Volume 13 Number 21 1985 Nucleic Acids Research

Location of the serine residue involved in the linkage between the terminal protein and the DNA of phage ϕ29

José M.Hermoso, Enrique Méndez[1], Fernando Soriano[1] and Margarita Salas

Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma, Canto Blanco, 28049 Madrid, and [1]Servicio de Endocrinología, Centro Ramón y Cajal, Carretera de Colmenar Viejo, Km 9,100, 28034 Madrid, Spain

Received 12 July 1985; Revised and Accepted 2 October 1985

ABSTRACT

B. subtilis phage Ø29 has a terminal protein, p3, covalently linked to the 5'ends of the DNA through a phosphodiester bond between a serine residue and 5'-dAMP. This protein acts as a primer in DNA replication by forming an initiation complex with the 5'-terminal nucleotide dAMP.

The amino acid sequence of the terminal protein, deduced from the nucleotide sequence of gene 3, showed the presence of 18 serine residues in a total of 266 amino acids. In this paper we have identified the serine involved in the linkage with the DNA as the residue 232, located close to the C-terminus of the molecule. This result was obtained by amino acid analysis of the peptide that remains linked to the DNA after proteinase K digestion of the terminal protein-Ø29 DNA complex and automated Edman degradation of the corresponding {^{125}I}-labeled tryptic peptide. Prediction of the secondary structure of the terminal protein suggested that the serine residue involved in the linkage with the DNA is placed in a β-turn, probably located on the external part of the molecule, as indicated by hydropathic values.

INTRODUCTION

B. subtilis phage Ø29 has a linear double-stranded DNA of about 18,000 base pairs with a terminal protein covalently linked to the 5'ends through a phosphoester bond between the OH group of a serine residue and 5'-dAMP, the terminal nucleotide at both DNA ends (1). The terminal protein p3 has been characterized as the product of gene 3 and plays a key role in the initiation of Ø29 replication that occurs at either DNA end by a mechanism of protein priming (reviewed in 2). A newly synthesized molecule of the terminal protein is placed at either end of the terminal protein-DNA complex, most probably by interaction with the parental terminal protein and/or a specific sequence at the ends of the DNA; in the presence of the viral DNA polymerase an initiation complex p3-dAMP is formed, that provides the 3'-OH

group needed for elongation. Adenovirus DNA also has a terminal protein covalently linked by a phosphodiester bond between serine and 5'-dCMP (3) and initiates replication by a similar protein priming mechanism (reviewed in 4). Besides phage Ø29 and adenovirus, a number of linear DNAs with terminal proteins have also been found in quite unrelated organisms such as plasmid pSLA2 from Streptomyces (5), mitochondrial DNA S1 and S2 from maize (6), E. coli phage PRD1 (7) and phage Cp-1 (8) and related phages (9) from Streptococcus pneumoniae. The terminal protein-DNA linkage, as determined in two cases, is a phosphodiester bond between tyrosine and 5'-dGMP for PRD1 (10) and between threonine and 5'-dAMP for Cp-1 (García, P., Hermoso, J.M., García, J.A., García, E., López, R. and Salas, M., submitted for publication).

The amino acid sequence of the terminal protein p3, deduced from the nucleotide sequence of gene 3 (11), disclosed the existence of 18 serine residues in a total of 266 amino acids. To better understand the protein priming mechanism we have determined the position in the terminal protein of the serine residue that is linked to the DNA (linking site). In this paper we show that the linking site is at position 232. The linking site has been determined also in adenovirus (12). Secondary structure predictions in a region surrounding the linking sites in Ø29 and adenovirus suggest that both serine residues lie in a β-turn preceded by an α-helix. In addition, the hydropathic profiles of those regions are similar and show that the linking sites of both Ø29 and adenovirus are probably placed at the external part of the molecules.

MATERIALS AND METHODS

Materials

Fungal proteinase K, chromatographically purified, pancreatic carboxypeptidases A and B, piperidine and cellulose thin layer chromatography plastic sheets were from Merck. Chloramine T, trypsin, TPCK treated, from bovine pancreas, polybrene, 3-iodotyrosine and the synthetic tetrapeptide pro-phe-gly-lys were from Sigma, sperm whale myoglobin from Beckman, guanidine hydrochloride from Schwarz/Mann and sodium iodine-{^{125}I}, 100 mCi/ml (3.76 Bq/ml) and {^{14}C} methylated protein mixture of low molecu-

lar weight markers from Amersham.

Purification and amino acid analysis of the proteinase K-resistant peptide linked to Ø29-DNA

DNA from the phage Ø29 delayed lysis mutant sus14(1242) was isolated by proteinase K digestion as described (13). The DNA (50-70 mg) was diluted to a concentration of 1 mg/ml with a buffer containing 0.5 M Tris-HCl, pH 8.0, 10 mM EDTA and 10 mM NaCl and extracted three times with the same volume of phenol equilibrated in this buffer. The DNA from the aqueous phases was precipitated with 2.5 volumes of ethanol, resuspended in the same buffer and phenol extracted and precipitated as before. The peptide-DNA complex was further purified by centrifugation as follows: 20 ml of peptide-DNA complex (0.25 mg/ml) in TES (50 mM Tris-HCl, pH 7.8, 10 mM EDTA, 0.1 M NaCl) were layered on top of 10 ml of a solution of CsCl in TES with a density of 1,45 g x cm^{-3} and 5 ml of the same solution of density 1,70 g x cm^{-3} and centrifuged during 25 hr. at 100,000 x g. Fractions containing DNA were pooled, dialyzed extensively against 0,1 x SSC (0.15 M NaCl, 0.015 M sodium citrate) and precipitated with 2,5 volumes of ethanol. Finally, the peptide-DNA complex was incubated, at a concentration of 1 mg/ml, with 1 M piperidine, freshly distilled, for 6 hr. at 37°C. The sample was lyophylized, dissolved in 10% formic acid and precipitated with two volumes of ethanol in the presence of 0.5 M ammonium acetate. Aliquots of the supernatant containing between 0.7 and 3.0 nmol of peptide were hydrolyzed overnight at 110°C with 5.7 M HCl, 0.05% 2-mercaptoethanol. The hydrolysates were analyzed in a Beckman 121-M B analyzer equipped with a Beckman integrator 126 data system.

Purification of {^{125}I}-labeled tryptic peptide linked to the DNA by reverse phase high performance liquid chromatography

Protein p3-DNA complex from phage Ø29 delayed lysis mutant sus14(1242) was prepared essentially as described (14). Phage treated with guanidinium hydrochloride at a DNA concentration of 0.25 mg/ml was layered on top of a CsCl gradient between densities 1.55 and 1.45 g x cm^{-3} in TES buffer and centrifuged for 19 hr. at 120,000 x g. DNA containing fractions were pooled and precipitated with ethanol.

Purified Ø29 terminal protein-DNA complex (100 µg) was labe-

led in the tyrosine residues with {^{125}I} (700 µCi) using the chloramine T method (15) as described (9), precipitated with ethanol, resuspended in 300 µl of 0.2 M N-methyl morpholine, pH 8.0, and digested with 30 µl of trypsin (2 mg/ml) by incubation at 37°C for 4 hr. The digested material was incubated for 5 min at 50°C in the presence of 0.1% SDS and layered on top of a CsCl linear gradient between densities 1.55 and 1.45 g x cm^{-3} in TES buffer and centrifuged for 4.5 hr. at 300,000 x g. Fractions sedimenting at the DNA position were pooled, dialyzed against 10 mM ammonium acetate and lyophylized. The peptide was released from the DNA by treatment with 1 M piperidine for 2 hr. at 37°C and the mixture was dissolved in 50 µl of 0.1% trifluoroacetic acid (TFA) and applied to a µ Bondapak C_{18} reverse phase column (3.9 x 150 mm, from Waters). The column was run at 37°C at a flow rate of 0.5 ml/min and developed with a linear gradient of 0-30% acetonitrile in 0.1% TFA. Fractions of 0.5 ml were collected for radioactivity measurements. The main radioactive peak was pooled for further analysis.

Sequence analysis of the {^{125}I}-tryptic peptide

A 0.64 ml sample containing 140,000 cpm of the {^{125}I}-peptide and 2.0 mg of sperm whale myoglobin were applied to the spinning cup of a Beckman 890C sequencer. Polybrene (3 mg) was mixed with the sample to avoid washout (16). Automated Edman degradation (17) was performed using the standard Beckman 0.1 M Quadrol program. The amino acid thiazolinones obtained in each cycle were used directly for radioactivity measurements. To check sequencing accuracy, one third of the amino acid thiazolinones corresponding to cycles 19, 20 and 21 were converted to phenylthiohydantoin (PTH) amino acids (18) and identified by thin-layer chromatography (19). Alanine, aspartic acid and valine derivatives were found, that correspond to the amino acids present at those positions in myoglobin.

Carboxypeptidase digestion of the {^{125}I}-tryptic peptide

Purified {^{125}I}-labeled tryptic peptide containing the linking site (1,000-3,000 cpm) was dried under vacuum together with 5 µg of the synthetic peptide pro-phe-gly-lys as carrier, dissolved in 10 µl of 10 mM sodium phosphate buffer, pH 7 and incubated at 30°C for the indicated times with 0.1 µg (0,04 U) of carboxypep-

tidase A and 0.06 µg (0,01 U) of carboxypeptidase B. The samples were applied immediately to thin layer chromatography plates and developed in n-butanol/acetic acid/water/pyridine (30:6:24:20). The radioactive digestion products were detected by autoradiography and those from the carrier peptide and the marker 3-iodotyrosine were identified by ninhydrin staining.

RESULTS

Identification of a peptide containing the linking site

Since proteolytic digestion of the Ø29 terminal protein-DNA complex yields a peptide covalently attached to the DNA and the linkage can be hydrolyzed by alkali (1), we have purified the peptide that remains attached to Ø29 DNA after proteinase K digestion by phenol extraction and CsCl gradient centrifugation. After alkaline hydrolysis with piperidine and ethanol precipitation to remove DNA, the material in the supernatant was hydrolyzed with HCl and the amino acid composition determined. Table 1 shows the amino acid composition of two independent peptide preparations in which a single serine residue was obtained. Glycine could not be detected due to ninhydrin reacting contaminants, probably DNA degradation products, that coelute at the correspon-

Table 1. Amino acid analysis of the proteinase K-resistant peptide linked to Ø29 DNA.

Residue	Observed[a]	Predicted[b]
Asp	2.16	2
Thr	0.95	1
Ser	1.01	1
Gly	N.D.	1
Glu	2.96	3
Val	1.07	1
Phe	0.95	1

[a] Average amino acid composition (mol/mol of peptide) obtained by three amino acid analysis of two independent peptide preparations. Amino acids present in amounts smaller than 0.1 mol/mol of peptide are not included.
[b] Amino acid composition of the corresponding peptide predicted from the nucleotide sequence (11) (see Figure 1).
N.D., not determined.

ding position in the amino acid analysis. The amino acid composition obtained better fits with a decapeptide around serine at position 232, either from position 228 to 237 or 229 to 238, since the presence of glutamic acid in positions 228 and 238 does not allow to determine the exact boundaries of the peptide (see Fig. 1). Attempts to determine the N-terminal amino acid, either by reaction with O-phtaladehyde (20) or dansyl-chloride (21,22) failed, probably due to contamination of the peptide with DNA degradation products containing amino groups.

Isolation and analysis of the {^{125}I}-labeled tryptic peptide linked to the DNA

According to the amino acid sequence of protein p3 (see Figure 1), serine residue 232 is included in a tryptic peptide of 25 amino acids (from position 222 to 246) that contains two other serine residues, at positions 223 and 226, and only one tyrosine at position 245 (position 24 in the peptide), and is the largest one of the molecule. To confirm the previous result, Ø29 terminal protein-DNA complex was labeled with {^{125}I} and digested with

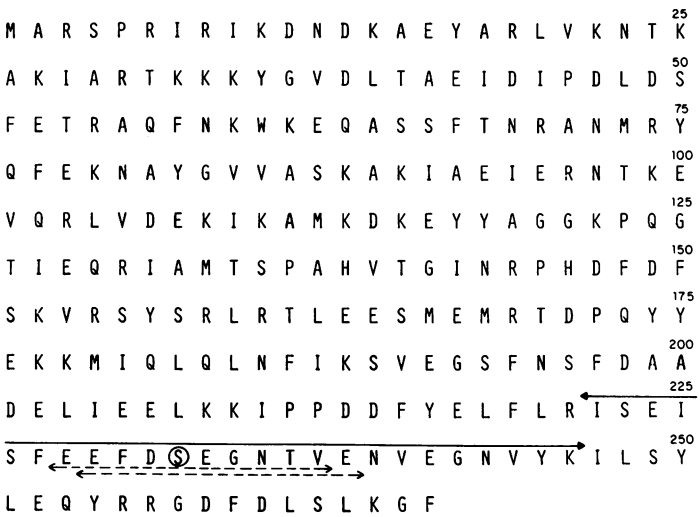

Fig. 1. Amino acid sequence of Ø29 terminal protein deduced from the nucleotide sequence of gene 3 (11). The serine residue through which the terminal protein is covalently attached to the DNA is encircled. (←- - ->) Possible peptides resistant to proteinase K digestion of the terminal protein-Ø29 DNA complex; (←——→) {^{125}I}-labeled tryptic peptide covalently linked to Ø29 DNA.

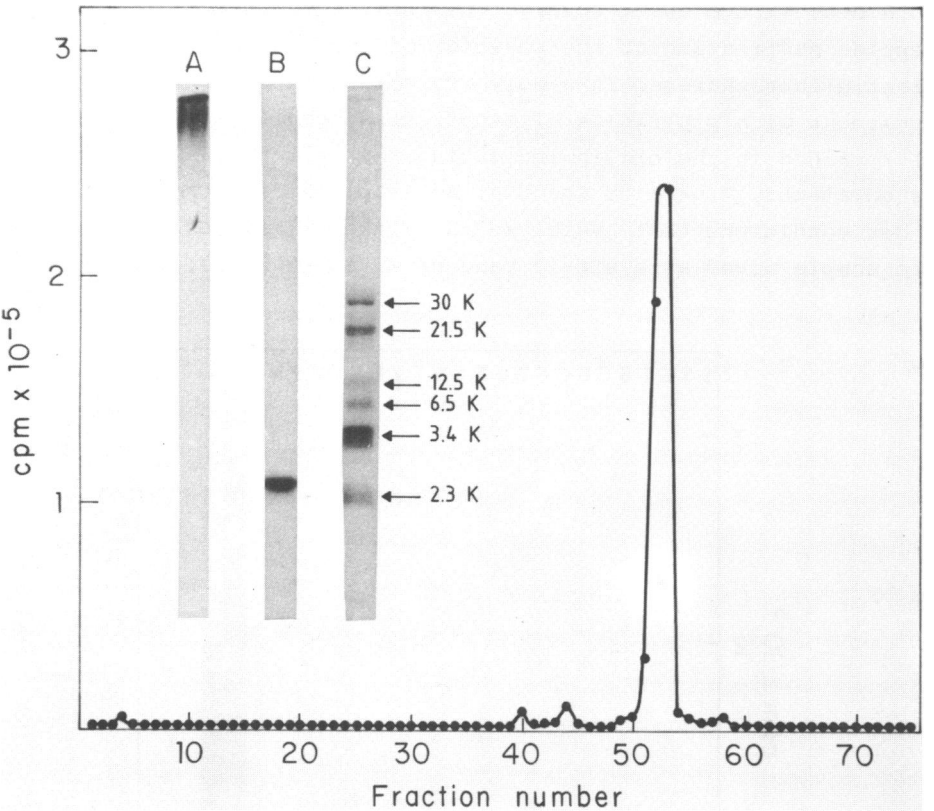

Fig. 2. Purification of {^{125}I}-labeled tryptic peptide containing the linking site by reverse phase chromatography. {^{125}I}-labeled tryptic peptide covalently linked to Ø29 DNA, purified through CsCl gradient centrifugation as described in Materials and Methods, was digested with 1 M piperidine for 2 hr. at 37ºC, applied to reverse phase high performance liquid chromatography column and eluted with a linear gradient of 0-30% acetonitrile as described in Materials and Methods. The inset shows samples electrophoresed in a 10-18% linear gradient of acrylamide gel containing 0.1% SDS and 7 M urea (23). A. Sample from the DNA-containing peak in the CsCl gradient. B. The same as A digested with piperidine. C. {^{14}C}-methylated low molecular weight markers : carbonic anhydrase, soybean trypsin inhibitor, cytochrome C, aprotinin and insulin, chain B and chain A.

trypsin. The {^{125}I}-labeled tryptic peptide linked to the DNA was isolated by centrifugation in a CsCl gradient. The peptide was released from the DNA by treatment with piperidine and purified by reverse phase high performance chromatography to remove the DNA. Figure 2 shows that a main peak of radioactivity was

obtained, with a 60% recovery. The inset shows the radioactive peptide sedimenting at the position of the DNA in the CsCl gradient electrophoresed in a polyacrylamide gel before and after treatment with piperidine. The untreated sample (Figure 2, lane A) remained in the origin indicating that all the radioactivity is covalently linked to the DNA. After piperidine treatment under conditions that hydrolyze the linkage (Figure 2, lane B), the sample moved as a single band of Mr about 2,600, very close

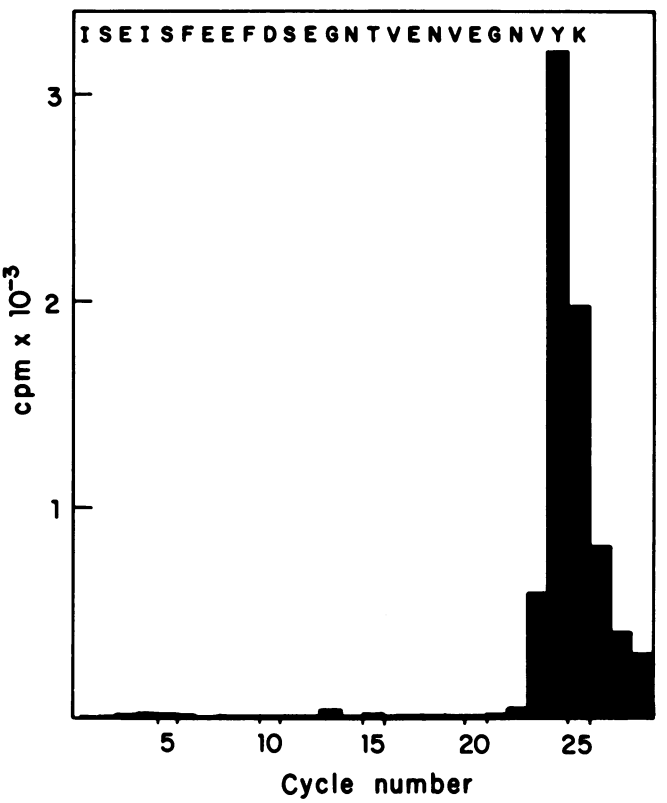

Fig. 3. Automated Edman degradation of {^{125}I}-labeled tryptic peptide containing the linking site. {^{125}I}-labeled tryptic peptide purified by reverse phase chromatography was sequenced together with sperm whale myoglobin as carrier in a Beckman 890 C sequencer using the standard Beckman 0.1 M Quadrol program. The radioactivity of the amino acid thiazolinones released in each cycle was determined. The amino acid sequence of the corresponding peptide (see Figure 1) is shown in the upper part.

to the theoretical Mr value (2,836) of the tryptic peptide containing serine at position 232.

Location of the {^{125}I}-labeled tyrosine in the tryptic peptide released from the DNA

To locate precisely the position of the {^{125}I}-labeled tyrosine in the tryptic peptide the latter was mixed with myoglobin as carrier and subjected to 28 automated Edman degradation cycles. The radioactivity released in each cycle is shown in Figure 3, where a maximum is seen in step 24, corresponding to the position of the tyrosine residue in the tryptic peptide. No other tryptic peptide in the whole molecule, even partially digested, contains a tyrosine residue in that position (see Figure 1).

In addition, the purified {^{125}I}-labeled tryptic peptide containing the linking site was digested with a mixture of carboxypeptidases A and B and analyzed by thin layer chromatography. Figure 4 shows that after 5 and 15 min of incubation at 30ºC (lanes B and C) only two radioactive spots are observed, one with the same mobility as the untreated peptide (lane A) and the other with that of the 3-iodotyrosine marker. After 60 min of incubation (lane D), all the radioactivity is released as 3-iodotyrosine. These results suggest that the labeled tyrosine residue lies close to the C-terminus of the peptide, in agreement

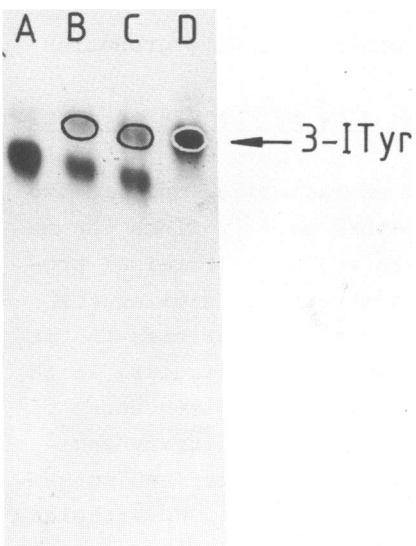

Fig. 4. Carboxypeptidase digestion of {^{125}I}-labeled tryptic peptide containing the linking site. {^{125}I}-labeled tryptic peptide purified by reverse phase chromatography was incubated at 30ºC with a mixture of carboxypeptidases A and B for 5 min (lane B), 15 min (lane C) and 60 min (lane D), applied immediately to cellulose thin layer plates and developed as indicated in Materials and Methods. Lane A, untreated sample. The position of 3-iodotyrosine marker is also shown.

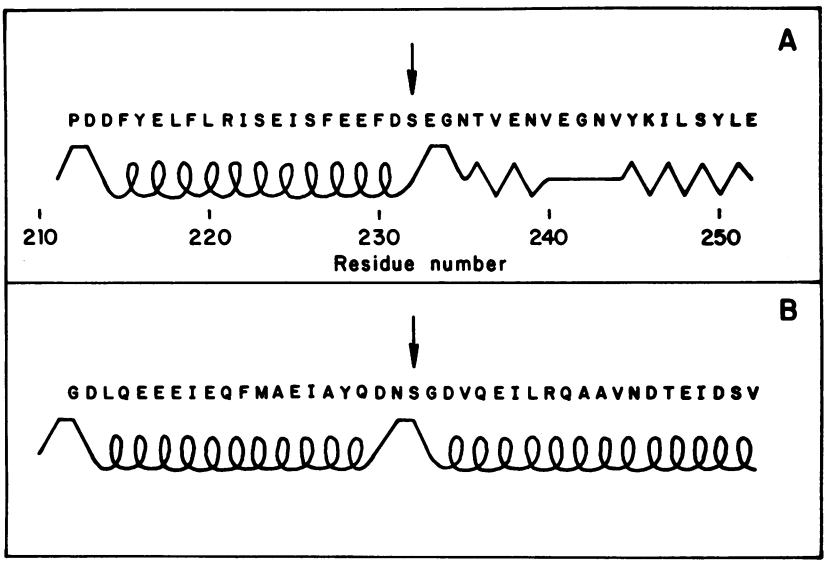

Fig. 5. Secondary structure prediction of regions surrounding the linking site in Ø29 and adenovirus-terminal proteins. Secondary structure using the Chou and Fasman method (24) was predicted for a 40 residues long region in which the linking site was centered in the terminal proteins from Ø29 (A) and adenovirus (B). Coil, α-helix; sawtooth pattern, β-sheet; straight line, random coil; arch, β-turn. The amino acid sequence for both regions, deduced from the nucleotide sequence (11,12), is also shown. The arrows indicate the position of the serine residue involved in the linkage.

with the sequencing data; otherwise, other labeled degradation products would have been obtained.

Secondary structure around the linking site of the Ø29 terminal protein

So far, the linking site has been characterized only in two cases, adenovirus (12) and phage Ø29 (this paper). Since the mechanism of initiation of replication in which the terminal protein plays a major role as a primer is strikingly similar for the two systems, we would expect structural similarities in these proteins reflecting their function and, in particular, in the region containing the linking site. The amino acid sequences of the adenovirus and Ø29 terminal proteins in an arbitrary environment from i-20 to i+20, i0 being the linking site, do not present any apparent sequence homology (see Figure 5). Nevertheless,

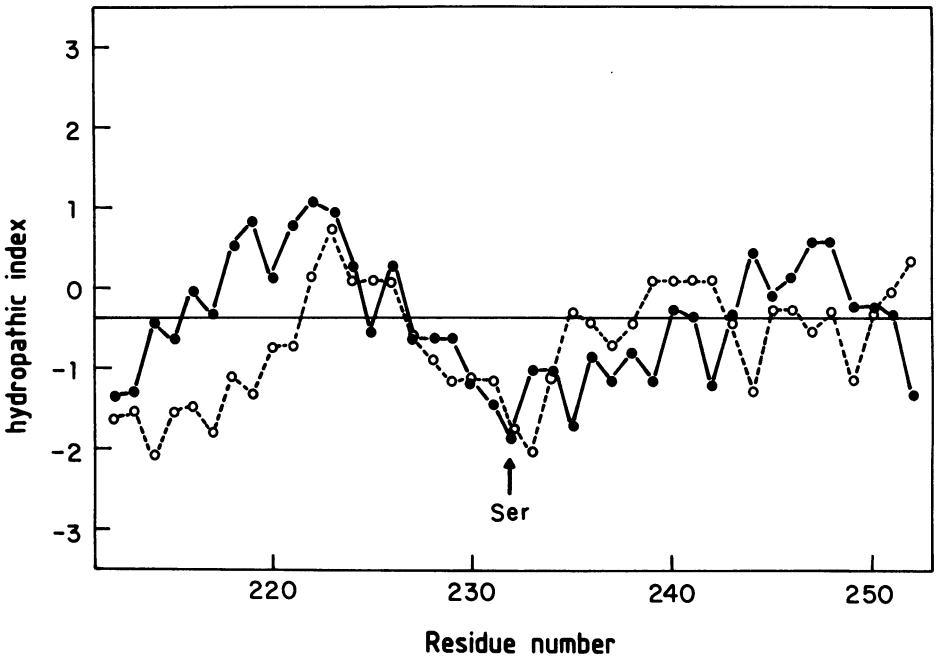

Fig. 6. Hydropathic index of a 40 residues long region including the linking site, using a span setting of nine according to Kyte and Doolittle (26). ●——●, Ø29 terminal protein; o---o, adenovirus terminal protein. The arrow indicates the location of the linking site. The residue number corresponds to the Ø29 terminal protein.

the amino acid composition of this region is very similar in both cases, containing about 25% acidic residues (aspartic acid, glutamic acid), about 3% basic residues (lysine, arginine) and about 25% non polar aliphatic large chain residues (valine, isoleucine, leucine). Figure 5 shows the predicted secondary structure for the i-20 to i+20 region in both proteins, using the Chou and Fasman method (24). The linking sites, for either Ø29 or adenovirus terminal proteins, are probably located in β-turns and both sites are preceded by α-helical regions.

Since β-turns are usually concentrated at the surface of the protein (25), the hydropathic values (26), that is the hydrophobic and hydrophylic tendencies of a given amino acid in a polypeptide chain, were calculated for the i-20 to i+20 region of the two terminal proteins. Figure 6 shows that the region containing

the linking site of both Ø29 and adenovirus terminal proteins has hydropathic values on the hydrophylic side of the midpoint line and therefore, as expected, the linking site is probably located on the surface of the molecules (26).

DISCUSSION

We have characterized two overlapping peptides containing the DNA-linking site : one was obtained by tryptic digestion of {^{125}I}-labeled terminal protein DNA-complex and contains three serine residues at positions 223, 226 and 232; the other, obtained by proteinase K digestion and characterized by amino acid analysis (see Table 1),fits well with the peptide containing serine 232, either from position 228 to 237 or 229 to 238, and rules out serine 223 as the DNA linking site. If serine 226 were the one linked to the DNA, the peptide that better fits with the amino acid composition is that between residues 226 and 243 (see Fig. 1), being the number of observed residues twice the value indicated in Table I. In this case, there are two discrepancies between the observed amino acid analysis (two threonine and six glutamic acid residues) and the predicted peptide (one threonine and five glutamic acid residues). Furthermore, it also seems unlikely that the DNA attached to serine 226 would protect from the proteinase K digestion only the protein moiety located at the C terminal site of this residue. Therefore, we conclude that the linking site of the terminal protein of phage Ø29 DNA is most likely the serine at position 232, 35 amino acids from the carboxyl terminus of the protein. This result is in agreement with experiments of *in vitro* mutagenesis of gene 3. When the C-terminal amino acids (leu-lys-gly-phe) or (lys-gly-phe) of the terminal protein were replaced by phe-asp-ala-val-val-tyr-lys-ser or by leu-ile-asp and the mutated proteins assayed *in vitro* for initiation complex formation, their specific activity was lower than that of the intact terminal protein (27). Deletion mutants in which the four C-terminal amino acids of the terminal protein were removed showed a decrease in activity and when the deletion was extended up to twenty amino acids, no activity was detected (unpublished results). All these results indicate that the C-terminal region is important for the activity of the ter-

minal protein.

Secondary structure predictions for the region around the linking site in Ø29 and adenovirus terminal proteins suggest that both serine residues lie in a β-turn located after an α-helix region. Hydropathic values have also been calculated and in both cases the linking site is probably on the surface of the molecule. Moreover, the hydropathic profile around such serine residues are similar for both proteins, thus suggesting similar surface topologies. The presence of a DNA molecule covalently linked to this site, or even dAMP in the initiation complex, may drastically change both the secondary and tertiary structure of the protein, reflecting the different roles played by the free and parental terminal proteins in the mechanism of initiation of replication of Ø29 DNA.

Although the free terminal protein plays most likely a structural role in the priming mechanism of initiation of Ø29 DNA replication, it must contain several recognition sites to fulfill its function. On the one hand, it interacts with the Ø29 DNA polymerase (unpublished results). On the other, it should bind to the parental terminal protein and/or a specific sequence at the Ø29 DNA ends and should contain a specific site through which dAMP is covalently linked giving rise to the p3-dAMP initiation complex. The linking site determined in this paper should be the acceptor site of dATP for the formation of the initiation complex. Modification of the terminal protein by site-directed mutagenesis should be of help in correlating the structure of the protein with its function as a primer in Ø29 DNA replication.

ACKNOWLEDGEMENTS

We are grateful to Dr. J. Gavilanes for advice on the structural predictions. This investigation has been aided by Research Grant 2 R01 GM27242-05 from the National Institutes of Health, by Grant nº 3325 from the Comisión Asesora para el Desarrollo de la Investigación Científica y Técnica and by a Grant from Fondo de Investigaciones Sanitarias.

REFERENCES

1. Hermoso, J.M. and Salas, M. (1980) Proc. Natl. Acad. Sci. USA 77, 6425-6428.

2. Salas, M. (1983) Curr. Top. Microbiol. Immunol. 190, 89-106.
3. Desiderio, S.V. and Kelly, T.J. (1981) J. Mol. Biol. 145, 319-337.
4. Stillman, B.W. (1983) Cell 35, 7-9.
5. Hirochika, H., Natamura, K. and Sakaguchi, K. (1984) EMBO J. 3, 761-766.
6. Kemble, R.J. and Thompson, D. (1982) Nucl. Acids Res. 10, 8181-8190.
7. Bamford, D., McGraw, T., MacKenzie, G. and Mindich, L.(1983) J. Virol. 47, 311-316.
8. García, E., Gómez, A., Ronda, C., Escarmís, C. and López, R. (1983) Virology 128, 92-104.
9. López, R., Ronda, C., García, P., Escarmís, C. and García, E. (1984) Mol.Gen. Genet. 197, 67-74.
10. Bamford, D.H. and Mindich, L. (1984) J. Virol. 50, 309-315.
11. Escarmís, C. and Salas, M. (1982) Nucl. Acids Res. 10, 5785-5798.
12. Smart, J.E. and Stillman, B.W. (1982) J. Biol. Chem. 257, 13499-13506.
13. Inciarte, M.R., Lázaro, J.M., Salas, M. and Viñuela, E. (1976) Virology 74, 314-323.
14. Peñalva, M.A. and Salas, M. (1982) Proc. Natl. Acad. Sci. USA 79, 5522-5526.
15. Greenwood, F.C., Hunter, W.H. and Glover, J.S. (1963) Biochem. J. 89, 114-123.
16. Tass, G.E., Beecher, J.F., Bell, M. and McKean, D.J. (1978) Anal. Biochem. 84, 622-627.
17. Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80-90.
18. Guyer, R.L. and Todd, C.W. (1975) Anal. Biochem. 66, 400-404.
19. Jeppson,J.O. and Sjoquirst, J. (1967) Anal. Biochem. 18, 264-269.
20. Méndez, E. (1982) Anal. Biochem. 127, 55-60.
21. Gray, W.R. (1967) Methods Enzymol. 11, 139-151.
22. Woods, K.R. and Wang, K.T. (1967) Biochim. Biophys. Acta 133, 369-370.
23. Hashimoto, F., Horigome, T., Kambayashi, M., Yoshida, K. and Sugano, H. (1983) Anal. Biochem. 129, 192-199.
24. Chou, P.Y. and Fasman, G.D. (1978) in Advances in Enzymol., Meister, E., Ed., Vol. XLVII, pp. 45-148 Wiley, New York.
25. Kuntz, I.D. (1972) J. Amer. Chem. Soc. 94, 4009-4012.
26. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
27. Mellado, R.P. and Salas, M. (1983) Nucl. Acids Res. 11, 7397-7407.