Mol Gen Genet (1987) 206:144–153

MGG

Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1

Wolf-Dieter Reiter, Peter Palm, Agnes Henschen, Friedrich Lottspeich, Wolfram Zillig, and Bernd Grampp
Max-Planck-Institut für Biochemie, D-8033 Martinsried, Federal Republic of Germany

Summary. Three structural proteins, VP1, VP2 and VP3, of the virus-like particle SSV1 of the thermoacidophilic archaebacterium *Sulfolobus* sp. strain B12 were purified. VP1 and VP3 are very hydrophobic and show a high degree of homology. They consist of 73 and 92 amino acid residues, respectively. The third protein, VP2, is extremely basic containing 29 basic amino acids but only 4 acidic ones in a total of 74 amino acid residues. The genes encoding these three proteins were mapped within the genome by comparison of N-terminal amino acid sequences with the SSV1 DNA sequence. The three genes are closely linked in the order VP1-VP3-VP2 and the coding strand is the same in all three genes. Ten nucleotides separate the stop codon for VP1 from the initiation codon for VP3 and one nucleotide separates the genes encoding VP3 and VP2. Duplicate putative ribosome binding sites are found upstream of the initiation codons for VP2 and VP3. The major coat protein VP1 does not start with a methionine residue and appears to be the result of proteolytic cleavage of a precursor molecule. Transcription of the region encoding VP1, VP2 and VP3 results in the formation of two mRNAs of 0.5 kb and 1.0 kb, the shorter one only encoding VP1, the longer one spanning all three genes. A 61 bp sequence encoding part of VP1 is exactly repeated within the gene for VP3 and these identical sequences are translated into stretches of identical amino acids in the two proteins. A function of this repeated DNA sequence beyond its coding properties is very likely.

Key words: Archaebacteria – *Sulfolobus* – Virus – Structural proteins – Gene organization

Introduction

It has been shown by comparative sequence analysis of conserved macromolecules that the prokaryotes can be divided into two very distinct groups: the eubacteria and the archaebacteria (Fox et al. 1980). The archaebacteria can in turn be subdivided into essentially two branches, one comprising the methanogens and the extreme halophiles and the other comprising extremely thermophilic archaebacteria (Woese and Olsen 1986). Certain features of archaebacteria appear more "eubacterial" in the former branch but more "eukaryotic" in the branch of extreme thermophiles (Zillig et al. 1985a).

In the case of eubacteria, the study of bacteriophages has significantly contributed to an understanding of the molecular biology of their hosts. Viruses that can serve as model systems for gene organization and gene expression in halophilic and methanogenic archaebacteria have been found in *Halobacterium* (Torsvik and Dundas 1974, 1980; Wais et al. 1975; Schnabel et al. 1982a; Pauling 1982; Daniels and Wais 1984; Vogelsang-Wenke and Oesterhelt 1986) and *Methanobrevibacter* (Baresi and Bertani 1984). Only one of these viruses, bacteriophage ΦH of *Halobacterium halobium,* has been characterized in detail at the molecular level (Schnabel et al. 1982a, b; Schnabel 1984a, b; Schnabel and Zillig 1984; Schnabel et al. 1984).

In the group of extremely thermophilic archaebacteria four viruses of *Thermoproteus* (Janekovic et al. 1983) and the virus-like particle SSV1 of *Sulfolobus* sp. strain B12 (Martin et al. 1984) have been described. SSV1 is the only extrachromosomal element of an extremely thermophilic archaebacterium, for which detailed molecular biological data have been obtained. SSV1 had originally been termed SAV1 (*Sulfolobus acidocaldarius* virus 1) because of the incorrect classification of its host as *Sulfolobus acidocaldarius* (Yeats et al. 1982). *Sulfolobus* B12 has indeed been shown to be a close relative of *S. solfataricus* (Zillig et al. 1985b).

A small number of SSV1 particles are spontaneously released from their host in late-logarithmic cultures. After UV irradiation of such a culture, however, large numbers of the lemon-shaped virus-like particles are released without apparent lysis of their host cells (Martin et al. 1984). The genome of SSV1 is a 15.46 kb plasmid (Yeats et al. 1982) that is packaged as positively supercoiled DNA (Nadal et al. 1986). Within the cells SSV1 DNA is present both as a free plasmid and in a site-specific integrated form (Yeats et al. 1982). The complete nucleotide sequence of SSV1 DNA has been determined (P. Palm and B. Grampp, unpublished data). Here we report the purification of three structural proteins from SSV1 and the characterization of the genes encoding these proteins.

Materials and methods

Materials. [α-^{35}S]dATP was from Amersham, *Escherichia coli* DNA polymerase (Klenow fragment) was from Pharmacia and T4 DNA ligase and pancreatic DNAase I were

Offprint requests to: W.-D. Reiter

from Boehringer. The low molecular weight protein standard was obtained from Bethesda Research Laboratories.

Bacterial strains and culture conditions. Growth conditions for *Sulfolobus* sp. strain B12 (formerly *Sulfolobus acidocaldarius* strain B12) were as described previously (Yeats et al. 1982).

Purification of SSV1. The UV irradiation of *Sulfolobus* B12 cultures and the preparation of a cell-free supernatant of the induced culture were carried out as described previously (Martin et al. 1984). SSV1 was precipitated from the culture medium by addition of polyethylene glycol 6000 to 10% final concentration. After 12 h at 4° C the precipitate was recovered by centrifugation for 1 h at 4° C in the 6 x 1,000 ml rotor of the WKF centrifuge G50K. The pellet was dissolved in 20 mM sodium acetate/acetic acid, pH 6.0, 10 mM $MgSO_4$ (AM buffer); 10 ml buffer were used for 1 l of the original culture. 20 ml of this concentrated SSV1 preparation were layered on top of a CsCl step gradient consisting of 9 ml of 13.5% (w/w) CsCl and 9 ml of 27% (w/w) CsCl in AM buffer. After centrifugation for 90 min at 20° C and 25,000 rpm in a Beckman SW27 rotor, almost pure SSV1 formed a band between the two CsCl solutions. This band was collected and subjected to CsCl equilibrium density gradient centrifugation as described (Martin et al. 1984), except that AM buffer was used to prepare the CsCl solution.

Purification of VP1. For preparation of VP1 6 vol. of a 3:2 mixture of chloroform and ethanol were added to 5 ml of a solution of SSV1 in 20 mM Tris-HCl, pH 8.0, 1 mM EDTA (TE buffer). The amount of SSV1 corresponded to a DNA concentration of about 1 mg/ml. The precipitated material (mostly DNA) was removed by centrifugation and the supernatant was dried in vacuo. The residue was dissolved in 100 µl of phenol equilibrated with TE buffer and the solvent was removed by lyophilization. Purification of this crude VP1 preparation by preparative SDS-polyacrylamide gel electrophoresis was performed as described by Schnabel et al. (1983). Visualization of bands and elution of the protein was essentially as in Schnabel et al. (1983), but 0.1% N-lauroylsarcosine was used instead of SDS. Two bands of apparent molecular weights of 13 kDa and 18 kDa were excised and eluted separately, but since these bands proved to be the same protein (see Results) both eluates were later combined. The eluate was extensively dialysed against 0.05% N-lauroylsarcosine and the detergent was extracted with diethylether after acidifying the solution with formic acid (1 M final concentration). The precipitate consisting of VP1 that was obtained upon removal of the detergent was collected by centrifugation.

Purification of VP2 and enrichment of VP3. Five millilitres of a preparation of SSV1 in TE buffer (concentration of SSV1 corresponding to 2 mg/ml DNA) were precipitated with 4 vol. ethanol. The precipitate was collected by centrifugation and resuspended in 10 ml of TE buffer. The suspension was heated to 95° C for 10 min and then cooled to room temperature. Five micrograms of pancreatic DNAase I were added to the viscous solution and the SSV1 DNA was digested for 5 h at 37° C. Almost all of the SSV1 protein remained insoluble at this step. It was recovered by centrifugation and dissolved in 0.5 ml of phenol saturated with TE buffer. After addition of 4 vol. ethanol the precipitate was recovered by centrifugation, washed with 70% ethanol and dried for 3 min in a vacuum desiccator. VP2 was extracted from this preparation ("fraction A") with 1 ml of a 3% solution of N-lauroylsarcosine at 95° C for 5 min. Removal of the detergent from the crude VP2 preparation was done as for VP1 (see above). For amino acid sequence analysis this preparation was further purified by preparative SDS-polyacrylamide gel electrophoresis as for VP1. As most of the VP2 remained in solution after extraction of the N-lauroylsarcosine, the preparation was dialysed against 20 mM ammonium acetate to remove non-volatile salts and then lyophilized.

For enrichment of VP3 protein fraction A (see above) was dissolved in 0.5 ml of phenol saturated with TE buffer and 6 vol. chloroform/ethanol (3:2) were added to remove most of the VP1. The resulting precipitate ("fraction B") contained VP3 as the major component.

SDS-polyacrylamide gel electrophoresis. SSV1 proteins were analysed on 15% SDS-polyacrylamide gels using the system of Laemmli (1970) or on 5%–25% gradient gels as described by Mirault et al. (1971). In some cases, especially for size determinations, SDS-urea gels as described by Tandy et al. (1983) were used.

DNA sequence analysis. All DNA sequencing was done by the chain termination method (Sanger et al. 1977) using the M13 cloning and sequencing technique (Sanger et al. 1980; Messing and Vieira 1982). The programs of Devereux et al. (1984) and Staden (1980) were used for computer-aided editing and comparison of sequences. The second largest *Eco*RI fragment of SSV1 DNA (Yeats et al. 1982) was circularized by ligation. Random cloning of fragments obtained by sonication of this DNA into the *Sma*I site of M13mp8 and sequencing of the clones were done according to Deininger (1983) following the protocol of Amersham International plc.

Protein sequence analysis. N-terminal amino acid sequences of purified VP1 and VP2 were determined in a prototype liquid-phase sequenator using a program optimized for peptide sequencing (Lottspeich et al. 1984). For determination of the N-terminal amino acid sequence of VP3 the procedure described by Aebersold et al. (1986) was used with some minor modifications. In short, a protein mixture containing this polypeptide (fraction B, see above) was separated on a 15% polyacrylamide gel and electrophoretically transferred to a sheet of derivatized glass fibre. The band of interest was cut out after staining with Coomassie Brilliant Blue R250 and directly sequenced in a gas phase sequenator (Applied Biosystems 470A). The phenyl-thiohydantoin amino acid derivatives were identified by an isocratic HPLC system as described by Lottspeich (1985).

Purification of Sulfolobus RNA. Frozen *Sulfolobus* cells were lysed in an SDS/urea buffer as described by Sather and Agabian (1985). The resulting viscous solution was repeatedly passed through a hypodermic syringe to shear DNA and then extracted three times with phenol/chloroform. The nucleic acid was precipitated with ethanol, collected by centrifugation and re-dissolved in 7 M guanidinium chloride, 0.025 M sodium citrate, pH 7.0. After repeated extraction of this solution with phenol/chloroform

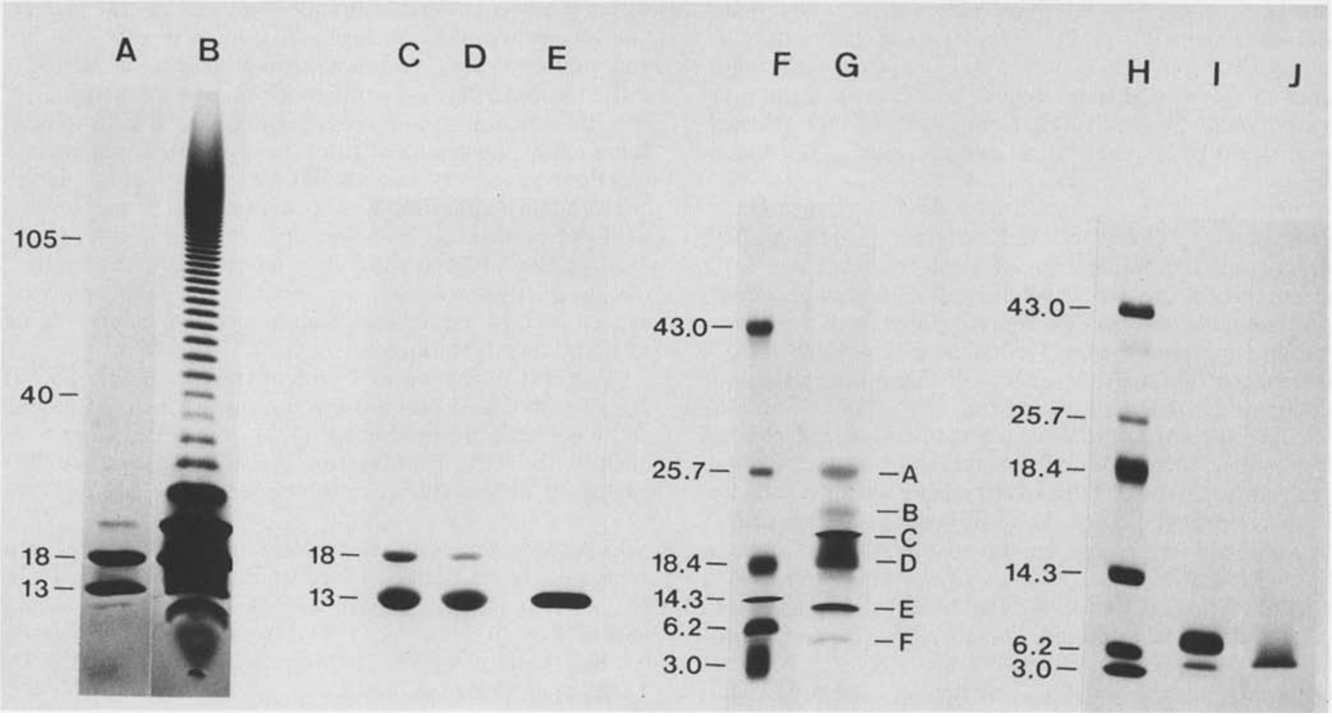


Fig. 1. SDS-polyacrylamide gel electrophoresis of SSV1 proteins. Proteins were separated on 5%–25% gradient gels (lanes A–E), 15% gels (lanes F, G) or 15% gels containing 6 M urea (lanes H, I, J). The numbers to the left of lanes A, C, F and H give apparent molecular weights in kilodaltons. The molecular weights of the SSV1 structural proteins calculated from DNA sequence data are 7.74 kDa (VP1), 8.61 kDa (VP2) and 9.80 kDa (VP3). Lane A, SSV1 protein soluble in 3:2 chloroform/ethanol (crude VP1). Lane B, same as lane A, but overloaded with protein. Lane C, protein eluted from the 13 kDa band of crude VP1 (see lane A) and re-applied to a gel. Lane D, protein eluted from the 18 kDa band of crude VP1 (see lane A) and re-applied to a gel. Lane E, VP2 after purification by preparative SDS-polyacrylamide gel electrophoresis . Lane G, fraction B of SSV1 protein mostly consisting of VP3. Lane I, crude VP2 obtained by detergent extraction of protein fraction A. The low molecular weight minor band probably corresponds to band F in lane G (a host-encoded DNA-binding protein). Lane J, VP1 after purification by preparative polyacrylamide gel electrophoresis. Lanes F, H, sizes markers

the RNA was again precipitated with ethanol. This crude RNA preparation was dissolved in 0.1 M Tris-HCl, pH 8.0, 0.01 M EDTA and caesium trifluoroacetate solution (Pharmacia Fine Chemicals) was added to a final density of 1.75 g/ml. The RNA was banded by centrifugation for 20 h at 65,000 rpm and 10° C in a Beckman VTi65 rotor. A flocculent band in the middle of the tube and all the solution below it were collected and the RNA was recovered by ethanol precipitation.

Preparation of strand-specific M13 probes. M13 clones containing fragments of SSV1 DNA obtained by sonication were radioactively labelled as described by Hu and Messing (1982).

Northern analysis of transcripts. Total *Sulfolobus* RNA (20 μg per lane) was denatured by incubation with glyoxal, separated on 1% agarose gels and transferred to nitrocellulose as described by Thomas (1983). Hybridization was performed essentially as described by Maniatis et al. (1982).

Results

Purification of SSV1

The procedure described by Martin et al. (1984) for the purification of the virus-like particles was simplified. For precipitation of SSV1 from the culture medium PEG 6000 was used instead of ammonium sulphate. A CsCl step gradient centrifugation followed by a CsCl equilibrium density gradient centrifugation was used to obtain pure SSV1.

Purification of SSV1 proteins

Since SDS-polyacrylamide gel electrophoresis of total SSV1 proteins only showed several blurred and overlapping bands (Martin et al. 1984), a fractionation procedure was employed that included detergents and organic solvents. Addition of a large excess of a 3:2 chloroform/ethanol mixture to a concentrated SSV1 preparation yielded a soluble fraction containing most of the SSV1 protein and a precipitate mainly consisting of DNA. After removing the precipitate by centrifugation and drying the supernatant in vacuo this protein fraction was insoluble in the same solvent mixture from which it had been recovered and it was also insoluble in concentrated solutions of strong detergents e.g. 10% SDS at 100° C. The protein residue was, however, soluble in phenol saturated with an aqueous buffer. Removal of the phenol by lyophilization gave rise to a preparation that readily dissolved in ionic detergents (e.g. 1% SDS) and in 3:2 chloroform/ethanol. Analysis by SDS-polyacrylamide gel electrophoresis showed the presence of essentially two bands corresponding to apparent molecular weights of 13 kDa and 18 kDa (Fig. 1, lane A). A ladder of more than 30 individual bands was observed when the gel was overloaded (Fig. 1, lane B). The 13 kDa and 18 kDa bands represented the same protein as shown by elution of the proteins from the excised bands and re-application of the eluates to an SDS-polyacrylamide gel. Both bands again yielded a 13 kDa major band and an 18 kDa minor band (Fig. 1, lanes C and D). The polypeptide forming these bands is the major protein component of the virus-like particles and was termed VP1 (viral protein 1). For the amino acid sequence analysis VP1 was purified by preparative SDS-polyacrylamide gel electrophoresis as described in Materials and methods.

A second protein component of SSV1 (VP2) could be selectively extracted with anionic detergents from an SSV1 protein preparation (fraction A) that was obtained after destruction of the SSV1 particles by addition of ethanol and removal of the DNA by nuclease digestion. The apparent molecular weight of VP2 from SDS-polyacrylamide gel electrophoresis was 13 kDa like that of the major band of VP1. VP2 could, however, be distinguished from VP1 as it stained dark blue with Coomassie Brilliant Blue R250 whereas VP1 stained faintly purple. For the amino acid sequence analysis preparative SDS-polyacrylamide gel electrophoresis was used to purify VP2 to homogeneity (Fig. 1, lane E).

After dissolving protein fraction A in phenol, a 3:2 chloroform/ethanol mixture was added, resulting in the precipitation of some protein but leaving VP1 in solution. The analysis of the precipitated protein (fraction B) by SDS-polyacrylamide gel electrophoresis indicated the presence of a major band with an apparent molecular weight of 18 kDa (band D in Fig. 1, lane G). This band, which was stained faintly purple by Coomassie Brilliant Blue R250, was completely obscured by the VP1 oligomers that dominated the protein pattern when unfractionated SSV1 protein was applied to a gel (Martin et al. 1984). The protein corresponding to this 18 kDa band was termed VP3.

Amino acid sequence analysis of the three SSV1 proteins

The N-terminal amino acid sequences of purified VP1 and VP2 were determined by Edman degradation in a liquid phase sequenator. Sixty-nine amino acid residues were identified for VP1 and 16 residues for VP2. The N-terminal amino acid residue of VP1 was glutamic acid and that of VP2 was methionine. Neither of the N-terminal amino acids was found to be blocked by covalent modification. As insufficient amounts of VP3 were available for purification by preparative gel electrophoresis, a different strategy for sequencing was employed. A protein fraction containing VP3 (fraction B, see above) was separated by SDS-polyacrylamide gel electrophoresis on a semi-preparative scale and transferred to a derivatized glass fibre sheet. The 18 kDa band mainly consisting of VP3 was cut out and an N-terminal amino acid sequence of four residues was determined in a gas phase sequenator. The N-terminal amino acid of this protein was a non-modified methionine residue. This sequence analysis showed that about 20% of the protein in the "VP3 band" (band D in Fig. 1, lane G) was contaminating VP1. Four other bands on the same gel were also analysed by this method. One band with an apparent molecular weight of 27 kDa (band A in Fig. 1, lane G) and another band with an apparent molecular weight of 23 kDa (band B in Fig. 1, lane G) both consisted of VP3 indicating that this protein forms aggregates similar to VP1. The third band (band E in Fig. 1, lane G) with an apparent molecular weight of 13 kDa was found to consist of a 1:1 mixture of VP1 and VP2. Finally a band of an apparent molecular weight of about 6 kDa (band F in Fig. 1, lane G) proved to be a mixture of at least two polypeptides that appeared not to be encoded by the SSV1 genome. At least one of these polypeptides is a host-encoded DNA-binding protein (R. Reinhardt, personal communication).

DNA sequence analysis and identification of the genes encoding VP1, VP2 and VP3

The nucleotide sequence of the complete SSV1 genome has been determined (P. Palm and B. Grampp, unpublished data) using the dideoxy chain termination method (for details see Materials and methods). The comparison of this DNA sequence with the N-terminal amino acid sequences of the three SSV1 structural proteins resulted in the unambiguous identification of the corresponding genes. Only one possible DNA sequence was found when all six possible reading frames of the SSV1 genome were screened for their ability to encode the N-terminal amino acid sequences of VP1, VP2 and VP3. The DNA sequence in this part of the genome was verified by using overlapping fragments throughout this region and by sequencing both strands.

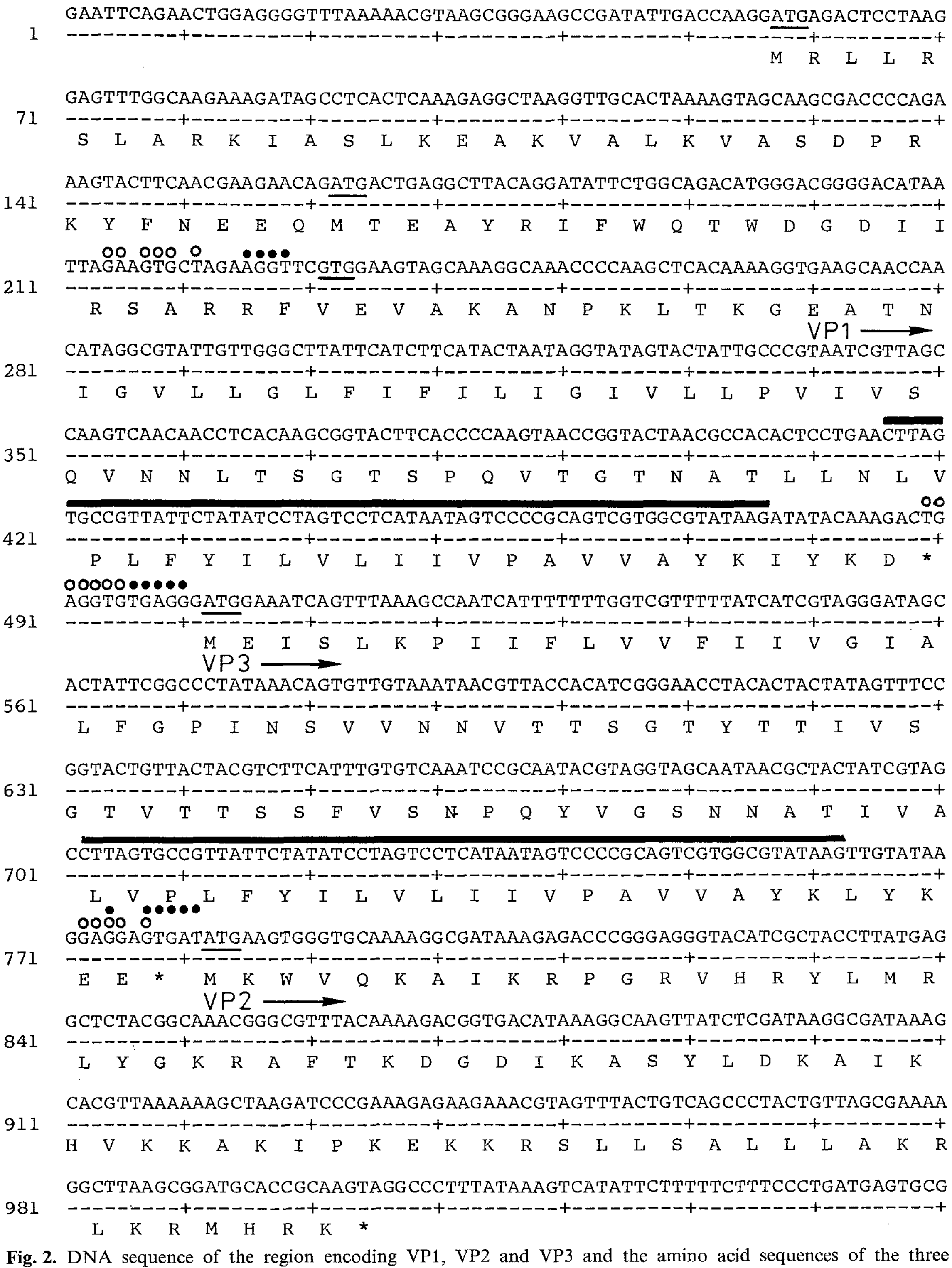
Size determination of the SSV1 proteins by gel electrophoresis

The apparent molecular weights of VP1, VP2 and VP3 as determined by SDS-polyacrylamide gel electrophoresis (15% gels or 5%–25% gradient gels) were 13 kDa (VP1), 13 kDa (VP2) and 18 kDa (VP3). In the protein preparation that was used for the sequence determination of VP3 (fraction B) two additional blurred bands with apparent molecular weights of 27 kDa and 23 kDa (bands A and B in Fig. 1, lane G) were also due to VP3 as shown by amino acid sequence analysis (see above). A dark blue band of an apparent molecular weight of 21 kDa (band C in Fig. 1, lane G) that was also seen in this preparation appeared to be due to aggregation between VP2 and VP3 (not shown).

The apparent molecular weights of VP1, VP2 and VP3 on SDS-polyacrylamide gels did not agree well with the sizes predicted from the DNA sequence (see below). Therefore VP1 and VP2 which were available in pure form, were also analysed on 15% SDS-polyacrylamide gels containing 6 M urea (Tandy et al. 1983). Using this system the apparent molecular weight of VP1 was 5.8 kDa (Fig. 1, lane J) and that of VP2 was 7.6 kDa (Fig. 1, lane I) which was in more reasonable agreement with the sizes deduced from the DNA sequence (73 amino acid residues for VP1 corresponding to 7.74 kDa; 74 residues for VP2 corresponding to 8.61 kDa).

Properties of VP2

According to DNA sequence data the SSV1 structural protein VP2 is 74 amino acid residues long and extremely basic (see Fig. 2 for the protein sequence). It contains 39% basic amino acids but only 5% acidic ones resulting in a very high positive net charge. The basic amino acids (17 lysines, 9 arginines and 3 histidines) are almost evenly distributed throughout the sequence. A short stretch of hydrophobic amino acid residues (...LLSALLLA...) is found close to the carboxy-terminus (Fig. 2). Several lines of evidence indicate



```
      GAATTCAGAACTGGAGGGGTTTAAAAACGTAAGCGGGAAGCCGATATTGACCAAGGATGAGACTCCTAAG
   1  ---------+---------+---------+---------+---------+---------+---------+
                                                              M  R  L  L  R

      GAGTTTGGCAAGAAAGATAGCCTCACTCAAAGAGGCTAAGGTTGCACTAAAAGTAGCAAGCGACCCCAGA
  71  ---------+---------+---------+---------+---------+---------+---------+
       S  L  A  R  K  I  A  S  L  K  E  A  K  V  A  L  K  V  A  S  D  P  R

      AAGTACTTCAACGAAGAACAGATGACTGAGGCTTACAGGATATTCTGGCAGACATGGGACGGGGACATAA
 141  ---------+---------+---------+---------+---------+---------+---------+
      K  Y  F  N  E  E  Q  M  T  E  A  Y  R  I  F  W  Q  T  W  D  G  D  I  I
         ○○ ○○○ ○   ●●●●
      TTAGAAGTGCTAGAAGGTTCGTGGAAGTAGCAAAGGCAAACCCCAAGCTCACAAAAGGTGAAGCAACCAA
 211  ---------+---------+---------+---------+---------+---------+---------+
        R  S  A  R  R  F  V  E  V  A  K  A  N  P  K  L  T  K  G  E  A  T  N
                                                                 VP1 ——►
      CATAGGCGTATTGTTGGGCTTATTCATCTTCATACTAATAGGTATAGTACTATTGCCCGTAATCGTTAGC
 281  ---------+---------+---------+---------+---------+---------+---------+
       I  G  V  L  L  G  L  F  I  F  I  L  I  G  I  V  L  L  P  V  I  V  S

      CAAGTCAACAACCTCACAAGCGGTACTTCACCCCAAGTAACCGGTACTAACGCCACACTCCTGAACTTAG
 351  ---------+---------+---------+---------+---------+---------+---------+
      Q  V  N  N  L  T  S  G  T  S  P  Q  V  T  G  T  N  A  T  L  L  N  L  V
                                                                          ○○
      TGCCGTTATTCTATATCCTAGTCCTCATAATAGTCCCCGCAGTCGTGGCGTATAAGATATACAAAGACTG
 421  ---------+---------+---------+---------+---------+---------+---------+
        P  L  F  Y  I  L  V  L  I  I  V  P  A  V  V  A  Y  K  I  Y  K  D  *
      ○○○○○●●●●●
      AGGTGTGAGGGATGGAAATCAGTTTAAAGCCAATCATTTTTTGGTCGTTTTTATCATCGTAGGGATAGC
 491  ---------+---------+---------+---------+---------+---------+---------+
                 M  E  I  S  L  K  P  I  I  F  L  V  V  F  I  I  V  G  I  A
                 VP3 ——►
      ACTATTCGGCCCTATAAACAGTGTTGTAAATAACGTTACCACATCGGGAACCTACACTACTATAGTTTCC
 561  ---------+---------+---------+---------+---------+---------+---------+
       L  F  G  P  I  N  S  V  V  N  N  V  T  T  S  G  T  Y  T  T  I  V  S

      GGTACTGTTACTACGTCTTCATTTGTGTCAAATCCGCAATACGTAGGTAGCAATAACGCTACTATCGTAG
 631  ---------+---------+---------+---------+---------+---------+---------+
      G  T  V  T  T  S  S  F  V  S  N  P  Q  Y  V  G  S  N  N  A  T  I  V  A

      CCTTAGTGCCGTTATTCTATATCCTAGTCCTCATAATAGTCCCCGCAGTCGTGGCGTATAAGTTGTATAA
 701  ---------+---------+---------+---------+---------+---------+---------+
        L  V  P  L  F  Y  I  L  V  L  I  I  V  P  A  V  V  A  Y  K  L  Y  K
         ●  ●●●●●
       ○○○○ ○
      GGAGGAGTGATATGAAGTGGGTGCAAAAGGCGATAAAGAGACCCGGGAGGGTACATCGCTACCTTATGAG
 771  ---------+---------+---------+---------+---------+---------+---------+
       E  E  *   M  K  W  V  Q  K  A  I  K  R  P  G  R  V  H  R  Y  L  M  R
                 VP2 ——►
      GCTCTACGGCAAACGGGCGTTTACAAAAGACGGTGACATAAAGGCAAGTTATCTCGATAAGGCGATAAAG
 841  ---------+---------+---------+---------+---------+---------+---------+
       L  Y  G  K  R  A  F  T  K  D  G  D  I  K  A  S  Y  L  D  K  A  I  K

      CACGTTAAAAAAGCTAAGATCCCGAAAGAGAAGAAACGTAGTTTACTGTCAGCCCTACTGTTAGCGAAAA
 911  ---------+---------+---------+---------+---------+---------+---------+
      H  V  K  K  A  K  I  P  K  E  K  K  R  S  L  L  S  A  L  L  L  A  K  R

      GGCTTAAGCGGATGCACCGCAAGTAGGCCCTTTATAAAGTCATATTCTTTTTCTTTCCCTGATGAGTGCG
 981  ---------+---------+---------+---------+---------+---------+---------+
        L  K  R  M  H  R  K  *
```

Fig. 2. DNA sequence of the region encoding VP1, VP2 and VP3 and the amino acid sequences of the three SSV1 proteins. The first six nucleotides of the DNA sequence correspond to the *Eco*RI site indicated in Fig. 4A. Initiation codons for VP2 and VP3 and the three possible start codons for a VP1 precursor protein are *underlined*. Putative ribosome binding sites are indicated by *small circles*. *Open circles* correspond to the first and *closed circles* correspond to the second possible alignment given in Fig. 5. The 61 bp directly repeated DNA sequence is indicated by *thick horizontal bars*

that VP2 is tightly bound to SSV1 DNA (W.-D. Reiter and R. Reinhardt, unpublished results).

Properties of VP1 and VP3

Both VP1 and VP3 are hydrophobic proteins consisting of 73 and 92 amino acid residues, respectively. As for VP2 the sizes of VP1 and VP3 were deduced from the DNA sequence, assuming that there are no introns within the genes and that the proteins are not processed at their carboxy-termini. An extreme hydrophobicity is especially apparent for VP1 that readily dissolves in 3:2 chloroform/ethanol.

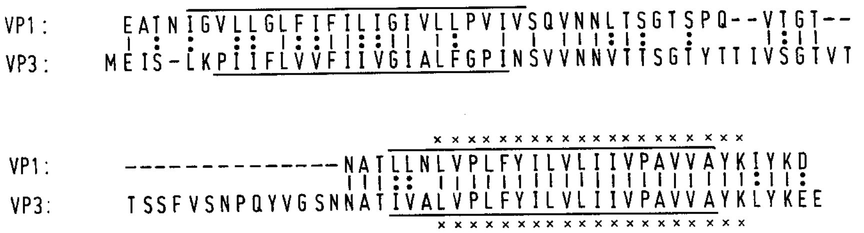

VP1:
EATNIGVLLGLFIFILIGIVLLPVIVSQVNNLTSGTSPQ--VTGT--
VP3:
MEIS-LKPIIFLVVFIIVGIALFGPINSVVNNVTTSGTYTTIVSGTVT
VP1:
--------------NATLLNLVPLFYILVLIIVPAVVAYKIYKD
VP3:
TSSFVSNPQYVGSNNATIVALVPLFYILVLIIVPAVVAYKLYKEE

Fig. 3. Homology between VP1 and VP3. *Vertical lines* between the aligned sequences indicate identical amino acid residues. Conservative exchanges are marked by *two dots*. The following amino acids were considered to be similar: (I, L, V, F), (S, T), (D, E). The extremely hydrophobic regions in VP1 and VP3 are indicated by *horizontal bars* and the amino acid residues that are encoded by the 61 bp direct DNA sequence repeat are marked by *crosses*

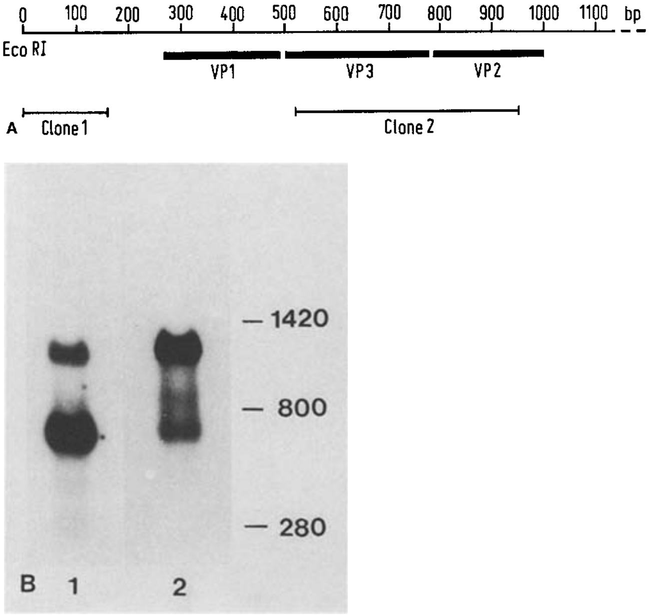


Fig. 4 A, B. Northern analysis of transcripts derived from the region encoding the SSV1 structural proteins. **A** Map position of M13 clones used as hybridization probes. The *Eco*RI site indicated on the left separates the second largest and the third largest *Eco*RI fragment of SSV1 DNA (see Yeats et al. (1982) for the restriction map of SSV1). **B** Autoradiograph of nitrocellulose filters containing total RNA from SSV1-induced cells. The filters were hybridized with clone 1 (lane 1) and clone 2 (lane 2). The numbers to the right of lane 2 give the sizes (in bases) of DNA markers run in parallel

The two proteins are highly homologous (Fig. 3). Charged amino acid residues are only found at the amino-termini and the carboxy-termini of the two polypeptides (Fig. 3). The central regions of both VP1 and VP3 consist of mostly hydrophilic but uncharged amino acids. This region is 18 amino acids longer in VP3 than in VP1, accounting for most of the size difference between the two proteins (73 residues in VP1 vs 92 residues in VP3). Two extremely hydrophobic regions consisting of about 20 amino acid residues each are located between the charged termini of VP1 and the central part of this polypeptide. A completely analogous situation is found in VP3 (Fig. 3). The alignment of the hydrophobic regions that are close to the amino-termini in the two proteins shows that there is a high degree of functional homology, i.e. amino acids in corresponding positions are similar in the two proteins and they are occasionally identical (Fig. 3). The hydrophobic regions close to the carboxy-termini of VP1 and VP3, however, are completely identical (Fig. 3). This surprising feature is the result of a direct DNA sequence repeat encoding this part of the two proteins (see below).

The organization of the genes encoding VP1, VP2 and VP3

All three genes map on the second largest *Eco*RI fragment of SSV1 DNA (Yeats et al. 1982) and they are closely linked in the order VP1-VP3-VP2 (Figs. 2 and 4A). The coding strand is the same for the three genes (Fig. 2). It can be safely assumed that the methionines that constitute the N-terminal amino acid residue in both VP2 and VP3 correspond to translational starts because there are no upstream initiation codons before a stop codon is encountered. An intergenic region of ten nucleotides separates the genes encoding VP1 and VP3, but there is only one nucleotide between the stop codon of VP3 and the initiation codon of VP2 (Fig. 2). For VP1 the translational start is not known since the N-terminal amino acid of this protein is a glutamic acid residue and proteolytic cleavage of a precursor molecule must therefore be assumed. Three possible initiation codons are located upstream of the triplet encoding the N-terminal glutamic acid before a stop codon is encountered (Fig. 2).

Ribosome binding sites

A comparison of the 3′-terminal sequence of *Sulfolobus* 16 S rRNA with the sequences upstream of the genes encoding VP2 and VP3 showed that there are two possible candidates for ribosome binding sites for each gene (Figs. 2 and 5). The anti-Shine-Dalgarno sequence of *Sulfolobus* sp. strain B12 has been determined (P. Palm and B. Grampp, unpublished data) and is identical to that of the related species *Sulfolobus solfataricus* (Olsen et al. 1985). For both genes one of the putative ribosome binding sites is located several bases upstream of the translational start codon and the other is directly adjacent to it (VP2) or one nucleotide upstream of it (VP3). As there is only a one nucleotide intergenic region between VP3 and VP2, one of the putative ribosome binding sites for VP2 is located within the coding region for VP3 and the other is centred around the VP3 stop codon (Figs. 2 and 5).

The analysis of putative Shine-Dalgarno sequences upstream of the three possible initiation codons for a VP1 precursor protein suggested that a GUG codon 13 amino acid residues upstream of the actual N-terminus is the most likely candidate for translation initiation. As for VP2 and VP3 there are two candidates for ribosome binding sites upstream of this GUG triplet (Figs. 2 and 5). These se-

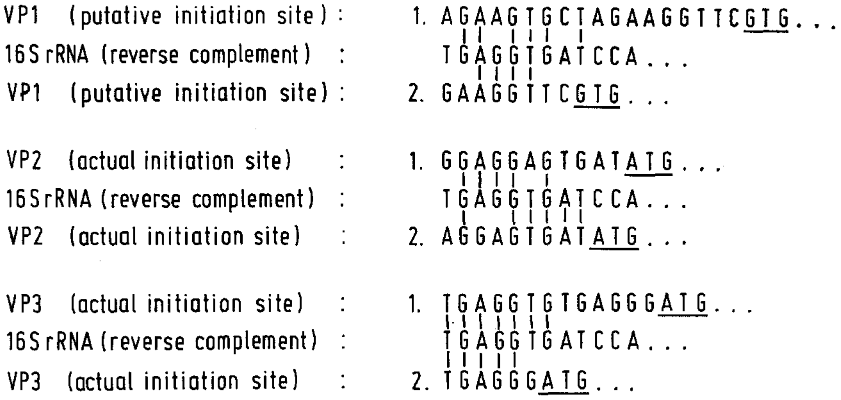

VP1 (putative initiation site) : 1. AGAAGTGCTAGAAGGTTCGTG...
16S rRNA (reverse complement) : TGAGGTGATCCA...
VP1 (putative initiation site) : 2. GAAGGTTCGTG...
VP2 (actual initiation site) : 1. GGAGGAGTGATATG...
16S rRNA (reverse complement) : TGAGGTGATCCA...
VP2 (actual initiation site) : 2. AGGAGTGATATG...
VP3 (actual initiation site) : 1. TGAGGTGTGAGGGATG...
16S rRNA (reverse complement) : TGAGGTGATCCA...
VP3 (actual initiation site) : 2. TGAGGGATG...

Fig. 5. Putative ribosome binding sites upstream of the initiation codons of VP2 and VP3 and upstream of a possible initiation codon for a VP1 precursor protein. For all three proteins two possible alignments (1. and 2.) with the reverse complement of the 3′-end of the *Sulfolobus* 16 S rRNA are shown. *Vertical lines* indicate identical nucleotides. The initiation codons are *underlined*

Table 1. Codon usage for the three SSV1 structural proteins

Amino acid	Codon	VP1	VP2	VP3	Total
Gly	GGG	–	1	1	2
Gly	GGA	–	–	1	1
Gly	GGU	3	1	2	6
Gly	GGC	2	1	1	4
Glu	GAG	–	1	2	3
Glu	GAA	1	–	1	2
Asp	GAU	–	1	–	1
Asp	GAC	1	2	–	3
Val	GUG	2	1	3	6
Val	GUA	4	1	4	9
Val	GUU	1	1	5	7
Val	GUC	4	–	4	8
Ala	GCG	1	4	1	6
Ala	GCA	2	1	2	5
Ala	GCU	–	1	1	2
Ala	GCC	1	1	1	3
Arg	AGG	–	3	–	3
Arg	AGA	–	1	–	1
Arg	CGG	–	2	–	2
Arg	CGA	–	–	–	–
Arg	CGU	–	1	–	1
Arg	CGC	–	2	–	2
Ser	UCG	–	–	1	1
Ser	UCA	1	1	2	4
Ser	UCU	–	–	1	1
Ser	UCC	–	–	1	1
Ser	AGU	–	2	2	4
Ser	AGC	2	–	1	3
Ile	AUA	7	3	5	15
Ile	AUU	–	–	1	1
Ile	AUC	3	1	6	10
Met	AUG	–	3	1	4
Lys	AAG	1	10	3	14
Lys	AAA	1	7	–	8
Thr	ACG	–	–	1	1
Thr	ACA	2	1	1	4
Thr	ACU	2	–	5	7
Thr	ACC	2	–	2	4
Asn	AAU	–	–	3	3
Asn	AAC	5	–	3	8
Cys	UGU	–	–	–	–
Cys	UGC	–	–	–	–
Tyr	UAU	2	1	3	6
Tyr	UAC	1	2	2	5
Leu	UUG	3	–	2	5
Leu	UUA	3	2	3	8
Leu	CUG	1	2	–	3
Leu	CUA	3	1	2	6
Leu	CUU	–	2	–	2
Leu	CUC	3	2	1	6
Phe	UUU	–	1	3	4
Phe	UUC	3	–	2	5
Gln	CAG	–	–	–	–
Gln	CAA	2	1	1	4
His	CAU	–	1	–	1
His	CAC	–	2	–	2
Pro	CCG	1	1	2	4
Pro	CCA	–	–	1	1
Pro	CCU	–	–	1	1
Pro	CCC	3	1	1	5
amber	UAG	–	1	–	1
ochre	UAA	–	–	–	–
opal	UGA	1	–	1	2
Trp	UGG	–	1	–	1

quences are two and nine nucleotides away from the putative initiation codon.

A 61 bp direct repeat in the genes encoding VP1 and VP3

A sequence of 61 nucleotides roughly corresponding to the coding region for the stretch of hydrophobic amino acids close to the carboxy-terminus of VP1 is exactly repeated within the gene for VP3 where it encodes the carboxy-terminal hydrophobic region of this protein (Fig. 2). Thus stretches of 20 amino acid residues ranging from positions 50 to 69 in VP1 and from positions 68 to 87 in VP3 are completely identical in the two proteins (Fig. 3). Short stretches of identical amino acids in corresponding positions of VP1 and VP3 are a general feature of these two proteins but these short stretches of perfect amino acid homology are not encoded by identical nucleotide sequences (Figs. 2 and 3).

The codon usage in VP1, VP2 and VP3

There is no apparent restriction in the codon usage for the three SSV1 proteins. Though the G + C content of SSV1 DNA is only 40%, not even a preference for A or U as the third base in codons was found. The codon usage in VP1, VP2 and VP3 is summarized in Table 1.

Transcription of the region encoding VP1, VP2 and VP3

As shown by Northern analysis, two mRNAs of 0.5 kb and 1.0 kb map within the region encoding the three SSV1 structural proteins (Fig. 4). A probe comprising sequences upstream of the VP1 gene (clone 1 in Fig. 4A) strongly hybridizes to both transcripts whereas a probe containing the coding region for VP3 and VP2 (clone 2 in Fig. 4A) strongly hybridizes to the longer RNA but only weakly to the smaller one (Fig. 4B). Clone 2 contains the "VP3 copy" of the 61 bp direct repeat and the hybridization of this clone to the 0.5 kb RNA is probably primarily due to the "VP1 copy" of this repeat being present in this small RNA. S1 nuclease protection analysis indicates that both RNAs have a common 5′-end mapping around nucleotide 48 (see Figs. 2 and 4A for numbering) and that the smaller RNA terminates within the T-rich region about 30 nucleotides downstream of the VP3 initiation codon (unpublished results from this laboratory). According to these mapping data the longer RNA terminates about 40–50 nucleotides downstream from the VP2 stop codon. From the relative intensities of the hybridization signals using clone 1 as a probe (Fig. 4) it can be concluded that the 0.5 kb RNA is several times more abundant than the 1.0 kb RNA.

Discussion

The organization of the genes encoding VP1, VP2 and VP3

The genes for the three SSV1 structural proteins are closely linked and coordinately expressed. In addition to a polycistronic messenger coding for all three proteins there is also an abundant monocistronic mRNA encoding VP1. Since VP1 is the major coat protein, this pattern of transcription appears to regulate the level of synthesis of the individual structural proteins. The organization of genes in polycistronic transcriptional units is a typical feature of eubacteria but not of eukaryotes. A close linkage of protein-encoding genes has also been found in methanogenic archaebacteria (Konheiser et al. 1984; Hamilton and Reeve 1985; Reeve et al. 1986) though a common transcript has only been demonstrated in one case (Konheiser et al. 1984).

Ribosome binding sites

The conservation of the 3′-end of 16 S rRNA between eubacteria and archaebacteria (Steitz 1978) and the faithful translation of genes from methanogenic archaebacteria in eubacteria (Wood et al. 1983; Bollschweiler and Klein 1982) strongly indicates that translation initiation is controlled by the same mechanism in both groups of organisms, i.e. by a complementarity between the 3′-end of 16 S rRNA and a sequence on the mRNA that is usually located several bases upstream of the initiation codon (Shine and Dalgarno 1974).

Two putative ribosome binding sites each have been found upstream of the initiation codons of VP2 and VP3. In both cases one of these sequences is located several bases upstream of the initiation codon and the other one is either directly adjacent to the AUG codon (VP2) or very close to it (VP3). It cannot be determined at this time whether this feature is of any importance for the initiation of translation in *Sulfolobus*. Since the genes encoding the SSV1 structural proteins are the first protein-encoding genes characterized for an archaebacterium belonging to the branch of extreme thermophiles, no data for comparison are available from closely related organisms. For methanogenic archaebacteria, however, the presence of putative ribosome binding sites has been investigated (Bollschweiler et al. 1985; Cue et al. 1985; Hamilton and Reeve 1985). In each case studied one single putative Shine-Dalgarno sequence was found upstream of the corresponding start codon.

A 61 bp direct DNA sequence repeat within the coding region

Identical sequences of 20 amino acid residues in the carboxy-terminal hydrophobic regions of VP1 and VP3 are encoded by identical DNA sequences of 61 nucleotides. It is conceivable that some evolutionary pressure favours the complete conservation of a certain region in two homologous viral proteins, but it is difficult to understand why the corresponding DNA sequence should also be completely conserved. Third base exchanges should not be of any selective disadvantage to the cell unless there is a restriction in the codon usage. An analysis of the codon usage in VP1, VP2 and VP3, however, showed that such a restriction does not exist in *Sulfolobus* B12 (Table 1).

The presence of a sequence repeat within a coding region inflicts severe limitations on the sequence and thus on the properties of the encoded protein(s). Apparently these limitations can be tolerated in the case of VP1 and VP3 as these two proteins are generally very homologous. Though it is very likely that this direct DNA sequence repeat has a function in addition to its coding properties, all discussion about such an additional function must remain speculative at the moment. A possible role in regulation of translation, in termination of transcription or in the processing of mRNAs cannot be ruled out though inverted rather than direct repeats are usually involved in these cases. A function of this repeat in the regulation of replication must be considered though this would probably interfere with gene expression. Another possible role of the direct DNA repeat is its involvement in the initiation of transcription. It is known that enhancer elements of some eukaryotic viruses contain long direct DNA sequence repeats that are usually in a tandem arrangement (Dorsch-Häsler et al. 1985). Though enhancers containing long direct repeats are usually found upstream of transcriptional starts, they can also be effective in downstream positions (for review see Gluzman 1985). Almost nothing is known as yet about the regulatory elements that govern the expression of protein-encoding genes in archaebacteria. It has been found, however, that there is a striking similarity between the eukaryotic nuclear RNA polymerases and the RNA polymerases of archaebacteria, especially those belonging to the branch of extreme thermophiles that includes *Sulfolobus* (Huet et al. 1983). Therefore it appears to be an intriguing possibility that some signal structures involved in the initiation of transcription are also shared by these two groups of organisms.

Acknowledgements. We thank C. Eckerskorn and M. Hofmann for their help with the electroblotting and H. Gross, H. Kaudewitz and G. Krauss for their assistance in protein sequencing. Thanks are also due to F. Gropp, F. Pfeifer and J. Trent for critical reading of the manuscript.

References

Aebersold RH, Teplow DB, Hood LE, Kent SBH (1986) Electroblotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. J Biol Chem 261:4229–4238

Baresi L, Bertani G (1984) Isolation of a bacteriophage for a methanogenic bacterium. Abstr Annu Meet Am Soc Microbiol 84:Abstract I 74

Bollschweiler C, Klein A (1982) Polypeptide synthesis in *Escherichia coli* directed by cloned *Methanobrevibacter arboriphilus* DNA. Zentralbl Bakteriol Mikrobiol Hyg [C] 3:101–109

Bollschweiler C, Kühn R, Klein A (1985) Non-repetitive AT-rich sequences are found in intergenic regions of *Methanococcus voltae* DNA. EMBO J 4:805–809

Cue D, Beckler GS, Reeve JN, Konisky J (1985) Structure and sequence divergence of two archaebacterial genes. Proc Natl Acad Sci USA 82:4207–4211

Daniels LL, Wais AC (1984) Restriction and modification of halophage S45 in *Halobacterium.* Curr Microbiol 10:133–136

Deininger PL (1983) Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. Anal Biochem 129:216–223

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dorsch-Häsler K, Keil GM, Weber F, Jasin M, Schaffner W, Koszinowski UH (1985) A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. Proc Natl Acad Sci USA 82:8325–8329

Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA, Wolfe RS, Balch WE, Tanner RS, Magrum LJ, Zablen LB, Blakemore R, Gupta R, Bonen L, Lewis BJ, Stahl DA, Luehrsen KR, Chen KN, Woese CR (1980) The phylogeny of prokaryotes. Science 209:457–463

Gluzman Y (ed) (1985) Eukaryotic transcription: The role of cis- and trans-acting elements in initiation. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Hamilton PT, Reeve JN (1985) Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii.* Mol Gen Genet 200:47–59

Hu N, Messing J (1982) The making of strand-specific M13 probes. Gene 17:271–277

Huet J, Schnabel R, Sentenac A, Zillig W (1983) Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. EMBO J 2:1291–1294

Janekovic D, Wunderl S, Holz I, Zillig W, Gierl A, Neumann H (1983) TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium *Thermoproteus tenax.* Mol Gen Genet 192:39–45

Konheiser U, Pasti G, Bollschweiler C, Klein A (1984) Physical mapping of genes coding for two subunits of methyl CoM reductase component C of *Methanococcus voltae.* Mol Gen Genet 198:146–152

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lottspeich F, Kellermann J, Henschen A, Rauth G, Müller-Esterl W (1984) Human low-molecular-mass kininogen. Amino-acid sequence of the light chain: homology with other protein sequences. Eur J Biochem 142:227–232

Lottspeich F (1985) Microscale isocratic separation of phenylthiohydantoin amino acid derivatives. J Chromatogr 326:321–327

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Martin A, Yeats S, Janekovic D, Reiter WD, Aicher W, Zillig W (1984) SAV1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J 3:2165–2168

Messing J, Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19:269–276

Mirault ME, Scherrer K, Hansen L (1971) Isolation of preribosomes from HeLa cells and their characterization by electrophoresis on uniform and exponential-gradient-polyacrylamide gels. Eur J Biochem 23:372–386

Nadal M, Mirambeau G, Forterre P, Reiter WD, Duguet M (1986) Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature 321:256–258

Olsen GJ, Pace N, Nuell M, Kaine BP, Gupta R, Woese CR (1985) Sequence of the 16S rRNA gene from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* and its evolutionary implications. J Mol Evol 22:301–307

Pauling C (1982) Bacteriophages of *Halobacterium halobium:* isolation from fermented fish sauce and primary characterization. Can J Microbiol 28:916–921

Reeve JN, Hamilton PT, Beckler GS, Morris CJ, Clarke CH (1986) Structure of methanogen genes. Syst Appl Microbiol 7:5–12

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143:161–178

Sather S, Agabian N (1985) A 5′ spliced leader is added in *trans* to both α- and β-tubulin transcripts in *Trypanosoma brucei.* Proc Natl Acad Sci USA 82:5695–5699

Schnabel H, Zillig W, Pfäffle M, Schnabel R, Michel H, Delius H (1982a) *Halobacterium halobium* phage *Φ*H. EMBO J 1:87–92

Schnabel H, Schramm E, Schnabel R, Zillig W (1982b) Structural variability in the genome of phage *Φ*H of *Halobacterium halobium.* Mol Gen Genet 188:370–377

Schnabel H (1984a) An immune strain of *Halobacterium halobium* carries the invertible L segment of phage *Φ*H as a plasmid. Proc Natl Acad Sci USA 81:1017–1020

Schnabel H (1984b) Integration of plasmid p*Φ*HL into phage genomes during infection of *Halobacterium halobium* R_1-L with phage *Φ*HL1. Mol Gen Genet 197:19–23

Schnabel H, Zillig W (1984) Circular structure of the genome of phage *Φ*H in a lysogenic *Halobacterium halobium.* Mol Gen Genet 193:422–426

Schnabel H, Palm P, Dick K, Grampp B (1984) Sequence analysis of the insertion element ISH1.8 and of associated structural changes in the genome of phage *Φ*H of the archaebacterium *Halobacterium halobium.* EMBO J 3:1717–1722

Schnabel R, Thomm M, Gerardy-Schahn R, Zillig W, Stetter KO, Huet J (1983) Structural homology between different archaebacterial DNA-dependent RNA polymerases analyzed by immunological comparison of their components. EMBO J 2:751–755

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342–1346

Staden R (1980) A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res 8:3673–3694

Steitz JA (1978) Methanogenic bacteria. Nature 273:10

Tandy NE, Dilley RA, Regnier FE (1983) High-performance liquid chromatographic purification of the hydrophobic ω subunit of the chloroplast energy coupling complex. J Chromatogr 266:599–607

Thomas PS (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Methods Enzymol 100:255–266

Torsvik T, Dundas ID (1974) Bacteriophage of *Halobacterium salinarium*. Nature 248:680–681

Torsvik T, Dundas ID (1980) Persisting phage infection in *Halobacterium salinarium* str. 1. J Gen Virol 47:29–36

Vogelsang-Wenke H, Oesterhelt D (1986) Halophage *Φ*N. In: Kandler O, Zillig W (eds) Archaebacteria 85. Gustav Fischer Verlag, Stuttgart New York, pp 403–405

Wais AC, Kon M, MacDonald RE (1975) Salt-dependent bacteriophage infecting *Halobacterium cutirubrum* and *H. halobium*. Nature 256:314–315

Woese CR, Olsen GJ (1986) Archaebacterial phylogeny: Perspectives on the urkingdoms. Syst Appl Microbiol 7:161–177

Wood AG, Redborg AH, Cue DR, Whitman WB, Konisky J (1983) Complementation of *arg*G and *his*A mutations of *Escherichia coli* by DNA cloned from the archaebacterium *Methanococcus voltae*. J Bacteriol 156:19–29

Yeats S, McWilliam P, Zillig W (1982) A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. EMBO J 1:1035–1038

Zillig W, Schnabel R, Stetter KO (1985a) Archaebacteria and the origin of the eukaryotic cytoplasm. Curr Top Microbiol Immunol 114:1–18

Zillig W, Yeats S, Holz I, Böck A, Gropp F, Rettenberger M, Lutz S (1985b) Plasmid-related anaerobic autotrophy of the novel archaebacterium *Sulfolobus ambivalens*. Nature 313:789–791

Communicated by A. Böck

Received August 22, 1986 / October 2, 1986