Amino Acid Sequence of the Small Core Protein from Bacteriophage ϕX174†

Daniel K. Freymeyer II, Peter R. Shank,‡ Marshall H. Edgell, Clyde A. Hutchison III,* and Thomas C. Vanaman

ABSTRACT: The amino acid sequence of the small core protein of bacteriophage ϕX174 has been determined by a combination of automated Edman degradation of the intact polypeptide and by analysis of tryptic and thermolytic peptides. The six lysyl and six arginyl residues of this 37-residue polypeptide are concentrated in two structurally homologous 12-residue segments of the sequence. The hydrophobic residues of valine, tryptophan, tyrosine, and phenylalanine are contained in the carboxyl-terminal nine residues of the protein, together with one of the two leucyl residues and two of the three glutaminyl residues. The single free carboxyl group in the protein is the α-COOH of the C-terminal phenylalanyl residue. The overall sequence of this small core protein suggests that it may function as a DNA-condensing protein. The protein sequence presented here corresponds exactly to the DNA base sequence of the cistron J region of the ϕX174 genome determined in another laboratory.

Bacteriophage ϕX174 is a small icosahedral virus composed of four major structural proteins encapsidating a circular single-stranded DNA molecule of 1.7×10^6 daltons (Sinsheimer, 1968). These four virion proteins are the products of the adjacent ϕX174 genes J, F, G, and H (Burgess and Denhardt, 1969; Gelfand and Hayashi, 1969; Mayol and Sinsheimer; 1970; Benbow et al., 1971, 1972; Barrell et al., 1976; Sanger et al., 1977). The gene G and H proteins comprise the spikes which protrude from the vertices of the icosahedron (Edgell et al., 1969). Jazwinski et al. (1975) have shown that the H protein is also involved in attachment, penetration, and viral DNA replication. Several minor protein species have been shown to be either permanently or transiently associated with the virion (Godson, 1971; Zuccarelli, et al., 1972; Weisbeek and Sinsheimer, 1972).

The fourth major virion protein is a low-molecular-weight basic protein (Poljak, 1968; Burgess and Denhardt, 1969; Suruda and Poljak, 1971). Recently, this protein has been found in the virion core after removal of the spikes (Shank et al., 1977). The function of this protein in the virus life cycle remains to be determined.

It was previously reported (Shank et al., 1976) that the low-molecular-weight core protein (referred to in this report as SCP[1]) has a very unusual amino acid composition, being rich in lysine, arginine, and glycine while being devoid of a number of amino acids including histidine and sulfur-containing amino acids. The complete amino acid sequence of this 37-residue ϕX174 polypeptide presented in this report suggests it could function as a DNA-condensing protein. The DNA nucleotide sequence of the ϕX174 genome has been determined in another laboratory (Sanger et al., 1977). This sequence contains a region between genes D and F which codes for the amino sequence presented in this paper. We therefore propose to define gene J as this region.

Materials

The small core protein (SCP), purified from ϕX174 *am3* as described by Shank et al. (1977), was homogeneous as judged by $NaDodSO_4$–polyacrylamide gel electrophoresis and by polyacrylamide gel electrophoresis performed on urea–acetic acid containing gels as described by Panyim and Chalkley (1969).

Special reagents used for amino acid analysis and sequence analysis were obtained from either Pierce Chemical Co. (Rockford, Ill.) or Beckman Instruments Co. (Palo Alto, Calif.) unless otherwise noted. Pyridine and *N*-ethylmorpholine (Eastman) were distilled after refluxing with ninhydrin (Hill and Delaney, 1967). Constant-boiling hydrochloric acid (5.7 N) was used for all acid-hydrolysis procedures.

Tos-PheCH_2Cl-treated trypsin, carboxypeptidases A and B (Worthington Biochemicals), thermolysin (A grade, Calbiochem), and aminopeptidase M (Rohm and Haas) were used without further treatment. Double-layered polyamide sheets (20 × 20 cm, Cheng-Chin) were used as supplied by Gallard-Schlesinger. Sephadex chromatographic media were prepared and used as described by Pharmacia. All other chemicals were standard reagent grade and were not further purified.

Methods

Amino Acid Analysis. Samples for amino acid analysis were prepared and analyzed as described by Pett et al. (1973). Cysteine and methionine were determined as cysteic acid and methionine sulfone, respectively, by analysis of hydrolysates of samples oxidized with performic acid (Hirs, 1967). Tryptophan was detected by the method of Erlich (Bennett, 1967) and quantified by the method of Scoffone et al. (1968). The

† From the Department of Bacteriology and Immunology, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27514 and the Department of Microbiology and Immunology, Duke University Medical Center, Durham, North Carolina 27710. *Received January 13, 1977.* This work was supported by National Institutes of Health Grants AI09044 and GM00125 (M.H.E.), GM21313 and AI70604 (C.A.H. III), CA18721 (T.C.V.), and by National Science Foundation Grant GB27597 (T.C.V.).

‡ Present address: Department of Microbiology, University of California, San Francisco, Calif. 94143.

* Correspondence should be directed to: Department of Bacteriology and Immunology, School of Medicine, University of North Carolina, Chapel Hill, N.C. 27514.

[1] Abbreviations used are: SCP, ϕX174 small core protein; $NaDodSO_4$, sodium dodecyl sulfate; Tos-PheCH_2Cl, (1-tosylamido-2-phenyl)ethyl chloromethyl ketone; Tp, tryptic peptide; Tl, thermolytic peptide; STp, succinyl tryptic peptide; Pth, phenylthiohydantoin; Tris, 2-amino-2-hydroxymethyl-1,3-propanediol; CM, carboxymethyl.

o-nitrophenylsulfenyl chloride derivative of SCP was freed of excess reagent by passing over a column of Sephadex G-10 in 50% (v/v) acetic acid prior to absorbance measurements at 362 nm, and the protein concentration was determined by amino acid analysis.

Peptide Purification. Operating conditions for column chromatographic separations of peptides including gradients and buffers used are listed with the appropriate figures. A portion of the effluent from each column was monitored with ninhydrin following alkaline hydrolysis using a modified Technicon-Autoanalyzer (Herman and Vanaman, 1975). Cysteic acid was added to the sample as an internal marker for alignment of the elution profile with fractions collected (Hill and Delaney, 1967). Appropriate fractions were combined and evaporated to dryness under reduced pressure at 40 °C. Peptides were dissolved in 50% (v/v) acetic acid and tested for purity by thin-layer chromatography on cellulose plates developed in pyridine–butanol–acetic acid–water (100:150:30:120). Peptides were detected on thin-layer plates by spraying with ninhydrin solution (2 mg/mL in 95% acetone) and developing at 100 °C for 5 min.

Enzymatic Digests

Trypsin Digestion. SCP (250 nmol) was dissolved in 2.0 mL of 0.1 M NH_4HCO_3, pH 8.5. Digestion was initiated by adding 30 μL of a 1 mg/mL solution of Tos-PheCH_2Cl-treated trypsin in 0.001 N HCl. After incubation for 2 h at 37 °C with stirring, an additional 30 μL of the trypsin solution was added and digestion continued for 2 h. The digestion mixture was shell frozen and lyophilized. The dried digest was dissolved in deionized H_2O, shell frozen, and lyophilized two additional times to ensure complete removal of ammonia.

Thermolysin Digestion. SCP (500 nmol) was dissolved in 1.0 mL of a buffer consisting of 0.1 M Tris-HCl, pH 7.5, 0.001 M $CaCl_2$. A 10-μL aliquot of a 1.0 mg/mL thermolysin solution in 2.5 mM $CaCl_2$, pH 7.5, was added and the digestion mixture incubated at 37 °C. After 2 h, another 10 μL of thermolysin solution was added and incubation continued for an additional 2 h. The digestion mixture was maintained at pH 7.5 throughout incubation by the addition of 1 M Tris base. After incubation, the digestion mixture was shell frozen and lyophilized.

Digestion with Exopeptidases. Carboxypeptidase A digests were performed on 5–10 nmol of peptide dissolved in 350 μL of 0.25 potassium phosphate, pH 7.65, by the addition of 50 μL of stock carboxypeptidase A (1.0 mg/mL; 50 units/mg). After 1 h at 37 °C, 100 μL of 1 N HCl was added, and the sample was dried at reduced pressure and dissolved in amino acid analyzer buffer (0.01 N HCl).

Carboxypeptidase B digests were performed by dissolving 5–15 nmol of peptide in 300 μL of 1.0 M $NaHCO_3$, pH 8.5, and then adding 2 μL of a 5 mg/mL solution of carboxypeptidase B. After digestion for 2 h at 27 °C, 100 μL of 1 N HCl was added and the sample was treated as described for carboxypeptidase A.

For aminopeptidase M digestion, 15 nmol of peptide was dissolved in 500 μL of 0.05 M *N*-ethylmorpholine hydrochloride, 1 mM $MgCl_2$, pH 8.0, and then a 25 μL aliquot of stock aminopeptidase M (1 mg/mL) was added. After incubation for 8 h at 37 °C, 100 μL of 1 N HCl was added and the sample was treated as described for carboxypeptidase A.

Succinylation of SCP. Three-hundred nanomoles of SCP was dissolved in 1.0 mL of deionized H_2O and adjusted to pH 8.5 with 1 N NaOH. This solution was chilled to 4 °C and an 80-fold excess of solid succinic anhydride was added in three equal portions at 20-min intervals. The solution was maintained at pH 8.5 with 1 N NaOH. After reaction, the succinylated protein was freed of salt on Sephadex G-15 in 0.5% (v/v) acetic acid, pooled, and lyophilized before digestion with trypsin. Succinylation was judged to be complete as carboxypeptidase B treatment of a sample of trypsin-digested succinyl SCP released only arginine.

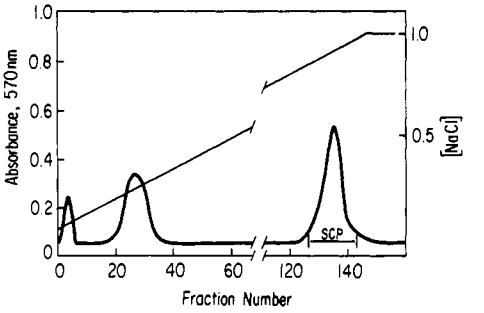


FIGURE 1: Purification of SCP on CM-Sephadex C-25 in 0.1 M NaOAc, pH 7.0. Samples of SCP in 5–10 mL of 0.1 M NaOAc, pH 7.0, were applied to a column (0.9 × 25 cm) of CM-Sephadex C-25 in the same buffer. The column was developed with a gradient formed using 300 mL each of 0.1 M NaOAc and 0.1 M NaOAc, 1.0 M NaCl as starting and limit buffers, respectively. The column was operated at 25 °C at a flow rate of 35 mL/h with continuous monitoring of the effluent for ninhydrin-reactive material as described under Methods. Fractions were collected at 6.0 min/tube.

Sequence Analysis. Amino-terminal residues were determined for purified peptides by the dansyl method as described by Hartley (1970). Automated Edman degradations were performed in a Beckman Model 890B Sequencer using the dimethylbenzylamine–protein methodology of Hermodson et al. (1972). Manual Edman degradations were performed as described by Peterson et al. (1972). Aminothiazolinone derivatives were converted to the corresponding Pth-amino acids as described by Niall (1974) and subsequently identified and quantified by gas chromatography (Pisano et al., 1972) and by amino acid analysis after back-hydrolysis of the Pth derivatives (Smithies et al., 1971). Subtractive Edman degradations were performed essentially as described by Konigsberg and Hill (1963).

Results

Amino Acid Composition of φX174 SCP. The amino acid composition of φX174 SCP, shown in Table I, was determined by analysis of material purified by ion-exchange chromatography on CM-Sephadex C-25 as shown in Figure 1. This composition agrees with our previously published results (Shank et al., 1977). The material eluting from this column in fractions 20–30 was a low-molecular-weight aspartic acid rich peptide. The presence of aspartic acid in acid hydrolysates of SCP prepared by gel filtration was invariably due to contamination by this peptide.

Based on an absorbance of 0.22 at 362 nm for a solution containing 0.045 μmol/mL of *o*-nitrophenylsulfenyl-SCP prepared as described under Methods, SCP contains 1.2 mol of tryptophan per mol of protein. Half-cystine was absent as judged by amino acid analysis performed on the performic acid oxidized protein (Hirs, 1967).

Amino and Carboxyl Terminus of SCP. Shank et al. (1977) previously identified serine as the amino-terminal residue of SCP using the dansyl technique. Results of automated Edman degradation and studies of the purified tryptic peptides (vide infra) confirmed that identification.

Digestion of SCP with carboxypeptidase A released 1 mol of phenylalanine per mol of SCP. No other amino acids were detected in the digest. This result, confirmed by exopeptidase

TABLE I: Composition of the Tryptic Peptides of ϕX174 SCP.[a]

Amino acid	SCP	Tp 1	Tp 2	Tp 3	Tp 4	Tp 5	Tp 6	Tp 7	Tp 8	Tp 9	Tp 10	Tp 11
Lys	6.03 (6)		1.1 (1)	0.1		1.0 (1)	1.1 (1)	0.1	0.1	2.0 (2)	0.8 (1)	1.0 (1)
His	0											
Arg	6.16 (6)	0.1	0.1	1.9 (2)	0.9 (1)		0.3	1.0 (1)	2.6 (3)		1.0 (1)	1.0 (1)
Asp	0											
Thr	1.13 (1)		0.9 (1)									
Ser	2.06 (2)				0.9 (1)		1.0 (1)	0.1	0.8 (1)	0.8 (1)		
Glu	3.12 (3)	2.0 (2)		1.1 (1)				0.1	1.2 (1)	0.1		
Pro	3.16 (3)			2.8 (3)				0.1	3.1 (3)			
Gly	8.10 (8)	2.0 (2)	1.0 (1)	1.2 (1)	1.1 (1)	1.0 (1)	0.1	1.0 (1)	2.2 (2)	1.1 (1)	1.1 (1)	0.8 (1)
Ala	2.10 (2)				1.1 (1)			1.0 (1)	1.1 (1)		1.0 (1)	
½-Cys[b]	0											
Val	0.93 (1)	1.0 (1)										
Met	0											
Ile	0											
Leu	2.06 (2)	0.9 (1)		1.0 (1)					1.0 (1)			
Tyr	0.92 (1)	1.0 (1)										
Phe	1.01 (1)	1.2 (1)										
Trp	1.20 (1)[c]	0.9 (1)[d]										
Total residues	37	9	3	8	4	2	2	3	12	4	4	3
% Yield		46	74	27	63	60	30	36	29	41	61	80
NH_2-Terminus		Leu	Gly	Pro	Ser	Gly	Ser	Gly	Ser	Ser	Lys	Gly

[a] Values not in parentheses are those determined, values in parentheses represent assumed residues/molecule. [b] Determined as cysteic acid (see text). [c] Determined by the procedure of Scoffone et al. (1968) as described in the text. [d] Determined by digestion with aminopeptidase M (see text).

TABLE II: Automated Edman Degradation of ϕX174 SCP.

Cycle	Residue	GC	Back-hydrolysis	nmol of amino acids[c] released per cycle
1	Ser	+[a]	−	
2	Lys	+	+	110
3	Gly	+	+	120
4	Lys	+	+	112
5	Lys	+	+	112
6	Arg	−	+	
7	Ser	+[a]	−	
8	Gly	+	+	65
9	Ala	+	+	49
10	Arg	−	+	
11	Pro	+	+	7.7
12	Gly	+	+	25
13	Arg	−	+	
14	Pro	+	+	8.7
15	Gln	+	+[b]	
16	Pro	+	−	2.4
17	Leu	+	−	3.4
18	Arg	−	+	
19	Gly	+	+	5.0
20		−	−	
21		−	−	
22	Gly	+	+	3.4

[a] Seen on run number 2 only. [b] Seen as glutamic acid. [c] For run number 1, as determined by gas chromatography.

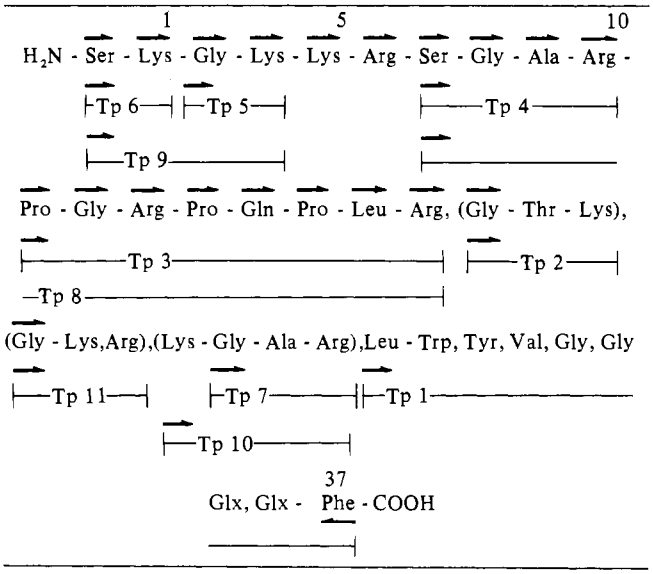
and composition studies of the C-terminal tryptic peptide Tp 1, indicated that the single phenylalanyl residue of SCP is at the carboxyl terminus of the molecule.

Automated Edman Degradation. The sequence of the amino-terminal 19 residues of SCP was established by a combination of automated Edman degradation on the intact protein and by analysis of tryptic peptides which will be described in a later section. Two separate automated Edman degradations, each of 25 cycles, were performed on intact SCP (400 nmol each) with a Beckman Model 890B sequencer. The fractions recovered from each cycle were converted to Pth-amino acids which were subsequently identified and quantified. The results of these determinations (Table II) allowed unequivocal identification of the Pth-amino acids produced at cycles 1–19 and cycle 22. As noted in Table II, Pth-serine was identified at cycles 1 and 7 only in the second determination owing to milder conversion conditions (2 min at 80 °C) used for these two cycles. In both determinations, the initial step yield was only 50%. However, the average repetitive yield from cycle 3 to cycle 12 was 86%. The overall repetitive yield from cycle 3 through cycle 22 was 83%.

Representative gas chromatography traces for cycles 2, 4, 5, 8, 19, and 22 are shown in Figure 2. Pth-lysine, detected after silylation, was the only Pth-amino acid found at cycles 2, 4, and 5. Similarly, Pth-glycine was the only amino acid detected at cycles 3 and 8. Due to an increase in background peaks in the gas chromatographic analysis, sequence assignments were no longer possible after cycle 22. Studies of the tryptic peptides from SCP described in the following sections confirmed the placement of residues derived from these sequencer runs and provide the basis for the remainder of the sequence.

Preparation and Characterization of Tryptic Peptides. Trypsin digests of ϕX174 SCP were chromatographed on Beckman AA-15 (Figure 3). Pooled fractions corresponding to the peaks in this elution profile contained pure peptides as judged by thin-layer chromatography. The amino acid compositions of eight unique peptides (Tp 1–7 and 11) and three peptides resulting from incomplete digestion (Tp 8–10) are shown in Table I. The unique peptides accounted for the composition of SCP except for two lysyl and one arginyl residues, both of which were present as free amino acids but not quantified in the tryptic digest. As noted in Figure 3, peptide Tp 11 was not eluted from the column until after extensive washing with 2.0 N pyridine–acetic acid, pH 5.0. The

TABLE III: Alignment of Tryptic Peptides and the Partial Amino Acid Sequence of SCP.[a]



[a] (⟝, below) Determined by the dansyl procedure on purified tryptic peptide; (⟝, below) determined by carboxypeptidase A digestion; (⟝, above) determined by automated Edman degradation (Table I).

amino-terminal residue of each tryptic peptide determined by dansylation is also shown in Table I.

The data obtained from automated Edman degradation of SCP and analyses of its tryptic peptides were used to deduce the partial sequence of SCP, as summarized in Table III. The peptides Tp 3–6, 8, and 9 were unequivocally assigned to the amino-terminal 18 residues of the sequence. The detection of glycines at cycles 19 and 22 in the sequencer runs placed Tp 2 and Tp 11 in the region of residues 19 through 24. However, their respective orders could not be determined from these data. The sequences of the carboxyl-terminal tryptic peptide (Tp 1) and residues 20 through 28 were established unequivocally as described below.

Sequence of Tp 1. Since Tp 1 contained no lysine or arginine and had the C-terminal phenylalanyl residue of SCP, it was the carboxyl-terminal tryptic peptide. The amino-terminal portion of Tp 1 was sequenced by manual Edman degradation of 60 nmol with direct identification of amino acids by gas chromatography of their phenylthiohydantoin derivatives. Pth-leucine (6.3 nmol) was detected at cycle 1 in a yield of 10%. The second cycle contained no identifiable Pth derivative. Cycle three contained 3.8 nmol of Pth-tyrosine. Cycle four contained 4.0 nmol of Pth-valine, while the fifth cycle contained 3.3 nmol of Pth-glycine. After cycle five, the amount of background increased to the point that any further sequence assignments were impossible. Since Tp 1 was shown to contain the single tryptophyl residue of SCP by both spot test with Erlich's reagent and by aminopeptidase M digestion (Table I), tryptophan was suspected at position 2 (Pth-glycine or Pth-glutamine at this cycle should have been detected).

In order to sequence the remainder of Tp 1, a thermolysin digest was performed on SCP and the resultant peptides resolved on Sephadex G-50 were as described in Figure 4. Pool Tl 7 was composed of two peptides, as judged by thin-layer chromatography, one of which contained phenylalanine. This peptide was purified by applying pool Tl 7 to a 0.9 × 15 cm column of Beckman AA-15 in 0.2 M pyridine–acetic acid, pH 3.1, and collecting the fraction eluted with 100 mL of this buffer. This fraction, designated Tl 7A, contained a single peptide, was isolated in 50% yield, and had the amino acid composition shown in Table IV. The presence of a single phenylalanyl residue in this peptide indicated that Tl 7A was the C-terminal thermolytic peptide. Other peptide pools from the thermolysin digest were not fully characterized.

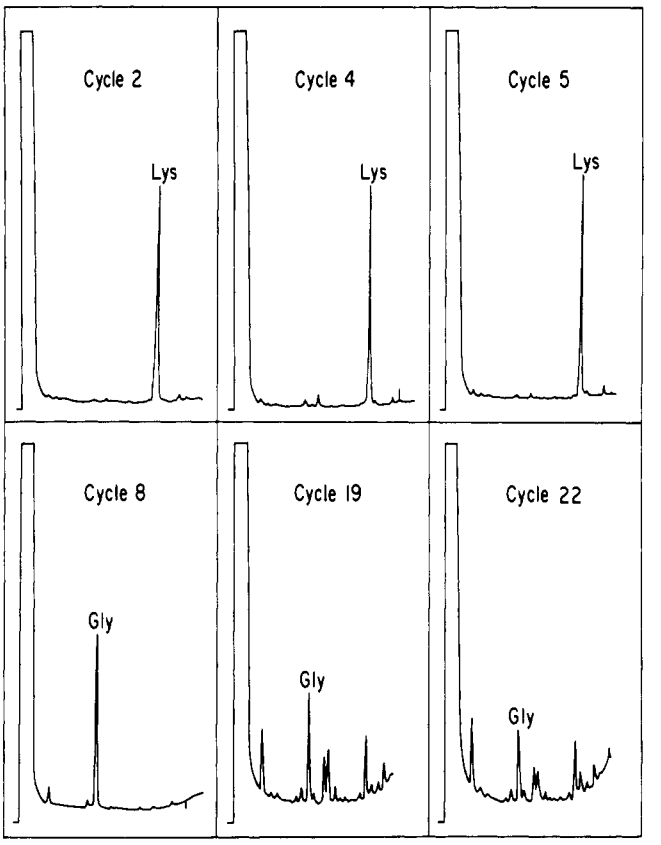


FIGURE 2: Gas chromatography of samples from automated Edman degradation. Fractions from the indicated cycles were treated as described under Methods prior to gas chromatographic analysis. The separations shown were obtained using Beckman 10% SP-400 with a linear temperature rise program consisting of 2 min at 165 °C, a linear rise of 110 °C in 16 and 4 min at the upper isothermal temperature (275 °C). The profiles shown for cycles 8, 19, and 22 were obtained without silylation while those for cycles 2, 4, and 5 were obtained after on column silylation with *N,O*-bis(trimethylsilyl)acetamide. The percent of sample injected at each cycle and the attenuation used were as follows: cycles 2, 4, 5, 5% at 1600; cycle 8, 8% at 1600; cycles 19, 22, 16% at 400.

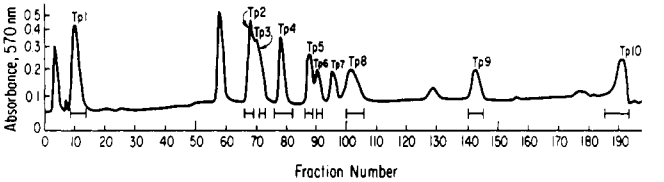


FIGURE 3: Chromatography of tryptic peptides. A trypsin digest of 250 nmol of ϕX 174 SCP dissolved in 2.0 mL of 50% (v/v) acetic acid was applied to a column (0.9 × 17 cm) of Beckman AA-15 equilibrated and eluted with 0.2 N pyridine–acetic acid, pH 3.1, as a starting buffer. Gradient elution was commenced after 1 h with a linear gradient formed using 100 mL each of 0.2 N pyridine–acetic acid, pH 3.1, and 2.0 N pyridine–acetic acid, pH 5.0, as starting and limit buffers, respectively. Peptides remaining on the column after gradient elution were recovered by elution with limit buffer for 3 h following gradient elution. The column was operated at a flow rate of 12 mL/h at 55 °C. Cysteic acid (1 μmol) was included in the sample as a marker for chart alignment. Column effluent was monitored continuously as described under Methods. Fractions were collected at 6.0 min/tube. The unlabeled peaks in the trace contained no amino acids and were presumably salt peaks.

TABLE IV: Edman Degradation of Peptide Tl 7A.[a]

Amino acid	Composition	Cycle 1	2	3	4	5
Glx	2.0	2.0	2.1	2.0	*1.7*	*1.1*
Gly	2.1	2.2	*1.6*	*1.0*	1.0	0.6
Val	0.9	*0.3*	0.0	0.0	0.0	0.0
Phe	0.9	0.9	0.9	0.9	0.9	0.9
	Residue identified	Val	Gly	Gly	Glx	Glx

[a] The values in italics are assumed to have dropped from the preceding cycle.

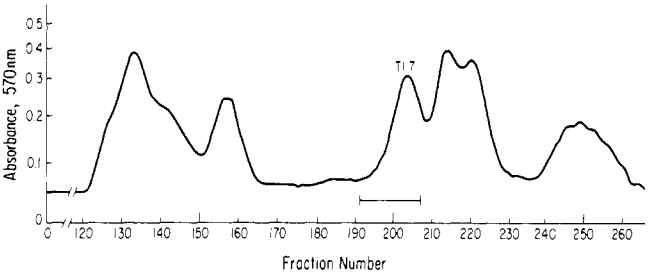


FIGURE 4: Chromatography of ϕX 174 SCP thermolysin digest. The peptides produced by digestion of ϕX 174 SCP with thermolysin as described under Methods were dissolved in 2.0 mL of 0.2 N acetic acid and separated by gel filtration on a column (2.0 × 140 cm) of Sephadex G-50 (fine) in the same buffer. E-dansyl-lysine (1 μmol) was added to the sample as a marker for chart alignment. The column was operated at a flow rate of 20 mL/h at room temperature with continuous monitoring of the effluent as described under Methods. Fractions were collected at 6.0 min/tube.

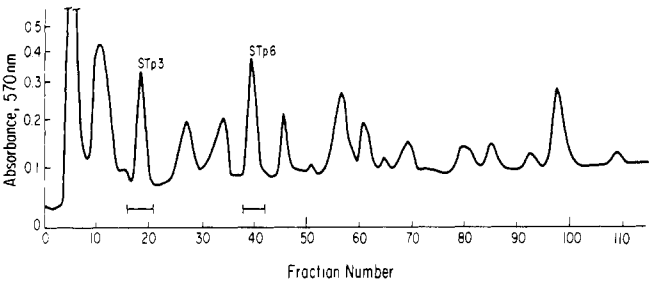
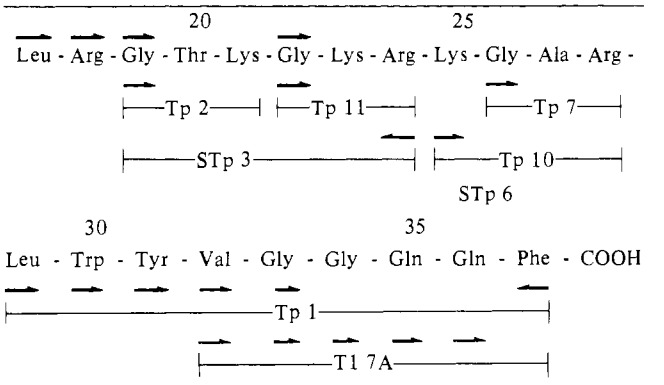


FIGURE 5: Chromatography of succinyl tryptic peptides. A trypsin digest of succinylated ϕX 174 SCP prepared as described under Methods was chromatographed on Beckman AA-15 exactly as described in the legend to Figure 3 except at a flow rate of 15 mL/h.

Peptide Tl 7A gave no detectable color when 10 nmol was spotted on filter paper and sprayed with Erlich's reagent. Aminopeptidase M treatment released no tryptophan as judged by amino acid analysis. These results confirmed the location of tryptophan at position 2 of Tp 1. Aminopeptidase M treatment released 1.5 mol of glutamine and only 0.3 mol of glutamic acid per mol of peptide. It was concluded that both glutamyl residues in Tl 7A were amidated. Subtractive Edman degradation gave the results shown in Table IV. The sequence of Tp 1 deduced from the above data is shown in Table V.

Sequence of Residues 20–28; Tp 2, 7, and 11. Since the three tripeptides Tp 2, 7, and 11 all have amino-terminal glycine, it was not possible to obtain a unique ordering of these peptides from the automated sequencer runs which placed glycine at residues 19 and 22. In order to obtain the necessary peptides to complete the sequence, whole SCP was succinylated. Succinylated SCP was digested with trypsin and resulting peptides were purified by chromatography on Beckman AA-15 as shown in Figure 5. Ten percent of each pool was

TABLE V: Amino Acid Sequence of Residues 19–37.[a]



[a] (⟝, above) Determined by automated Edman degradation (see Table I); (⟝, below) determined by the dansyl technique or by manual Edman degradation as described in the text; (⟝, below) determined by digestion with carboxypeptidase as described in the text.

taken for amino acid analysis. Pool STp 3 which contained a single peptide based on thin-layer chromatography had the following composition: 1.90 (2) lysine, 1.06 (1) arginine, 0.93 (1) threonine, 2.26 (2) glycine. STp 3, isolated in 50% yield, was placed in the region 20–28 by virtue of the fact that it contained the only threonyl residue of SCP. In addition, STp 3 had the exact composition of Tp 2 plus Tp 11. Treatment of STp 3 with carboxypeptidase B released 1.0 mol of arginine per mol of peptide. Therefore, Tp 11 must represent the carboxyl-terminal three residues of STp 3 and the sequence of residues 19–24 (STp 3) must be that shown in Table V. This was further substantiated by the fact that the peptide STp 6, which had the amino acid composition 0.85 (1) lysine, 1.05 (1) arginine, 1.28 (1) glycine, 0.98 (1) alanine, was isolated in 67% yield. STp 6 had the same composition as tryptic peptide Tp 10 which contained an amino-terminal lysine (Table I). The composition and molar yields of Tp 10 and Tp 7 (Table II) indicated that Tp 10 represented Tp 7 with an amino-terminal lysine. Therefore, Tp 10 was placed in the sequence at positions 25–28 by elimination as seen in Table V. Assignment of residues 20–28 based on the above data completed the sequence of SCP as shown in Figure 6.

Discussion

Data presented here establish the sequence of ϕX174 SCP. The two sequencer runs yielded consistent results. The serines at cycles 1 and 7 were detected in the second run using the milder conversion conditions mentioned under Results. Assignments made by gas chromatographic analysis were confirmed in most cases by back-hydrolysis of Pth derivatives with subsequent amino acid analysis. Automated Edman degra-

dation was limited owing to the sequence of residues 11 through 16 (-Pro-Gly-Arg-Pro-Gln-Pro-) where a substantial reduction in yield was experienced. The assignment of glycyl residues at positions 19 and 22 (Figure 2) was, however, unequivocal as Pth-glycine was absent at cycle 18 and present in only minor amounts at cycles 20, 21, and 23.

Data from analyses of tryptic peptides confirmed the position of residues obtained in the sequencer runs and provided the basis for much of the remainder of the sequence. The characterization of these peptides shown in Table II indicated that the complete sequence of SCP was accounted for in the unique tryptic peptides plus one arginyl and two lysyl residues. Peptides Tp 1, 2, and 11 were completely released by trypsin digestion and isolated in yields of 46–80%. The remainder of the unique tryptic peptides (Tp 3–7) were also present in incomplete digestion products.

The total yield of Tp 3, including that present as the partial digestion product Tp 8, was 56%. Similarly, the yield of Tp 4 plus that present as Tp 8 represented a 92% yield of the tetrapeptide Tp 4. The amino-terminal four residues Ser-Lys-Gly-Lys were isolated in peptides Tp 5, Tp 6, and Tp 9, the total yield of Ser-Lys (Tp 6) being 71% and Gly-Lys (Tp 5) 101%. The yield of Tp 7 plus that present as the partial digestion product Tp 10 was 97%. Based on the amino acid compositions of the overlap peptides, Tp 9 and Tp 8, the pairs of peptides Tp 6 and 5, and Tp 4 and 3 were grouped together. Dansyl end-group analysis of these peptides positioned the Tp 6 amino terminal to Tp 5 and the Tp 4 amino terminal to Tp 3. Since the amino terminus of SCP was shown to be serine, either Tp 9 or Tp 8 had to be the amino-terminal tryptic peptide. The presence of Pth-lysine at cycle 2 in the sequencer run indicated that Tp 9 was at the NH_2 terminus.

The most difficult region of SCP to sequence was the region of residues 19 through 28. Completion of the sequence of other regions of SCP indicated residues 19–28 were composed of three unique peptides (Tp 2, 7, and 11) and the overlapping peptide Tp 10. From amino acid composition, dansyl end-group analysis, and isolation of STp 6, Tp 10 was shown to be a combination of Tp 7 with amino-terminal lysine. Consequently, the glycyl residues at positions 19 and 22 could not come from Tp 10; therefore, by elimination Tp 10 must represent residues 25 to 28. In order to establish the order of Tp 2 and 11, SCP was succinylated and the peptide STp 3 was isolated. This peptide represents an overlap of Tp 2 and 11. Since succinylation blocks trypsin cleavage at lysine but not arginine, the tryptic peptides must be joined in the succinyl peptide with the lysine internal and STp 3 must have the sequence -Gly-Thr-Lys-Gly-Lys-Arg- (Tp 2–Tp 11). This was confirmed by the fact that carboxypeptidase B released arginine from STp 3. The isolation of the succinyl tryptic peptide STp 6 when considered with the above data rules out the presence of a lysine but not an arginine between position 24 and 25 or 28 and 29 in the sequence. We feel confident, however, that all of the arginyl residues in SCP have been positioned at other places in the sequence. In addition, B. G. Barrell (see Sanger et al., 1977) has obtained a complete nucleotide sequence in the region of the ϕX174 genome extending from the C terminus of cistron D to the N-terminal region of cistron F. This independently derived nucleotide sequence across the entire gene is completely consistent with our amino acid sequence and adds additional evidence that the amino acid sequence in the regions where we have not isolated overlap peptides is correct.

Benbow et al. (1972) suggested the existence of a ϕX174 gene immediately preceding cistron F, which they called cistron J. Cistron J was genetically defined by a single nonsense mutant (*am* 6). However, recent marker rescue experiments

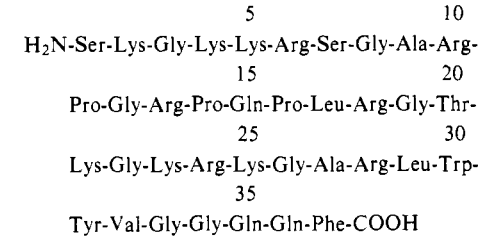


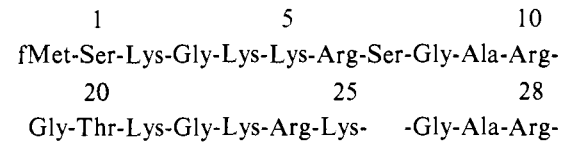
FIGURE 6: The complete amino acid sequence of ϕX174 SCP.

(Weisbeek, 1976) indicate that *am* 6 is really a gene E mutant as earlier work had suggested (Hutchison and Sinsheimer, 1966; Benbow et al., 1971). In the absence of a genetic mutation which unambiguously defines cistron J, we propose to redefine J as the ϕX174 gene which encodes the amino acid sequence of the SCP.

The amino acid sequence of the ϕX174 SCP (Figure 6) has a number of interesting features. Lysine, arginine, and glycine comprise 20 of the 37 residues. The spacing of the basic residues is reminiscent of the sequence of the amino terminal portion of eucaryotic histones thought to represent the DNA-binding domain of these proteins (DeLange and Smith, 1971). The hydrophobic residues phenylalanine, tyrosine, tryptophan, leucine, valine, and glutamine are concentrated in the nine carboxyl-terminal residues of SCP. The α-carboxyl group of the C terminus represents the only negative charge of the molecule.

SCP is devoid of histidine, aspartic acid, asparagine, isoleucine, and sulfur-containing amino acids. This is particularly important, since [^{35}S]methionine or [^{3}H]histidine (Jazwinski et al., 1975), often used to label proteins, would not label the SCP.

Another interesting feature of SCP is the presence of two regions of homologous amino acid sequence. Residues 1 to 10 and 20 to 28 can be aligned as follows:



Although tandem duplications may have been involved in the evolution of SCP, it is also possible that the function of this protein (e.g., binding DNA) requires a repeating linear sequence of basic residues separated by an eight-residue segment. Inspection of the DNA nucleotide sequence (Sanger et al., 1977) has not allowed us to decide between convergent and divergent mechanisms for the evolution of these two homologous segments.

Although more work will be necessary to determine the role of SCP in the virus life cycle, the sequence of this protein suggests it would be well suited to function in condensing the single-stranded DNA in the virion. The sequence -Pro-X-X-Pro-X-Pro- (residues 11–16) occurs in center of the two basic regions of SCP. A similar sequence occurring in the much larger basic protein of myelin may allow the formation of an extended linear conformation (Eylar, 1973). Such a folded linear structure in the SCP would produce a hairpin molecule containing two opposing basic regions and a very hydrophobic tail structure which would have obvious advantages for close packing of DNA. The data of other workers, in fact, suggest that the SCP can bind to DNA (Linney et al., 1972). Siden and Hayashi (1974) have detected a DNA–protein complex formed when normal virus maturation is blocked in which 70% of the protein is SCP. These results further support the view

that SCP may function by condensing the single-stranded DNA of the ϕX174 virion.

Acknowledgments

We thank Farida Sharief and Pearl Cole for excellent technical assistance. Special thanks are extended to Dr. B. G. Barrell for comunicating his results prior to publication.

References

Barrell, B. G., Air, G. M., and Hutchison, C. A. III (1976), *Nature (London) 264,* 34.

Benbow, R. M., Hutchison, C. A. III, Fabricant, J. D., and Sinsheimer, R. L. (1971), *J. Virol. 7,* 549.

Benbow, R. M., Mayol, R. F., Picchi, J. C., and Sinsheimer, R. L. (1972), *J. Virol. 10,* 99.

Bennett, J. C. (1967), *Methods Enzymol. 11,* 330.

Burgess, A. B., and Denhardt, D. T. (1969), *J. Mol. Biol. 44,* 377.

DeLange, R. J., and Smith, E. L. (1971), *Annu. Rev. Biochem. 40,* 279.

Edgell, M. H., Hutchison, C. A. III, and Sinsheimer, R. L. (1969), *J. Mol. Biol. 42,* 547.

Eylar, E. H. (1973), in Proteins of the Nervous System, Schneider, D. I., Ed., New York, N.Y., Ravon Press, 27.

Gelfand, D. H., and Hayashi, M. (1969), *J. Mol. Biol. 42,* 547.

Godson, G. N. (1971), *J. Mol. Biol. 57,* 541.

Hartley, B. S. (1970), *Biochem. J. 119,* 805.

Herman, A. C., and Vanaman, T. C. (1975), *Anal. Biochem. 63,* 550.

Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry 11,* 4493.

Hill, R. L., and Delaney, R. (1967), *Methods Enzymol. 11,* 339.

Hirs, C. H. W. (1967), *Methods Enzymol. 11,* 197.

Hutchison, C. A. III, and Sinsheimer, R. L. (1966), *J. Mol. Biol. 18,* 429.

Jazwinski, S. M., Lindberg, A. A., and Kornberg, A. (1975), *Virology 66,* 283.

Konigsberg, W., and Hill, R. J. (1962), *J. Biol. Chem. 237,* 2547.

Linney, E. A., Hayashi, M. N., and Hayashi, M. (1972), *Virology 50,* 381.

Mayol, R. F., and Sinsheimer, R. L. (1970), *J. Virol. 6,* 310.

Niall, H. D. (1974), *Methods Enzymol. 27,* 942.

Panyim, S., and Chalkley, R. (1969), *Arch. Biochem. Biophys. 130,* 337.

Peterson, J. D., Nerhlich, S., Dyer, P. E., and Steiner, D. F. (1972), *J. Biol. Chem. 247,* 4866.

Pett, D. M., Vanaman, T. C., and Joklik, W. K. (1973), *Virology 52,* 174.

Pisano, J. J., Bronzert, T. J., and Brewer, H. B., Jr. (1972), *Anal. Biochem. 45,* 43.

Poljak, R. J. (1968), *Virology 35,* 185.

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 (London) 265,* 687.

Scoffone, E., Fontana, A., and Rocchi, R. (1968), *Biochemistry 7,* 971.

Shank, P. R., Hutchison, C. A. III, and Edgell, M. H. (1977), *Biochemistry 16* (preceding paper in this issue).

Siden, E. J., and Hayashi, M. (1974), *J. Mol. Biol. 89,* 1.

Sinsheimer, R. L. (1968), *Progr. Nucleic Acid Res. Mol. Biol. 8,* 115.

Smithies, O., Gibson, D., Fanning, E. M., Goddfliesh, R. M., Gilman, J. G., and Gallantyne, D. C. (1971), *Biochemistry 10,* 4912.

Suruda, A. J., and Poljak, R. J. (1971), *Virology 46,* 164.

Weisbeek, P. J. (1976), *Virology 72,* 61.

Weisbeek, P. J., and Sinsheimer, R. L. (1972), *Proc. Natl. Acad. Sci. U.S.A. 71,* 3054.

Zuccarelli, A. J., Benbow, R. M., and Sinsheimer, R. L. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 1905.