Arch Virol (1995) 140: 1033–1047

Archives of Virology

68

Analysis of the open reading frames of the main capsid proteins of actinophage VWB

J. Anné[1], P. Fiten[2], L. Van Mellaert[1], B. Joris[3], G. Opdenakker[2], and H. Eyssen[1]

[1] Laboratory of Bacteriology and [2] Laboratory of Molecular Immunology, Rega Institute, K. U. Leuven, Leuven, [3] Center for Protein Engineering, University of Liège, Sart-Tilman-Liège, Belgium

Accepted January 24, 1995

Summary. The nucleotide sequence of a 6 kb fragment encoding the main late proteins (p14, p38 and p24) of actinophage VWB was obtained. Sequence comparison of the encoded proteins with those filed in databases indicated that the phage VWB main late proteins were all novel. A search for special motifs revealed that p14 (13.3 kDa) has a P-loop sequence commonly found in ATP- and GTP-binding proteins. This observation might indicate that p14 is important for ATP-driven DNA translocation during encapsidation of VWB phage DNA into the phage head. Furthermore, the polypeptide ORF2 (26.9 kDa) has an unusual primary structure consisting of 3 stretches of acidic amino acid residues and a glycine/arginine rich C-terminal end. From comparison with other proteins including the bacteriophage T4 prohead core component and from the data of special motif analysis the ORF2 gene product is probably involved in prohead core formation.

Introduction

Actinophages are viruses which infect members of the order *Actinomycetales* and related organisms such as proprionobacteria and corynebacteria. All actinophages described so far share characteristics of having a double stranded DNA genome that is packaged in a tailed icosahedral phage particle, most often of the morphological type B1 [1]. Actinophages are important for fundamental and applied research, e.g. phage typing [2], contamination in the fermentation industry [3], studies to characterize in vivo restriction [4]. Nevertheless, characterization of actinophages is in most instances limited to host range determination and the classification of some physical and morphological properties. Reports on the molecular biological analysis are rare. Whereas for mycobacteriophage L5 the complete genome has been sequenced [5], only the well-studied φC31 phage has been so far characterized in more detail among

Streptomyces phages [6]. A more extensive knowledge of different actinophages including their sequence analysis is of much interest for fundamental research including (quasi)-species phage description or studies on phage evolution, as well as for applied genetics, such as for the development of phage-based integrative vectors.

Phage evolution is suggested to be a combination of factors including alterations caused by point mutations [7] and the exchange of gene modules between not necessarily closely related phages [8], occurring through genetic exchange within common hosts [9]. Sequence comparison of capsid and other phage proteins and the structural organization of genes in actinophages could provide insight into their phylogenetic relationship.

In a previous article [10] we reported already on the molecular characterization of actinophage VWB. The protein pattern of this temperate phage (47.3 kb) which is infectious to *Streptomyces venezuelae* ETH14630 and *Streptomyces exfoliatus* ATCC12672, showed at least 17 different protein bands. Three major proteins were observed, estimated 16.5, 27.2 and 43 kDa in size. Using oligonucleotides synthesized on the basis of the N-terminal amino acid sequences, the genes coding for these major late proteins were shown to be clustered.

In this report the nucleotide sequence of the main capsid proteins of a *Streptomyces* phage are described for the first time. For phage VWB, they are located on a 6.0 kb *Sph*I fragment. Following DNA sequencing 10 complete open reading frames were identified. The amino acid sequences of the deduced polypeptides were compared with different protein databases, codon usage and protein motifs were also investigated.

Materials and methods

Media, bacteria, phage and plasmids

Phage VWB was propagated on *Streptomyces venezuelae* ETH14630 (ATCC40755) and its DNA was isolated as previously described [11]. For sequence analysis, restriction fragments were subcloned into pUC19 and multiplied in *Escherichia coli* JM109 [12] grown in Luria Broth (LB) medium supplemented with ampicillin at 50 µg/ml.

Subcloning and DNA purification

The approximate 6 kb *Sph*I/C fragment [13] to be sequenced was digested with the restriction endonucleases *Apa*I, *Ava*I, *Bss*HII, *Hin*dII, *Nco*I, *Pst*I, *Pvu*II, *Sac*I, and *Xma*I (obtained from BioLabs, Boehringer Mannheim or Pharmacia and used according to the manufacturers' instructions). In this manner a range of fragments of size suitable for sequencing was obtained. The restriction fragments, eventually blunted, were subcloned in pUC19 and ligated with T4 DNA ligase. Blunting occurred by *Pol*Ik (Klenow-polymerase) or T4 DNA polymerase as described [14]. All DNAs prepared for sequencing were purified by CsCl-ethidium bromide equilibrium centrifugation.

Phage DNA cloning and sequencing

The DNA sequence was determined by the dideoxy chain termination method [15] using the AutoRead sequencing kit (Pharmacia), and 7-deaza-2'-guanosine triphosphate to minimize GC compressions. Primers were the 5'-fluorescein labeled M13 universal and reverse primers and oligonucleotides complementary to internal sequences of the *Sph*I/C fragment. The latter were synthesized on an ABI DNA synthesizer (Applied Biosystems Inc., Foster City, USA). These oligonucleotides were 5'-labeled with fluorescein amidite Fluoro Prime (Pharmacia). Fragments were separated on a 6% polyacrylamide-gel [acrylamide:bisacrylamide (19:1) (BioRad)] with 7.0 M urea in TBE buffer (60 mM Tris, 50 mM boric acid, 1.5 mM EDTA) at 1500 V (35 W, 35 mA).

Computer analyses of the DNA and protein sequences

The sequencing data were processed using the A.L.F. Manager and PC Gene (Genofit, IntelliGenetics, Inc.) computer programs and by FRAME analysis [16]. Protein coding sequences were identified by virtue of the similarity of their codon usage to a codon frequency table or on the basis of their composition (usually G or C) in the third position of each codon [CODON PREFERENCE software of GCG package and ANALYSEQ]. [17]. The codon frequency table was computed from *Streptomyces coelicolor hrdB* gene for the principal sigma factor [18], using a scan window of 45 nucleotides [19].

To look for similarities with other sequences and for the presence of any special motifs, searches through the nucleic acid (EMBL version 36) and protein (SWISSPROT version 27 and PIR version 38) sequence databases were performed by using the procedure of Pearson and Lipman (FASTA and TFASTA softwares, GCG package) [19]. The presence of protein sequence motifs was carried out by comparing the deduced amino acid sequences to the patterns defined in the PROSITE dictionary version 10 (MOTIFS software of the GCG package).

Preparation and analysis of proteins by SDS-PAGE

Phage proteins along with a calibration kit for molecular mass determination (GIBCO-BRL) were subjected to SDS-PAGE under reducing conditions in 12.5% (w/v) gels (0.7 mm thick) [20] using a Midget system (Pharmacia). After electrophoresis, proteins were visualized in the gel by staining with Coomassie Blue R-250 [21].

Northern blot analysis

S. venezuelae cultures infected with 10^{10} p.f.u. of phage VWB (m.o.i. $\sim$ 1) were grown for 5 h. Mycelium of 100 ml cultures was harvested by centrifugation (4000 $\times$ g, 10 min) and total RNA was extracted as described [22]. Fifty µg samples of RNA in 20 µl loading buffer were applied to a 1.2% (w/v) agarose-gel containing 0.66 M formaldehyde in MOPS-buffer [20 mM 3-(N-morpholino) propane sulfonic acid, 1 mM EDTA, 5 mM sodium acetate, pH 7.0] and electrophoretically separated at 10 V/cm. RNAs were blotted on a Gene Screen plus membrane (Dupont – NEN Research Products, Boston, MA) using a Vacugene blotting system (Pharmacia). For the reversal of the formaldehyde reaction the membrane was baked at 80 °C for 2 h. VWB transcripts were detected by hybridizing the fixed mRNAs with phage VWB fragment *Sph*I/C of 6.0 kb [13], radioactively labeled with ^{32}P-α-dATP using a random primed DNA labeling kit (Boehringer Mannheim). Subclones were used for fine mapping.

Nucleotide sequence accession number

The nucleotide sequence data of the *Streptomyces venezuelae* phage VWB ORFs was deposited in the EMBL data library and has the accession number X72092.

Results

Sequencing strategies and open reading frame analysis

The sequencing strategy of the 6,051 bp fragment is outlined in Fig. 1. The high GC content of the DNA (in average 71.4% GC) made sequencing a difficult task, particularly in regions rich in GC-there are 33 regions (20–30 bp long) with a GC content of 90–92%. Therefore, sequencing was repetitively carried out in forward and reverse sense and sequence gels were often run at elevated temperature (up to 60 °C) to avoid or diminish band compressions in the gel. For several areas, in which case the forward or reverse M13 primer gave ambiguous results on sequencing, VWB-specific primers were applied.

The result of the nucleotide sequence analysis is shown in Fig. 2. Accuracy of the sequence was checked, e.g. by mapping analysis, by confirmation of restriction endonuclease cleavage sites found by sequence analysis and by ORF analysis. This was carried out by subjecting the nucleotide sequence to computer

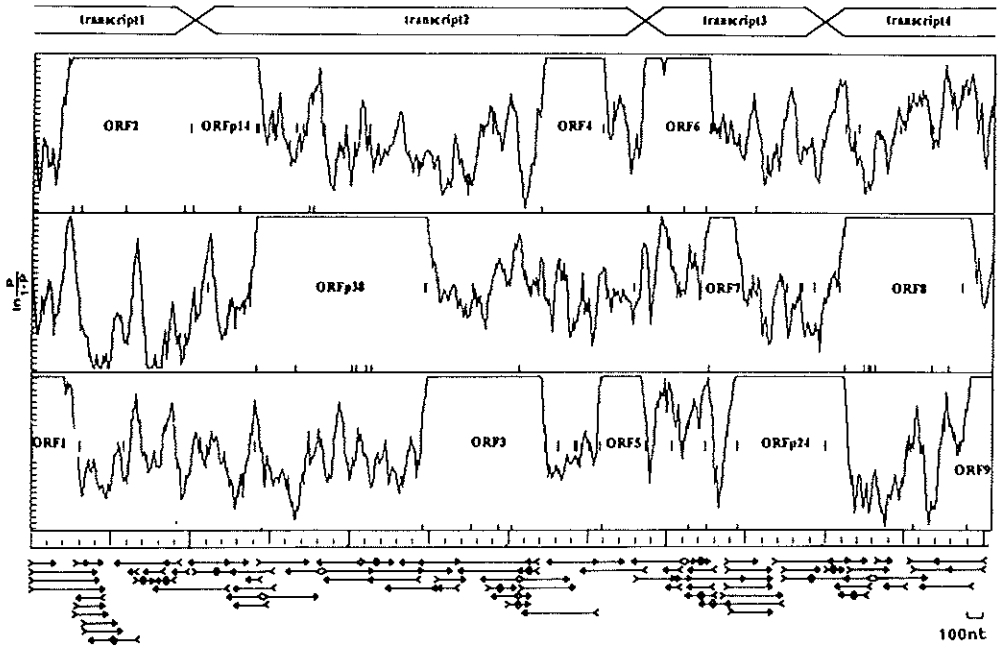

Fig. 1. Strategy for sequencing the 6,051 bp *Sph*I/*Bam*HI fragment of actinophage VWB and codon usage analysis of the sequenced DNA fragment. P expresses the probability of coding. The transcripts identified following Northern analysis – see Fig. 3b – are also indicated. ATG codons are represented by vertical bars at the base line (⊥); stop codons are indicated by vertical bars (|). Sequences obtained with VWB-specific primers are indicated(—○—)

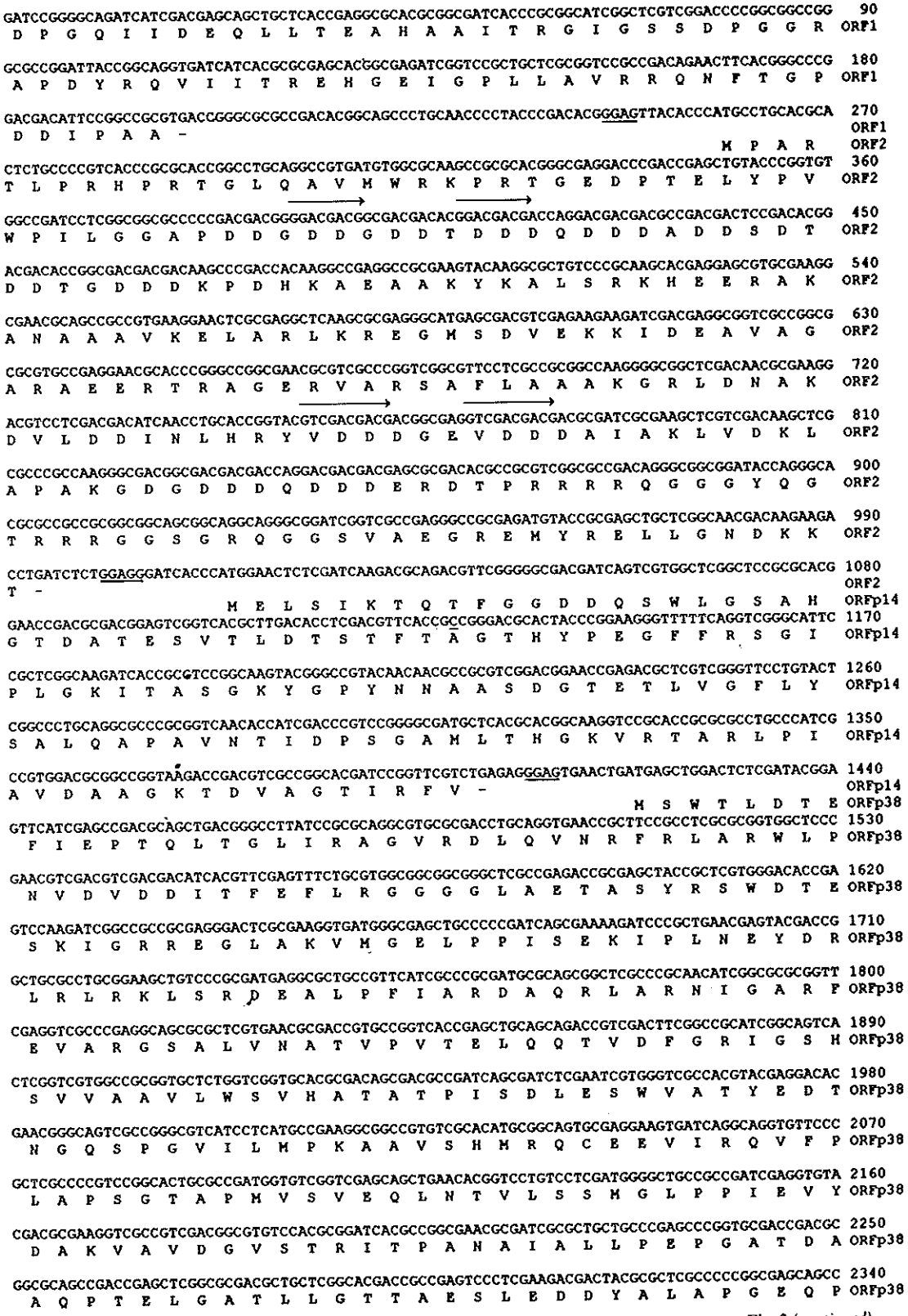

Fig. 2 (*continued*)

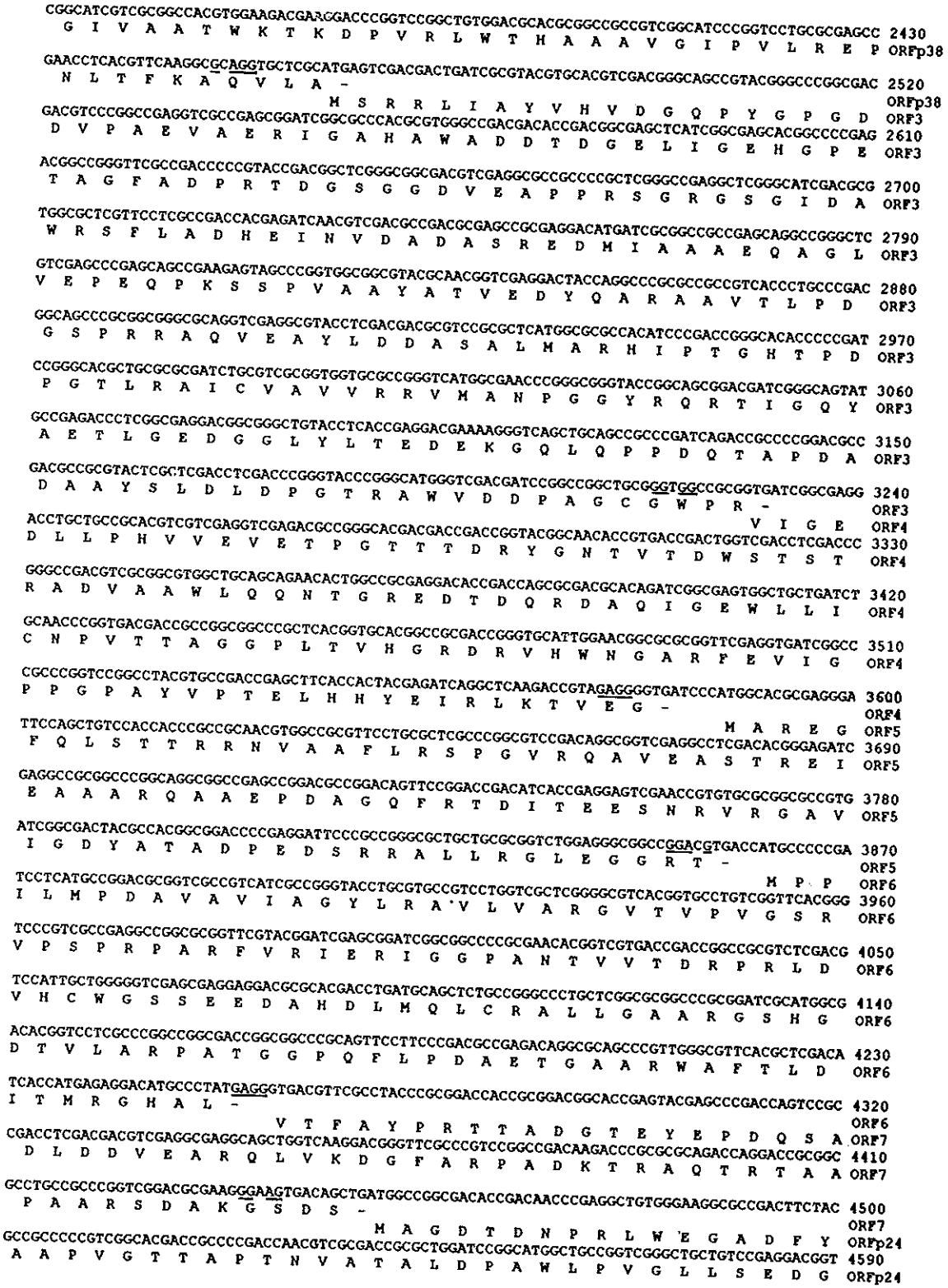

Fig. 2 (*continued*)

```
GCGAGCGAGTCCCGCGACCAGGACTCGACCGACTTCTACGCGTGGGGCGGCGTGCTCGTTCGCACGGCGAAGAGCAAGCACAAGCGGCAG 4680
 A  S  E  S  R  D  Q  D  S  T  D  F  Y  A  W  G  G  V  L  V  R  T  A  K  S  K  H  K  R  Q  ORFp24
                                  ------------>  ------------>

ATCGTCGTCACGTGCCTCGAAGAGAACCTGGTCGTGTTCGGGCTCGTCAACCCGGGCTCGTCCGCGGTGACCGCGACCGGGGTCACCACG 4770
 I  V  V  T  C  L  E  E  N  L  V  V  F  G  L  V  N  P  G  S  S  A  V  T  A  T  G  V  T  T  ORFp24

CGCACGGTCAAGGTTCCGAAGGCGGACCCGCGCGCGTTCGTCCTCGAACTGCGGGACGGCGCGGTGAAGAAGCGCCGTGTGATCCCGAAG 4860
 R  T  V  K  V  P  K  A  D  P  R  A  F  V  L  E  L  R  D  G  A  V  K  K  R  R  V  I  P  K  ORFp24

GGCGAGGTCGAGTCGGTTGGCGAGGTCACCCTGTCGGACAGTGCGCTCACCGCGTACGAACTGACGATCACGATCTACCCGGCGGCCGAC 4950
 G  E  V  E  S  V  G  E  V  T  L  S  D  S  A  L  T  A  Y  E  L  T  I  T  I  Y  P  A  A  D  ORFp24
                  ------------>  <------------

GGCACGCTCTACCTCGACATCACGGACGACCCGCAGGCGGTCGTCACCCCCTGACGGCGCTCCGGTGCCGTCCCTCTGCCGGCGGGCGGC 5040
 G  T  L  Y  L  D  I  T  D  D  P  Q  A  V  V  T  P  -                                        ORFp24

CCGATCTGCGCGGGTCCGGGCCGCCCGTCTCTCACCAGTCCGTTGACCCGCGCACGCTCGAAGGAGACCCGCGCCCATGACCACTCAGCC 5130
                                                                                  M  T  T  Q  P  ORF8

GAAGAAGACCGTTCGCCGGCCGGACGACAAGCCGTTCGACTTCAACCTCGACGCCGTCGAGGCCGAGGTCGACCTGACCCCGTTCGTGTT 5220
  K  K  T  V  R  R  P  D  D  K  P  F  D  F  N  L  D  A  V  E  A  E  V  D  L  T  P  F  V  F  ORF8

CCAGCACAGCAGCCGCCGCTGGACGATGCAGCACCTGCAGGCGCTCAACATCATGCCCCTCATGGTGGCGGCGCAGAGCGGCGACGCCGC 5310
  Q  H  S  S  R  R  W  T  M  Q  H  L  Q  A  L  N  I  M  P  L  M  V  A  A  Q  S  G  D  A  A  ORF8

GGCGATGCTCGGCGCGTTCCGCGAGGCACTCGGGAAGCAGTGGGCCGAGTTCGAGGCGATCGGGCTGCCGCAGTACAAGCTGCAGCCGCT 5400
  A  M  L  G  A  F  R  E  A  L  G  K  Q  W  A  E  F  E  A  I  G  L  P  Q  Y  K  L  Q  P  L  ORF8

GTTCGAGGCGTACAGCAAGCACTGCGGTACGACCGGGGGAATCGGAGGCCTCGCCCACCTCCTGACGGAGCACGGCGAGGCAGTCGAGGC 5490
  F  E  A  Y  S  K  H  C  G  T  T  G  G  I  G  G  L  A  H  L  L  T  E  H  G  E  A  V  E  A  ORF8

CGACCTGCGCGAGACCTACGGCGTGCGTCTTAGCGATCTGTACCGCCGGGACGCCGCCGGCCGGCCGCTGCTCACATGGCGCGAGCTCGG 5580
  D  L  R  E  T  Y  G  V  R  L  S  D  L  Y  R  R  D  A  A  G  R  P  L  L  T  W  R  E  L  G  ORF8

CGTGTACGTGCGGCAGCTGCCACGCACCGCCCGGACCCGGATCGCGCTCGGCGACAACGACGGGCTGTGGGGACTGACCGAGCACCTGCT 5670
  V  Y  V  R  Q  L  P  R  T  A  R  T  R  I  A  L  G  D  N  D  G  L  W  G  L  T  E  H  L  L  ORF8

CGCGATGAACCTCGACGCCGCGCACATCGCCAACTGGCAGCGCGCGAACGAGGGACTCAAGTCACACGAGCAGAGCCGGCCCCCCGAACC 5760
  A  M  N  L  D  A  A  H  I  A  N  W  Q  R  A  N  E  G  L  K  S  H  E  Q  S  R  P  P  E  P  ORF8

GATCACACGGCCCGGCATGGCACGCCGGAAGAAGATCACAAGCGCCGAGCTGCTCGCCCACCGCGAGCGGGCCCGCGCCCATCTCCACGC 5850
  I  T  R  P  G  M  A  R  R  K  K  I  T  S  A  E  L  L  A  H  R  E  R  A  R  A  H  L  H  A  ORF8

GGTCGCGGCGTAGCCCGGTGCCGCCCACACCGGACGGGAGGGCGGCACATGGCTACAGCCGCGGTCGAGGTCGGCGTCGGGTACGTATCG 5940
  V  A  A  -                                                                                  ORF8
                                                  M  A  T  A  A  V  E  V  G  V  G  Y  V  S  ORF9
GTCGTCCCGAGCGCCCGGGGATTCGCCGAGGACCTGCAGCGGCAGATCGGCCGGCCGACGGTGCAGGTCGGCGCGGAGATCGGCGAGCAG 6030
 V  V  P  S  A  R  G  F  A  E  D  L  Q  R  Q  I  G  R  P  T  V  Q  V  G  A  E  I  G  E  Q  ORF9

TCCGGGCAGGCGGCGAGCCGC 6051
 S  G  Q  A  A  S  R  ORF9
```

Fig. 2. DNA sequence and predicted amino acid sequences of the encoded proteins. The gene name is given at the right of each amino acid sequence (single letter code) line and the position of the most right nucleotide is given at the right of each nucleotide sequence line. The sequence is numbered starting at the second base of the recognition sequence of *Bam*HI located near one of the *Sph*I sites defining the 6 kb fragment. The putative Shine-Dalgarno sequence is underlined and the inverted repeats creating a potential hairpin loop (nt 4985) and most prominent direct repeats are indicated

analysis for positional base preference [codon preference (GCG)]. From this scan 10 complete (and 2 partial) putative ORFs could be detected (Fig. 1) including ORFp14, p24 and p38, coding for the most prominent proteins of VWB phage particles (Fig. 3a) [10]. The translation of the nucleotide sequence of ORFp14, p24 and p38 corresponds with the formerly identified N-terminal amino acid sequence, except for amino acid residue 20 of p38, which should read as alanine instead of glutamic acid according to the nucleotide sequence.

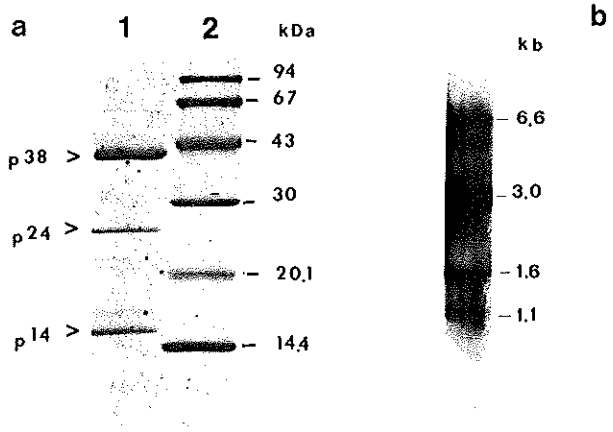

Fig. 3. a SDS-PAGE (15% gel) of the phage VWB proteins visualised by Coomassie stain. *1* VWB phage proteins at a total concentration of 1.8 µg: *2* protein molecular weight standard (Pharmacia) (1 µg/protein). b Northern blot analysis of RNA isolated from *S. venezuelae* infected with phage VWB

The comparison between the nucleotide sequence and the N-terminal amino acid sequence also indicates that the N-terminal methionine is removed for p24 and p38, but not for p14. The estimated size of the predominant phage proteins on SDS-PAGE was 16.5 kDa, 27.2 kDa and 43 kDa (Fig. 3a). Sequence analysis gave M_r values of 13,386 Da for p14; 23,630 Da for p24 and 37,694 Da for p38, indicating that the protein size of the main proteins were accurate by SDS-PAGE analysis. Computer analysis for codon preference indicates that ORFp24 might be longer (until nt 5118 instead of nt 5001) (Fig. 1). However, repeated nucleotide sequence analysis revealed the proposed sequence as correct. In addition, as discussed in the next section, there is a putative terminator at the end of ORFp24.

Although only for ORFp14, p24 and p38 the proteins have been identified on gel, it is most likely that also the other ORFs (Table 1, Fig. 2) code for proteins, because the sequence in front of the putative ATG or GTG start codons exhibits considerable complementarity to the 3'end of the 16S rRNA of *S. lividans* [23].

Transcripts and translational initiation

As mentioned above, computer analysis indicated that the 6 kb *Sph*I/C fragment contains 10 putative ORFs. With the exception of the spacing between ORF1 and ORF2 (56 nt) and between ORFp24 and ORF8 (113 nt), the ORFs are positioned with very little interspace (Fig. 2). In several cases, there are even overlapping start and stop codons, a phenomenon also observed with other bacteriophages [24, 25]. The packed organization could suggest that the ORFs might be expressed as a single operon. However, DNA/RNA hybridization experiments using RNA isolated from *S. venezuelae* cultures, infected

Table 1. Properties of the analysed bacteriophage VWD predicted ORFs and their deduced polypeptide sequences

Gene	Start	Stop	M_r (kDa)		pI[d]	[G+C][e]	GC3s[f]	Stop
	at nt number[a]		predicted[b]	on gel[c]				
ORF2	258	995	26.93	n.i.	4.47	0.721	0.942	UGA
ORFp14	1014	1403	13.38	16.5	5.3	0.678	0.929	UGA
ORFp38	1418	2464	37.69	43	4.34	0.701	0.941	UGA
ORF3	2461	3231	27.13	n.i.	4.05	0.749	0.940	UGA
ORF4	3228	3581	12.93	n.i.	4.66	0.705	0.947	UGA
ORF5	3586	3858	9.78	n.i.	7.21	0.748	0.910	UGA
ORF6	3861	4256	13.91	n.i.	9.08	0.723	0.963	UGA
ORF7	4259	4450	7.05	n.i.	4.55	0.722	0.937	UGA
ORFp24	4450	5004	23.63	27.2	4.90	0.691	0.933	UGA
ORF8	5117	5862	27.76	n.i.	9.49	0.701	0.918	UAG

[a] Numbering started as indicated in Fig. 2
[b] As calculated from the nucleotide-deduced protein sequence
[c] As experimentally determined by SDS-PAGE
[d] Isoelectric point as calculated from the protein sequence
[e] Relative G/C-content
[f] Proportion of G+C at third codon positions

with phage VWB and subsequently incubated for 5 h (the end of the rise period of phage VWB) [10], and the 6 kb *Sph*I/C DNA fragment as probe, showed the presence of 4 transcripts (6.6; 3; 1.6 and 1.1 kb in size) (Figs. 1 and 3b). Using different subfragments of *Sph*I/C as probe indicated that ORFp14, ORFp38 (probably together with ORF3, 4 and 5) appeared as one transcript of circa 3 kb, while ORF6, 7 and p24 are comprised within another transcript (1.1 kb). Transcripts with the size 6.6 and 1.6 kb contained ORF2 and ORF8, respectively. A search for promoters on the basis of the compiled DNA sequences associated with streptomycete promoters [26] did, however, not indicate any apparent consensus promoter sequence. This observation could suggest, therefore, that these genes might still be in a single operon but that the mRNA is cleaved at discrete sites post-transcriptionally. A search for potential hairpin loops revealed the presence of a unique hairpin centered around nt 4985 with a stem of 7 bp and a free energy value $\Delta G = 13$ kcal/mol. This hairpin loop may serve as a transcription terminator, since it is located at the end of the 1.1 kb transcript. Alternatively, and maybe more likely due to the low free energy value of this inverted repeat, it may serve as an mRNA processing signal. In the late operon of phage λ, where all transcription is known to originate from a single promoter (and where the mRNA is certainly processed, even though the details of that processing are not well defined), there are 4 or 5 inverted repeats near gene boundaries suggested to be involved in RNA

processing by ribonuclease III [24]. These loops contain the sequence CCGCC and thus resemble the one reported here.

Analysis of the sequences upstream the putative ORFs showed in most instances an apparent Shine-Dalgarno sequence (GGAGG) located 7 to 10 nt upstream the ATG or GTG initiation codon (Fig. 2). This is in agreement with the average distance (8.5 nt) for *Streptomyces* genes [26], suggesting that the putative ORFs code for functional polypeptides.

The initiation codon was in all instances ATG, except for ORF4 and ORF7 in which case GTG was the most likely initiation codon. No UUA_{Leu} codon was detected within the reading frames confirming the observation that this codon seems to be more or less confined to a small number of genes involved in antibiotic production and differentiation [27]. In the same line, the UGA stop codon is most frequently used (Table 1).

Protein sequence homology and motifs

A database search for the presence of any special motifs and similiarities with other sequences that might exist in the putative ORFs of phage VWB has been carried out. However, no significant similarity has been found.

The analysis of amino acid sequence of p14 for the presence of protein sequence motifs revealed a phosphate-binding loop from amino acid residue 114 to 121. This P-loop motif is commonly found in ATP-and GTP-binding proteins as – (A/G)XXXXGK(T/S) – X being any amino acid residue [28]. Some ATP-binding proteins show, in addition or alternatively, a less conserved site, designated B-motif. This B-motif consists of an aspartate residue preceded by four hydrophobic amino acids [29]. In p14 the P-loop has the sequence – AVDAAGKT and no B-motif is present.

Deduced amino acid sequence of ORF2 exhibits a particularly striking bias in its amino acid composition. This putative protein contains three stretches of acidic residues (20 out of 29 amino acids, 7 of 9 and 9 of 12, respectively) and a carboxy terminal region rich in glycine and arginine (G/R, 8R + 9G out of 23 residues, Fig. 4a). A search in protein databases (PIR and SWISSPROT) for aspartic and/or glutamic rich proteins resulted in two interesting observations. On the one hand, a similar organization as ORF2 protein with three stretches of acidic residues and a glycine/arginine-rich domain has been found in nucleolins, but nucleolin contains between the acidic N-terminal and glycine/arginine rich C-terminal domains an additional domain which is responsible for rRNA binding. Biochemical evidence indicates that nucleolins interact with both nucleolar chromatin and rRNA and might participate in the nucleocytoplasmic translocation of rRNA [31]. On the other hand, for bacteriophage T4 an acidic stretch and an arginine (+lysine)-rich tail has been described for the gene products 67 and 68, respectively. Gene *67* encodes the precursor of the internal peptide II (gp67) [32] and gene *68* the 17 kDa prohead core protein of bacteriophage T4 [33]. The amino acid sequence of gp67, formerly called PIP [(80 amino acid residues) [32] shows a bipartite orga-

a

```
  1 MPARTLPRHPRTGLQAVMWRKPRTGEDPTE
 31 LYPVWPILGGAPDDGDDGDDTDDDQDDDAD
 61 DSDTDDTGDDDKPDHKAEAAKYKALSRKHE
 91 ERAKANAAAVKELARLKREGMSDVEKKIDE
121 AVAGARAEERTRAGERVARSAFLAAAKGRL
151 DNAKDVLDDINLHRYVDDDGEVDDDAIAKL
181 VDKLAPAKGDGDDDQDDDERDTPRRRRQGG
211 GYQGTRRRGGSGRQGGSVAEGREMYRELLG
241 NDKKT
```

ac1, ac2, ac3, G/R

b

```
103 LARLKREGMSDVEKKIDEAVAGARAEERTRAGERVARSAFLAAAKGRLDN 152
    :. :|.::: .. | :.|||:| ||. : .:.:.       :| |:. | :
  5 IEAIKSNDLVAARKLFAEAMA.ARTIDLIKEEK.......IAIARNFLIE 46

153 AKDVLDDINLHRYVDDDGEVDDDAIAKLVDKLAPAKGDGDDDQDDDE 199
    :.:. |: :     |:|:| .||  .|       |::::.|:::|||
 47 GEEPEDEDE.....DEDDEDSDDKDDK.......KDEDSDEDEDDE  80
```

c

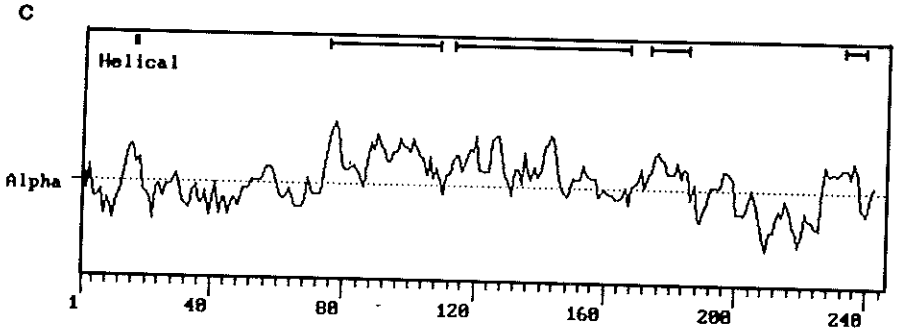

Fig. 4. a Deduced amino acid sequence of ORF2. Boxes labeled ac1, ac2 and ac3 denote three acidic stretches containing 20 acidic residues out of 29 (20/29), 7/9 and 9/12, respectively. Horizontal arrowheads mark the boundaries of the C-terminal glycine-arginine (G/R) domain. b Alignment of the translation product of ORF2 with the prehead core component gp68 of bacteriophage T4. The upper line consists of the amino acids 103–199 from ORF2. The bottom line represents the amino acids 5–80 of gp 68. Bar between two lines indicates identical amino acid residues; double and single dots show a high and low similarity of the amino acid residues, respectively. c α-Helix prediction for gene product of ORF2 computed according to the programme of Novotny and Auffray [30]

nization. The N-terminal part (residues 1 to 46) has a normal amino acid composition, which becomes degraded by the T4 head protease during head maturation. Only the C-terminal part which is acidic (27 acidic residues – either Glu or Asp – out of 34 amino acid residues) is found as the internal peptide II in the mature phage head. It has been shown that gp67 has a crucial role in assembly of a correct scaffolding core and in size determination of T4 [34]. Furthermore, similarly to gp67, scaffolding proteins of several other bacteriophages have a nonrandom charge distribution with negatively charged residues concentrated near the N-terminus [25], and nearly all of them are quite highly charged. E.g., T4 gp67, 45.9% charged amino acids [33], T4 gp68, 33.3% [34], P22 scaffolding protein gp8, 31.7% [25], T4 gp22, 35.7% [35], φ29 gp7, 40.2% [36]. Intriguing is that also the gene product of ORF2 has 34.7% of charged amino acids (D, E, H, K, R). It is believed that the clustering of charged residues in the essential core protein (for T4) may have some function in core assembly. Also striking is that all of the sequenced scaffolding proteins have high α-helix prediction. Similarly, the gene product of ORF2 has a high α-helix prediction (Fig. 4c). An alignment of the putative protein of ORF2 with the bacteriophage T4 prohead core component coded for by gene *67* shows a similarity of 61.84% (Fig. 4b), which might indicate a similar function for ORF2 and gene *67*.

Discussion

Over 3000 tailed phages have been described. They are extremely variable in dimension and physico-chemical properties. Comparative analysis of viral sequences could lead to a better understanding of virus evolution. We have sequenced a 6 kb *Sph*I fragment of the actinophage VWB, which contains the coding sequences of the main structural proteins [10]. Comparison of DNA and protein sequences indicated that these sequences in the 6 kb fragment are all novel and that, as a consequence, phage VWB belongs to a class of phages, of which as yet no DNA sequences have been reported.

The putative proteins identified, in addition to ORFp14, p24 and p38, probably are also head and/or tail proteins. This supposition is made on the basis of a survey of genetic maps of several bacteriophages. These maps show for all phages analysed that head and tail proteins are clustered at one end of the phage genome. Excellent examples are T7 [37], bacteriophage lambda [24], P22 [25], P2 [38], mycobacteriophage L5 [5]. This obviously functionally clustered arrangement of genes has even led to the idea that genes are clustered into functional modules which could be exchanged among different phages [9].

Of most interest is the analysis of the protein sequence motifs, indicating that p14 contains a phosphate-binding loop (P-loop motif). Proteins with ATPase activity have been commonly found in double-stranded DNA bacteriophages to effect translocation of the viral DNA into a preformed protein shell [39, 40]. Although the P-loop motif might be a false positive, the finding of the nucleotide-binding motif in p14 could indicate that this protein is involved in

encapsidation of VWB phage DNA into the phage head. Further investigations are necessary to assess the functionality of the nucleotide-binding site motif of p14.

In addition, from the analysis of ORF2 and its comparison with other bacteriophage proteins – in particular with the bacteriophage T4 scaffolding proteins – it can be concluded that the gene product of ORF2 is probably a scaffolding protein of VWB. Gene *68* (of T4) is located directly downstream of gene *67*. It codes for a 17 kDa protein with an acidic amino terminal half and a basic carboxy terminus rich in arginine and lysine. This 17 kDa protein is believed to be involved in head shape determination [33]. Gene product ORF2 has a carboxy terminus rich in glycine and arginine. It could be that for ORF2 lysine is replaced by arginine, since codons for lysine are very poor in G/C. Taking all these data together and considering the unusual primary structure of ORF2, this gene could be a fusion of 2 genes with the same functionality in the scaffolding process as gene *67* and *68* of bacteriophage T4.

The sequencing data of actinophage VWB confirm that despite the existence of a high similarity in gene organization between different dsDNA bacteriophages, there is surprisingly little or no sequence similarity between proteins which have highly analogous functions. Therefore, to get a better insight in phage evolution much more phage sequences need to become available.

Acknowledgements

These studies were supported by a grant from the Belgian Nationaal Fonds voor Wetenschappelijk Onderzoek (N.F.W.O.) (Project 3.0018.90). The DNA sequencing facilities (PF&GO) were supported by the "Vereniging voor Kankerbestrijding", the Cancer Research Foundation of the General Savings and Retirement Fund (A.S.L.K.) and the N.F.W.O. We thank Dr. Mervyn Bibb for the FRAME analysis of the sequence and discussion, and Dr. A. Van Aerschot and G. Schepers for primer synthesis. Dominique Brabants is acknowledged for fine editorial work. LVM is a postdoctoral fellow of IWT (Flemish Institute for the Promotion of Scientific and Technological Research in Industry). GO is senior research associate of the NFWO.

References

1. Ackermann HW, Berthiaume L, Jones LA (1985) New actinophage species. Intervirology 23: 121–130
2. Korn-Wendisch F, Schneider J (1992) Phage typing – a useful tool in actinomycete systematics. Gene 115: 243–247
3. Ogata S (1980) Bacteriophage contamination in industrial processes. Biotech Bioeng 22: 177–193
4. Hahn DR, McHenney MA, Baltz RH (1990) Characterization of FP22, a large streptomycete bacteriophage with DNA insensitive to cleavage by many restriction enzymes. J Gen Microbiol 136: 2395–2402
5. Hatfull GF, Sarkis GJ (1993) DNA sequence, structure and gene expression of mycobateriophage L5: a phage system for mycobacterial genetics. Mol Microbiol 7: 395–405

6. Lomovskaya ND, Chater KF, Mkrtumian NM (1980) Genetics and molecular biology of *Streptomyces* phages. Microbiol Rev 44: 206–229
7. Studier FW (1979) Relationships among different strains of T7 and among T7-related bacteriophages. Virology 95: 10–76
8. Schneider J, Kutzner HJ (1989) Distribution of modules among the central regions of the genomes of several actinophages of *Faenia* and *Saccharopolyspora*. J Gen Microbiol 135: 1671–1678
9. Casjens S, Hatfull G, Hendrix R (1992) Evolution of dsDNA tailed bacteriophage genomes. Semin Virol 3: 383–397
10. Anné J, Van Mellaert L, Decock B, Van Damme J, Van Aerschot A, Herdewijn P, Eyssen H (1990) Further biological and molecular characterization of actinophage VWB. J Gen Microbiol 136: 1365–1372
11. Anné J, Wohlleben W, Burkardt HJ, Springer R, Pühler A (1984) Morphological and molecular characterization of several actinophages isolated from soil which lyse *Streptomyces cattleya* or *S. venezuelae*. J Gen Microbiol 130: 2639–2649
12. Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene 33: 103–119
13. Anné J, Verheyen P, Volckaert G, Eyssen H (1985) A restriction endonuclease map of *Streptomyces* phage VWB. Mol Gen Genet 200: 506–507
14. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbor
15. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
16. Bibb MJ, Findlay PR, Johnson MW (1984) The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30: 157–166
17. Staden R (1986) The current status and portability of our sequence handling software. Nucleic Acids Res 14: 217–231
18. Tanaka K, Shina T, Takahashi H (1991) Nucleotide sequence of genes *hrdA*, *hrdC* and *hrdD* from *Streptomyces coelicolor* A3(2) having similarity to *rpoD* genes. Mol Gen Genet 229: 334–340
19. Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395
20. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685
21. Chrambach A, Reisfeld RA, Wyckoff M, Zaccari J (1967) A procedure for rapid and sensitive staining of protein fractionated by polyacrylamide gel electrophoresis. Anal Biochem 20: 150–154
22. Hopwood DA, Bibb MJ, Chater KF, Kieser TJ, Bruton CJ, Kieser HM Lydiate DJ, Smith CP, Ward JM, Schrempf H (1985) Genetic manipulation of Streptomyces: a laboratory manual. John Innes Foundation, Norwich
23. Bibb MJ, Cohen SN (1982) Gene expression in *Streptomyces*: Construction and application of promoter-probe plasmid vectors in *Streptomyces lividans*. Mol Gen Genet 187: 265–277
24. Sanger F, Coulson AR, Hong GF, Hill DF, Pedersen GB (1982) Nucleotide sequence of bacteriophage λ DNA. J Mol Biol 162: 729–773
25. Eppler K, Wyckoff E, Goates J, Parr R, Casjens S (1991) Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging. Virology 183: 519–538
26. Strohl WR (1992) Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res 20: 961–974

27. Leskiw BK, Mah R, Lawlor EJ, Chater KF (1993) Accumulation of *bld*A-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3(2). J Bacteriol 175: 1995–2005
28. Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop – a common motif in ATP- and GTP-binding proteins. TIBS 15: 430–434
29. Higgins CF, Gallagher MP, Mimmack ML, Pearce SR (1988) A family of closely related ATP-binding subunits from prokaryotic and eukaryotic cells. Bioassays 8: 111–116
30. Novotny J, Auffray C (1984) A program for prediction of protein secondary structure from nucleotide sequence data: application to histocompatibility antigens. Nucleic Acids Res 12: 243–255
31. Maridor G, Krek W, Nigg EA (1990) Structure and developmental expression of two abundant non-ribosomal nucleolar proteins. Biochim Biophys Acta 1049: 126–133
32. Völker TA, Gafner J, Bickle TA, Showe MK (1982) Gene *67*, a new, essential bacteriophage T4 head gene codes for a prehead core component, PIP. J Mol Biol 161: 479–489
33. Keller B, Senstag C, Kellenberger E, Bickle T (1984) Gene *68*, a new bacteriophage T4 gene which codes for the 17K prohead core protein is involved in head size determination. J Mol Biol 179: 415–430
34. Keller B, Maeder M, Becker-Laburte C, Kellenberger E, Bickle TA (1986) Amber mutants in gene *67* of phage T4. Effects on formation and shape determination of the head. J Mol Biol 190: 83–95
35. Marusich E, Mesyanzhinov V (1989) Nucleotide and deduced amino acid sequences of bacteriophage T4 gene *22*. Nucleic Acids Res 17: 8865
36. Vlcek C, Paces V (1986) Nucleotide sequence of the late region of *Bacillus* phage φ29 completes the 19285-bp sequence of φ29 genome. Comparison with the homologous sequence of phage PZA. Gene 46: 215–225
37. Dunn JJ, Studier FW (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the location of T7 genetic elements. J Mol Biol 166: 477–535
38. Linderoth NA, Ziermann R, Haggard-Ljungquist E, Christie GE, Calender R (1991) Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2. Nucleic Acids Res 19: 7202–7214
39. Guo P, Peterson C, Anderson D (1987) Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage φ29. J Mol Biol 197: 229–236
40. Black LW (1989) DNA packaging in dsDNA bacteriophages. Annu Rev Microbiol 43: 267–292

Authors' address: Dr. J. Anné, Rega Institute, Minderbroedersstraat 10, B-3000 Leuven, Belgium.

Received October 26, 1994