Volume 10 Number 19 1982 Nucleic Acids Research

# Nucleotide sequence of the early genes 3 and 4 of bacteriophage $\phi$29

Cristina Escarmís and Margarita Salas

Centro de Biología Molecular (CSIC-UAM), Facultad de Ciencias, Universidad Autónoma, Canto Blanco, Madrid-34, Spain

Received 12 August 1982; Accepted 14 September 1982

ABSTRACT

The nucleotide sequence of an early region of the Ø29 genome has been determined. The sequenced region includes genes 3 and 4, which code for the protein covalently linked to the 5' ends of Ø29 DNA and the protein involved in the control of late transcription, respectively. The position and nature of the mutations of mutants sus3(91) and sus4(56) has also been determined.

## INTRODUCTION

A novel mechanism for the initiation of replication of linear DNAs having a protein covalently bound to the 5' ends, namely bacteriophage Ø29 DNA and adenovirus DNA, involves the priming of DNA chain elongation by a deoxynucleotide covalently attached to the terminal protein (1-4). In both cases, bacteriophage Ø29 and adenovirus, the priming protein is covalently bound to the termini of replicating DNA in vivo (5-10). The above model has been supported recently by the demonstration of the formation in vitro of a covalent complex between the Ø29 terminal protein p3 and dAMP, the 5' terminal nucleotide at both ends of Ø29 DNA (11)and between the precursor of the adenovirus terminal protein (pTP) and dCMP, the 5' terminal nucleotide at the ends of adenovirus DNA (12-15, P.C. Van der Vliet, personal communication).

Fragment Hind III G from the left early region of Ø29 DNA has been cloned and identified as containing gene 3, except the last 15 base pairs, and gene 4 (16). Gene 3 codes for the 5' terminal protein (17) and gene 4 for the protein involved in the control of late transcription (18).

As a first step to understand the relation between structure and function of the 5' terminal protein of Ø29 DNA, it is important to know the sequence of the gene coding for protein p3. We



0305-1048/82/1019-5785$ 2.00/0

report the sequence of an early region of the Ø29 genome, namely the Hind III G fragment plus 286 adjacent nucleotides to the left, containing genes 3 and 4 and probably the beginning of gene 2. We have also determined the site and nature of the mutation of mutants sus3(91) and sus4(56) (19). Features of the nucleotide sequence and of the derived amino acid sequences are discussed.

## MATERIALS AND METHODS

### Nucleoside triphosphates and enzymes

{$\gamma$-$^{32}$P}ATP (carrier free) was prepared according to an unpublished procedure of I. Kennedy (C. Weissmann, personal communication) or obtained ($\sim$3000 Ci/mmol) from the Radiochemical Centre (Amersham). {$\alpha$-$^{32}$P}dATP ($\sim$400 Ci/mmol) was from the Radiochemical Centre (Amersham). Polynucleotide kinase and restriction enzymes were from New England BioLabs; alkaline phosphatase from calf intestine and DNA polymerase (Klenow fragment) were from Boehringer Mannheim.

### Preparation of Ø29 DNA and restriction fragments

The DNA was extracted from the phage Ø29 delayed-lysis mutant sus14(1242) (20) or from mutants sus3(91) or sus4(56) (19) as described (21).

DNA restriction fragments were separated by electrophoresis on 3.5, 5 or 8% polyacrylamide gels in 0.1 M Tris-borate, pH 8.3, 2 mM EDTA (TBE). The DNA was extracted from the gel and purified as described (21). In some experiments, labelled DNA was extracted from the gel by electroelution into a dialysis bag.

### Labelling of the DNA restriction fragments and sequencing

Labelling at the 5' ends with polynucleotide kinase and {$\gamma$-$^{32}$P}ATP was as described (21). The 5' recessed ends were first converted to flush ends using the 3' exonuclease activity of DNA polymerase (Klenow fragment) and then labelled with polynucleotide kinase. Labelling at the 3' ends with DNA polymerase (Klenow fragment) and {$\alpha$-$^{32}$P}dATP was as described (21). The labelled DNA restriction fragments were either cut with a second restriction enzyme or the DNA bands were separated as described (22).

DNA sequencing was done essentially by the Maxam and Gilbert method (22), except that the G + A reaction was carried out by treatment of the DNA with formic acid (23). The slab gels (20%

or 8% polyacrylamide) were 400 x 300 x 0.5 mm and had 10 mm slots. The 8% polyacrylamide gels were fixed immediately after the run with 10% acetic acid and dried as described (24) except that the gel was transferred to a 3 MM paper and dried under vacuum. Autoradiography was carried out with intensifying screens at -70°C.

Computer analysis of DNA and protein sequences was made in a PDP-11/45 minicomputer from Digital Equipment Corp. with a DOS/BATCH operating system.

## RESULTS AND DISCUSSION

### Sequence of the region of phage Ø29 DNA that codes for proteins p3 and p4

Genes 3 and 4 have been mapped in the restriction fragment Hind III G (see Fig. 1) (16). Therefore, we sequenced the Hind III G fragment and some neighbouring regions following the strategy detailed in Fig. 1. All restriction sites that were used to label the DNA were sequenced through from other labelled sites. The sequence of the two DNA strands was determined.

The sequence of the Hind III G fragment (1150 nucleotides) and 286 nucleotides from the right end of the Hind III B fragment is given in Fig. 2. Only the sequence of the nonsense strand (18), i.e. equivalent to the mRNA sequence, is given, written in the sense of translation and numbered starting at the right end of the Hind III G fragment. In Fig. 3 the distribution of initiation (AUG) and stop codons in the three possible reading frames is shown. We can see three open reading frames. In frame 1 there is one starting at nucleotide 102 and ending at nucleotide 477 which would be translated into a protein of 125 amino acids and another one starting at nucleotide 1182 and whose stop codon is not included in the sequenced region. No open reading frame exists in frame 2. In frame 3 there is an open reading frame starting at nucleotide 368 and ending at nucleotide 1166, which could code for a protein of 266 amino acids. This open reading frame partially overlaps with that in frame 1 coding for a 125 amino acids protein. The protein of 266 amino acids, deduced from the nucleotide sequence, will have a molecular weight of 31049 daltons that is in good agreement with the estimated molecular weight of 27 000 for the terminal protein p3 in SDS polyacrylami-

de gels (17,29,30) as well as with that of one of the proteins synthesized in vivo from cloned Hind III G fragment identified as protein p3 since it was not synthesized from a cloned fragment from a sus3 mutant (16). In addition, preliminary sequence analysis of the N-terminal amino acids of protein p3 labelled with $^{3}$H-arginine, indicated that arginine was present in the second, fifth and seventh position (Wittmann-Liebold,B., unpublished results) which will agree with the protein sequence deduced from the DNA sequence if the F-Met had been eliminated in the mature protein.

The protein of 125 amino acids, deduced from the nucleotide sequence, will have a molecular weight of 15 133, in good agree-

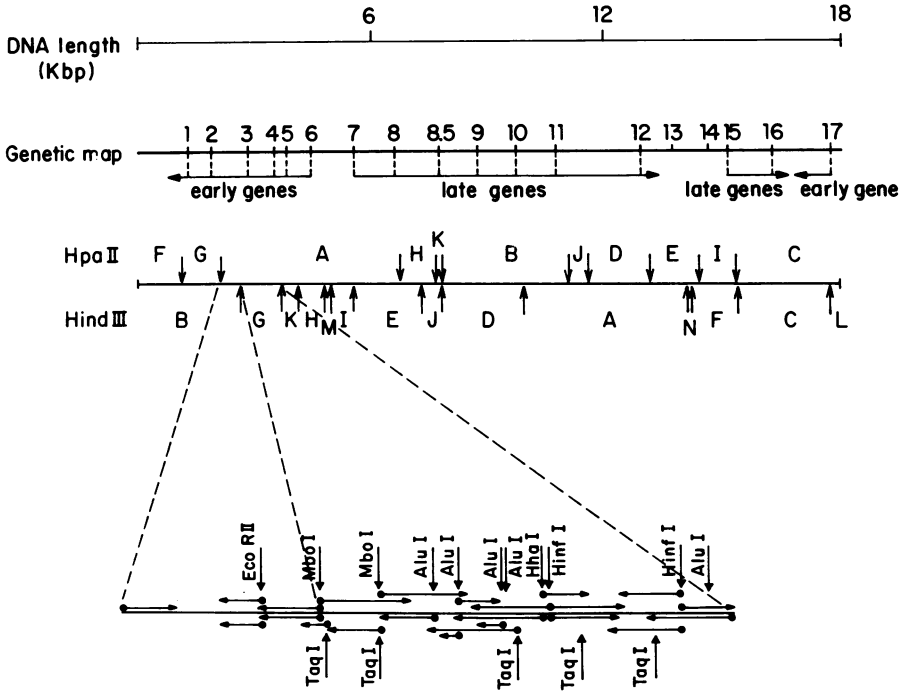



Fig. 1. Genetic and physical map of Ø29 DNA. The genetic map was adapted from Mellado et al. (25) and Reilly et al. (26), taking into account that gene 6 is contained in fragment Hind III H (27) and genes 3 and 4 are contained mostly in fragment Hind III G (this paper and 16). The arrows indicate the direction and extent of transcription from early and late promoters (18). The physical map is from our unpublished results and from Yoshikawa and Ito (28). DNA sequencing strategy (bottom). The horizontal arrows indicate the direction and extent of sequence determination.

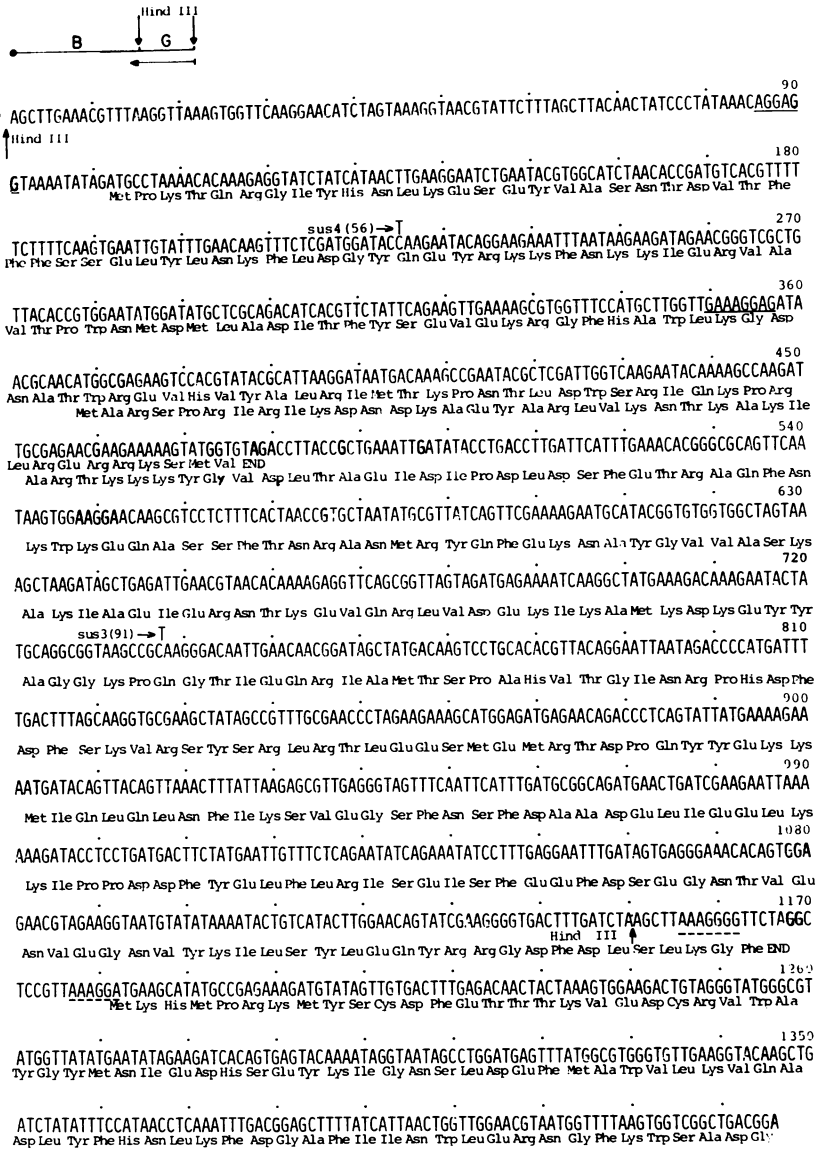



Fig. 2. Nucleotide sequence of fragment Hind III G and part of Hind III B. The deduced amino acid sequence for proteins p4, p3 and part of p2 is given and the probable Shine-Dalgarno sequence for proteins p4 and p3 is underlined. The two more likely Shine-Dalgarno sequences for protein p2 are underlined with discontinuous lines (see text). The position of the mutation in mutants sus4(56) and sus3(91) is also shown.

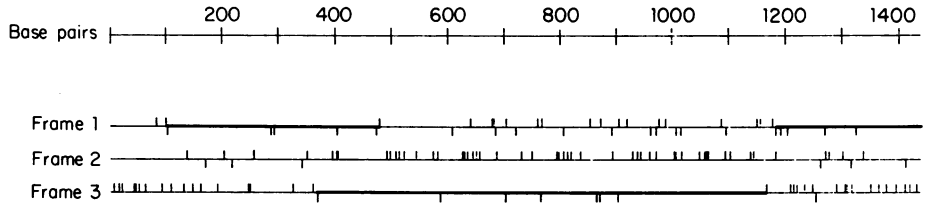



Fig. 3. Distribution of initiation codons (AUG), and stop codons in the three possible reading frames of the sequence shown in Fig. 2. Upper vertical bars are for nonsense codons and underneath vertical bars are for ATG triplets. Heavy lines show the open reading frames.

ment with the estimated molecular weight of 12 500 for the protein translated *in vivo* from the cloned Hind III G fragment, identified as p4 since it was not produced when the fragment cloned was from a *sus*4 mutant (16).

The open reading frame starting at nucleotide 1182 could code (due to its position in the genome) for the amino terminal region of protein p2.

DNA sequence analysis of mutants sus3(91) and sus4(56)

To further identify the sequences of the two proteins, of 266 and 125 amino acids, as those of p3 and p4, respectively, the sequence of the corresponding region from the DNA of mutants *sus*3(91) and *sus*4(56) was determined. For gene 3 the wild-type and *sus*3 sequences differed in only one nucleotide. At position 737 (Fig. 2) in the wild-type sequence a C-G pair is present, while in the mutant sequence a T-A pair is found (Fig. 4). This mutation gives rise to the change of a CAA triplet (Gln) into a nonsense UAA triplet (see Fig. 2). For gene 4, the wild-type and *sus*4 sequences differed in only one base, at position 225 (Fig. 2), in which a C-G pair is present in the wild-type DNA and a T-A pair in the mutant DNA (Fig. 4). This mutation gives rise, as in the case of mutant *sus*3(91), to the change of a CAA triplet (Gln) into a nonsense UAA triplet (see Fig. 2). The *sus*4(56) mutation is located in a portion of p4 that does not overlap with p3, therefore will not affect protein p3 function. Since both mutants were obtained by treatment of the phage with hydroxylamine (19), a C-G → T-A transition is the expected change.

During the preparation of this manuscript Yoshikawa and Ito published the nucleotide sequence of the first 5708 bp from the left end of Ø29 DNA (28). The nucleotide sequence assigned there to gene 3, from nucleotides 3686 to 2886 corresponds exactly to the sequence that we demonstrate is that of gene 3. On the contrary, the sequence reported in this paper for gene 4 does not correspond to the reading frame tentatively assigned to gene 4 by Yoshikawa and Ito (28) but to the open reading frame extending from nucleotides 3952 to 3575.

As mentioned above, both mutants sus3(91) and sus4(56) have ochre mutations. This was unexpected, since they are suppressed by the B. subtilis suppressor strains $su^{+3}$ and $su^{+44}$ thought to be suppressors of ochre and amber mutations, respectively (31). We do not have an explanation for this result at present, al-

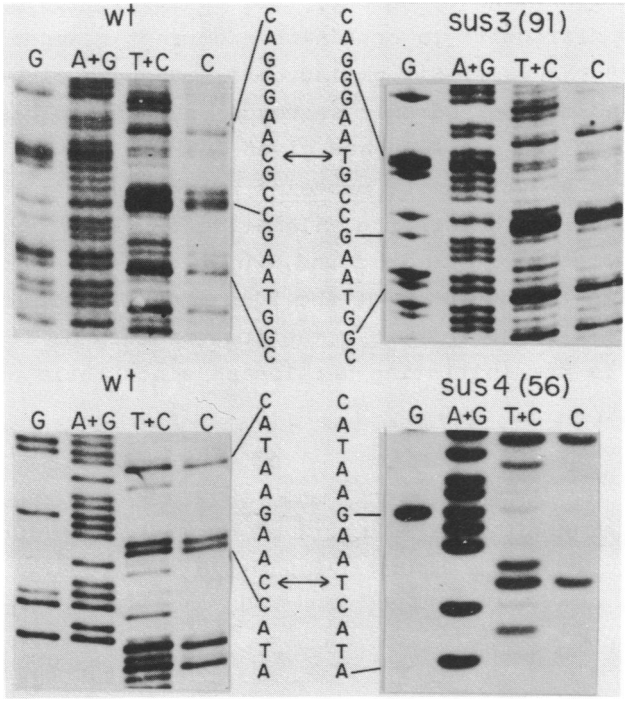



Fig. 4. Sequence analysis of nucleotides 727-744 in the nonsense strand of wild-type and mutant sus3(91) DNAs (upper part) and of nucleotides 221-233 in the nonsense strand of wild-type and mutant sus4(56) DNAs (lower part).

though it suggests that if the rules for the suppression of nonsense mutations stablished for E. coli are valid in B. subtilis, the two B. subtilis suppressor strains $su^{+3}$ and $su^{+44}$ are both ochre suppressors. Obviously the sequencing of Ø29 mutants genetically classified as ochre (31) will be necessary. On the other hand, a possible role played in suppression by bases adjacent to the nonsense triplet (32,33) has to be kept in mind.

Features of the nucleotide sequence

A close examination of the DNA sequence shown in Fig. 2 revealed the existence of eleven different octanucleotides repeated twice, one octanucleotide repeated three times, three nonanucleotides and three decanucleotides repeated twice. A high number of repetitions is also found in the sequence of the 5708 nucleotides at the left end of Ø29 DNA. These repetitions are statistically highly unprobable, and we do not know the significance of them, if any. Nevertheless, they could provide an explanation for the difficulty in obtaining a correct reannealing of denatured Ø29 DNA (21). Six palindromic sequences eight nucleotides long and one palindromic sequence ten nucleotides long are also present in the DNA sequence given in Fig. 2.

Shine-Dalgarno sequences for proteins p3 and p4

Ten nucleotides before the initiating triplets for p3 and p4 a Shine-Dalgarno sequence is found, which shows a strong complementarity with the sequence of the 3' end of B. subtilis 16S rRNA (27) (Fig. 5). For the hypothetical protein p2 it is not clear which is the initiating methionine since there are 3 me-

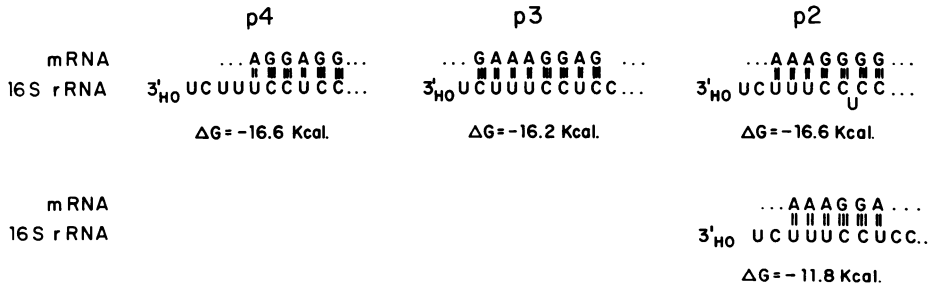



Fig. 5. Possible Shine-Dalgarno sequences for proteins p4, p3 and p2. The free energies of formation of the Shine-Dalgarno pairing were calculated following the rules of Tinoco et al.(34).

thionines close to the possible $NH_2$ terminal end. Taking into account both the free energy of formation of the Shine-Dalgarno pairing and the spacing between the Shine-Dalgarno region and the initiating codon (35) we think that the more probable Shine-Dalgarno sequences for protein p2 are the ones shown in Fig. 5 and Fig. 2. The first sequence, AAAGGGG shows a high degree of complementarity with the 3' end of 16S rRNA but the spacing is very high, 19 nucleotides. The second sequence, AAAGGA shows less complementarity with the 16S rRNA but the spacing is 8, closer to the optimal (35). The possible Shine-Dalgarno sequence for the initiation at the third methionine, AGAAAG, does not show very good complementarity with the 16S rRNA and in addition it is too close to the initiator AUG. The free energies of formation of the Shine-Dalgarno pairing for proteins p4, p3 and p2 (see Fig. 5) are within the range of those calculated for other initiation sites recognized by B. subtilis ribosomes (27,35) and support the hypothesis that an extensive mRNA-rRNA interaction is a requirement for efficient translation by B. subtilis ribosomes (35).

Genes 3 and 4 have an overlapping region of 109 nucleotides (see Fig. 2). As it is the case with other bacteriophages, the same DNA sequence is used to code for two different proteins in two different frames, being the Shine-Dalgarno sequence of gene 3 part of gene 4.

Relative codon usage for proteins p3 and p4

The relative codon usage for proteins p3 and p4 is shown in Fig. 6. Taking into account only those amino acids which are used more than six times in proteins p3 and p4 together, a clear preference exists for U over C as the third base of the triplet in the amino acids which are encoded by 4 codons (quartets), being the preference less strong in the amino acids which are encoded by 2 codons (duets). For the amino acids coded for by a triplet with an A or a G as the third base there is not a clear pattern of preferences, except for threonine, arginine, glutamic acid and serine where the A is used more than the G. These bias cannot be explained only by the base composition of the DNA (T/C = 1.68; A/G = 1.45) since in the case of quartets the average use of U over C is 5.25, and the average use of A over G (for Thr, Arg, Glu and Ser) is 3.65. A similar bias can be observed

in the relative codon usage by 13 mRNAs sequenced from different double-stranded DNA phages (36) with some exceptions. The relative use of quartets and duets for amino acids that can be encoded by 6 codons is similar to that observed in the 13 mRNAs of double stranded DNA phages (36). Only one codon is not used in p3 and p4 : UCG for serine, which is also not frequently used in the 13 mRNAs of double-stranded DNA phages (36). These observed bias in the choice of the third base in a codon in proteins p3 and p4 cannot be explained by dinucleotide preferences (37) since the observed dinucleotide frequency in the sequence reported here agrees with the theoretical one calculated from the nucleotide composition. It remains unclear which is the significance of these bias in the relative codon usage (36-38).

Features of the sequences of proteins p3 and p4

According to Hopp and Woods the point of highest local average hydrophilicity, is located in or immediately adjacent to an antigenic determinant (39). In Fig. 7 we show the hydrophilicity profile of proteins p3 and p4, taking for each point the average value of an hexapeptide (i-2, i-1, i, i+1, i+2, i+3). For p3 the hexapeptide with highest average hydrophilicity is 95-Glu-Arg-Asn-Thr-Lys-Glu. For p4 the hexapeptide with higher average hydrophilicity is 118-Arg-Glu-Arg-Arg-Lys-Ser. The antigenic determinant predicted for p3 is located in the amino terminal moiety of the molecule, in such a position that the peptide produced by

| p3 | U | | C | | A | | G | | |
|---|---|---|---|---|---|---|---|---|---|
| U | PHE | 9 | SER | 1 | TYR | 9 | CYS | 0 | U |
| | | 6 | | 2 | | 4 | | 0 | C |
| | LEU | 5 | | 4 | | | TRP | 1 | A |
| | | 4 | | 0 | | | | | G |
| C | | 2 | | 5 | HIS | 1 | | 5 | U |
| | LEU | 1 | PRO | 1 | | 1 | ARG | 1 | C |
| | | 2 | | 1 | GLN | 3 | | 4 | A |
| | | 2 | | 1 | | 7 | | 3 | G |
| A | | 7 | | 1 | ASN | 8 | SER | 5 | U |
| | ILE | 2 | THR | 2 | | 5 | | 6 | C |
| | | 9 | | 8 | LYS | 11 | ARG | 5 | A |
| | MET | 7 | | 1 | | 16 | | 1 | G |
| G | | 3 | | 8 | ASP | 10 | | 6 | U |
| | VAL | 1 | ALA | 2 | | 8 | GLY | 1 | C |
| | | 4 | | 4 | GLU | 19 | | 2 | A |
| | | 4 | | 5 | | 8 | | 2 | G |

| p4 | U | | C | | A | | G | | |
|---|---|---|---|---|---|---|---|---|---|
| U | PHE | 4 | SER | 2 | TYR | 3 | CYS | 0 | U |
| | | 3 | | 0 | | 4 | | 0 | C |
| | LEU | 1 | | 3 | | | TRP | 4 | A |
| | | 5 | | 0 | | | | | G |
| C | | 0 | | 1 | HIS | 2 | | 1 | U |
| | LEU | 3 | PRO | 0 | | 1 | ARG | 0 | C |
| | | 0 | | 1 | GLN | 3 | | 3 | A |
| | | 0 | | 2 | | 0 | | 1 | G |
| A | | 0 | | 0 | ASN | 3 | SER | 2 | U |
| | ILE | 2 | THR | 1 | | 4 | | 0 | C |
| | | 3 | | 4 | LYS | 4 | ARG | 4 | A |
| | MET | 5 | | 3 | | 8 | | 2 | G |
| G | | 2 | | 2 | ASP | 5 | | 2 | U |
| | VAL | 3 | ALA | 0 | | 1 | GLY | 0 | C |
| | | 1 | | 4 | GLU | 9 | | 2 | A |
| | | 2 | | 0 | | 0 | | 0 | G |

Fig. 6. Relative codon usage for proteins p3 and p4.

the mutant sus3(91) will still have it and, therefore, should react with anti-p3 serum as it happens to be the case (Mellado, R.P. and Salas, M., unpublished results). The predicted antigenic determinant for p4 is located at the carboxyl end of the protein.

The secondary structure of proteins p3 and p4 was predicted using a computerized program according to Chou and Fasman (40) with slight modifications (41) and is shown in Fig. 8. It can be observed that both proteins are highly ordered : p4 shows 44% amino acids involved in α-helices and 23% amino acids involved in β-strands; p3 shows 67% amino acids in α-helices and 5% in β-strands. In the same figure the distribution of charged and hydrophobic residues along p3 and p4 is also shown. Both proteins, p3 and p4 are quite hydrophobic, 24-37% more than the average

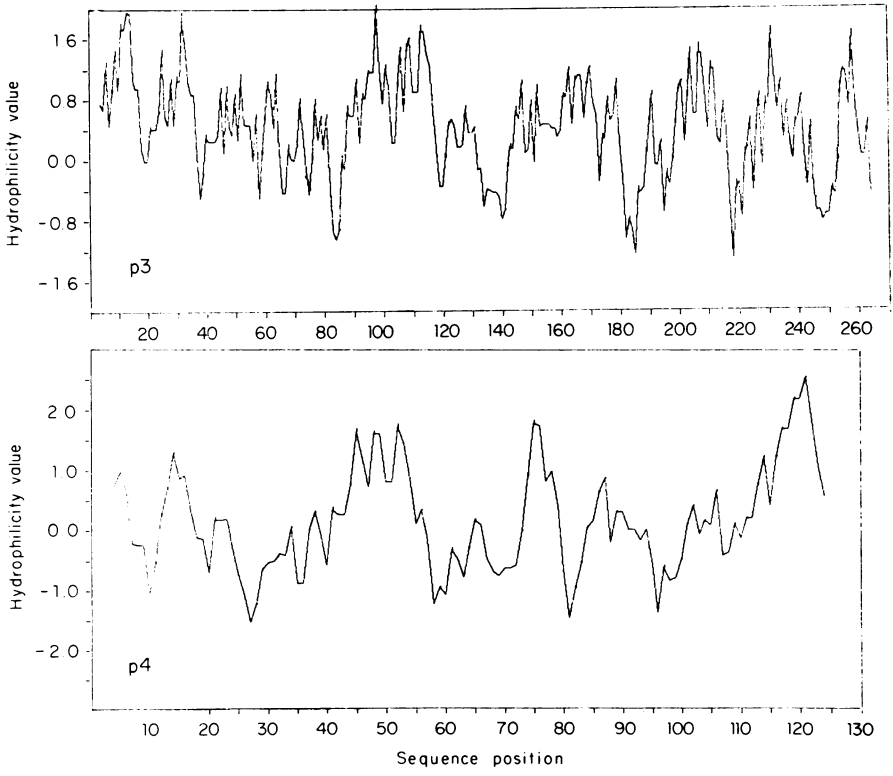



Fig. 7. Hydrophilicity profile of proteins p3 and p4. The hexapeptide average value (see text) is plotted versus position along the amino acid sequence.

for 185 other proteins (42). Protein p4 is, in addition, very basic. Protein p3 shows a cluster of basic amino acids at the amino end, then the number of positive charges diminishes in the middle of the molecule and near the carboxyl terminus a cluster of negative charges is present. We still do not know which serine residue binds to the initiating dAMP. We could speculate that a good hydrophobic environment to bind the nucleotide (43) is present near the carboxyl terminus and that any other part of the molecule with positive charges could bind to the DNA strand that will serve as template.

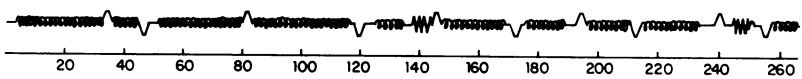
The sequence of the adenovirus 5' terminal protein has also been reported recently (44). We have not found any similarity

p3

Residue distribution

| Amino acids | 20 | 40 | 60 | 80 | 100 | 120 | 140 | 160 | 180 | 200 | 220 | 240 | 266 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hydrophobic (%) | 25 | 25 | 30 | 20 | 25 | 35 | 20 | 30 | 35 | 35 | 45 | 30 | 42 |
| Charged (%) | 45 | 40 | 35 | 30 | 35 | 45 | 20 | 40 | 40 | 15 | 45 | 35 | 31 |
| Net charge | +3 | +6 | -3 | +2 | +1 | +1 | +2 | +4 | -2 | -1 | -5 | -5 | 0 |

Predicted secondary structure




p4

Residue distribution

| Amino acids | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 125 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hydrophobic (%) | 30 | 30 | 40 | 50 | 30 | 40 | 50 | 30 | 30 | 50 | 30 | 26 |
| Charged (%) | 30 | 30 | 10 | 30 | 50 | 30 | 20 | 50 | 20 | 40 | 20 | 53 |
| Net charge | +3 | -1 | -1 | -1 | +3 | +1 | -2 | +1 | 0 | +2 | 0 | +6 |

Predicted secondary structure

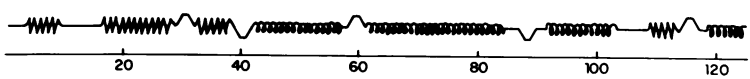



Fig. 8. Distribution of hydrophobic residues (Met, Val, Ile, Leu, Phe, Tyr and Trp) and charged groups (Arg, Lys, His, Asp and Glu) along the sequence of proteins p3 and p4. The lower portions summarize the predicted secondary structure for the two proteins using the Chou and Fasman method (40) (see text) : coil, α-helix; sawtooth pattern, β-strand; straight line, aperiodic structure, arch, β-turn.

between that sequence and the sequence of p3. For the adenovirus terminal protein the position of the serine residue that binds to the dCMP is known and in agreement with the above speculation it is located in a hydrophobic environment (44).

ACKNOWLEDGEMENTS

We are grateful to M. Caballero for carrying out the computer work. We also thank J.M. Rojas for his contribution in the initial experiments of this work. This investigation has been aided by research grant 1 R01 GM27242-03 from the National Institutes of Health and by grants from the Comisión Asesora para el Desarrollo de la Investigación Científica y Técnica and Fondo de Investigaciones Sanitarias.

REFERENCES

1 Inciarte, M.R., Salas, M. and Sogo, J.M. (1980) J. Virol.34, 187-199.
2 Mellado, R.P., Peñalva, M.A., Inciarte, M.R. and Salas, M. (1980) Virology 104, 84-96.
3 Harding, N.E. and Ito, J. (1980) Virology 104, 323-338.
4 Rekosh, D.M.K., Russell, W.C., Bellet, A.J.D. and Robinson, A.J. (1977) Cell 11, 283-295.
5 Sogo, J.M., García, J.A., Peñalva, M.A. and Salas, M. (1982) Virology 116, 1-18.
6 Kelly, T.J.Jr. and Lechner, R.L. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 721-728.
7 Robinson, A.J., Bodnar, J.W., Coombs, O.H. and Pearson, G.D. (1979) Virology 96, 143-158.
8 Stillman, B.W. and Bellett, A.J.D. (1979) Virology 93, 69-79.
9 Van Weilink, P.S., Naaktgeboren, N. and Sussenbach, J.S.(1979) Biochim. Biophys. Acta 563, 89-99.
10 Reiter, T., Fütterer, J., Weingärtner, B. and Winnacker,E.L. (1980) J. Virol. 35, 662-671.
11 Peñalva, M.A. and Salas, M. (1982) Proc. Natl. Acad. Sci.USA, in press.
12 Lichy, J.H., Horwitz, M.S. and Hurwitz, J. (1981) Proc. Natl. Acad. Sci. USA 78, 2678-2682.
13 Pincus, S., Robertson, W. and Rekosh, D. (1981) Nucl. Acids Res. 9, 4919-4938.
14 Challberg, M.D., Ostrove, J.M. and Kelly, T.J.,Jr. (1982) J. Virol. 41, 265-270.
15 Tamanoi, F. and Stillman, B.W. (1982) Proc. Natl. Acad. Sci. USA 79, 2221-2225.
16 Mellado, R.P. and Salas, M. (1982), accompanying paper.
17 Salas, M., Mellado, R.P., Viñuela, E. and Sogo, J.M. (1978) J. Mol. Biol. 119, 269-291.
18 Sogo, J.M., Inciarte, M.R., Corral, J., Viñuela, E. and Salas, M. (1979) J. Mol. Biol. 127, 411-436.

19 Moreno, F., Camacho, A., Viñuela, E. and Salas, M. (1974) Virology 62, 1-16.
20 Jiménez, F., Camacho, A., de la Torre, J., Viñuela, E. and Salas, M. (1977) Eur. J. Biochem. 73, 57-72.
21 Escarmís, C. and Salas, M. (1981) Proc. Natl. Acad. Sci. USA 78, 1446-1450.
22 Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
23 Cooke, N.E., Coit,D., Weiner, R.I., Baxter, J.D. and Martial, J.A. (1980) J. Biol. Chem. 255, 6502-6510.
24 Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H. and Delius, H. (1980) Proc. Natl. Acad. Sci. USA 77, 6376-6380.
25 Mellado, R.P., Moreno, F., Viñuela, E., Salas, M., Reilly,B. E. and Anderson, D.L. (1976) J. Virol. 19, 495-500.
26 Reilly, B.E., Nelson, R.A. and Anderson, D.L. (1977) J. Virol. 29, 367-377.
27 Murray, Ch.L. and Rabinowitz, J.C. (1982) J. Biol. Chem.257, 1053-1062.
28 Yoshikawa, H. and Ito, J. (1982) Gene 17, 323-335.
29 Carrascosa, J.L., Viñuela, E. and Salas, M. (1973) Virology 56, 291-299.
30 Hawley, L.A., Reilly, B.E., Hagen, E.W. and Anderson, D.L. (1973) J. Virol. 12, 1149-1159.
31 Mellado, R.P., Viñuela, E. and Salas, M. (1976) Eur. J. Biochem. 65, 213-223.
32 Bossi, L. and Roth, J.R. (1980) Nature 286, 123-127.
33 Taniguchi, T. and Weissmann, C. (1978) J. Mol. Biol. 118, 533-565.
34 Tinoco, I.,Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
35 McLaughlin, J.R., Murray, Ch.L. and Rabinowitz, J. (1981) J. Biol. Chem. 256, 11283-11291.
36 Grantham, R., Gautier, C., Govy, M., Jacobzone, M. and Mercier, R. (1981) Nucl. Acids Res. 9, r43-r74.
37 Nussinov, R.J. (1981) J. Mol. Biol. 149, 125-131.
38 Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Holemans, F.,Raeymackers, A., Van den Berghe, A., Volckaert, G. and Ysebaert, M. (1976) Nature 260, 500-507.
39 Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824- 3828.
40 Chou, P.Y. and Fasman, G.D. (1978) in Advances in Enzymol., Meister, E., Ed., Vol. XLVII, pp. 45-148 Wiley, New York.
41 Shoji, S., Parmelee, D.C., Wade, R.D., Kumar, S., Ericsson, L.H., Walsh, K.A., Neurath, H., Long, G.L., Demaille, J.G., Fischer, E.H. and Titani, K. (1981) Proc. Natl. Acad. Sci. USA 78, 848-851.
42 Dayhoff, M.O., Dayhoff, R.E. and Hunt, L.T. (1976) In Atlas of Protein Sequence and Structure, Dayhoff, M.O., E., Vol.V, Suppl. II, pp. 301-310 National Biomedical Research Foundation, Silver Spring, Maryland.
43 Rossman, M.G., Liljas, A., Branden, C.I. and Banaszak, I.J. (1975) in The Enzymes, Boyer, P.D., Ed. Vol. XI, 3rd ed, pp. 61-102 Academic Press, New York.
44 Smart, J.E. and Stillman, B.W. (1982) J. Biol. Chem., in press.