Microbiology (1999), **145**, 1661–1670 Printed in Great Britain

Genomic structure of phage B40-8 of *Bacteroides fragilis*

Montserrat Puig and Rosina Gironés

Author for correspondence: Montserrat Puig. Tel: +34 93 4021491. Fax: +34 93 4110592.
e-mail: mpuig@porthos.bio.ub.es

Department of
Microbiology, Faculty of
Biology, University of
Barcelona, Avgda
Diagonal, 645-08028
Barcelona, Spain

Very few data are available on the molecular biology of *Bacteroides fragilis* bacteriophages, which have been considered in several studies as indicators of faecal contamination. Phage B40-8, initially isolated from an urban sewage sample using a strain of *B. fragilis* (HSP40) isolated from a clinical specimen, was chosen in this study as a prototype for morphological and molecular studies. Like most of the phages infective for *B. fragilis*, B40-8 belongs to the *Siphoviridae* family. Its genome has been found to be a double-stranded DNA molecule, of approximately 51·7 kb, containing a rather low percentage (38·9 mol%) of G+C. The ends of the molecule appeared not to be cohesive but permuted, with a terminal redundancy of 7·3%. A genomic map was constructed. Three major proteins (MP) out of 15 peptides in the SDS-PAGE profile were selected for N-terminal sequencing. From these data, degenerate probes were designed to locate the ORFs in the genomic map. Immunodetection by electron microscopy revealed that MP1 and MP3 were structural proteins of the phage head and that MP2 was a constituent of the tail. A genomic library of the phage was prepared, and a clone including the MP2 ORF was identified and sequenced.

Keywords: *Bacteroides fragilis*, phage B40-8, genomics

INTRODUCTION

Bacteroides is one of the most abundant anaerobic bacterial genera in the gut of warm-blooded animals (10^7–10^{10} cells per g faeces) (Eggerth & Gagnon, 1933). *Bacteroides fragilis* is the anaerobic species most frequently isolated from clinical samples (Holland *et al.*, 1977), particularly from infected human soft tissues and anaerobic bacteraemias (Holdeman *et al.*, 1976; Polk & Kasper, 1977). Previous studies have attempted to understand the physiology and pathogenicity of *B. fragilis*, and some genes have been cloned and expressed in order to identify virulence factors (Russo *et al.*, 1990; Ono *et al.*, 1994), regulatory proteins (Rasmussen & Kovacs, 1993) or products related to the antibiotic resistance of bacteria (Rashtchian & Booth, 1981; Smith, 1985; Rasmussen *et al.*, 1990). In these studies phage λ-based vectors were used, and results were not always successful due to incompatibility of their expression regulation systems. One possible way to deal with this may be the use of *B. fragilis* phage genetic elements in vector construction, which will require detailed characterization of individual phages. At present, very few data related to *Bacteroides* phage molecular biology are available.

There are more than 100 descriptions of *Bacteroides* phages in the literature but information on genome length, mol% G+C, hybridization patterns (Kory & Booth, 1986) and restriction profiles (Kai *et al.*, 1985) is available for only seven phage isolates of *B. fragilis*.

The group of phages that infect *B. fragilis* strain HSP40 have been proposed as a potential model for the study of viral contamination of faecal origin (Tartera & Jofre, 1987; IAWPRC Study Group on Health Related Water Microbiology, 1991; Havelaar, 1993). Furthermore, strain HSP40 selectively detects phages of human origin while it is resistant to infection by most of the phages present in animal faeces (Tartera & Jofre, 1987; Puig *et al.*, 1997). Thus, the presence of *B. fragilis* HSP40 phages is informative of the origin of faecal contamination.

The present study is focused on the molecular characterization of phage B40-8 of *B. fragilis* HSP40. B40-8 was selected as phage prototype for this study based on the following: (i) it is a *Siphoviridae* family member, like

---

Abbreviations: AFM, atomic force microscopy; CAP, calf alkaline phosphatase; MP, major protein; TEM, transmission electron microscopy.

The GenBank/EMBL/DDBJ accession number for the sequence reported in this paper is AF074719.

most *B. fragilis* phages reported in the literature (Keller &Traub, 1974; Booth *et al.*, 1979; Kai *et al.*, 1985; Kory & Booth, 1986); (ii) it also has the most frequently isolated morphology of *B. fragilis* phages present in natural waters (Tartera & Jofre, 1987; Lasobras *et al.*, 1997); (iii) it has been isolated from an urban sewage sample on *B. fragilis* strain HSP40 (Tartera & Jofre, 1987); and (iv) it infects the collection strain *B. fragilis* ATCC 25285. This study describes the structure of the genome and the localization of the major proteins in the genomic restriction map of this phage.

METHODS

Phage, bacterial strains and media. Phage B40-8 was initially isolated from an urban sewage sample by Tartera & Jofre (1987) and it was cultured on *Bacteroides fragilis* strain HSP40 (ATCC 15477) as previously described (Tartera & Jofre, 1987). *Escherichia coli* XL-1 Blue MRF′ (Stratagene) was used as the host for molecular cloning. *Bacteroides fragilis* was cultured with *Bacteroides* phage recovery medium (BPRM) (Tartera *et al.*, 1992). Luria–Bertani (LB) broth (per litre: 1 g tryptone, 0·5 g yeast extract, 0·5 g NaCl) or LB agar plates supplemented with ampicillin (100 μg ml^{-1}) and tetracycline (12·5 μg ml^{-1}) were used for growth of *E. coli* containing vectors.

Concentration and purification of B40-8 phage. Highly concentrated and purified phage stocks were prepared following the method described for λ phage (Sambrook *et al.*, 1989), slightly modified. A single lytic plaque was suspended in 400 μl phage buffer (19·5 mM Na_2HPO_4, 22 mM KH_2PO_4, 85·5 mM NaCl, 1 mM $MgSO_4$, 0·1 mM $CaCl_2$) and incubated for at least 4 h at 4 °C. The total volume of the phage suspension was added to a 30 ml bacterial culture, growing in BPRM (3×10^8 c.f.u. ml^{-1}). Infection was induced by incubating the mixture in hermetic bottles at 37 °C for 16–20 h. Partially clarified bacterial lysates were treated with chloroform (0·3 vol.); the supernatant recovered was centrifuged at 12100 ***g*** for 10 min and the cell debris was discarded. Later, 1 litre of bacterial culture (3×10^8 c.f.u ml^{-1}) was seeded with the phage solution obtained and the mixture was cultured in anaerobic conditions (500 ml bottles in GasPack), at 37 °C, for 16–20 h. Complete lysis of bacteria was achieved by adding 1 M NaCl. Subsequent addition of polyethylene glycol 6000 (100 g l^{-1}) permitted the concentration of viral particles. Absorbed phages were collected by centrifugation at 17700 ***g*** for 10 min at 4 °C, and pellets were resuspended in 30 ml SM buffer [100 mM NaCl, 10 mM $MgSO_4$, 50 mM Tris/HCl (pH 7·5), 0·01 % gelatin] per litre of initial culture. Concentrated solutions were purified on two successive CsCl gradients (as described by Sambrook *et al.*, 1989) and dialysed against 50 mM NaCl, 10 mM $MgSO_4$, 50 mM Tris/HCl (pH 7·5).

Transmission electron microscopy (TEM). Approximately 10^9–10^{10} p.f.u. of purified B40-8 were deposited on copper grids coated with Formvar film and carbon. They were negatively stained with 2 % phosphotungstic acid (pH 7·0) for 2 min. Particles were viewed with a Hitachi H600 AB electron microscope at a 75 kV accelerating voltage. Scanned images (Epson GT-8000) were analysed by the IMAT program (Scientific and Technical Services, University of Barcelona) and the dimensions and morphology of the viral particles were recorded.

Atomic force microscopy (AFM). Aliquots containing 3×10^6 to 3×10^7 p.f.u. of purified B40-8 in two different buffers [40 mM HEPES/KOH (pH 7·6), 10 mM $MgCl_2$ and 10 mM NaCl, 5 mM Tris/HCl (pH 8·0), 10 mM $MgCl_2$] were placed on mica supports. Samples were dried at room temperature for 5 min, washed in deionized water and finally dried under liquid argon. The viral particles were observed by a Nanoscope III atomic force microscope (Extended Multimode AFM, Digital Instruments) using tapping mode.

DNA extraction, restriction analysis and mol % G + C content. Purified phage particles were incubated at 65 °C for 3 min with 0·25 % SDS at a final volume of 500 μl. After cooling the sample at room temperature, 250 μl 7·5 M ammonium acetate was added. Samples were centrifuged at 4500 ***g*** for 15 min to sediment the proteins and separate the supernatant containing nucleic acids (the pellet was kept at −20 °C and used for protein analysis). DNA was concentrated by ethanol precipitation and resuspended in 50–100 μl deionized water.

Viral DNA was digested using several restriction enzymes (see Table 2), following the instructions of the supplier. Fragments were analysed in 0·7–2·5 % agarose gels by conventional electrophoresis. Alternatively, to improve the sizing of the fragments over 10 kb, PFGE based on the hexagonal electrode array developed by Chu *et al.* (1986) was applied. Conditions of PFGE were 1·2 % agarose gel (SeaKem Gold, FMC) in 0·15 × TBE buffer (1 × TBE is 89 mM Tris, 89 mM boric acid, 2 mM EDTA). Samples were run at 220 V, at 10 °C for 8 h, in a CHEF-DR II system (Bio-Rad). Electric-field direction was changed every 0·6 s.

The G + C content of B40-8 DNA was quantified as follows. Phage DNA (1 mg ml^{-1}) was heat-denatured and digested with nuclease P1 (Sigma) (Tamaoka & Komagata, 1984). Terminal phosphate groups were eliminated from nucleotides using calf alkaline phosphatase (CAP, Boehringer Mannheim). The hydrolysate and an external standard solution (for all four nucleosides) were then subjected to reverse-phase HPLC [Chromspher C_{18} column, Chrompack; mobile phase 0·1 M $NH_4H_2PO_4$ (pH 4·0)/acetonitrile (20:1, v/v) solution]. G + C percentages were calculated from relative peak areas.

Terminal structure of DNA. The presence of cohesive ends was checked by comparing the restriction patterns of phage DNA with and without a previous treatment with T4 DNA ligase. After heat inactivation of the enzyme (70 °C, 15 min), samples were digested separately with *Hin*dIII or several other endonucleases. Digested products were heated at 70 °C and kept on ice until being loaded in agarose gels. Phage λ-DNA was used as control.

Two additional methods based on the terminal labelling of DNA with digoxigenin-11-dUTP (Boehringer Mannheim, *The DIG System User's Guide for Filter Hybridization*, 1995) and [γ-^{32}P]ATP (as described by Sambrook *et al.*, 1989) were applied to determine the structure of terminal sequences in the genome. Total phage DNA was later enzyme-digested and results were analysed as follows. (1) Digoxigenin labelling: restriction products were run in vertical 0·8 % agarose gels and subsequently transferred to Hybond + nylon membranes (Amersham); labelled fragments were visualized by the enzymic method supplied by Boehringer Mannheim. (2) [γ-^{32}P]ATP labelling: vertical 0·8 % agarose gels containing well-separated restriction products were dried and autoradiographed using Curix RP2 films (Agfa), at different exposure times depending on the sample signal.

Nuclease BAL31 analysis. BAL31 requires free DNA ends as substrate. Based on this principle, we identified the genomic regions that were first degraded by BAL31. After treating 1 μg B40-8 DNA with 3 units of the enzyme, according to supplier's recommendations, the reaction was stopped, at 4 min intervals, by ethanol precipitation of DNA. Samples were digested using restriction endonucleases and band disappearance was visualized by agarose gel electrophoresis.

Southern blotting and hybridization analysis. The fragments produced by *Hin*dIII digestion of B40-8 DNA were isolated from agarose gels or obtained from genomic libraries, and then labelled with digoxigenin-11-ddUTP by using a random primer kit (Boehringer Mannheim). After electrophoresis, the restriction digests were transferred to Hybond+ nylon membranes (Amersham) and Southern blotting was performed at high stringency [5 × SSC, 0·1 % sodium lauroylsarcosine, 2 % blocking reagent (Boehringer Mannheim), 0·2 % SDS, 50 % formamide, at 45 °C], with probe concentrations of 2–3 ng ml^{-1}. Filters were washed three times in 0·1 × SSC, 0·1 % SDS at 68 °C, and subsequently developed using chemiluminescent detection (Boehringer Mannheim).

Recombinant DNA techniques. All *Hin*dIII-digested restriction fragments of B40-8 DNA, except H1 and H6′, were cloned into CAP-treated pBluescript II SK+ vector (Stratagene), using standard protocols (Sambrook *et al.*, 1989). Ligated plasmids were used to transform *E. coli* XL-1 Blue MRF′ (Stratagene). The 15 kb (H1) *Hin*dIII fragment was subcloned in the same vector using internal restriction by *Sau*3AI.

DNA sequencing. DNA sequencing was performed by applying ABI Prism dye terminator cycling sequencing (Perkin Elmer) and a Gene Amp PCR Systems 2400 (Perkin Elmer). In standard 10 μl sequencing reactions, 250–500 ng recombinant DNA was analysed using 3·2 pmol M13 (GTAAAACGACGGCCAGT) or reverse-M13 (GGAAACAGCTATGACCATG) primers (Stratagene).

The software package from the University of Wisconsin, Genetic Computer Group (Genetics Computer Group, 1991) was used for the analysis of sequence data. Data searches for DNA and the amino acid sequences of the putative protein similarity were performed with the program FASTA on the GenBank, EMBL and SWISS-PROT databases.

Protein analysis. SDS-PAGE was performed according to Laemmli (1970). Approximately 10^{10} p.f.u. of CsCl-purified phage preparation was boiled for 10 min in SDS-PAGE sample buffer (Laemmli, 1970), followed by electrophoretic analysis in gels containing 10 % (w/v) polyacrlylamide. Gels were stained with Coomassie Brillant Blue R250, with silver salts (Sambrook *et al.*, 1989), or both. Low-molecular-mass marker proteins (Pharmacia) were used to estimate virion protein sizes.

Anti-B40-8 serum production. The anti-B40-8 serum was produced by MedProbe, as follows. Two rabbits were immunized by subcutaneous injection of 10^9 p.f.u. of purified B40-8 phage particles in Freund's incomplete adjuvant. Two subsequent doses were supplied at 2- and 4-week intervals, respectively. Serum obtained from the first and second bleed was tested for specific antibodies by immunoblotting. Shortly after the second bleed, a total blood volume of 45 ml was collected. B40-8 phage antiserum could be diluted 1:5000 without loss of signal when tested against virion proteins on PVDF filters (Millipore) and developed by enzyme-conjugated secondary antibody (Pharmacia) at a dilution of 1/2000.

Specific antibodies and immunogold electron microscopy. Phage proteins were fractionated on SDS-polyacrylamide (10 % acrylamide) gels and electrophoretically transferred to PVDF sheets by using transfer buffer (48 mM Tris, 39 mM glycine, 0·0375 % SDS, 20 % methanol) and a Trans-blot SD semi-dry electrophoretic transfer cell (Bio-Rad). Blotted proteins were incubated in TST-3 % BSA buffer [TST buffer is 50 mM Tris/HCl (pH 10·2), 500 mM NaCl, 0·5 % Tween20] for 30 min. A 1/1000 antiserum dilution was added immediately afterwards, and the incubation was left to proceed for 16–20 h at room temperature. The bound antibodies were visualized with alkaline-phosphatase-labelled anti-rabbit IgG using the enhanced chemiluminescent system (Boehringer Mannheim) according to the manufacturer's instructions.

Antibodies were prepared as described by Harlow & Lane (1988), but on nitrocellulose membranes (Bio-Rad), from 1/500 antiserum dilution. Antibody preparations for immunogold labelling experiments were dialysed against deionized water, and then samples were concentrated to a volume of approximately 25 μl in a Speed-Vac vacuum system. Volumes of 5–10 μl of antibody preparation were used for labelling experiments.

Approximately 10^{10} p.f.u. of purified phages were attached to a Formvar-carbon film on 400-mesh gold grids, and exposed to anti-B40-8 serum and specific antibody solutions for 30 min (Hayat, 1989). Grids were washed three times in PBS (pH 7·4) with 3 % BSA and subsequently suspended for 15 min on drops of 1/60 dilution of 5 nm colloidal-gold-conjugated protein A (Amersham), in the same buffer. After three additional washes in water, samples were negatively contrasted by 2 % phosphotungstic acid (pH 6·5) for visualization by electron microscopy (Philips 8000 MT apparatus, 100 kV).

N-terminal sequencing of proteins. Horizontal sections of PVDF sheets containing blotted phage major proteins were excised using phage protein lanes stained with Coomassie Brillant Blue as markers. The N-terminal sequence of these proteins was kindly determined by T. Ellebaek Petersen (Aarhus University, Denmark), by Edman degradation on an Applied Biosystems protein sequencer.

Degenerate probes. Degenerate oligonucleotides were designed on the basis of high-frequency codon tables established according to *B. fragilis* plasmid DNA (EMBL database, M72418), since such data are not available for phage DNA. A 22-mer mix covering 384 coding possibilities was 5′-labelled with [γ-^{32}P]ATP and used as a probe in Southern blotting analysis (as described above) to locate major-protein ORFs on the phage genomic map.

RESULTS

Phage particle morphology

B40-8 particles showed icosahedral heads and long, non-contractile tails when visualized by both TEM (Fig. 1a) and AFM (Fig. 1b). The dimensions of the structural elements were estimated using IMAT image analysis and Nanoscope software, respectively (Table 1). A spherical body with a similar diameter to the phage tail was visualized in phage samples analysed by AFM (Fig. 1b).

Size and structure of phage DNA

The genome of B40-8 is a single double-stranded DNA linear molecule. Its size was calculated as 51·7 ± 0·2 kb, according to the sizes of fragments generated by digesting total DNA with different restriction endonucleases. The size of the fragments over 10 kb was estimated more precisely using PFGE. HPLC analysis

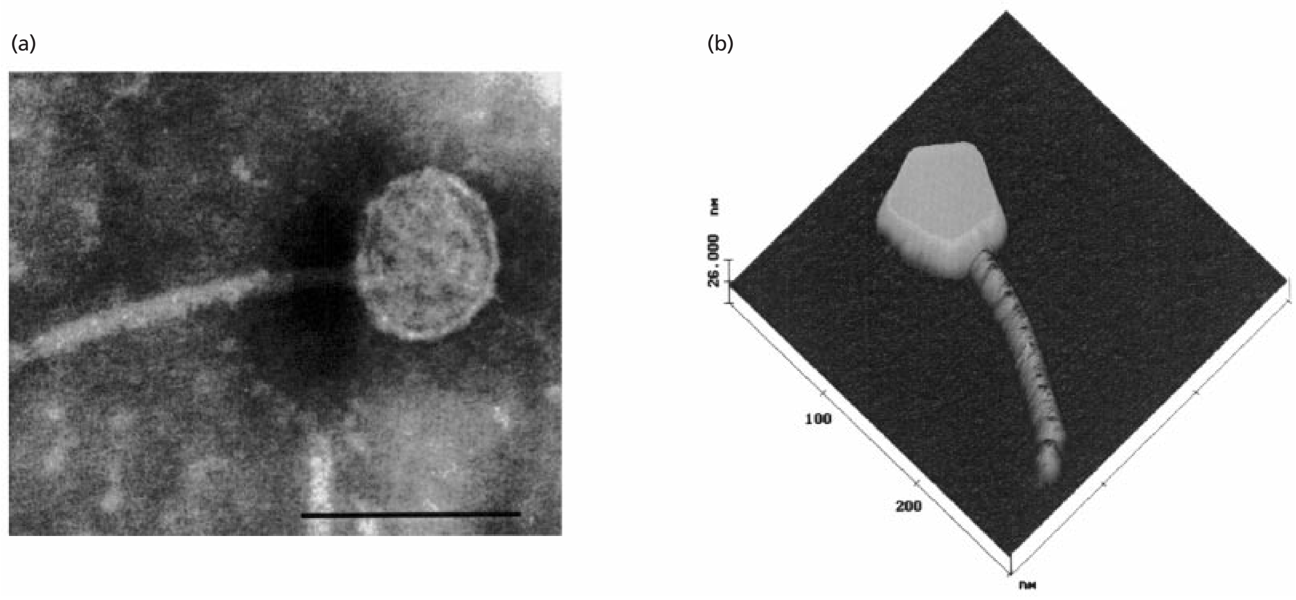

Fig. 1. (a) Electron micrograph of negatively stained phage B40-8. Bar, 100 nm. (b) Image of phage B40-8 obtained by AFM, using NaCl/Tris/$MgCl_2$ buffer.

Table 1. Phage head and tail measures by TEM and AFM using IMAT and Nanoscope software

The values are expressed in nm.

Parameter	TEM			AFM		
	Range	Mean	SD	Range	Mean	SD
Head diameter	58·8–68·2	62·8	3·7	87·1–92·5	90·0	2.8
Tail length	148·7–164·0	156·8	7·1	136·5–155·3	146·1	9·86
Tail width	8·8–9·6	9·3	0·4	19·2–21·3	20·9	1·54

showed a G+C DNA content of 38·9 mol%. No modified bases were detected in a DNA base analysis of the B40-8 phage by HPLC. The structure of DNA ends was studied enzymically and by end-labelling. The ends were not ligated after ligase and heat treatment, thus differing from the control pattern of phage λ, and suggesting the absence of cohesive motifs in the B40-8 genome. Non-radioactive and radioactive terminal labelling followed by restriction analysis showed weak labelling in some of the bands or a smearing pattern when the phage DNA was analysed. These results and further evidence described below indicate permuted ends on the DNA molecule.

Physical map

Several experiments were required to construct a physical map with sites for six restriction endonucleases, packaging initiation sites (*pac*) and location of major-protein ORFs (Fig. 2). Single and double digestions permitted preliminary mapping of the B40-8 genome, and hybridization tests were performed at high stringency using *Hin*dIII-generated fragments as probes, to detect homologous sequences that overlapped in the fragments. Upon single digestions with the six endonucleases selected, a pattern of subfragments was repeatedly detected (Fig. 3). Bands *Bgl*II-B3, *Sal*I-S1, *Bcl*I-C4, *Bst*EII-E5 and *Stu*I-T2 (the last not shown) showed lower intensities in the ethidium-bromide-stained agarose gels (marked with white arrowheads in Fig. 3). The submolar fragment for the *Hin*dIII digestion (H6′, 2·8 kb) had the same molecular mass as fragment H6, and was consequently not seen as a submolar band on the agarose gels. The presence of terminal redundancy was verified by hybridizing the submolar fragments E5 and B3, and also H6′, isolated from agarose gel and labelled with digoxigenin-11-dUTP, to DNA extracted from virions. The hybridization to these

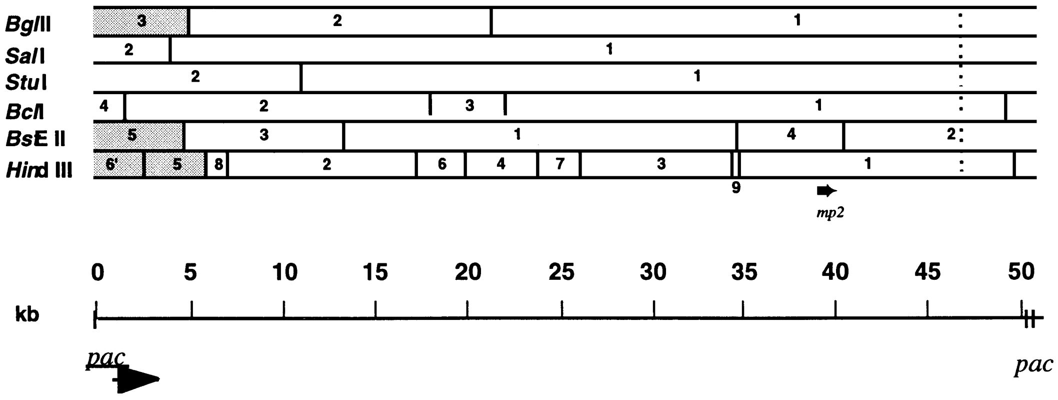

Fig. 2. Restriction map of B40-8 DNA. The circularly permuted B40-8 DNA molecule has been drawn linear, and shows cleavage sites for six restriction endonucleases. Restriction fragments sizes are as follows (in kb). *Bgl*II fragments: B1, 29·5; B2, 16·9; B3, 5·3. *Sal*I fragments: S1, 47·4; S2, 4·3. *Stu*I fragments: T1, 40·2; T2, 11·5. *Bcl*I fragments: C1, 27·6; C2, 16·1; C3, 3·9; C4, 1·9. *Bst*EII fragments: E1, 21·7; E2, 10·5; E3, 8·6; E4, 5·9; E5, 5·0. *Hin*dIII fragments: H1, 15·0; H2, 10·5; H3, 8·3; H4, 3·9; H5, 3·2; H6/H6′, 2·8; H7, 2·4; H8, 1·2; H9, 0·4. The approximate location of packaging initiation regions (*pac*) are marked: pac (on the left side) is the most frequent cleavage site and the dotted line indicates initiation of terminal redundancy. Both *pac* regions (on the right side) are secondary cleavage sites for first headful packaging. The arrow indicates the sense of packaging. MP1, MP2 (see also the orientation of *mp2* in the map) and MP3 ORFs were located by Southern blotting on the H1 fragment. The shaded restriction fragments were used as probes to establish the *pac* site position.

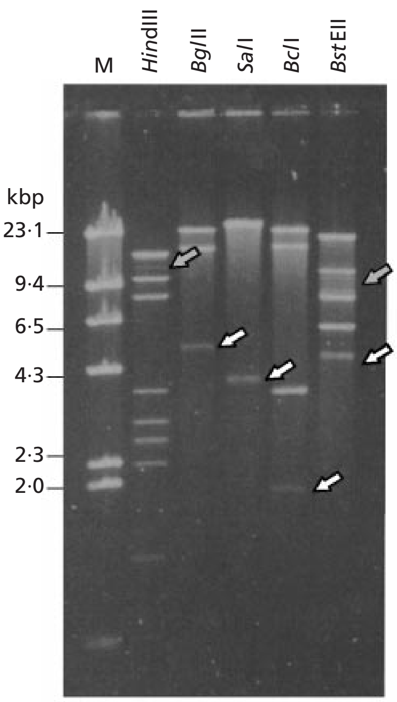

Fig. 3. Single enzymic digestions of B40-8 DNA. Submolar bands (marked with white arrowheads) are present in all lanes except in the *Hin*dIII pattern, where the submolar band (H6′) is at the same position as H6. Submolar bands correspond to the fragments flanked by the most frequent cleavage site (pac in Fig. 2) for the first round of DNA packaging and the nearest restriction site for the endonuclease used. Very weak bands are indicated by grey arrowheads and represent less frequently generated fragments by the same mechanism, but with an origin in secondary *pac* sites (*pac* in Fig. 2). Lane M, size standards (*Hin*dIII digest of coliphage-λ-DNA).

digestions gave a signal with full-length restriction fragments as well as as derived submolar (*pac*) fragments. The H5 fragment was also used as probe; it did not show a hybridization signal either on fragment H1 or on C1. These results proved the existence of a terminal redundancy of 3·8 kb (7·3 %). This redundancy corresponds to the size of H6′ (2·8 kb) plus an extra 1·0 kb region situated on the right of the map that should be the repeated end of H5. We think that this small region may not be detected in the agarose gels because it would be present in very low quantities (first round of packaging), and also because it probably contains the secondary *pac* sites.

Restriction profiles were scanned and intensity of bands was measured by densitometry (IMAT software). Submolar bands showed an intensity of 40–50 % of the values expected according to their size (bands marked with white arrowheads in Fig. 3), or even less (bands marked with grey arrowheads in Fig. 3). In addition, the *Hin*dIII-H6 band showed a higher intensity than 100 %, confirming the presence of a double band in that position. Secondary *pac* sites could explain the submolar fragments that appeared between bands H1 and H2 in the *Hin*dIII digestion, and between E2 and E3 in the *Bst*EII digestion (Fig. 3). These bands of fluffy appearance were present at a ratio of less than 0·3 in the restriction patterns of the phage B40-8 DNA. BAL31 nuclease DNA end-degradation combined with enzymic restriction gave additional information on the terminally located fragments and the relative position of contiguous regions on the map. The sequential shortening expected for the fragments exposed early to the enzyme was seen as the disappearance of the following fragments at 4 min of BAL31 treatment: E2 and E5 in *Bst*EII digestion; C1

Table 2. Number of cleavage sites for different restriction endonucleases in the B40-8 DNA

Restriction endonuclease	Cleavage sequence	No. of restriction sites in B40-8 DNA	
		Expected*	Observed
*Sau*3AI	↓GATC	182	> 25†
*Alu*I	↓AGCT	182	> 26†
*Acc*I	GT↓(A/C)(G/T)AC	45	> 25†
*Dra*I	TTT↓AAA	41	> 29†
*Hin*dIII	A↓AGCTT	17	9
*Bcl*I	T↓GATCA	17	4
*Bgl*II	A↓GATCT	17	2
*Cla*I	ATC↓GAT	17	2
*Eco*RI	G↓AATTC	17	2‡
*Stu*I	AGG↓CCT	7	1
*Bst*EII	GG↓TNACC	7	4
*Pvu*II	CAG↓CTG	7	3
*Sal*I	G↓TCGAC	7	1
*Xba*I	T↓CTAGA	17	0
*Bam*HI	G↓GATCC	7	0
*Kpn*I	GGTAC↓C	7	0
*Pvu*I	CGA↓TCG	7	0
*Xho*I	C↓TCGAG	7	0
*Pst*I	CTGCA↓C	7	0
*Sma*I	CCC↓GGG	3	0
*Bgl*I	GCCNNNN↓NGGC	3	0

* Calculated according to Powell & Davidson (1986).

† Number of bands detected in 1·5 % agarose gels. Some of these bands could contain more than one fragment; therefore these numbers do not represent cleavage sites.

‡ Smearing was detected in the profile.

and C4 in the *Bcl*I profile. In *Hin*dIII digestion, the H1 band lost intensity at 4 min of treatment while the same effect on the H5 band was not seen for a further 8 min (data not shown).

A lack of capacity to digest DNA was observed for many restriction endonucleases. Moreover, all the enzymes with cleavage sites recognized those specific sequences at lower frequencies than expected (assuming random bases in the molecule, and calculating the number of expected sites according to Powell & Davidson, 1986) (Table 2). In general, restriction sites with 6 and 4 bases and a high content (over 66·6 %) of adenines and thymines were identified more frequently than those with lower contents. Eight out of 22 enzymes tested that digested phage λ DNA did not cut the genome of B40-8. Thus, B40-8 DNA is refractory to digestion by a large number of endonucleases.

The relative position of the gene encoding the tail major protein (*mp*2) on the map was determined by Southern blotting as described below. The map shown in Fig. 2 starts at the *pac* site and it is orientated 5′–3′ according to the positive sense of the MP2 ORF.

Genomic libraries

The phage DNA was cloned in recombinant vectors and transformed into *E. coli*. A library of *Hin*dIII fragments was prepared containing clones for 8 of the 10 restriction fragments (the exceptions are H1 and H6′). H1 is a 15 kb fragment and because of this size it is difficult to clone in a pBluescriptII SK+ vector. The terminal 2·8 kb submolar fragment (H6′) was not cloned. Inserts for other fragments in the library were characterized and oriented by double digestion, using the expected internal cleavage sites. The H1 region was subcloned in pBluescriptII SK+ using *Sau*3AI. Some of the fragments were selected by Southern blotting hybridization and further characterized.

Major proteins and their location on phage particle

SDS-PAGE analyses were performed and a pattern of 15 peptides was found, with molecular masses ranging from over 97·4 kDa to 25 kDa (Fig. 4). Three major proteins (MP) were selected from Coomassie-Brillant-Blue-stained gels according to their relative amount: MP1 (46·0 kDa), MP2 (28 kDa) and MP3 (62·5 kDa)

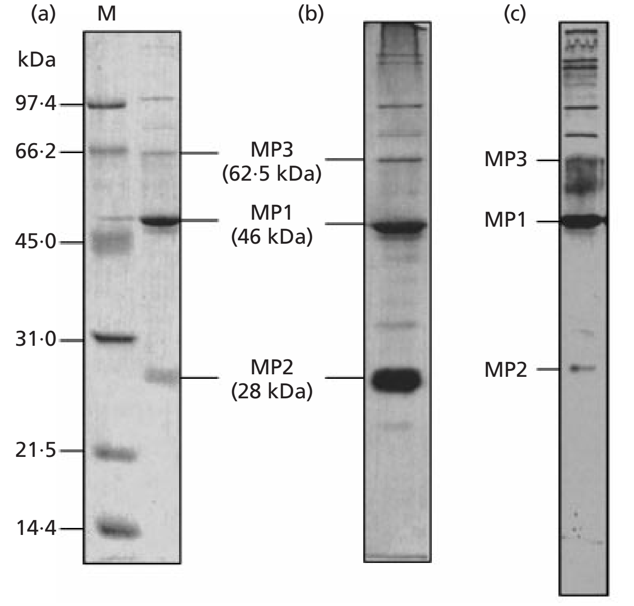

Fig. 4. (a, b) Phage B40-8 protein profiles analysed by SDS-PAGE and stained with Coomassie Brillant Blue (a) or $AgNO_3$ (b). Lane M is low-range molecular mass marker (Bio-Rad). (c) Western blotting of total phage proteins using an anti-B40-8 serum.

(Fig. 4). Immunogold electron microscopy using specific antibodies selected from total anti-B40-8 serum showed the location of the major proteins on B40-8 particles. Purified anti-MP1 and anti-MP3 antibodies specifically react with phage heads (Fig. 5b and Fig. 5d, respectively), MP1 appearing as the major constituent. Anti-MP2 antibodies attached to the phage tail showed low colloidal-gold labelling on this structure (Fig. 5c). Western blotting studies were necessary to verify the immunoreactivity of less concentrated proteins and to reveal a direct correlation between the antibody reactivity and the relative amount of each major protein in the gels (see Fig. 4).

Location of ORFs of major proteins

The 20 amino acid N-terminal sequence for each of the three structural peptides was compared with data in the SWISS-PROT database. A certain level of similarity was observed for MP2 and MP3 but not for MP1. A similarity of 47 % was found between the MP2 N-terminal sequence and structural peptide P7 of *Pseudomonas phaseolicola* phage ϕ6. The MP3 N-terminal sequence shared 42·1 % similarity with a replicative Vg15 protein belonging to *E coli* phage ϕ80.

Degenerate oligonucleotides (underlined in Fig. 6) were

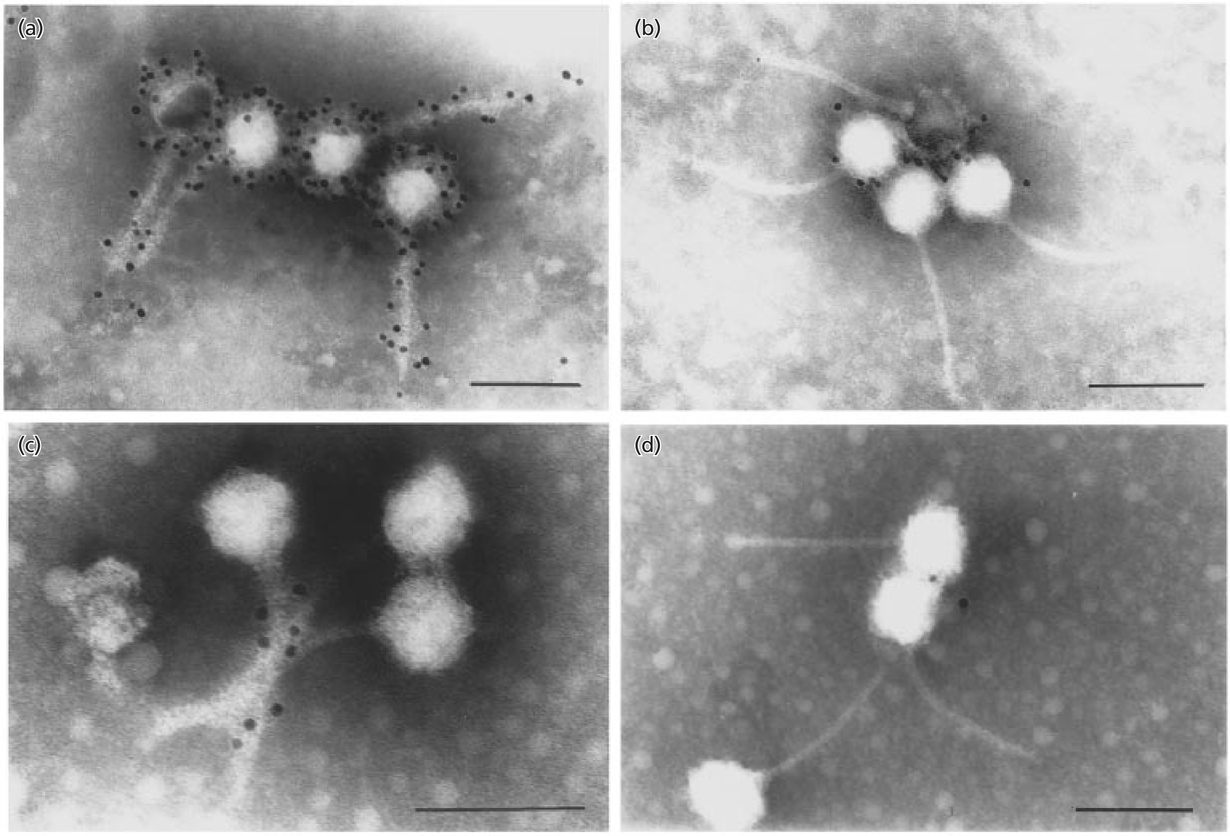

Fig. 5. Immunogold electron microscopy of B40-8 virions. Phage were probed with control anti-B40-8 serum (a), and with specific anti-MP1 (b), anti-MP2 (c) and anti-MP3 (d) antibodies. All the samples were labelled with gold-conjugated protein A. Negative stain with potassium phosphotungstate (pH 6·5). Bars, 100 nm.

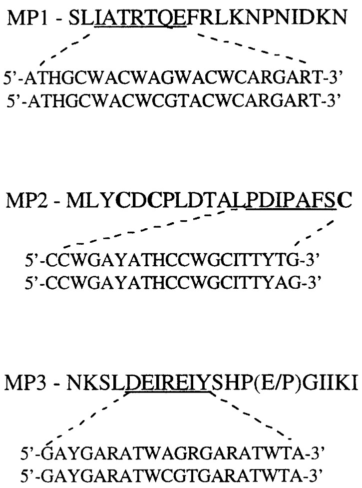

Fig. 6. Limited N-terminal amino acid sequence of major proteins determined by Edman degradation. Degenerate probe sequences derived from seven of the protein residues (underlined) are also indicated. The three Cs shown in bold in the MP2 sequence were obtained by translation of the nucleotide sequence.

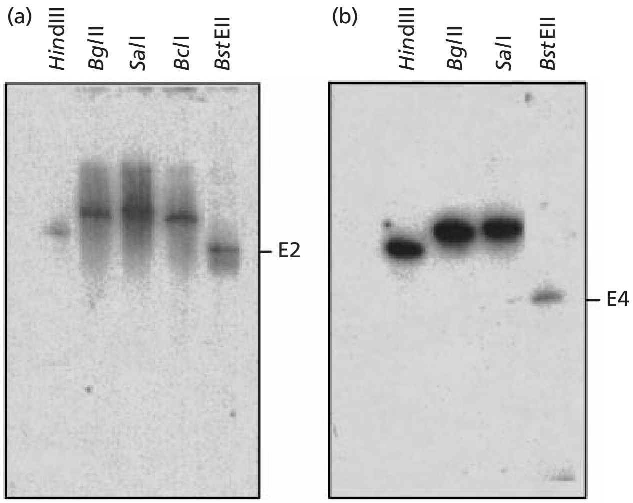

Fig. 7. Southern blotting of mapped B40-8 DNA restriction fragments with MP1 (a), and with MP2 and MP3 (b) degenerate probes (the latter two probes gave identical hybridization patterns).

designed based on N-terminal sequence data for the major proteins and high codon frequency tables, and were used as probes to localize the ORF of each protein in the genome. Southern blotting hybridization analyses were conducted against (i) total DNA digested with different restriction enzymes (Fig. 7), and (ii) fragment H1 digested with *Sau*3AI using degenerate probes, at high stringency (data not shown). MP1 ORF seemed to be located in the H1, B1, S1, C1 and E2 regions (Fig. 7a) while MP2 and MP3 degenerate probes gave identical hybridization patterns: H1, B1, S1 C1 and E4 (according to the genomic restriction map, Fig. 7b).

Characterization of MP2

The MP2 ORF was further characterized by sequencing both strands of the insert contained in the clone pBSKSa-22. The 850 nt sequence obtained (GenBank AF074719) showed one of the 20 base combinations of degenerate oligomer mixture used as a probe, corresponding to the codon reading for residues PDIPAFS of the MP2 N-terminal sequence (positions 158–177 in the sequence of the insert). Translation of the entire sequence resulted in a 194 amino acid polypeptide, considering as start codon an ATG triplet located 14 amino acids upstream of the first methionine obtained by protein sequencing. A consensus ribosome-binding site sequence (RBS) (GAGG; Watson *et al.*, 1987) was located 9 nt upstream of the translation start ATG.

The low level of similarity found between the N-terminal protein sequence of MP2 and structural protein P7 of the phage ϕ6 (according to searches in the SWISS-PROT database) was not improved when the entire nucleotide sequence of the insert or the MP2 DNA/protein sequences were analysed by the same process (GenBank and SWISS-PROT database, respectively). Nor were similarities with other sequences found.

A *Bst*EII restriction site was located at position 739–745 of the MP2 nucleotide sequence and was identified as the cleavage site between fragments E2 and E4 (see Fig. 2) by Southern blotting with specific probes flanking the digestion point. The restriction genetic map of B40-8 has been orientated 5′–3′ according to the positive sense of the structural MP2 ORF.

DISCUSSION

Phage B40-8 was previously described as a member of the *Siphoviridae* family (Tartera, 1986). It shows the most common morphology of *B. fragilis* phages isolated from environmental samples (Tartera & Jofre, 1987; Booth *et al.*, 1979; Lasobras *et al.*, 1997). Image analysis of multiple TEM micrographs gives more accurate measures of the phage structures than previous descriptions. AFM data from the phage capsid and tail showed significant deviation from the TEM data images, which is due to the convolution effect generated by AFM tapping mode. This phenomenon has been reported in other studies on phage fibres, heads and tails (Ikai *et al.*, 1993), and DNA (Bustamante *et al.*, 1994; Thundat *et al.*, 1994). Globular structures were also observed at the terminal part of the phage tail. These structures were probably generated by the physical pressure of the microscopy tip on the sample, producing some degree of degradation at the distal section of the tails.

The size and composition of phage DNA, double stranded, 51·7 kb in length and with a G+C content of 38·9 mol%, are similar to the values calculated by Kory & Booth (1986) for three different isolates of *B. fragilis* phages. The paucity of cleavage sites of the phage DNA may be the reason for the bias of *B. fragilis* phage DNAs against cleavage by a number of restriction endonucleases. In accordance with Powell & Davidson (1986) we suggest that this effect is an evolutionary response of these phages to restriction–modification systems of their bacterial hosts. The similarity found between the G + C content of phage DNA (38·9 mol%) and host DNA (41 mol%; Johnson, 1978) is further evidence of the evolutionary pressure acting on phage–bacteria systems.

Similar restriction patterns were observed when the B40-8 genome was compared with other similar-sized *B. fragilis* phage DNA (Kory & Booth, 1986), suggesting a high level of DNA homology between isolates of the same host. Additional DNA–DNA similarity and sequencing experiments are at present under way in order to test this hypothesis.

A physical map of the B40-8 genome has been established, containing positions for six different restriction endonuclease cutting sites. Unique cleavage sequences for *Sal*I and *Stu*I are candidate regions for use in polylinkers of phage-based vector constructions. To date, genetic studies on *Bacteroides* species have been conducted using phage λ-derived vectors and constructions based on plasmid DNA isolated from *Bacteroides* species (Anderson *et al.*, 1984; Guthrie *et al.*, 1985; Russo *et al.*, 1990; Ono *et al.*, 1994).

The presence of phage DNA restriction digests of submolar quantities of overlapping fragments provides a strong indication that packaging of B40-8 DNA occurs on a precursor concatemeric molecule according to a headful mechanism (Jackson *et al.*, 1978). This type of pattern has been previously described as a characteristic for genomes with terminal redundancies belonging to phages using this model of encapsidation (Jackson *et al.*, 1978; Black, 1989; Alatossava & Klaenhammer, 1991; Hahn *et al.*, 1991; Klieve *et al.*, 1991; Inal *et al.*, 1996, Johnsen *et al.*, 1996). The submolar fragments in the profile are defined as those in which the *pac* site is located (Johnsen *et al.*, 1996). We could hypothesize that B40-8 DNA packaging seems to start with a higher frequency at a site-specific cleavage point (*pac*) (generating submolar fragments H6′, B3, S2, T2, C4 and E5 in the respective endonuclease digestion patterns) and less frequently in at least two additional *pac* sites (adjacent to one end of the fragments that appear as weak bands in *Hin*dIII and *Bst*EII digestions). Genome structure is circularly permuted and shows a terminal redundancy of 3·8 kb, representing 7·3% of the total molecule. Little information is available on other phages infecting *Bacteroides* species. A circularly permuted genome with redundant ends has been described for the only mapped *Bacteroides* phage genome reported in the literature, ϕBrb01 of *B. ruminicola* (Klieve *et al.*, 1991).

Three major proteins out of a pattern of 15 bands have been further characterized. Hybridizations with degenerate oligonucleotides suggest that all three proteins are coded in a 15 kb *Hin*dIII fragment: the ORF for MP1 could be located on fragment E2 while the ORFs for MP2 and MP3 appeared to be on fragment E4. However, these data have been proved by sequence analysis only for MP2. This result may indicate a putative cluster of structural genes similar to the genetic grouping shown by other phages (Campbell, 1996). The clone containing the tail major protein gene, *mp2*, was studied in more detail by DNA sequencing, identifying the entire ORF, and some genetically important elements in the preceding region. The gene encodes a 208 amino acid sequence (23·1 kDa protein). Both the presence of the 20 N-terminal residues and the nucleotide combination of the degenerate probe designed were confirmed. The putative RBS sequence for this gene at 47 nucleotide positions downstream of the transcription origin shows a consensus motif (GAGG) common in bacteria and similar to that reported for *cepA* gene of *Bacteroides fragilis* (Rogers *et al.*, 1994). This clone could be used in future studies of the expression of the MP2 protein in *B. fragilis*.

This work is the first characterization at a molecular level of a phage of *B. fragilis*. The data reported here may be of interest in the field of genetic engineering of both the phage and the host bacteria for future studies and also for the development of molecular tests for the detection of *B. fragilis* phages in the environment.

ACKNOWLEDGEMENTS

This study was partially supported by research grant GRQ 94-1073 and 1995 SGR 00415, Generalitat de Catalunya. Montserrat Puig was a fellow of the Catalan Government. We thank Dr Joan Jofre for scientific discussions and support, and Dr Mads G. Johnsen for technical and scientific advice. We also thank Serveis Científico-tècnics (Universitat de Barcelona) and Dr Mª José Prieto for their help in TEM and AFM, and image analysis.

REFERENCES

Alatossava, T. & Klaenhammer, T. R. (1991). Molecular characterisation of three small isometric-headed bacteriophages which vary in their sensitivity to the lactococcal phage resistance plasmid pTR2030. *Appl Environ Microbiol* **7**, 1346–1353.

Anderson, B. J., Bills, M. M., Egerton, J. R. & Mattick, J. S. (1984). Cloning and expression in *Escherichia coli* of the gene encoding the structural subunit of *Bacteroides nodosus* fimbriae. *J Bacteriol* **60**, 748–754.

Black, L. W. (1989). DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol* **3**, 267–292.

Booth, S. J., Van Tassell, R. L., Johnson, J. L. & Wilkins, T. D. (1979). Bacteriophages of *Bacteroides*. *Rev Infect Dis* **1**, 325–334.

Bustamante, C., Erie, D. A. & Keller, D. (1994). Biochemical and structural applications of scanning force microscopy. *Curr Opin Struct Biol* **4**, 750–760.

Campbell, A. M. (1996). Bacteriophages. In *Fundamental Virology*, 3rd edn, pp. 455–475. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott-Raven.

Chu, G., Vollrath, D. & Davis, R. W. (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **34**, 1582–1585.

Eggerth, A. H. & Gagnon, B. H. (1933). The *Bacteroides* of human feces. *J Bacteriol* **5**, 389–413.

Genetics Computer Group (1991). *Program Manual for the GCG Package*, version 7.0, April 1991. Madison: University of Wisconsin.

Guthrie, E. P., Shoemaker, N. B. & Salyers, A. (1985). Cloning and expression in *Escherichia coli* of a gene coding for a chondroitin lyase from *Bacteroides thetaiotaomicron*. *J Bacteriol* **64**, 510–515.

Hahn, D. R., McHenney, M. A. & Baltz, R. H. (1991). Properties of the streptomycete temperate bacteriophage FP43. *J Bacteriol* **73**, 3770–3775.

Harlow, E. & Lane, D. (1988). Purifying antibodies from immunoblots. In *Antibodies: a Laboratory Manual*, pp. 471–510. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Havelaar, A. H. (1993). Bacteriophages as models of human enteric viruses in the environment. *ASM News* **9**, 614–619.

Hayat, M. A. (1989). *Colloidal Gold. Principles, Methods and Applications*, vol. 2, pp. 243–284. New York: Academic Press.

Holdeman, L. V., Good, I. J. & Moore, W. E. C. (1976). Human fecal flora: variation in bacterial composition within individuals and a possible effect of emotional stress. *Appl Environ Microbiol* **1**, 359–375.

Holland, J. W., Hill, E. O. & Altemeier, W. A. (1977). Numbers and types of anaerobic bacteria isolated from clinical specimens since 1960. *J Clin Microbiol* **5**, 20–25.

IAWPRC Study Group on Health Related Water Microbiology (1991). Bacteriophages as model viruses in water quality control. *Water Res* **5**, 529–545.

Inal, J. M., Karunakaran, V. & Jones, D. R. (1996). *Bacillus thuringensis* subsp. *aizawai* generalized transducing phage *ϕ*HD248: restriction site map and potential for fine-structure chromosomal mapping. *Microbiology* **42**, 1409–1416.

Ikai, A., Yoshimura, K., Arisaka, F., Ritani, A. & Imai, K. (1993). Atomic force microscopy of bacteriophage T4 and its tube–baseplate complex. *FEBS Lett* **26**, 39–41.

Jackson, E. N., Jackson, D. A. & Deans, R. J. (1978). *Eco*RI analysis of bacteriophage P22 DNA packaging. *J Mol Biol* **18**, 365–388.

Johnsen, M. G., Appel, K. F., Madsen, P. L., Vogensen, F. K., Hammer, K. & Arnau, J. (1996). A genomic region of lactococcal temperate bacteriophage TP901-1 encoding major virion proteins. *Virology* **18**, 306–315.

Johnson, J. L. (1978). Taxonomy of the *Bacteroides*. I. DNA homologies among *Bacteroides fragilis* and other saccharolytic *Bacteroides* species. *Int J Syst Bacteriol* **8**, 245–256.

Kai, M., Watanabe, S., Furuse, K. & Ozawa, A. (1985). *Bacteroides* bacteriophages isolated from human feces. *Microbiol Immunol* **9**, 895–899.

Keller, R. & Traub, N. (1974). The characterisation of *Bacteroides fragilis* bacteriophages recovered from animal sera: observations on the nature of *Bacteroides* phage carrier cultures. *J Gen Virol* **4**, 179–189.

Klieve, A. V., Gregg, K. & Bauchop, T. (1991). Isolation and characterisation of lytic phages from *Bacteroides ruminicola* ss *brevis*. *Curr Microbiol* **3**, 183–187.

Kory, M. M. & Booth, S. J. (1986). Characteristics of *Bacteroides fragilis* bacteriophages and comparison of their DNAs. *Current Microbiol* **4**, 199–203.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **27**, 680–685.

Lasobras, J., Muniesa, M., Frías, J., Lucena, F. & Jofre, J. (1997). Relationship between the morphology of bacteriophages and their persistence in the environment. *Water Sci Technol* **5**, 129–132.

Ono, T., Akimoto, S., Kinouchi, T., Kataoka, K. & Ohnishi, Y. (1994). Cloning and expression of the *Bacteroides fragilis* YCH46 neuraminidase gene in *Escherichia coli* and *Bacteroides uniformis*. *FEMS Microbiol Lett* **21**, 153–158.

Polk, B. F. & Kasper, P. L. (1977). *Bacteroides fragilis* subspecies in clinical isolates. *Ann Intern Med* **6**, 569–571.

Powell, I. B. & Davidson, B. E. (1986). Resistance to *in vitro* restriction of DNA from lactic streptococcal bacteriophage c6A. *Appl Environ Microbiol* **1**, 1358–1360.

Puig, A., Jofre, J. & Araujo, R. (1997). Bacteriophages infecting various *Bacteroides fragilis* strains differ in their capacity to distinguish human from animal faecal pollution. *Water Sci Technol* **5**, 11–12.

Rashtchian, A. & Booth, J. (1981). Stability in *Escherichia coli* of an antibiotic resistance plasmid from *Bacteroides fragilis*. *J Bacteriol* **46**, 121–127.

Rasmussen, B. A. & Kovacs, E. (1993). Cloning and identification of a two-component signal-transducing regulatory system from *Bacteroides fragilis*. *Mol Microbiol* **7**, 765–776.

Rasmussen, B. A., Gluzman, Y. & Tally, F. P. (1990). Cloning and sequencing of the class B *β*-lactamase gene (*ccr*A) from *Bacteroides fragilis* TAL3636. *Antimicrob Agents Chemother* **4**, 1590–1592.

Rogers, M. B., Bennett, T. K., Payne, C. M. & Smith, C. J. (1994). Insertional activation of *cep*A leads to high-level *β*-lactamase expression in *Bacteroides fragilis* clinical isolates. *J Bacteriol* **76**, 4376–4384.

Russo, T. A., Thompson, J. S., Godoy, V. G. & Malamy, M. H. (1990). Cloning and expression of the *Bacteroides fragilis* TAL2480 neuraminidase gene, *nanH*, in *Escherichia coli*. *J Bacteriol* **72**, 2594–2600.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Smith, C. J. (1985). Development and use of cloning systems for *Bacteroides fragilis*: cloning of a plasmid-encoded clindamycin resistance determinant. *J Bacteriol* **64**, 294–301.

Tamaoka, J. & Komagata, K. (1984). Determination of DNA base composition by reversed-phase high-performance liquid cromatography. *FEMS Microbiol Lett* **5**, 125–128.

Tartera, C. (1986). *Estudi dels bacteriòfags de Bacteroides fragilis com a indicadors de contaminació fecal vírica*. Tesi doctoral, Universitat de Barcelona.

Tartera, C. & Jofre, J. (1987). Bacteriophages active against *Bacteroides fragilis* in sewage-polluted waters. *Appl Environ Microbiol* **3**, 1632–1637.

Tartera, C., Araujo, R., Michel, T. & Jofre, J. (1992). Culture and decontamination methods affecting enumeration of phages infecting *Bacteroides fragilis* in sewage. *Appl Environ Microbiol* **8**, 2670–2673.

Thundat, T., Allison, D. P. & Warmack, R. J. (1994). Stretched DNA structures observed with atomic force microscopy. *Nucleic Acids Res* **2**, 4224–4228.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Argetsinger Steitz, J. & Weiner, A. M. (editors) (1987). *Molecular Biology of the Gene*, vol. 1, pp. 409–411. Menlo Park, CA: Benjamin/Cummings Publishing Co.

Received 17 November 1998; revised 2 March 1999; accepted 24 March 1999.