Complete nucleotide sequence of the gene encoding bacteriophage E endosialidase: implications for K1E endosialidase structure and function

Graham S. LONG,*‡ J. Mark BRYANT,* Peter W. TAYLOR† and J. Paul LUZIO*§

*University of Cambridge, Department of Clinical Biochemistry, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QR, U.K., and †Research Centre, Ciba Pharmaceuticals, Wimblehurst Road, Horsham, West Sussex RH12 4AB, U.K.

Bacteriophage E specifically recognizes and infects strains of *Escherichia coli* which display the α-2,8-linked polysialic acid K1 capsule. Bacteriophage E endosialidase, which is thought to be responsible for initial absorption of the phage to the host bacterium, was purified, and the N-terminal amino acid sequences of the polypeptide monomer and cyanogen bromide fragments were determined. Synthetic oligonucleotide probes were designed from the N-terminal amino acid sequences and used to identify restriction fragments of bacteriophage E DNA encoding the endosialidase. The primary nucleotide sequence of the bacteriophage E endosialidase gene contains an open reading frame encoding a 90 kDa polypeptide which is processed to give a mature 74 kDa protein. The native enzyme is probably a trimer of identical 74 kDa subunits. In the bacteriophage E genome the K1E endosialidase open reading frame is preceded by a putative upstream promoter region with homology to a bacteriophage SP6 promoter. A central region of 500 amino acids of the deduced protein sequence of the K1E endosialidase was found to have 84% identity to K1F endosialidase. Both endosialidases contain two copies of a sialidase sequence motif common to many bacterial and viral sialidases. These sequences flank the region of greatest identity between the two endosialidase forms, which suggests that this central domain is involved in binding and hydrolysis of the polysialic acid substrate.

INTRODUCTION

Bacteriophage E is a member of the PK1A–PK1E family of phages; these phages were isolated originally from European sewage to aid in the clinical identification of *Escherichia coli* K1 infections [1], which can result in high mortality rates in cases of neonatal meningitis [2–4]. Bacteriophage K1E endosialidase is thought to be the protein responsible for initial binding to host bacteria by specifically recognizing and hydrolysing the α-2,8-linked poly-*N*-acetylneuraminic acid (polysialic acid; PSA) carbohydrate polymers of the K1 glycocalyx [5]. α-2,8-linked PSA is also expressed on the cell surfaces of several other pathogenic bacteria [3,6] and of various tumour cells and cell lines [7–12]. It has been proposed that K1E endosialidase could be used in the diagnosis and therapy of K1 meningitis, septicaemia or bacteraemia due to the enzyme's high specificity for hydrolysing α-2,8-sialosyl linkages [13]. PSA has been suggested as an oncodevelopmental marker in tumours of the human kidney and neuroendocrine tissues [8,10], and also may contribute to the invasive and metastatic potential of some tumours [7–9,14].

There have been several reports of endosialidases associated with K1 coliphages and which are specific for α-2,8-polysialosyl linkages [15–18]. Petter and Vimr [19] have described the primary nucleotide sequence and deduced amino acid structure for endosialidase from bacteriophage K1F, a related strain isolated from the North American continent. The purification and preliminary characterization of bacteriophage E endosialidase was documented by Tomlinson and Taylor [5]. In the present study we report the sequencing of the gene and the deduced amino acid structure for K1E endosialidase.

EXPERIMENTAL

Materials

All reagents were purchased from Sigma Chemical Co., Poole, Dorset, U.K., unless stated otherwise. Restriction enzymes and DNA-modifying enzymes were purchased mainly from New England Biolabs Inc., Beverly, MA, U.S.A., and Northumbrian Biologicals Ltd., Cramlington, Northumbria, U.K.

Bacteriophage E propagation and assay

Stocks of K1-specific bacteriophage E (A.T.C.C. no. 40221) were propagated using *E. coli* (LP1674/A.T.C.C. no. 53351; serotype O7:K1) as host, essentially as described previously [5,20] using standard top agar overlay techniques [21]. Five 9 cm diameter culture plates typically yielded approx. 50 ml of stock phage preparation with a titre of $\sim 10^{10}$–10^{11} plaque-forming units (p.f.u.)·ml^{-1}. Liquid cultures of *E. coli* were used to propagate bacteriophage E for endosialidase preparation. As a method for stock phage production this was less efficient than the top agar

Abbreviations used: CBB, Coomassie Brilliant Blue; DTT, dithiothreitol; GST, glutathione S-transferase; IPTG, isopropyl β-D-thiogalactopyranoside; PSA, polysialic acid (poly-*N*-acetylneuraminic acid); PVDF, polyvinylidene difluoride; TBA, thiobarbituric acid; X-Gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside.

‡ Present address: University of California San Diego, UCSD Cancer Center, School of Medicine, 9500 Gilman Drive, MC-0063 La Jolla, CA 92093, U.S.A.

§ To whom correspondence should be addressed.

The nucleotide sequence data reported have appeared in the EMBL and GenBank Nucleotide Sequence Databases under the accession numbers X78310 and K1EPGP90.

technique, and yielded approx. 10 ml of phage preparation with a titre of $\sim 10^9$–10^{10} p.f.u. $\cdot$ ml^{-1} from 2 litres of culture.

Purification of bacteriophage E DNA and K1E endosialidase

Phage suspensions were combined with caesium chloride to a final concentration of 1.37 g $\cdot$ ml^{-1} in 38 ml capacity polyallomer centrifuge tubes (Beckman Instruments Inc., Spinco Division, Palo Alto, CA, U.S.A.) and centrifuged for 16 h at 206000 *g* and 4 °C using a Beckman Vti50 rotor and a Beckman LB-50 ultracentrifuge. Bacteriophage E particles equilibrated in a single band equivalent to a density of 1.47 g $\cdot$ ml^{-1}. Ultracentrifugation was repeated twice, the recovered phage suspension was dialysed against 25 mM Tris/HCl (pH 7.5) to remove CsCl, and stocks were frozen in liquid nitrogen. Phage purity was assessed by SDS/PAGE under reducing conditions, which yielded 11 protein bands as observed previously [5].

Bacteriophage E DNA was purified essentially according to a technique commonly used for λ phage [21]. K1E endosialidase was purified as previously described [5,20].

Analytical assays

K1E endosialidase was assayed by measuring the release of AcNeu from K1 polymer purified from liquid cultures of *E. coli* K1 LP1674 by an adaptation of a phenol/water extraction method for lipo- and acidic polysaccharides [22]. Release of AcNeu was quantified using the thiobarbituric acid (TBA) assay of Horgan [23], a modification of the Aminoff [24] assay.

Aliquots of enzyme preparations (100 μl) were added to 100 μl of freshly prepared K1 substrate (4 mg $\cdot$ ml^{-1} in 50 mM acetate buffer, pH 6.5) and incubated at 37 °C for 60 min in a shaking water bath. The enzyme was inactivated by the addition of 200 μl of 25 mM periodic acid/0.125 M HCl and incubated at 37 °C for 30 min. Excess periodate was reduced by the addition of 160 μl of 2 % (w/v) sodium arsenite solution in 0.5 M HCl. Then 1.6 ml of 0.1 M TBA (adjusted to pH 9.0 with 1.0 M NaOH) was added, and the samples were mixed and heated at 100 °C for 10 min. The pink TBA–AcNeu chromophore was extracted into 1.5 ml of acidified n-butanol [5 % (v/v) conc. HCl], and the absorbance was measured at 549 nm against a blank containing no AcNeu. Readings were corrected for a substrate blank containing no enzyme, and converted to equivalent amounts of sialic acid release by using pure AcNeu as standard. One unit of endosialidase activity was defined as that amount of enzyme which released 1 μmol of sialic acid equivalent from LP1674 K1 polymer in 1 min at 37 °C and pH 6.5.

Protein was determined by using the Pierce bicinchoninic acid microtitre plate assay with BSA as standard (purchased from Pierce, Rockford, IL, U.S.A.).

Amino acid sequence analysis

N-terminal microsequencing of proteins was conducted by the Microchemicals Division at the Agriculture and Food Research Council (A.F.R.C.) Research Institute, Babraham, U.K. Purified K1E endosialidase was electrophoresed using the SDS/PAGE system of Laemmli [25] and electrophoretically transferred on to a ProBlott polyvinylidene difluoride (PVDF) membrane (Applied Biosystems Ltd., Warrington, Cheshire, U.K.). CNBr digestion of purified K1E endosialidase was achieved by mixing 250 μg (in 250 μl of 50 mM Tris/HCl, pH 7.5) with 25 μl of freshly prepared 10 mg $\cdot$ ml^{-1} CNBr in 70 % (v/v) formic acid and incubating for 3 h at 37 °C in the dark. A further 50 μl of CNBr solution was added and the mixture was incubated for a further 3 h at 37 °C. The reaction was terminated by the addition of 3 ml of distilled water, and the digestion mixture was lyophilized. Samples were run on 12.5 %-acrylamide SDS/PAGE gels, peptides were transferred on to ProBlott PVDF membrane and bands were visualized by staining with Coomassie Brilliant Blue (CBB; R-250; Bio-Rad Laboratories Ltd., Hemel Hempstead, Herts., U.K.) and de-staining with 50 % (v/v) methanol. CBB-stained bands were excised from the ProBlott PVDF membrane and subjected to automated Edman degradation on an Applied Biosystems protein sequencer using an on-line phenylthiohydantoin amino acid analyser.

Cloning of bacteriophage E endosialidase

Unless otherwise stated, all procedures used were as described by Sambrook et al. [21].

Degenerate oligonucleotide probes were designed with reference to *E. coli* codon usage tables [26], prepared using an automated Applied Biosystems PCR-MATE model 391 DNA synthesizer and 5′ end-labelled with [γ-^{32}P]ATP (Amersham International, Amersham, Bucks., U.K.) using T4 polynucleotide kinase. The radiolabelled oligonucleotide probes were hybridized to restriction enzyme digests of bacteriophage E DNA, electrophoresed in agarose gels and transferred to a Hybond-N nylon membrane (Amersham International). Bacteriophage E DNA fragments reacting with the probes were identified by autoradiography, purified from NA grade agarose gels (Pharmacia Biosystems Ltd., Milton Keynes, Bucks., U.K.) and ligated into Bluescript SK$^+$ (Stratagene Inc., La Jolla, CA, U.S.A.) using T4 DNA ligase (NEB Inc.). Transformations of *E. coli* Epicurian SURE cells (Stratagene Inc.) with Bluescript SK$^+$ were conducted by an electroporation method [27] using a Bio-Rad Gene Pulser and Pulse Controller, or alternatively high-efficiency *E. coli* JM109 competent cells (Promega Inc., Madison, WI, U.S.A.) were transformed by heat shock at 42 °C for 60 s [28]. Clones transformed with recombinant plasmid were identified by growing on 2TY (appendix A.3 in [21])/ampicillin agar plates and using a mixture of 0.1 M 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 0.1 M isopropyl β-D-thiogalactopyranoside (IPTG) to allow blue–white colour selection of colonies. Double-stranded DNA sequencing was conducted using the Sequenase Version 2.0 sequencing kit from United States Biochemical Corporation (Cleveland, OH, U.S.A.) and a model SA sequencing apparatus from BRL Life Technologies Inc. (Gaithersburg, MD, U.S.A.). Sequencing was facilitated by the technique of nested deletions [21] or by using synthetic oligonucleotide primers prepared by British Bio-technology Products Ltd. (Abingdon, Oxon., U.K.), or as above.

A degenerate oligonucleotide probe, Probe 4 [5′-TAC-(T)CAC(T)CAGGGT(G)GAC(T)GTG(T)GCG(C)CC-3′], was derived from the CNBr fragment of K1E endosialidase with the longest unambiguous amino acid sequence, and was the least degenerate of five probes which were designed using the partial amino acid sequences obtained from the CNBr fragments. A 1.9 kb *Bgl*II restriction digest fragment of genomic bacteriophage E DNA was identified as potentially encoding endosialidase sequence by Southern blot analysis using ^{32}P-radiolabelled Probe 4. *Bgl*II and *Bam*HI restriction endonucleases generate cohesive protruding ends with the same sequence, and this enabled the ligation of the 1.9 kb *Bgl*II fragment into the *Bam*HI site of the Bluescript SK$^+$ cloning vector (Promega Inc.). Plasmid miniprep DNA from a clone transformed with the resultant recombinant vector (K2) yielded a DNA sequence which encoded a deduced protein sequence containing a stretch of sequence identical to that of the CNBr fragment used to design Probe 4.

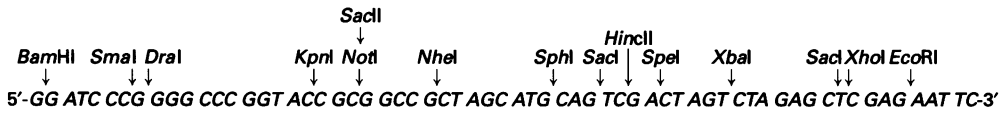

Figure 1 Polylinker used in the construction of the pGEX-3PLX vector

The open reading frame which is translated is indicated by the codon triplets shown, and the restriction enzyme sites are shown.

Probe 6 (5′-GATCTTGGTCTAATCCCT-3′), a non-degenerate oligonucleotide 18-mer, was synthesized using the sequence at the 5′ end of clone K2. This probe identified one of two *Sin*I digest fragments of genomic bacteriophage E DNA (see Figure 3, band b) which ran as a singlet equivalent to about 3.3 kb. It was possible to verify that this fragment coded for DNA sequence upstream of the 5′ end of clone K2 by digesting the K2 insert DNA with *Sin*I. The result of this digest showed that there were at least three *Sin*I sites in the K2 insert DNA, the largest fragment being 1.1 kb in size. Attempts to clone the *Sin*I fragments by filling in the ends using the Klenow fragment of T4 DNA polymerase followed by blunt-ended ligation proved unsuccessful. An alternative approach was to ligate with one end sticky and the other end blunt. Since restriction analysis of bacteriophage E DNA showed that there were only two *Bgl*II sites in the whole genome, the gel-purified *Sin*I fragments were digested with *Bgl*II, and the fragment containing the Probe 6 recognition sequence and the *Bgl*II site yielded two fragments of 2.1 kb and 1.1 kb. The 2.1 kb *Sin*I × *Bgl*II digest fragment was cloned into Bluescript SK$^+$ by ligation of the *Bgl*II end to a *Bam*HI end, followed by end-filling using the Klenow fragment of T4 DNA polymerase and ligating the resultant blunt ends together to circularize the plasmid. The resultant clone Y10 was found to contain an open reading frame encoding the N-terminus of K1E endosialidase by comparison with the N-terminal amino acid sequence of the ~ 76 kDa enzyme subunit. Overlapping sequence was obtained for clones Y10 and K2 in both the 5′ and 3′ directions, and the positions of open reading frames were determined by codon preference and positional base preferences analysis [29,30].

In vitro translation of the N-terminal domain of K1E endosialidase

Recombinant plasmid DNA was purified from clone Y10 and linearized by cleavage of the unique *Eco*RI site, and 5′ capped RNA was transcribed using SP6 RNA polymerase and an mCAP mRNA capping kit (Stratagene Inc.). *In vitro* translation reactions (25 μl) using 0.1 μg of RNA transcript, 20 μCi of [^{35}S]methionine and a rabbit reticulocyte lysate system were carried out according to the manufacturer's instructions (Promega). Confirmation that the SP6 RNA polymerase and the *in vitro* translation system were functional was obtained by running a positive control alongside. The control plasmid was a linearized SV64–carboxypeptidase E construct with an upstream SP6 promoter region [31].

Construction and *in vitro* translation of a plasmid containing the complete gene encoding K1E endosialidase

A fragment of K1E DNA of 1892 bp containing the complete K2 insert was excised from clone K2 using *Eco*RI and *Xba*I. This was directionally cloned into the vector pGEM-11z (Promega) cut with the same restriction enzymes, thus placing a *Sac*I site 3′ of the K2 insert. A 707 bp *Sac*I/*Avr*II fragment was excised from this new construct. This 707 bp fragment encoded the predicted C-terminal 114 amino acids of the endosialidase and the 3′ untranslated region of K1E DNA. It was ligated into the 3253 bp product of a *Sac*I/*Avr*II digest of the Y10 construct. The resulting plasmid (Y10-707) contained only the extreme 5′ and 3′ regions of the originally cloned K1E DNA in a Bluescript SK$^+$ vector, effectively lacking the central 2975 bp of the predicted endosialidase open reading frame. A 2975 bp fragment derived from an *Avr*II digest of total K1E DNA was ligated into construct Y10-707 digested with *Avr*II. The resulting construct in Bluescript SK$^+$ (clone F15) contained the full-length endosialidase gene previously encoded in clones K2 and Y10, and the gene was sequenced using the Sequenase 2.0 sequencing kit (USB Corp.). Recombinant plasmid DNA from clone F15 was purified and translated in an S30 *E. coli* lysate system (Promega) according to the manufacturer's instructions. Translation products were analysed by SDS/PAGE followed by autoradiography.

Construction of a glutathione S-transferase (GST) fusion protein for the C-terminal region of K1E endosialidase

The insert of clone K2 encoding the C-terminus of K1E endosialidase was excised from the recombinant plasmid using *Eco*RI and *Not*I. This fragment was directionally subcloned into the same restriction sites of a modified pGEX-3X [32] vector (termed pGEX-3PLX, for polylinker insertion). pGEX-3X was obtained from Pharmacia Biosystems Ltd., and pGEX-3PLX was a gift from Dr. John C. Hutton, Department of Clinical Biochemistry, University of Cambridge. pGEX-3PLX encodes GST 5′-proximal to the polylinker shown in Figure 1, which was ligated into the original *Eco*RI site. The open reading frame which is translated is indicated by the codon triplets shown and the restriction enzyme recognition sites are indicated. Liquid cultures of transformants which possessed a vector insert of the correct size were grown to mid-exponential phase and induced to express the GST fusion protein by addition of 0.1 mM IPTG (final concentration) [32].

Sequence analysis and database searches

DNA and protein sequence comparisons with various databases were carried out using the FASTA [33] and Prosearch [34] programmes. Sequence manipulation and analysis was performed using the molecular biology software packages of Staden [29,30,35] and the Genetics Computer Group Ltd./GCG [36].

RESULTS

Purification and characterization of endosialidase

Bacteriophage K1E endosialidase was purified to apparent homogeneity in milligram quantities, as previously described [5,20]. Typically, 2.5 mg of enzyme with specific activity 1600 munits·mg^{-1} was obtained from 2 litres of bacterial culture, representing a yield of ~ 40% with 800-fold purification [20]. SDS/PAGE showed that the holoenzyme ran in 10% gels as a

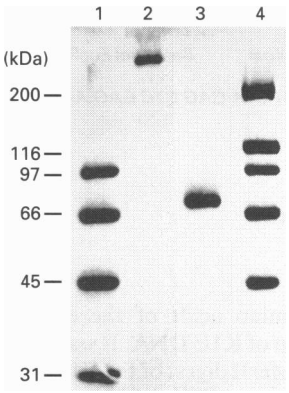

Figure 2 SDS/PAGE analysis of purified bacteriophage E endosialidase

Samples were electrophoresed in a 7.5–20% acrylamide gradient gel using a Laemmli buffer system [25]. All samples were heated at 100 °C for 5 min in the presence of DTT, except for one enzyme sample which was incubated at 37 °C for 5 min in the presence of DTT. Lane 1, low-molecular-mass markers; lanes 2 and 3, 5 μg of purified K1E endosialidase incubated at 37 °C and 100 °C respectively; lane 4, high-molecular-mass markers.

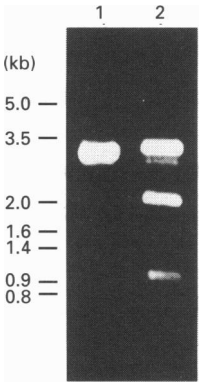

Figure 4 Identification of a DNA fragment encoding the N-terminus of bacteriophage E endosialidase

Lane 1, gel-purified 3.2 and 3.4 kb fragments from a *Sin*I digest of genomic bacteriophage E DNA which run as a single band; lane 2, *Bgl*II digest of both *Sin*I fragments, which allowed the resolution and purification of a 2.1 kb band for cloning. The DNA fragments were resolved by electrophoresis in a 0.7% agarose gel and visualized by staining with ethidium bromide. Molecular size markers are indicated.

single band of greater than 200 kDa if the sample was not pre-heated in the presence of SDS, but as a single band of ~ 76 kDa if the sample was pre-heated for 5 min at 100 °C (Figure 2). This pattern was not affected by reduction with dithiothreitol (DTT), and agrees with previous observations [5]. The additional 38 kDa band observed by Tomlinson and Taylor [5] was not seen in the purest preparations used in the present study. The K_m of the purified K1E endosialidase was 10.6 μM, assuming that the mean molecular mass of the K1 PSA polymer preparation was equivalent to an oligosaccharide of 175 units of sialic acid (~ 54 kDa).

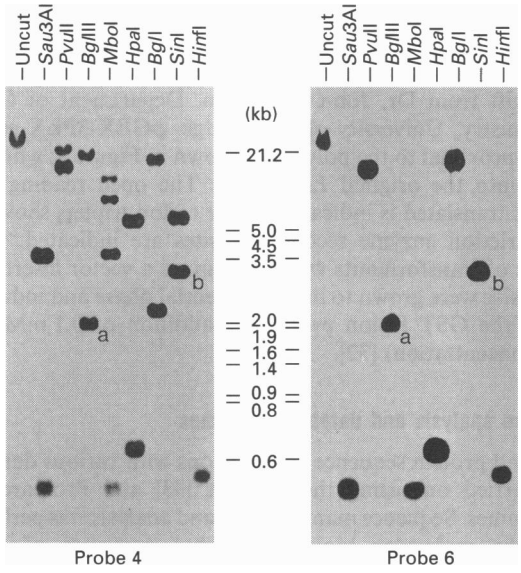

Figure 3 Southern blot analysis of restriction digests of genomic DNA from bacteriophage E

Samples were digested with various restriction enzymes as indicated and run in a 0.7% agarose gel. DNA fragments were transferred to a Hybond nylon membrane and probed with ^{32}P-labelled oligonucleotide probes. The same Southern blot membrane was probed first with Probe 4, and then with Probe 6 following regeneration. Owing to the degeneracy of Probe 4, it recognized two DNA fragments in each lane, whereas the non-degenerate Probe 6 recognized only one. Bands a and b represent a 1.9 kb *Bgl*II fragment and a 3.2 kb *Sin*I fragment respectively, used in the cloning of K1E endosialidase. The positions of molecular size markers are shown between the panels.

Cloning and sequencing of endosialidase

The N-terminal amino acid sequence was obtained for the purified 76 kDa enzyme subunit, but it was apparent that there was substantial blockage of the N-terminus (> 95%). There were clear assignments for the first six residues (Met-Ile-Gln-Arg-Leu-Gly), after which there were several ambiguities in the sequence obtained. Therefore sequences for three CNBr digest fragments of the enzyme subunit were determined.

The least degenerate oligonucleotide probe (Probe 4) designed from the CNBr fragment sequences was used in Southern blot analysis of restriction digests of genomic bacteriophage E DNA (Figure 3). A 1.9 kb *Bgl*II fragment identified with this probe was subcloned into Bluescript SK$^+$ to give a recombinant (K2) which, on sequencing, was found to encode the C-terminus of K1E endosialidase. A non-degenerate oligonucleotide probe (Probe 6) was prepared using the sequence at the 5′ end of the K2 insert. This probe was used to identify a 3.3 kb *Sin*I fragment of bacteriophage E DNA (Figure 3) and subsequently a 2.1 kb *Sin*I × *Bgl*II fragment (Figure 4) which, on subcloning into Bluescript SK$^+$ (to give the recombinant Y10) and sequencing, was found to encode the N-terminus of K1E endosialidase. The positions within the complete deduced K1E endosialidase open reading frame of the obtained sequences of the CNBr fragments were determined (Figure 5). A full-length construct, F15, was prepared and sequenced across the *Bgl*II site, confirming that the Y10 and K2 clones did contain two adjoining fragments which together comprised the entire endosialidase open reading frame. Sequence comparisons with various databases showed similarities within the deduced open reading frame to K1F endosialidase [19] (K1Fenan; Figure 6), and to several bacterial sialidases over the sialidase sequence motif/'Asp-box' regions ([37,38]; Figure 5). In addition, a 58-amino-acid-region of the N-terminal domain of K1E endosialidase showed 36% identity with residues 87–144 of pre-neck appendage protein (late protein GP12) of bacteriophage 29, a coliphage which also possesses host capsule

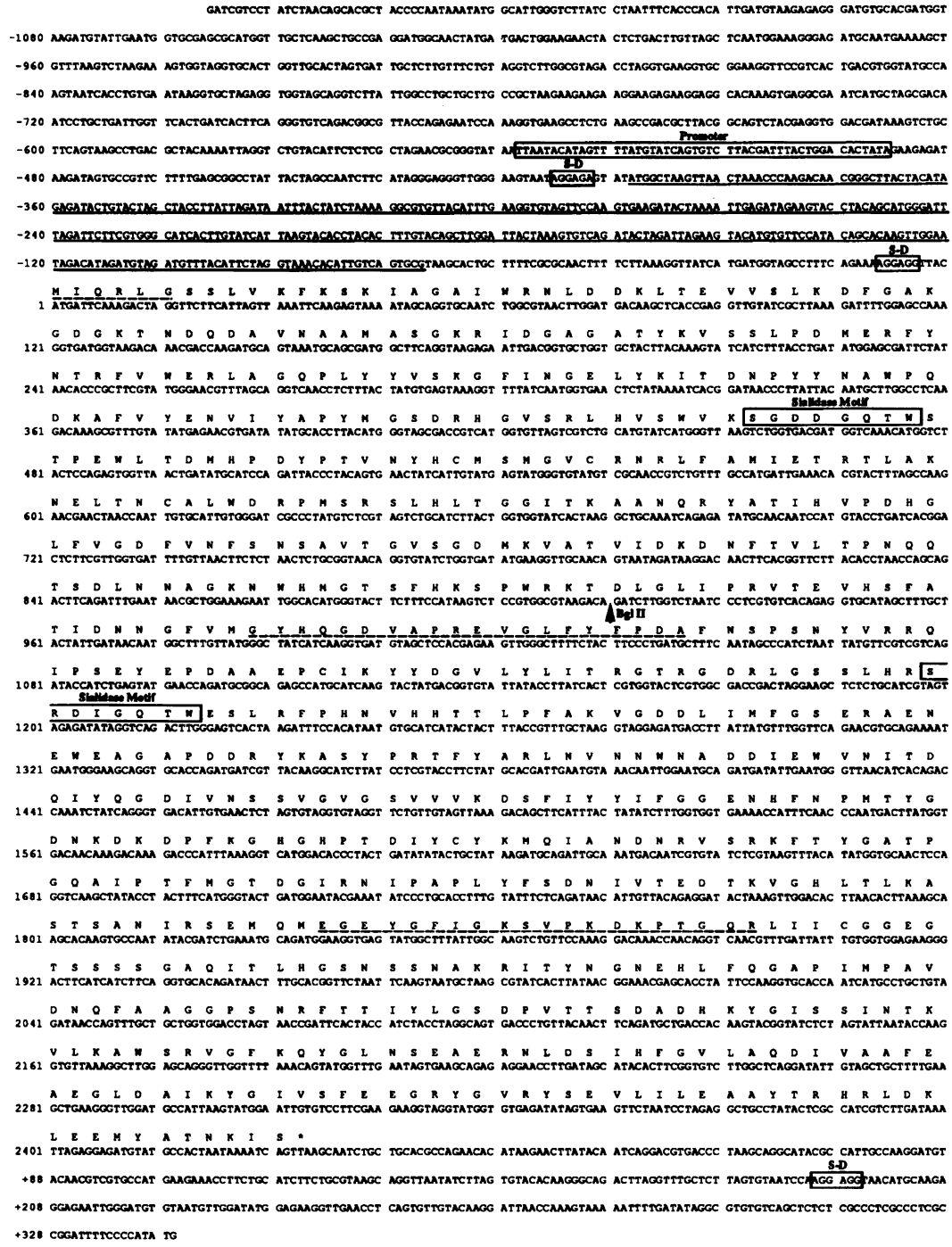

Figure 5 Primary nucleotide sequence and deduced amino acid sequence of bacteriophage E endosialidase

An SP6-like promoter region is shown boxed, as are putative Shine–Dalgarno sites (S-D) and two sialidase motif sequences. A potential open reading frame is shown by an unbroken underline (−403 to −71) which is upstream of the open reading frame coding for endosialidase (1–2433). Amino acid sequence obtained by N-terminal Edman degradation analysis is shown by a broken underline. The sequence of CNBr 1 (amino acid residues 330–349) contained two unknown assignments at residues 330 and 334, one uncertain assignment at residue 335 and four incorrect assignments at residues 332, 339, 342 and 338. CNBr 2 (residues 613–632) contained five ambiguous assignments at residues 615, 617–619 and 622, and six incorrect assignments at residues 614, 621, 625, 627, 629 and 632. All amino acids for CNBr 3 (residues 330–344) were correctly assigned, except for residue 330. This sequence was used to design Probe 4. An arrowhead depicts the *Bgl*II restriction site (915) which was used in the cloning of the endosialidase. The *Avr II* sites used in the construction of a plasmid containing the complete gene encoding K1E endosialidase (F15) are found at nucleotides −884 and 2091. The Y10 clone referred to in the text contained 2094 bp of K1E DNA (nucleotides −1179 to 915). Similarly, the K2 clone contained nucleotides 916 to +345.

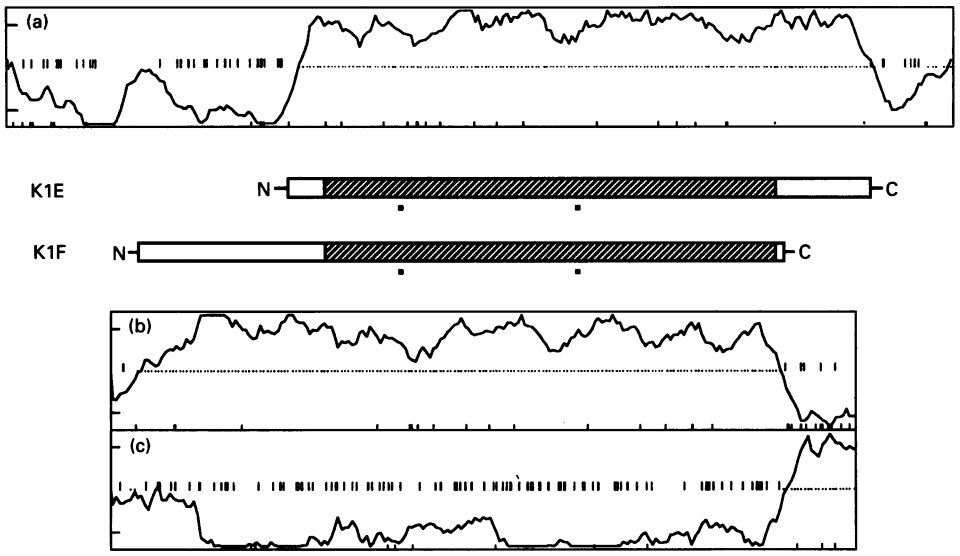

Figure 6 Comparison of K1E endosialidase and K1F endosialidase

Panels (**a**), (**b**) and (**c**) represent predictions of which reading frame of a nucleotide sequence is most likely to encode protein, using the positional base preferences method of Staden [30]. The horizontal lines on the left-hand side of each box show the expected score line that may be achieved if the sequence codes for protein (top), and the expected score line for sequence not coding for protein (bottom). Vertical lines across the middle of each panel depict stop codons, and smaller vertical lines along the bottom depict potential start codons. A broken line across the middle of a panel represents the region of a sequence with the highest score of all three frames and which therefore is most likely to code for protein. Panel (**a**) shows the open reading frame for K1E endosialidase, and panels (**b**) and (**c**) show open reading frames for K1F. The hatched regions of each endosialidase form share ~ 80% identity over 2200 bp; ■ represents the position of each sialidase motif.

depolymerizing activity [39]. Contained within this region was a sequence matching the consensus sequence of the 'P-loop' motif [40]. Also referred to as the 'A' consensus sequence [41], this motif is commonly found in proteins that bind nucleotides, especially GTP and ATP, and possesses the consensus sequence (Ala or Gly)-Xaa-Xaa-Xaa-Xaa-Gly-Lys-(Thr or Ser) [40]. A repeated sequence, Arg-Leu-Gly-Ser-Ser-Leu, was also noted starting at amino acid residues 4 and 392.

The potential open reading frame (shown underlined from nucleotide −403 to −71 in Figure 5) located upstream of the deduced K1E endosialidase open reading frame was not observed to have significant identity to any sequences in the EMBL database, and there is no indication that it might be involved in endosialidase function.

In vitro translation

A putative promoter with 75% identity to the bacteriophage SP6 promoter [42] was observed upstream of the deduced open reading frame for K1E endosialidase and an additional open reading frame encoding a protein with a predicted molecular mass of ~ 12 kDa (Figure 5). To demonstrate the ability of the putative SP6-like promoter to direct transcription, an *in vitro* transcription system in the presence of SP6 RNA polymerase was used to obtain a clone Y10 RNA transcript for *in vitro* translation by a rabbit reticulocyte lysate system. In this system the Bluescript T7 promoter would not be functional without the addition of the appropriate RNA polymerase. A translation product of 34 kDa was obtained, in agreement with the predicted size for the N-terminal fragment of K1E endosialidase encoded by the Y10 transformant (Figure 7). A protein of 12 kDa representing the first deduced open reading frame 3′ of the promoter sequence was not apparent, but this is explained by the absence of internal methionine residues, since [^{35}S]methionine was used as the radiolabel.

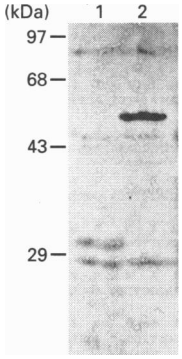

Figure 7 Autoradiograph of ^{35}S-labelled methionine *in vitro* translation products run in a 10% acrylamide SDS/PAGE gel: clone Y10

Lane 1 shows a 34 kDa polypeptide corresponding to the N-terminal fragment of K1E endosialidase encoded by clone Y10. The 55 kDa band seen in the positive control lane (lane 2) is the translation product derived from an SV64–carboxypeptidase E construct [31]. The relative positions of molecular mass markers are shown.

An *E. coli* lysate was used to provide a coupled transcription/translation system in which the full-length endosialidase construct, F15, was translated. This system was chosen because a prokaryotic lysate is better suited to the transcription and translation of bacteriophage DNA, although it is not possible to specify which promoter was utilized. A translation product of

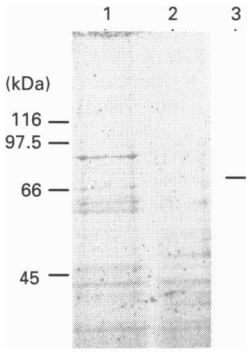

Figure 8 Autoradiograph of ^{35}S-labelled methionine *in vitro* translation products run in a 10% acrylamide SDS/PAGE gel: clone F15

Lane 1 shows a 90 kDa polypeptide corresponding to the K1E endosialidase encoded by clone F15; lane 2 shows a control translation with no DNA template added; lane 3 shows the position to which a sample of non-recombinant 76 kDa endosialidase migrated on the same SDS/PAGE gel. The relative positions of molecular mass markers are shown.

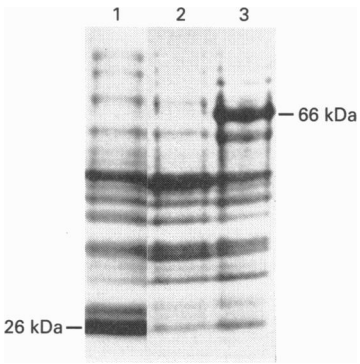

Figure 9 SDS/PAGE analysis of whole-cell lysates of pGEX fusion protein transformants

Samples were heated at 100 °C for 5 min in the presence of DTT and electrophoresed in a 15% acrylamide gel using a Laemmli buffer system [25]. Lane 1, transformant of pGEX-3PLX without DNA insert, induced with IPTG to express GST (26 kDa band); lane 2, transformant of pGEX-3PLX with K2 insert (clone N10), non-induced with IPTG; lane 3, N10 induced with IPTG to express GST–3′-domain fusion protein of K1E endosialidase (66 kDa band).

90 kDa was obtained (Figure 8), in accordance with the size of the polypeptide encoded by the endosialidase open reading frame and clearly migrating more slowly than the 74 kDa non-recombinant endosialidase.

The translation products were assayed for endosialidase activity using the TBA assay [23] and found to give no measurable amount of AcNeu release. There was no evidence that the 90 kDa translation product aggregated to give a higher-molecular-mass polymeric form of the protein.

Expression of the C-terminal fragment of endosialidase as a GST fusion protein

The transformant N10 shown in Figure 9, obtained by transferring the DNA from clone K2 into pGEX-3PLX, gave rise to a fusion protein product of 66 kDa. Given that GST alone has a molecular mass of 26 kDa, this indicated that a 40 kDa peptide was encoded by the K2 insert. Sequencing of plasmid DNA from selected transformants which encoded fusion protein confirmed the identity of the K2 insert and established that the insert was in the correct reading frame with respect to the predicted K1E endosialidase translation.

DISCUSSION

The identification and sequencing of the bacteriophage K1E endosialidase gene has shown the presence of two copies of the sialidase motif or 'Asp-box' [37,38], Ser-Xaa-Asp-Xaa-Gly-Xaa-Thr-Trp, seen in a variety of bacterial and viral sialidases specific for α-2,3- and α-2,6-sialosyl linkages, and also in some non-sialidase proteins [38]. In addition, comparison of the deduced protein sequence of K1E endosialidase with that of α-2,8-specific K1F endosialidase [19] shows a high degree of identity which is unbroken over 532 residues. Both endosialidases possess two sialidase motifs which flank the region of greatest similarity, and it is highly likely that the amino acid epitopes involved in substrate binding and catalysis are contained within this region. In addition to the two sialidase motifs, a 'P-loop' motif [40] is present towards the N-terminus of the protein. This motif is also present in K1F endosialidase but with the substitution of Lys-44 to Val-44. This lysine is generally considered to be crucial to the nucleotide binding activity of the 'P-loop'. Petter and Vimr [19] did not comment on the presence of sialidase motifs within the K1F sequence, and suggested from Southern blotting experiments that the C-terminal sequence of K1F would not be encoded by the K1E gene due to the failure of a K1F C-terminal-specific oligonucleotide probe to bind to K1E DNA. Comparison of the K1F probe sequence with that of the K1E endosialidase gene revealed that there were sufficient mismatches to impair hybridization. In fact, the identity between the K1F and K1E DNA sequences extends beyond the predicted C-terminus of K1F endosialidase, suggesting that there may have been a single base addition during the evolution of K1F to give a shortened C-terminal domain. Figure 6 shows the positional base preference plots for two reading frames of K1F endosialidase which indicate that, at the locus of the stop codon, the probability of continuing sequence encoding protein is highest in a different reading frame. The region of greatest difference between the genes encoding the α-2,8-endosialidases is at the 5′ end, where the K1F form possesses an extra N-terminal domain which has identity with the tail proteins of T3 and T7 bacteriophages and which is not present in K1E endosialidase. It is this region which may confer the greater stability of the K1F form reported by Petter and Vimr [19].

Bacteriophage K1F endosialidase gene transcription is postulated to be under the control of a promoter which has homology to the consensus sequence of phage promoters recognized by coliphage T7 RNA polymerase [19]. An interesting observation was that SP6 RNA polymerase, originating from bacteriophage SP6 which is specific for strains of *Salmonella typhimurium*, was able to recognize and effectively transcribe an RNA product using the putative coliphage K1E promoter sequence.

A discrepancy exists between the molecular mass of the amino acid sequence of K1E endosialidase deduced from the gene sequence and that obtained by SDS/PAGE of the purified enzyme. The enzyme was previously reported as having a subunit molecular mass of 74 kDa [5]. The present study suggests a polypeptide of ~ 76 kDa, yet the gene sequence predicts a size of 90 kDa and the *in vitro* transcription/translation product migrates at 90 kDa on SDS/PAGE. The most likely explanation of this discrepancy is that the 90 kDa product of the endosialidase open reading frame is post-translationally processed to give a mature 76 kDa protein. This processing activity appears to be

absent in the cell-free *E. coli* translation system. It is of interest to note that, whereas transcription/translation of the 5′ end of the predicted K1E endosialidase gene gave a polypeptide of the predicted size by SDS/PAGE analysis, expression of the 3′ end of the gene as a GST fusion protein also resulted in a shorter form than that predicted from the DNA sequence. This indicates that post-translational processing occurs in the C-terminal region of the protein.

If the molecular mass of the mature monomer is 76 kDa, representing a processed form of the translation product, then the ~200 kDa form observed in SDS/PAGE gels without heating would represent a native trimer. Thus the structure of K1E endosialidase is probably a non-covalently linked homotrimer of 76 kDa subunits, similar to that predicted for K1F endosialidase [19].

The cell-free translation product of the endosialidase open reading frame is inactive when assayed using the TBA assay [23]. This may be attributed to the absence of the proposed post-translational processing of the 90 kDa translation product to a 76 kDa mature protein in the translation system. Other explanations cannot be excluded; the protein may be misfolded, inhibitors of the endosialidase may be present, or the activity may be below the detection limits of the assay. Clearly, if the study of sequence motifs and amino acids which may be involved in catalysis is to be pursued, an active form of expressed endosialidase must be sought.

Mutant strains of bacteriophage, derived from the original PK1A–PK1E-related strains [1], bind to the PSA K1 coat polymer of *E. coli* K1 strains without hydrolysing the ketosidic linkages [43]. Thus it appears that amino acid residues separate from those involved in catalysis are involved in the binding of the PSA substrate molecule. Additional evidence suggesting that the substrate binding and catalytic sites may be distinct comes from the observation that an antiserum which neutralizes K1F and K1E endosialidase activities by inhibition of PSA binding does not affect the hydrolysis of the shorter-chain oligomers of colominic acid [19]. It has been speculated that the sialidase motifs may play a functional role in sialidase activity [37], such as binding of substrates or co-substrates, or formation of active centres [38]. However, evidence from studies on the crystal structure of *Salmonella typhimurium* LT2 sialidase complexed with the competitive inhibitor 2-deoxy-2,3-dehydro-*N*-acetyl-neuraminic acid predicts that the 'Asp-box' motifs occur on the outside of the structure, remote from the active site [44]. These data suggest that these motifs do not play a functional role in the catalytic domain, and it is speculated that they are involved in secretion and/or protein folding [44]. In the case of K1E endosialidase, it is possible that direct comparison between the DNA sequences of bacteriophage K1E endosialidase and inactive mutant forms which bind PSA may enable the identification of amino acid residues that are essential for catalysis.

We thank Dr. Pat Barker of the Microchemicals Division, A.F.R.C. Research Institute, Babraham, U.K., for his advice and provision of N-terminal microsequencing. The modified fusion protein expression vector pGEX-3PLX was a gift from Dr. John C. Hutton of the Department of Clinical Biochemistry, University of Cambridge, Cambridge, U.K. The SV64-carboxypeptidase E vector was used with the kind permission of Professor L. D. Fricker of the Department of Molecular Pharmacology, Albert Einstein College of Medicine, New York, NY, U.S.A. We also thank Dr. Steve Tomlinson, Department of Pathology, New York University Medical Centre, New York, NY, U.S.A., for valuable advice, discussion and provision of reagents. J.M.B. holds a Medical Research Council collaborative studentship. This work was funded by Ciba Pharmaceuticals, Horsham, U.K.

Received 23 December 1994/10 March 1995; accepted 24 March 1995

REFERENCES

1 Gross, R. J., Cheasty, T. and Rowe, B. (1977) J. Clin. Microbiol. **6**, 548–550
2 Robbins, J. B., McCracken, G. H., Gotschlich, E. C., Ørskov, F., Ørskov, I. and Hanson, L. A. (1974) N. Engl. J. Med. **290**, 1216–1220
3 Bhattacharjee, A. K., Jennings, H. J., Kenny, C. P., Martin, A. and Smith, I. C. P. (1975) J. Biol. Chem. **250**, 1926–1932
4 Ørskov, I., Ørskov, F., Jann, B. and Jann, K. (1977) Bacteriol. Rev. **41**, 667–710
5 Tomlinson, S. and Taylor, P. W. (1985) J. Virol. **55**, 374–378
6 Adlam, C., Knight, J. M., Mugridge, A., Williams, J. M. and Lindon, J. C. (1987) FEMS Microbiol. Lett. **42**, 23–25
7 Livingston, B. D., Jacobs, J. L., Glick, M. C. and Troy, F. A. (1988) J. Biol. Chem. **263**, 9443–9448
8 Roth, J., Zuber, C., Wagner, P., Taatjes, D. J., Weisgerber, C., Heitz, P. U., Goridis, C. and Bitter-Suermann, D. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2999–3003
9 Roth, J., Zuber, C., Wagner, P., Blaha, I., Bitter-Suermann, D. and Heitz, P. U. (1988) Am. J. Pathol. **133**, 227–240
10 Figarella-Branger, D. F., Durbec, P. L. and Rougon, G. N. (1990) Cancer Res. **50**, 6364–6370
11 Heitz, P. U., Komminoth, P., Lackie, P. M., Zuber, C. and Roth, J. (1990) Verh. Dtsch. Ges. Pathol. **74**, 376–377
12 Moolenaar, C. E., Muller, E. J., Schal, D. J., Figdor, C. G., Bock, E., Bitter-Suermann, D. and Michalides, R. J. (1990) Cancer Res. **50**, 1102–1106
13 Taylor, P. W. (1987) U.S. Pat. 4695541
14 Roth, J. and Zuber, C. (1990) Lab. Invest. **62**, 55–60
15 Kwiatkowski, B., Boschek, B., Thiele, H. and Stirm, S. (1982) J. Virol. **43**, 697–704
16 Finne, J. and Makela, P. H. (1985) J. Biol. Chem. **260**, 1265–1270
17 Hallenbeck, P. C., Vimr, E. R., Yu, F., Bussler, B. and Troy, F. A. (1987) J. Biol. Chem. **262**, 3553–3561
18 Pelkonen, S., Pelkonen, J. and Finne, J. (1989) J. Virol. **63**, 4409–4416
19 Petter, J. G. and Vimr, E. R. (1993) J. Bacteriol. **175**, 4357–4363
20 Long, G. S., Taylor, P. W. and Luzio, J. P. (1993) in Polysialic Acid: From Microbes to Man (Roth, J., Rutishauser, U. and Troy, F. A., II, eds.), pp. 137–144, Birkhäuser Verlag, Basel, Switzerland
21 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) in Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor
22 Westphal, O. and Jann, K. (1965) Methods Carbohydr. Chem. **5**, 83–91
23 Horgan, I. E. (1981) Clin. Chim. Acta **116**, 409–415
24 Aminoff, D. (1961) Biochem. J. **81**, 384–392
25 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
26 Holm, L. (1986) Nucleic Acids Res. **14**, 3075–3087
27 Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988) Nucleic Acids Res. **16**, 6127–6145
28 Hanahan, D. (1983) J. Mol. Biol. **166**, 557–580
29 Staden, R. and McLachlan, A. D. (1982) Nucleic Acids. Res. **10**, 141–156
30 Staden, R. (1990) Methods Enzymol. **183**, 163–180
31 Fricker, L. D., Adelman, J. D., Douglass, J., Thompson, R. C., von Strandmann, R. P. and Hutton, J. C. (1989) Mol. Endocrinol. **3**, 666–673
32 Smith, D. B. and Johnson, K. S. (1988) Gene **67**, 31–40
33 Pearson, W. R. (1990) Methods Enzymol. **183**, 63–98
34 Coulson, A. F. W., Collins, J. F. and Lyall, A. (1987) Comput. J. **30**, 420–424
35 Staden, R. (1990) Methods Enzymol. **183**, 193–211
36 Devereux, W. J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **16**, 6127–6145
37 Roggentin, P., Rothe, B., Kaper, J. B., Galen, J., Laurisuk, L., Vimr, E. R. and Schauer, R. (1989) Glycoconj. J. **6**, 349–353
38 Rothe, B., Rothe, B., Roggentin, P. and Schauer, R. (1991) Mol. Gen. Genet. **226**, 190–197
39 Reiger, D., Freund-Molbert, E. and Stirm, S. (1975) J. Virol. **15**, 964–975
40 Saraste, M., Sibbald, P. R. and Wittinghofer, A. (1990) Trends Biochem. Sci. **15**, 430–434
41 Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982) EMBO J. **1**, 945–951
42 Melton, D. A., Kreig, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984) Nucleic Acids Res. **12**, 7035–7057
43 Pelkonen, S., Aalto, J. and Finne, J. (1992) J. Bacteriol. **174**, 7757–7761
44 Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. and Taylor, G. L. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 9852–9856