme fragments *Hae*III 8 and 9 which are located in the gene AB region and which correspond to the junction of ϕX174 fragments *Hin*dIII 3 and *Hin*dIII 8 and include most of ϕX174 *Hin*dIII 8.

MATERIALS AND METHODS

Chemicals. Ethidium bromide was purchased from Sigma Chemical Co. and optical grade CsCl from the Harshaw Chemical Co. DEAE Sephadex A-25 was obtained from Pharmacia (Canada) Ltd, Bio-Gel A from Bio-Rad Laboratories and $H_3{}^{32}PO_4$ from NEN (Canada) Ltd. [γ-^{32}P]ATP was prepared as described by Schendel and Wells[12]. Reagents for electrophoresis, acrylamide and bis (N,N'-methylene-bis (acrylamide)) were Eastman products. DEAE-cellulose and cellulose for thin-layer chromatography were products of Macherey, Nagel and Co., purchased from Brinkmann Instruments (Canada) Ltd. Cellulose acetate strips were purchased from Schleicher and Schuell Inc., Keene,N.H., and cellogel strips (manufactured by Chemtron, Milan, Italy) from Mandel Scientific Co.

Enzymes. Snake venom phosphodiesterase and bacterial alkaline phosphatase (electrophoretically pure) were obtained from ICN (Canada) Ltd. Alkaline phosphatase was heat treated to remove phosphodiesterases[13], snake venom phosphodiesterase was pretreated at 37°C for 3 hr at pH 3.6 to remove 5' nucleotidase activity[14]. Pancreatic deoxyribonuclease A was purified as described by Junowicz and Spencer[15] and exonuclease I according to Lehman and Nussbaum[16]. Polynucleotide kinase and T4 endonuclease IV were prepared from *Escherichia coli* B infected with bacteriophage T4amN82[17]. Polynucleotide kinase was isolated according to Richardson[18] with the modifications described by Harbers *et al.*[6]

T4 endonuclease IV was isolated by the method described by Sadowski and Bakyta[19] up to the first phosphocellulose column fractionation step. Each fraction from the phosphocellulose column was assayed for the presence of T4 endonuclease IV using radioactively labeled (^{32}P or ^{3}H) S13 DNA as substrate in the presence of exonuclease I, essentially as described by

Sadowski and Bakyta[19]. The fractions that contained T4 endonuclease IV activity were then assayed again using the same substrate but incubation was for 30-40 minutes instead of 10 minutes and exonuclease I was omitted from the reaction mixture. This assay modification located those fractions containing endogenous 3' exonuclease activity. The exonuclease activity was found in the last third of the fractions with T4 endonuclease IV activity eluted from the phosphocellulose column. The T4 endonuclease IV fractions containing little or no 3' exonuclease activity were further purified by DEAE-cellulose chromatography as described by Sadowski and Bakyta[19]. The fractions which did not adhere to the DEAE cellulose column were then refractionated twice on phosphocellulose. The T4 endonuclease IV preparation from the last phosphocellulose column, when assayed by the 'minus' exonuclease I assay described above showed a definite lag before the appearance of acid soluble radioactivity. This was taken as a strong indication of little or no 3' exonuclease contamination in the preparation. The preparation was dialyzed against 50 mM Tris-HCl buffer (pH 7.6) in 50% glycerol and stored at -20°C.

Restriction endonuclease *Hae*III was isolated from *Hemophilus aegyptius* (ATCC strain IIII6) by the method of Smith and Wilcox[20] with the modifications described previously[7].

Preparation and isolation of restriction fragments of S13 DNA. Purification of ^{32}P-labeled bacteriophage S13 replicative form DNA (RF DNA) was performed as described by Schekman *et al.*[21] up to the phenol extraction and isopropanol precipitation step. The precipitated material was then dissolved in 50 mM Tris-HCl, 20 mM EDTA (pH 8.0) and applied to a Bio-Gel A-15m column which was eluted with the same buffer. The void volume containing the S13 RF DNA was collected and RF I DNA separated from RF II DNA by centrifugation in a CsCl, ethidium bromide gradient. The purification was completed as described by Grosveld *et al.*[7] The Bio-Gel column modification replaced the sucrose gradient step in the original procedure. Digestion of S13 RF DNA with restriction endonuclease *Hae*III was as described by Middleton *et al.*[22]

Gel electrophoresis of restriction fragments was performed as described elsewhere[7]. The separated restriction fragments were eluted from gels by soaking the crushed gel fragments in 0.6 M NaCl for at least 4 hours at 0°C. The extracted gel was removed by filtration through a fine grade sintered glass filter. The fragment extract was applied to a column of hydroxyapatite (1 x 1 cm) which was washed with 0.1 M potassium phos-

phate buffer (pH 7.0), then eluted with 1.0 M potassium phosphate buffer (pH 7.0). The restriction fragments were desalted by adsorption to DEAE-cellulose, eluted with 1.0 M triethylammonium bicarbonate (TEAB), pH 8.0, and repeatedly evaporated to dryness.

Depurination of restriction fragments. The isolated restriction fragments were hydrolyzed with 2% diphenylamine in 67% formic acid at 30°C for 18 hr[23]. The hydrolysate was extracted three times with ether, dried and fractionated by ionophoresis-homochromatography[24], as modified for pyrimidine oligonucleotides[25].

Strand separation and T4 endonuclease IV hydrolysis. The solution of restriction fragments chosen for sequence analysis was heated at 100°C for 2 min and applied directly to a cellogel strip. Electrophoresis conditions were identical to those for oligonucleotide separation[6]. The restriction fragment strands (4-6 µg) were eluted from the strips, dried, and dissolved in 50 µl water. Twenty µl of 0.1 M Tris-HCl buffer (pH 8.5) containing 50 mM $MgCl_2$ and 25 mM mercaptoethanol were added, followed by 20-30 µl of T4 endonuclease IV (20-30 units). The mixture was incubated at 37° or 45° for 12 hrs. The reaction was stopped by adsorption of the oligonucleotides to DEAE-Sephadex A-25, elution by 1.0 M TEAB (pH 8.0) and evaporation of the eluate to dryness.

5'-^{32}P-Terminal labeling of T4 endonuclease IV oligonucleotides. The oligonucleotides released by T4 endonuclease IV were incubated with 20 µg of alkaline phosphatase in 100 µl of 50 mM Tris-HCl, 10 mM $MgCl_2$ buffer (pH 8.9) at 55°C for 1 hr. The alkaline phosphatase was inactivated by treatment with alkali[26], the neutralized solution transfered to a tube containing 10^8 dpm of dried [γ-^{32}P]ATP, and 20 µl of polynucleotide kinase (0.01 units) and 1 µl of 1.4 M β-mercaptoethanol added. The reaction mixture was incubated for 2 hrs at 37°C, then applied directly to a 1 x 0.5 cm DEAE-Sephadex A-25 column. The excess ATP was eluted with 0.2 M TEAB (pH 8.0), the 5'-labeled oligonucleotides eluted with 1.3 M TEAB, collected and evaporated to dryness.

Fractionation of T4 endonuclease IV oligonucleotides. The 5' ^{32}P-labeled oligonucleotides were separated by ionophoresis-homochromatography[24] as described in detail elsewhere for pyrimidine oligonucleotides[6] except that the DEAE-cellulose plates were 20 x 40 cm and the eluate a 2% solution of yeast RNA partially degraded by alkali for 12 hours[27].

5' ^{32}P-labeled oligonucleotides separated by ionophoresis-homochromatography were visualized by radioautography, the spots were

scraped from the thin layer plate and the oligonucleotides eluted with 1.0 M TEAB (pH 8.0) and evaporated to dryness.

Partial digestion of oligonucleotides with snake venom phosphodiesterase. This was performed as described by Ling[25]. The hydrolysis products were separated by ionophoresis-homochromatography[24,6] and the oligonucleotides visualized by radioautography.

RESULTS

Since the sequence methods to be used were dependent on the measurement of a 5' ^{32}P-terminal label the specificity, exonucleolytic and phosphatase activities of the T4 endonuclease IV preparations were investigated. To determine the specificity of hydrolysis, S13 DNA was degraded with T4 endonuclease IV, the resulting oligonucleotides 5' ^{32}P-labeled then hydrolyzed to 5' mononucleotides. Separation of the 5' mononucleotides is shown in Fig. 1a. More than 95% of the ^{32}P label occurred in the

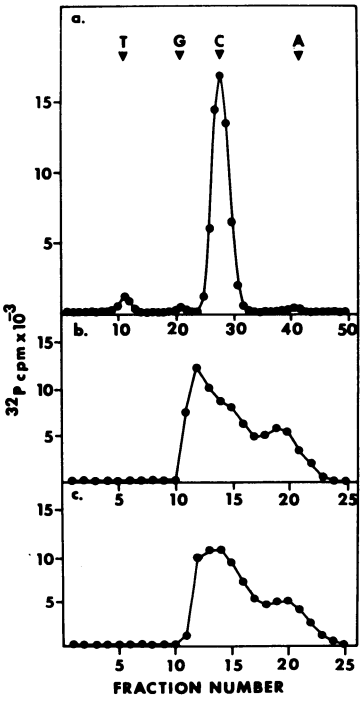


Figure 1
(a) 5' terminal nucleotide analysis of the oligonucleotides released by T4 endonuclease IV. The oligonucleotides were 5'^{32}P-labeled with [γ-^{32}P]ATP and polynucleotide kinase. Excess ATP was removed by chromatography on Sephadex G-25[28] and the oligonucleotides degraded to 5' mononucleotides with DNase A and snake venom phosphodiesterase[15]. The 5' mononucleotides were separated by Dowex-50 chromatography[29].
(b) Sephadex G-25 chromatography of oligonucleotides released by T4 endonuclease IV.
(c) Sephadex G-25 chromatography of oligonucleotides released by T4 endonuclease IV redigested with T4 endonuclease IV.

cytidylic acid fraction indicating that the T4 endonuclease IV preparation is highly specific in producing oligonucleotides with a 5' terminal C

residue. The 3' terminal specificity was deduced from the sequence analyses (see Discussion). To test the enzyme preparation for exonuclease and phosphatase activity, 5' ^{32}P-labeled T4 endonuclease IV oligonucleotides were redigested with T4 endonuclease IV, the enzyme inactivated by heat denaturation for 3 min at 100°C and the mixture chromatographed on Sephadex G-25 (Fig. 1c). The control experiment prior to redigestion with T4 endonuclease IV is shown in Fig. 1b. Comparison of the results shows a slight shift of ^{32}P label into shorter oligonucleotides based on the broadening of the larger peak (fractions 10-17). There was no change in the amount of label in fractions 18-25 relative to that in fractions 10 to 17 in either experiment. This indicates that the T4 endonuclease IV preparation was essentially free of phosphatase and exonuclease activity. Hydrolysis of uniformly ^{32}P-labeled S13 DNA with T4 endonuclease IV in the 'minus' exonuclease I assay (see Methods) did indicate the presence of a very small amount of 3' exonuclease activity.

Strand separation of fragments *Hae*III 8 and 9 and identification of strand polarity. The separation of uniformly ^{32}P-labeled strands of fragments *Hae*III 8 and 9 by electrophoresis on cellogel strips is shown in Figure 2.

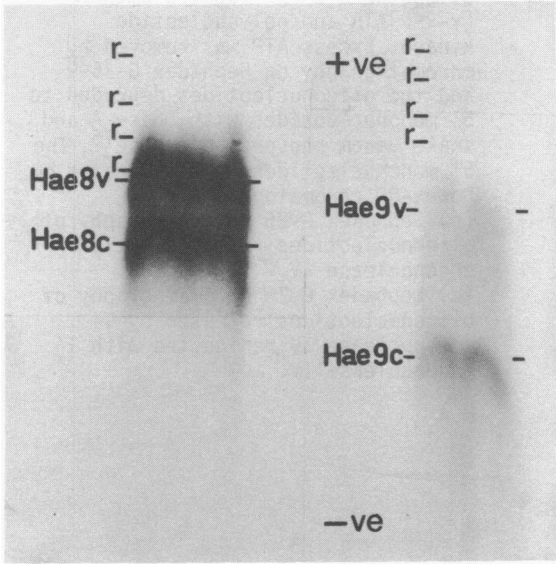


Figure 2. Radioautograms of the strand separations of *Hae*III 8 and 9 on cellogel strips. r indicates the positions of red marker dyes, v viral and c, complementary strands. Electrophoresis was for 90 mins, migration was approximately 20 to 25 cms from the origin.

To determine which strands were viral and which complementary in origin, they were eluted from the cellogel and hybridized to S13 viral DNA. The hybridization mixtures were fractionated on Bio-Gel A-15m columns. The

chromatographic profiles presented in Fig. 3 show that the electrophoretically slower moving strands of both *Hae*III 8 and 9 were eluted in the void volume having hybridized to the viral DNA, while the electrophoretically faster moving strands were retained on the column. Thus the slower moving strands are complementary strand in origin, and the faster moving strands viral.

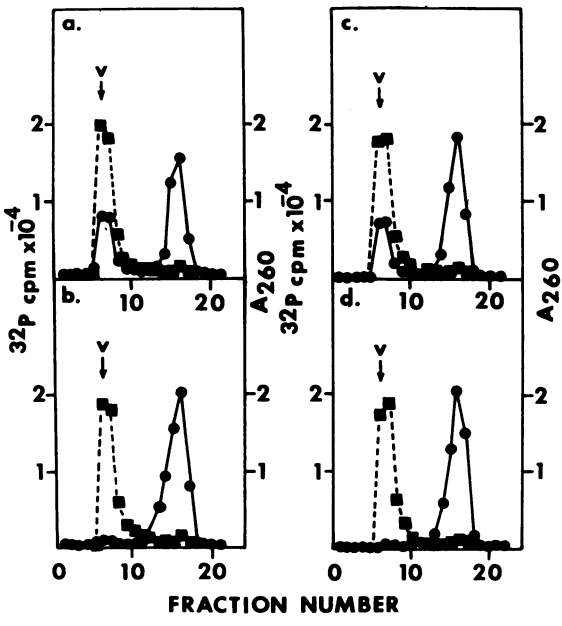


Figure 3. Hybridization of the ^{32}P-labeled separated strands of *Hae*III 8 and 9 to viral S13 DNA. Each of the separated strand preparations was mixed with 200 µg of S13 DNA and hybridized for 36 hrs at 60°C in a volume of 200 µl of 3 x SSC. The mixtures were applied to a 15 ml Bio-Gel A-15m column and eluted with 3 x SSC. a) *Hae*III 8 slow strand, b) *Hae*III 8 fast strand, c) *Hae*III 9 slow strand, d) *Hae*III 9 fast strand. ●—● ^{32}P-label of the separated strands; ■---■, absorbance at 260 nm of the viral DNA.

Sequence analysis of oligonucleotides released by T4 endonuclease IV.

Fig. 4 shows radioautograms of ionophoretic-homochromatographic separations of some of the 5' ^{32}P-labeled oligonucleotides released by T4 endonuclease IV from the separated strands of *Hae*III 9. The patterns show the divergences expected from products derived from separated DNA strands and indicate that strand separation is adequate since there is no obvious cross contamination. There were minor differences only in fingerprint patterns of degradations performed at 37°C compared to ones at 45°C[30] with no obvious shifts from longer to shorter oligonucleotides. Incubations at both temperatures were used to obtain oligonucleotides for sequence analyses. The separated 5'-^{32}P-labeled oligonucleotides were scraped from the TLC plates, each partially degraded with snake venom phosphodiesterase and the

products separated by ionophoresis-homochromatography. Radioautographs of some of the partial digest separations are shown in Fig. 5. The successive loss of nucleotides from the 3' end of an oligonucleotide results in a chromatographic fingerprint with characteristic differences in angles and distances between the particular oligonucleotide spots, dependent on the 3' terminal base which has been removed[31,32]. In general the nucleotides

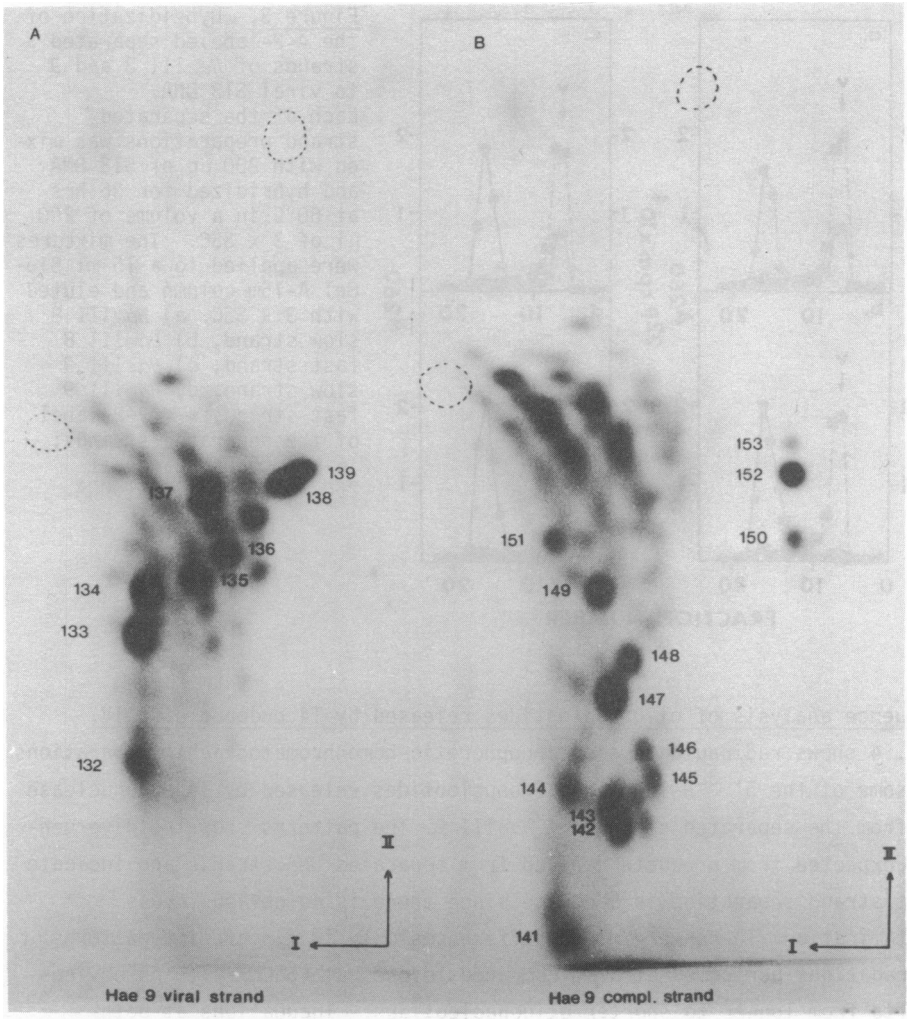


Figure 4. Radioautograms of two dimensional separations of the oligonucleotides released by T4 endonuclease IV from (a) viral and (b) complementary strands of *Hae*III 9. The dotted circles indicate the yellow dye marker. Dimension I, electrophoresis; Dimension II, homochromatography.

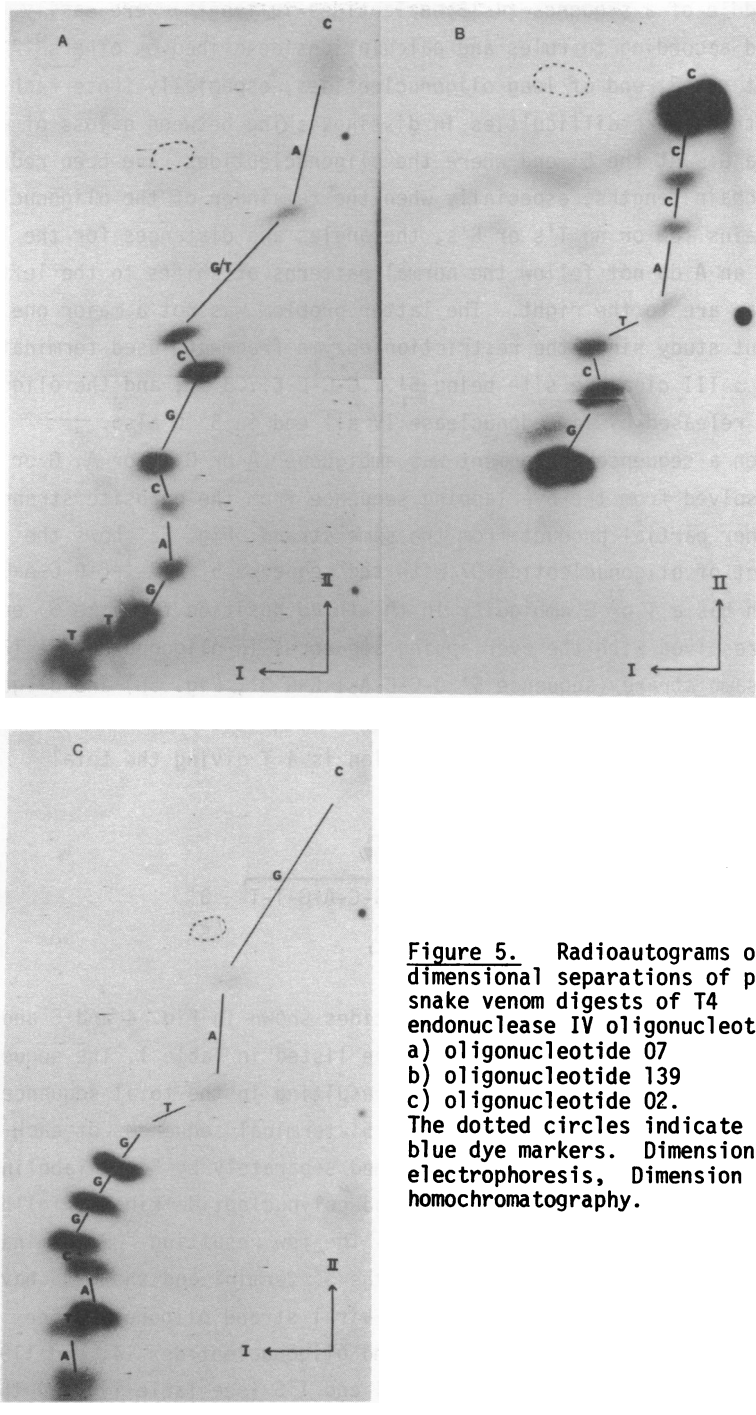
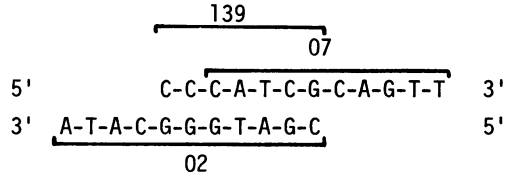


Figure 5. Radioautograms of two dimensional separations of partial snake venom digests of T4 endonuclease IV oligonucleotides.
a) oligonucleotide 07
b) oligonucleotide 139
c) oligonucleotide 02.
The dotted circles indicate the blue dye markers. Dimension I, electrophoresis, Dimension II, homochromatography.

in the middle of a sequence (4-15 nucleotides in length) were easily identified according to rules and calculations described by others[31,32]. However at the 3' end of long oligonucleotides, especially those rich in G and T, there were difficulties in distinguishing between a loss of an A *versus* a G. At the 5' end where the oligonucleotides have been reduced to short chain lengths, especially when the remainder of the oligonucleotide contains few or no T's or G's, the angles and distances for the loss of a C or an A do not follow the normal patterns of shifts to the left, but instead are to the right. The latter problem was not a major one in the present study since the restriction enzyme fragments used terminate in 5'C, the *Hae*III cleavage site being 5'...G-G↓C-C...3'[33], and the oligonucleotides released by T4 endonuclease IV all end in 5' C also.

When a sequence assignment was ambiguous (A or G, C or A, G or T) it was resolved from the overlapping sequence from the opposite strand or from another partial product from the same strand. Fig. 5a shows the fingerprint of oligonucleotide 07 with the sequence 5' C-A-$\frac{T}{G}$-C-G-C-A-G-T-T 3' which has a T or G ambiguity in the third position from the 5' end. This was resolved with the overlapping sequences in oligonucleotide 139 from the same strand (sequence 5' C-C-C-A-T-C-G 3', Fig. 5b) and oligonucleotide 02 from the complementary strand (sequence 5' C-G-A-T-G-G-G-C-A-T-A 3', Fig. 5c). The ambiguous position is a T giving the total sequence:



The sequences of the oligonucleotides shown in Fig. 4 and 5 and from other T4 endonuclease IV digests are listed in Table 1. The sequences were ordered on the basis of overlaps, resulting in the total sequence for *Hae*III 8 and 9 as shown in Fig. 6. The 5' terminal sequences of each restriction enzyme fragment were confirmed separately by 5'^{32}P-labeling of the separated strands with [γ-^{32}P]ATP and polynucleotide kinase, followed by degradation with T4 endonuclease IV. The few resulting ^{32}P-terminally labeled oligonucleotides were all from the 5' termini and shown to have the sequences corresponding to *Hae*III 9 viral strand oligonucleotides 06, 34 and 134, *Hae*III 8 complementary strand oligonucleotides 14, and 119 and viral strand oligonucleotides 7, 12, 111 and 115 (see Table 1). Further

Table 1 Sequences of the oligonucleotides released by T4 endonuclease IV from the viral and complementary strands of S13 *Hae*III 8 and 9.

*Hae*III 8 viral strand oligonucleotides

1.	------T-G-C-A-T-T-T-T-A-T-G
2 & 103.	C-G-C-C-G-C-C-A-G-T-T-A-A-A-T-A-G
3 & 107.	C-A-G-G-T-A-G-C-G-T-T-G-A
3.	C-A-T-T-T-T-A-T-G-C-G-G-A
4 & 104.	------C-A-T-T-T-T-A-T-G
5 & 109.	C-G-G-G-T-A-C-G-C-A-A-T
6.	C-A-G-G-T-A-G-C-G-T-T-G
7.	C-C-G-T-C-T-T-C-A-T-T-T
8.	C-A-A-A-A-T-A-C-G
9 & 112.	C-T-T-T-G-G-T-C-G-T
10 & 113.	C-G-G-G-T-A-C-G
11 & 114.	C-A-C-T-T-C-C-T-T
12 & 115.	C-C-G-T-C-T-T
13a.	C-G-G-T-G
105.	C-A-C-T-T-C-C-T-T-C-A-G-G-T-A-G
106.	C-C-C-T-A-A-C-T-T-T-G-G-T-C-G-T
108.	C-A-A-A-A-T-A-C-G-T-G-G
110.	C-A-A-A-A-T-A-C-G-T-G
	C-G-T-T-G-A-C-C-C-T-A-A
111.	C-C-G-T-C-T-T-C-A-T-T-T-C

*Hae*III 8 complementary strand oligonucleotides

14 & 119.	C-C-A-C-G-T-A-T-T-T-T-G-C-A-A-G
15.	C-G-C-T-A-C-C-T-G-A-A-G-G-A-A
16 & 120.	C-T-A-T-T-T-A-A-C-T-G-G-C-G-G
17.	C-C-A-C-G-T-A-T-T-T-T-G-C-A-A
	C-A-T-G-G-A-A-A-T-G-A-A-G-A
18.	C-A-C-C-G-C-A-T-G-G-A-A-A-T
	C-G-A-C-G-A-C-C-A-A-A-G-T-T
19.	C-G-T-A-T-T-T-T-G-C-A-A-G
20 & 121.	C-C-G-C-A-T-A-A-A-A-T-G
21 & 123.	C-C-A-C-G-T-A-T-T-T-T-G
22.	C-C-G-C-A-T-A-A-A-A-T
23.	C-G-A-T-T-G-C-G-T-A
24.	C-A-T-A-A-A-A-T-G
25.	C-G-A-T-T-G-C-G-T
26 & 126.	C-C-C-G-A-C-G-A
27.	C-G-T-A-T-T-T-T-G
28 & 127.	C-T-A-T-T-T-A-A
28 & 127.	C-T-G-G-C-G-G
29 & 128.	C-G-A-C-G-A-C
31.	C-G-T-A-C-C
122.	C-G-A-T-T-G-C-G-T-A-C-C-C-G
124.	C-G-A-T-T-G-C-G-T-A

*Hae*III 9 viral strand oligonucleotides

32.	C-A-C-G-T-T-C-T-G-G-T-T-G-G
33 & 133 & 05.	C-A-G-G-A-C-G-C-T-T-T-T-T
34 & 134 & 06.	C-C-T-T-A-T-G-G-T-T-A
35 & 135 & 07.	C-A-T-C-G-C-A-G-T-T
36 & 136.	C-A-G-T-A-T-G-C-C
37 & 137 & 09.	C-A-G-T-A-T-G
37 & 137 & 09.	C-A-G-T-T-C-G
39 & 138 & 08.	C-T-A-C-A-C-G
139.	C-C-C-A-T-C-G
132.	C-C-T-T-A-T-G-G-T-T-A-C-A-G-T-A-T-G
140.	C-A-G-G-A-C-G

*Hae*III 9 complementary strand oligonucleotides

41 & 143.	C-A-G-A-A-C-G-T-G-A-A-A-A-A-G-C-G-T
42.	C-C-T-G-C-G-T-G-T-A-G-C-G-A-A
43.	C-A-G-A-A-C-G-T-G-A-A-A-A-A
47.	C-T-G-T-A-A-C-C-A
48 & 153.	C-A-C-A-A-C-C-A-A
49 & 152 & 03.	C-A-A-C-C-A-A-C
50 & 153 & 04.	C-A-A-C-C-A-A
52.	C-G-A-T-G-G-G-C
53.	C-T-G-T-A-A-C
144.	---------T-A-G-C-G-A-A-C-T-G
147 & 01.	C-T-G-T-A-A-C-C-A-T-A-A-G-G
149 & 02.	C-G-A-T-G-G-G-C-A-T-A

The numbers 01 to 09, 1 to 53 and 103 to 153 refer to oligonucleotides from three different experiments. The underlined sequences are the 5' termini (see text). Sequences without a number are minor sequences present together with the sequence directly above.

confirmation of the sequence was by pyrimidine tract analyses. Radioautograms of uniformly ^{32}P-labeled pyrimidine oligonucleotides released by formic acid diphenylamine hydrolysis of fragments *Hae*III 8 and 9 and separated by ionophoresis-homochromatography are shown in Fig. 7a, b. The longest pyrimidine oligonucleotide present in *Hae*III 8 is C_4T_4 (nucleotides 38-45) and in *Hae*III 9, C_2T_5 (nucleotides 165-171).

The total sequence (Fig. 6) has two sequence assignments based on assumptions due to lack of satisfactory overlaps. Base 34 is the 5' end C in the complementary strand sequences of oligonucleotides 20, 22 and 121, while base 36 is the 5' end C of the viral strand oligonucleotides 11, 105 and 114. Since the first two bases at the 3' end of oligonucleotide 15 (complementary strand) could not be assigned with certainty, the only evidence for A in position 35 is provided by a minor sequence in oligonu-

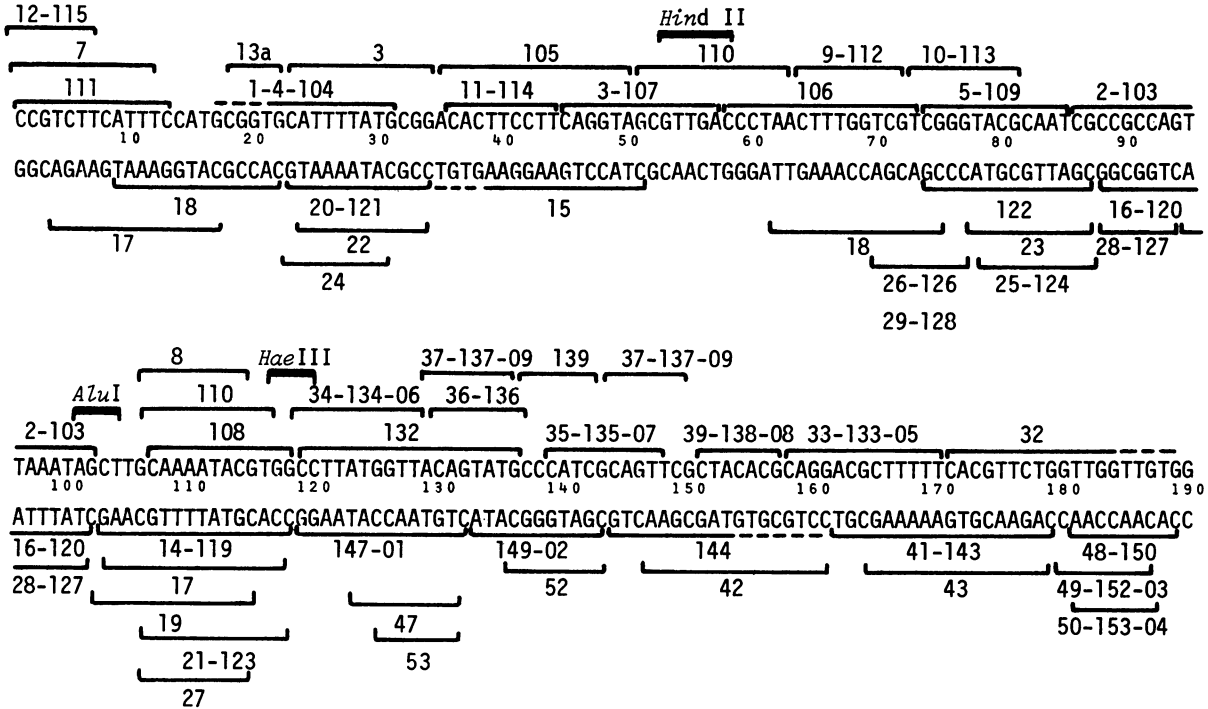


Figure 6. The sequence of restriction fragments *Hae*III 8 and 9. The brackets and numbers refer to the T4 endonuclease IV oligonucleotides listed in Table 1. The heavy brackets indicate restriction endonuclease cleavage sites.

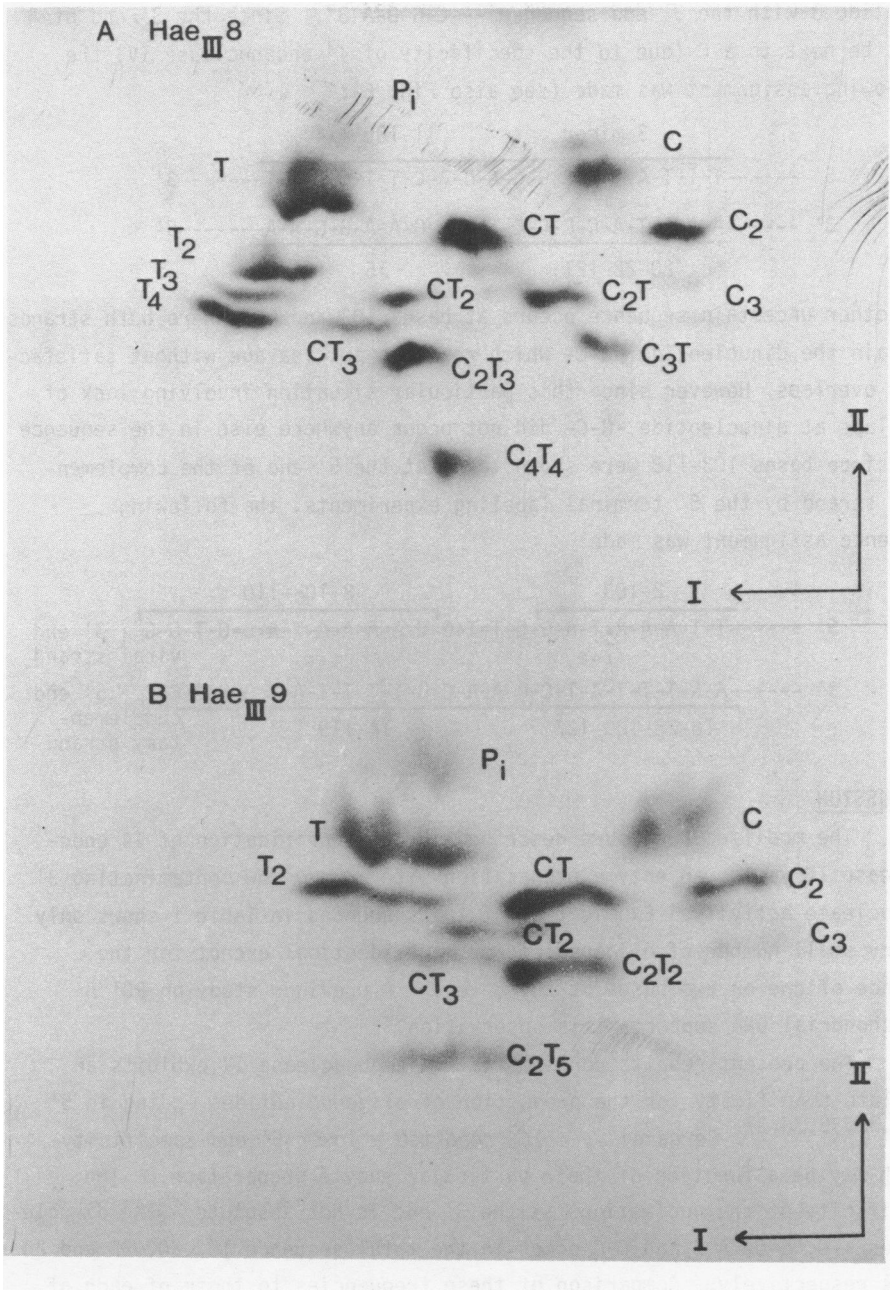
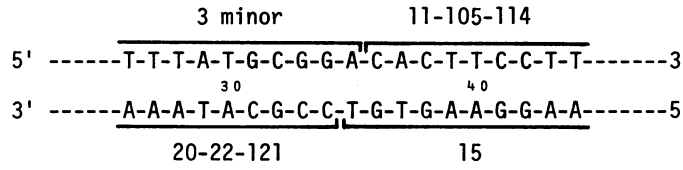
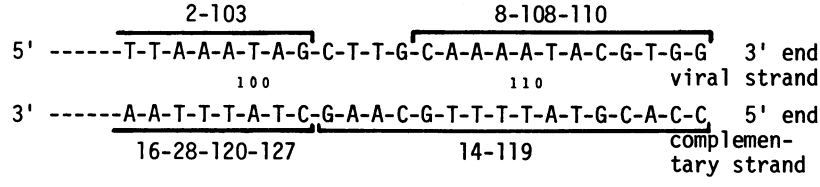


Figure 7. Radioautograms of pyrimidine cluster analyses of a) *Hae*III 8, b) *Hae*III 9. Pi is inorganic phosphate. Dimension I, electrophoresis; Dimension II, homochromatography.

cleotide 3 with the 3' end sequence ---C-G-G-A 3'. Since the 3' end of A must be next to a C (due to the specificity of T4 endonuclease IV) the following assignment was made (see also Fig. 6):



The other uncertain sequence occurs at bases 102 and 103 where both strands contain the dinucleotide -G-C- which resulted in cleavage without satisfactory overlaps. However since this particular situation involving lack of overlaps at dinucleotide -G-C- did not occur anywhere else in the sequence and since bases 103-118 were shown to be at the 5' end of the complementary strand by the 5' terminal labeling experiments, the following sequence assignment was made:



DISCUSSION

The modified procedure described for the purification of T4 endonuclease IV yields an enzyme preparation with a very low contaminating 3' exonuclease activity. Examination of the sequences in Table I shows only a very small number of oligonucleotides are identical except for the absence of one or two bases at the 3' end. A previous study on RD1 A mitochondrial DNA supports this observation[34].

The present results confirm that T4 endonuclease IV exhibits an absolute specificity for the production of oligonucleotides ending in 5' pC[19,32,35,36,37]. Bernardi *et al.*[38] reported a lower 5' end specificity which may be a function of their particular enzyme preparation. The specificity of the nucleotides at the 3' end is not absolute. The dinucleotides T-C, G-C, A-C and C-C occur in the total sequence 15, 30, 25 and 23 times respectively. Comparison of these frequencies to those of each of the four bases occurring at the 3' ends of the sequenced oligonucleotides which were obtained in high yield, shows that T4 endonuclease IV exhibits a preference for the cleavage of T-C and G-C bonds, over A-C bonds. The

C-C bonds are cleaved very infrequently. These results are in good agreement with previously published sequence data[32,36,37].

The DNA sequencing method used does not require the DNA to be uniformly ^{32}P-labeled. The strand separation can be performed after *in vitro* ^{32}P-labeling of the purified restriction fragments and carrier DNA or RNA is not required. After T4 endonuclease IV hydrolysis the termini of the oligonucleotides can be labeled with polynucleotide kinase[18] as in this study, or by terminal transferase[39]. The limiting factor in the procedure is the strand separation which depends on the difference in GT/AC content of the strands to be separated. Maniatis *et al.*[37] separated the strands of two restriction fragments, 30 and 45 base pairs in length, with 20% and 10% differences in GT content between the strands respectively. The present study extends the electrophoretic separation method to fragments 118 bases in length with 8% differences in GT content between the strands.

Examination of the sequence reveals an *Alu*I site at nucleotides 101-104 inclusive. The *Alu*I map of S13 RF DNA has not been determined. The existence of a *Hin*dII site in *Hae*III 8[7] is confirmed at positions 53-48 inclusive. Smith and Sinsheimer[40] investigated the 5' end sequences of *in vitro* transcription products from φX174 RF DNA and reported that the gene B transcript has the 5' end sequence AUCG(C). The transcript hybridized to φX174 *Hin*d fragment 8[11], which corresponds to S13 *Hin*d fragment 6, which in turn overlaps S13 *Hae*III 8 and 9[7]. The sequence ATCGC occurs twice in the *Hae*III 8 and 9 sequence (Fig. 6) at nucleotides 84-88 and 140-144. A high degree of symmetry occurs in the sequence prior to nucleotides 140-144 and there is also a sequence homologous to the canonical promoter site sequence T-A-T-Pu-A-T-Pu[41] at position 128-134. Thus the region sequenced may contain the promoter site for the transcription of gene B, assuming the 5' end of the S13 transcript is identical to that from φX174.

The genetic maps of S13 and φX174 are identical[5] and the gene A proteins of the two phages coelectrophorese on SDS gels and presumably are the same size[9]. Since φX174 *Hae*III 9 and 10 map entirely in gene A[42] and correspond to S13 *Hae*III 8 and 9[7] the two S13 fragments sequenced should contain the genetic code for part of the S13 gene A protein. Examination of the three possible reading frames in the viral sequence (Fig. 8) reveals that two of the three frames contain stop codons. Nonsense triplets occur at positions 49-51, 55-57, 61-63 and 100-103 in one frame and

at position 96-98 in the second. The amino acid sequence corresponding to the third reading frame is presented in Fig. 8. Twenty-seven of the 67

```
GGC CGT CTT CAT TTC CAT GCG GTG CAT TTT ATG CGG ACA CTT CCT TCA
                                    10
gly arg leu his phe his ala val his phe met arg thr leu pro ser

GGT AGC GTT GAC CCT AAC TTT GGT CGT CGG GTA CGC AAT CGC CGC CAG
          20                                              30
gly ser val asp pro asn phe gly arg arg val arg asn arg arg gln

TTA AAT AGC TTG CAA AAT ACG TGG CCT TAT GGT TAC AGT ATG CCC ATC
                              40
leu asn ser leu gln asn thr trp pro tyr gly tyr ser met pro ile

GCA GTT CGC TAC ACG CAG GAC GCT TTT TCA CGT TCT GGT TGG TTG TGG CCN
    50                                        60
ala val arg tyr thr gln asp ala phe ser arg ser gly trp leu trp pro
```

Figure 8. The sequence of the viral strands of *Hae*III 8 and 9 and the postulated protein sequence of gene A protein. The lines indicate nonsense codons (see text).

codons end in T, 15 in G, 15 in C and 7 in A. A similar unusual T distribution was observed by Sanger[43] in the codons in the F and G proteins of ϕX174 DNA. He suggested that the presence of U in the 3rd position may have an advantageous effect in translation, a possibility first noted by Denhardt and Marvin[44]. Other features of the predicted amino acid sequence are the five proline residues occurring at amino acid positions 15, 21, 41, 47 and 65 (Fig. 8). These would tend to prevent helix structure formation in this part of the protein molecule. Five arginine residues occur in a 7 amino acid sequence stretch in positions 25-31. This potentially highly-positively charged area of the protein is adjacent to a neutral area due to the aspargine and glutamine residues at positions 29, 32, 34, 37 and 38. If the protein does not have any helical content (due to the proline residues) the highly positively charged area would create a very efficient site for DNA binding. The A protein has a nickase function [45] and therefore must have a DNA binding site located somewhere in its structure. The unusual arginine distribution with its potential high positive charge would be an excellent candidate for this DNA binding site.

ADDENDUM

Since the completion of this paper the sequence of ϕX174 DNA has

been reported[46]. The ϕX174 sequence differs at positions 25, 44 and 64 (Fig. 6). The changes at 25 and 64 do not alter the predicted protein sequence. The change at position 44 would result in a threonine in ϕX174 instead of the serine in S13, a change which should not affect greatly the structure or the function of the protein. These results are in accord with the similarities expected of the two DNA's in this area[1,2,3].

ACKNOWLEDGEMENTS

FGG was the holder of a studentship from the Medical Research Council of Canada and the Conseil de la Recherche en Santé du Québec. The work was supported by grant MT 1453 from the Medical Research Council of Canada to JHS.

REFERENCES

1. Zahler, S.A. (1958) J. Bacteriol. 75, 310-315.
2. Tessman, E.S. and Shleser, R. (1963) Virology 19, 239-240.
3. Jeng, Y., Gelfand, D., Hayashi, M., Shleser, R. and Tessman, E.S. (1970) J. Mol. Biol. 49, 521-526.
4. Baker, R. and Tessman, I. (1967) Proc. Natl. Acad. Sci. USA 58, 1439-1445.
5. Benbow, R.M., Hutchison, C.A. III, Fabricant, J.D. and Sinsheimer,R.L. (1971) J. Virology 7, 549-558.
6. Harbers, B., Delaney, A.D., Harbers, K. and Spencer, J.H. (1976) Biochemistry 15. 407-414.
7. Grosveld, F.G., Ojamaa, K.M. and Spencer, J.H. (1976) Virology 71, 312-324.
8. Godson, G.N. (1976) Virology 75, 263-280.
9. Hayashi, M.N. and Hayashi, M. (1974) J. Virology 14, 1142-1151.
10. Tessman, I., Tessman, E.S., Pollock, T.J., Borrás, M.-T., Puga, A. and Baker, R. (1976) J. Mol. Biol. 103, 583-598.
11. Smith, L.H. and Sinsheimer, R.L. (1976) J. Mol. Biol. 103, 699-710.
12. Schendel, P.F. and Wells, R.D. (1973) J. Biol. Chem. 248, 8319-8321.
13. Garen, A. and Levinthal, C. (1960) Biochim. Biophys. Acta 38, 470-483.
14. Sulkowski, E. and Laskowski, M.Sr. (1971) Biochim. Biophys. Acta 240, 443-447.
15. Junowicz, E. and Spencer, J.H. (1970) Biochemistry 9, 3640-3648.
16. Lehman, I.R. and Nussbaum, A.L. (1964) J. Biol. Chem. 239, 2628-2636.
17. Warner, H.R. and Lewis, N. (1966) Virology 29, 172-175.
18. Richardson, C.C. (1965) Proc. Natl. Acad. Sci. USA 54, 158-165.
19. Sadowski, P.D. and Bakyta, I. (1972) J. Biol. Chem. 247, 405-412.
20. Smith, H.O. and Wilcox, V.W. (1970) J. Mol. Biol. 51, 379-391.
21. Schekman, R.W., Iwaya, M., Bromstrup, K. and Denhardt, D.T. (1971) J. Mol. Biol. 57, 177-199.
22. Middleton, J.H., Edgell, M.H. and Hutchison, C.A. III (1972) J. Virol. 10, 42-50.
23. Burton, K. (1967) in Methods in Enzymology, Vol. 12A, pp. 222-224, Academic Press, New York.
24. Brownlee, G.G. and Sanger, F. (1969) Eur. J. Biochem. 11, 395-399.
25. Ling, V. (1972) J. Mol. Biol. 64, 87-102.

26. Ho, N.W.Y. and Gilham, P.T. (1973) Biochim. Biophys. Acta 308, 53-58.
27. Jay, E., Bambara, R., Padmanabhan, R. and Wu, R.(1974) Nucleic Acid Res. 1, 331-353.
28. Delaney, A.D. and Spencer, J.H. (1976) Biochim. Biophys. Acta 435, 269-281.
29. Blattner, F.R. and Erickson, H.P. (1967) Anal. Biochem. 18, 220-227.
30. Ziff, E.B., Sedat, J.W. and Galibert, F. (1973) Nature, New Biol. 241, 34-37.
31. Bambara, R., Jay, E. and Wu, R. (1974) Nucl. Acids Res. 1, 1503-1520.
32. Galibert, F., Sedat, J. and Ziff, E. (1974) J. Mol. Biol. 87, 377-407.
33. Murray, K. and Morrison, A. unpublished results, reference in Wu, R., Bambara, R. and Jay, E. (1974) CRC Crit. Rev. Biochem. 2, 455-512.
34. Mol, J.N.M., Borst, P., Grosveld, F.G. and Spencer, J.H. (1974) Biochim. Biophys. Acta 374, 115-128.
35. Sadowski, P.D. and Hurwitz, J. (1969) J. Biol. Chem. 244, 6192-6198.
36. Proudfoot, N.J. (1976) Nucleic Acids Res. 3, 1811-1821.
37. Maniatis, T., Ptashne, M., Bachman, K., Kleid, D., Flashman, S, Jeffrey, A. and Maurer, R. (1975) Cell 5, 109-113.
38. Bernardi, A., Maat, J., de Waard, A. and Bernardi, G. (1976) Eur. J. Biochem. 66, 175-179.
39. Roychoudhury, R., Jay, E. and Wu, R. (1976) Nucleic Acid. Res. 3, 101-116 and 863-877.
40. Smith, L.H. and Sinsheimer, R.L. (1976) J. Mol. Biol. 103, 711-735.
41. Pribnow, D. (1975) Proc. Natl. Acad. Sci. USA 72, 784-788.
42. Weisbeek, P.J., Vereijken, J.M., Baas, P.D., Jansz, H.S. and van Arkel, G.A. (1976) Virology 72, 61-71.
43. Sanger, F. (1975) Proc. Royal Soc. London 191, 317-333.
44. Denhardt, D.T. and Marvin, D.A. (1969) Nature 221, 769-770.
45. Henry, T.J. and Knippers, R. (1974) Proc. Natl. Acad. Sci. USA 71, 1549-1553.
46. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A. III, Slocombe, P.M. and Smith, M. (1977) Nature 265, 687-695.