ELSEVIER

Journal of Chromatography B, 685 (1996) 91–104

JOURNAL OF CHROMATOGRAPHY B: BIOMEDICAL APPLICATIONS

Purification and characterization of a recombinant hepatitis E protein vaccine candidate by liquid chromatography–mass spectrometry

C. Patrick McAtee[a,*], Yifan Zhang[a], Patrice O. Yarbough[a], Thomas R. Fuerst[a], Kathy L. Stone[b], Susan Samander[b], Kenneth R. Williams[b]

[a]*Genelabs Technologies, Inc., 505 Penobscot Drive, Redwood City, CA 94063, USA*

[b]*The W.M. Keck Foundation Biotechnology Resource Laboratory and Howard Hughes Medical Institute, Yale University, New Haven, CT 02115, USA*

Received 27 December 1995; revised 14 March 1996; accepted 14 March 1996

Abstract

A protein with a molecular mass of approximately $62 \cdot 10^3$, derived from open reading frame 2 (ORF-2) of the hepatitis E virus (HEV; Burma strain), was expressed in a baculovirus expression vector and purified to homogeneity. The recombinant 62 kDa protein appeared to be a doublet, as determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). Tryptic digestion in conjunction with laser desorption mass spectrometry (LD–MS) and sequence analysis of the tryptic peptides indicated that the amino terminus was blocked, although no proteolytic degradation occurred. The determined internal sequences of peptides were in agreement with the predicted ORF-2 protein. Reversed-phase liquid chromatography coupled to electrospray mass spectrometry (LC–MS) resolved the doublet proteins into two major components with molecular masses of 56 548.5 and 58 161.4. Confirmation of the amino terminus of the molecule by LD–MS post-ion decay enabled us to tentatively assign the carboxyl terminus of each species at residues 540 and 525. Sequencing of the intact protein by automated carboxyl terminal sequencing confirmed that the carboxyl terminus was truncated and that the sequence assignment predicted by LC–MS was correct.

Keywords: Hepatitis E virus protein vaccine; Open reading frame 2 protein

1. Introduction

The hepatitis E virus (HEV) has been identified as the cause of hepatitis in several epidemics and in sporadic cases in Asia, Africa, Europe, Mexico and in the Indian subcontinent [1–3]. Transmission is usually by water contaminated with feces, although there is some evidence of person-to-person transmission. HEV has been previously shown to passage in several non-human primates, including cynomolgus macaques, with typical liver enzyme elevations and recovery of 27- to 34-nm virus-like particles from the feces of infected experimental animals, similar to those seen in clinical specimens [4]. The development of a vaccine to HEV has been hampered by the inability to culture the virus in sufficient quantities. Molecular cloning and sequence analysis of the HEV genome has revealed three open reading frames [5].

*Corresponding author.

0378-4347/96/$15.00
PII S0378-4347(96)00143-0

ORF1 is predicted to encode a polyprotein containing five domains, of which a methyltransferase and an RNA-dependent RNA polymerase have been tentatively identified [6]. ORF2 encodes a capsid protein with a predicted molecular mass of approximately $72 \cdot 10^3$. The function of ORF3 remains elusive. However, ORF3 encodes immunodominant epitopes that are recognized by acute phase and convalescent sera of HEV-infected patients [7].

We have previously reported the cloning and nucleotide sequence of an overlapping, contiguous set of cDNA clones, representing the entire genome of the HEV Burma strain [5]. Upon subcloning and expression of the designated capsid open reading frame using a baculovirus expression system, we observed the conversion of the full-length ORF-2 expressed protein from an insoluble 73 kDa protein to a soluble protein of approximately 62 kDa [8,9]. The 62 kDa protein was identified as an endoproteolytic cleavage product of the ORF-2 protein, in which the N-terminus of the 62 kDa protein was found to match the ORF-2 sequence starting at residue 112 in the ORF-2 protein sequence [8,9]. The precise coding sequence of this 62 kDa protein was cloned and expressed using a recombinant baculovirus expression system. The recombinant 62 kDa protein (r62-kDa) showed antigenic and structural properties similar to the processed form [8,9].

In this study, we report the purification of the r62-kDa protein and its preliminary biochemical characterization. Establishment of process development and product identity parameters will facilitate the potential application of this protein as a diagnostic reagent and vaccine candidate. This r62-kDa protein has been shown to be capable of eliciting a protective immune response in primates following heterologous challenge with infectious HEV [10].

2. Experimental

2.1. Reagents

DEAE EMD 650(S) was obtained from E. Merck Separations. Sephacryl S-100 and S-1000 were obtained from Pharmacia Biotech. Poros Q/F and HQ/F chromatography columns were obtained from PerSeptive Biosystems. All buffer ingredients were obtained from AMRESCO and all proteins were purified on a Waters 650E chromatography workstation, with data management provided by Millenium 2010 software.

2.2. Cell culture and expression conditions – Spodoptera frugiperda (Sf9)

Cells in spinner flasks were maintained at 27°C in Grace's insect medium supplemented with 5% (v/v) fetal bovine serum, 50 μg/ml of gentamycin and 0.1% Pluronic F-68. All culture ingredients were obtained from Gibco/BRL and cells were cultured according to protocols described by the manufacturer (Invitrogen). Viable cells with a density of $2 \cdot 10^6$/ml were pelleted by centrifugation. The cell pellet was resuspended in one tenth of the original volume of medium containing the recombinant virus, BBIII-r62-kDa, at a multiplicity of infection of two plaque forming units (PFU) per cell. Recombinant virus BBIII-62-kDa was constructed previously to express the recombinant 62 kDa protein [8]. Infection was carried out for 1 h, without stirring. The infected cells were diluted to the original density with fresh medium and maintained at 27°C for two to seven days, with agitation (75–95 rpm).

2.3. Purification of the 62 kDA protein

Sf9 cells infected with recombinant baculovirus BBIII-62-kDa were lysed in phosphate buffered saline (PBS) containing protease inhibitors, by two passes through a Microfluidics M-110S microfluidizer at 14 000 p.s.i. (1 p.s.i.=6894.76 Pa) The crude lysate was centrifuged at 10 000 *g* for 30 min at 4°C and the supernatant was decanted into a beaker. Solid dithiotreitol (DTT) was added to the lysate supernatant to a final concentration of 50 m*M*. This material was dialyzed overnight against 10 m*M* Tris pH 8.6–50 m*M* NaCl–1 m*M* DTT, followed by 10 m*M* Tris pH 8.6–50 m*M* NaCl. The r62-kDa protein was found to be completely soluble in the lysis buffer, with quantitative recovery of the molecule in the cell lysis supernatant. The dialysate was prefiltered through a Millipore 0.22-μm filter and then loaded directly onto a DEAE EMD 650(S) column (Merck) that had been equilibrated in 10 m*M* Tris pH 8.6–50 m*M* NaCl. The r62-kDa protein was

eluted in a linear gradient of 50–500 m*M* NaCl over 15 column volumes, at a superficial linear velocity of 100 cm/h. The r62-kDa protein was further purified and buffer-exchanged on a 60-cm Sephacryl S-100 column that had been equilibrated in 25 m*M* Tris pH 7.2. r62-kDa-Containing fractions were chromatographed on a Poros HQ/F strong anion-exchange column with the r62-kDa protein eluting in a NaCl gradient of 0–1 *M* NaCl. The Poros HQ/F column was operated at a superficial linear velocity of 5000 cm/h. If necessary, the r62-kDa was buffer-exchanged by tangential cross flow diafiltration using a spiral wound cartridge (Millipore Prep TFF).

Buffer conductivities and pH were monitored with a Radiometer Copenhagen CDM 83 conductivity meter and a PHM reference standard pH meter, respectively. All buffer constituents were either of biotechnology or USP grade and were determined to be essentially pyrogen-free by limulus amebocyte lysate assay.

2.4. Amino acid analysis

Amino acid analysis was carried out on a Beckman Model 6300 ion-exchange instrument following a 16 h hydrolysis at 115°C in 100 μl of 6 *M* HCl, 0.02% phenol plus 2 nmol norleucine. Following hydrolysis, the samples were dried on a Speedvac and the resulting amino acids were dissolved in 100 μl of sample buffer (Beckman) containing 2 nmol homoserine, with the homoserine acting as a second internal standard to independently monitor the transfer of the sample onto the analyzer [11]. The instrument was calibrated with a 2 nmol mixture of amino acids and was operated according to the manufacturer's specifications.

2.5. In-gel enzymatic digestion

In-gel enzymatic digestion was carried out according to Williams and Stone [12], using perfusion in an approximately 1:5 (enzyme weight–substrate weight) ratio of modified trypsin (Promega) and digestion for 24 h at 37°C. The resulting peptides were reduced/carboxylmethylated, extracted with 0.1% trifluoroacetic acid (TFA), 60% acetonitrile (CH_3CN) and then subjected to hydrolysis/amino acid analysis so that the amount and density (μg protein/mm^3) could be determined, with both of the parameters serving as valuable criteria to judge the probability of success of the impending digest.

2.6. Peptide analysis of the 62 kDa protein

Reversed-phase HPLC was carried out on a Hewlett-Packard 1090 HPLC system equipped with an ISCO Model 2150 peak separator and a 25-cm Vydac C_{18} (5 μm, 300 Å) column equilibrated with 98% buffer A (0.06% TFA) and 2% buffer B (0.052% TFA, 80% CH_3CN), as described [12]. Peptides were then eluted with the following gradient program: 0–60 min (2–37% B), 60–90 min (37–75% B) and 90–105 min (75–98% B) and were detected by their absorbance at 210 nm. Aliquots of digests in the 25–250 pmol range were fractionated on a 2.1-mm I.D. column and were eluted at a flow-rate of 0.15 ml/min. Fractions were collected in capless Eppendorf tubes.

2.7. Laser desorption mass spectrometry

Aliquots (3 μl) of peptides isolated via reversed-phase HPLC were added on top of a 1-μl of α-cyano-4-hydroxycinnamic acid (α-CHCA) matrix solution that was spotted onto a new target [13]. Mixing of the matrix was accomplished by repeatedly pulling the sample into and expelling it from a micropipette. The samples were then allowed to air dry at room temperature. To avoid cross-contamination, all targets were used only once. The α-CHCA matrix solution was prepared at a concentration of 10–20 mg/ml in 40% CH_3CN–0.1% TFA and was used after being vortex-mixed and allowed to stand for a few minutes. Matrix solutions were stored for a maximum of two days at −20°C. The calibrants used for external calibration of peptides were gramicidin S (*m/z* 1142.5) and insulin (*m/z* 5734.5). Both calibrants were stored at −20°C as 10 pmol/μl stocks in either 50% CH_3CN–0.1% TFA or in 0.1% TFA. LD–MS was carried out on a VG/Fisons TofSpec mass spectrometer that was operated in the positive linear ion mode at an accelerating voltage of 25 kV and was equipped with a nitrogen laser (337 nm) and a 0.65-m linear flight tube. Routinely, 30 shots were averaged for each spectrum with three-to-six spectra

acquired for each sample. The predicted masses were based upon the average isotypic, singly protonated mass and the expected mass accuracy was about ±0.25%. Post-source decay sequencing was done in the reflectron mode, with the ladder sequencing calibrated to a 2 pmol sequence of adrenocorticotropic hormone fragment (ACTH clip fragment 18–39).

2.8. Protein/peptide sequencing

Sequencing was carried out on either an Applied Biosystem 470A or 477 that was equipped with on-line HPLCs for the identification of the resulting phenylthiohydantoin (Pth) amino acid derivatives. Prior to sample application, 25 pmol of a sixteen residue internal sequencing standard peptide with the formula: [norleucine-(succinyl-lysine)$_4$]$_3$-norleucine was first spotted onto the sequencing filter [14]. The 470A and 477 instruments were operated according to the manufacturer's recommendations and 3 pmol Pth standards were routinely used. All sequences were searched via the BLAST Network Service operated by the National Center for Biotechnology Information.

2.9. Liquid chromatography–mass spectrometry (LC–MS)

Purified 62 kDa protein was chromatographed on a Vydac C_{18} reversed-phase microbore column (150× 1 mm) using an ABI Model 410B dual syringe pumping system. The flow-rate was maintained at 50 μl/min and elution was achieved using a linear gradient (over 10 column volumes) from 0.1% aqueous TFA to 0.1% TFA in acetonitrile. A Carlo Erba Phoenix 20 CU pump was used to deliver a mixture of methoxyethanol and isopropanol (1:1, v/v) at 50 μl/min, which was combined with the column eluent in a post-column mixing chamber. An in-line flow splitter was used to restrict flow to the mass spectrometer to approximately 10 μl/min. Detection was performed immediately following elution from the column at 214 nm, using an ABI 759A variable-wavelength detector. Mass spectrometric detection was achieved following post column solvent addition and flow splitting by a VG BioQ triple quadrupole mass spectrometer. Spectra were recorded in the positive ion mode using electrospray ionization. Calibration of the instrument was performed in the range m/z 500–2000 by using direct injection analysis of myoglobin. Spectra were recorded at 1.5 s intervals and a drying gas of nitrogen was used to aid evaporation of the solvent. The capillary voltage was maintained at approximately 4 kV, with a source temperature of 60°C.

2.10. Automated C-terminal sequence analysis

Protein samples were applied to Zitex membranes pretreated with isopropanol and inserted into inert Kel-F columns. The sequencer column was installed into a Hewlett Packard G1009A sequencer for chemical coupling and cyclization. The coupled peptidylthiohydantoin and cyclyzed product was cleaved to the C-terminal thiohydantoin-amino acid residue and the shortened peptide using an alkaline salt of trimethylsilanolate (KOTMS). The derivatized sample was analyzed by an HP1090 liquid chromatograph with filter photometric detection at 269 nm using a Hewlett-Packard specialty (25 cm×2.1 mm) reversed-phase PTH analytical HPLC column. A 39-min binary gradient (solvent A, phosphate buffers, pH 2.9; solvent B, acetonitrile) utilizing alkyl sulfonate as an ion pairing agent was developed. Thiohydantoin-amino acid standards at 100 pmol were used to standardize the analysis.

2.11. SDS–PAGE/Western blotting

SDS–PAGE was performed according to the tricine SDS–PAGE procedure [15]. Western blotting was performed as previously described [16]. A rabbit polyclonal antibody directed to the carboxyl-terminal region of ORF-2, 1L6, was used to evaluate the purified protein.

2.12. Indirect enzyme-linked immunosorbent assay (ELISA), "sandwich assay"

A double antibody sandwich technique was used for determination of the concentration of the 62 kDa protein in purification streams. Anti-1L6 antibody, a polyclonal rabbit antibody directed against the carboxyl terminal portion of the ORF-2 protein was purified by Protein G Sepharose affinity chromatography and coated onto microtiter plates at a con-

centration of 1.6 μg/ml per well in carbonate–bicarbonate buffer. After washing and blocking, purified 62 kDa antigen (quantitated by amino acid analysis) or process samples were added to the plate and diluted in a series of two-fold dilutions. After incubation of the antigen and washing, a second antibody, anti-SG3 conjugated to biotin, was added to the plate and incubated. The SG3 antigen corresponds to the carboxyl terminal 328 amino acids of the 62 kDa protein [17]. This incubation was followed by a final incubation with streptavidin-conjugated horseradish peroxidase. After final washing, substrate was added and the plates were read at an absorbance of 490 nm.

Protein concentrations of process samples were established by the Bradford assay (Pierce).

3. Results

3.1. Expression of recombinant 62 kDa protein in insect cells

The DNA sequence from amino acid 112 to the carboxyl terminal amino acid 660 was cloned into a baculovirus expression vector pBluBacIII, as previously described [8]. A methionine codon was incorporated at the 5′ end to ensure proper translation initiation. Following transfection and four rounds of plaque purification, viral stocks and cell lysates were prepared from ten individual plaque isolates.

3.2. Purification of recombinant 62 kDa protein

3.2.1. Preparation of cell lysate

Frozen Sf9 cells infected with recombinant baculovirus BBIII-62-kDa were resuspended in PBS containing protease inhibitors and were lysed. The cell lysate was processed according to the methods detailed in Section 2.

3.2.2. DEAE EMD 650(S) chromatography

The dialysate containing the 62 kDa protein was chromatographed on a DEAE EMD 650(S) column (10×2 cm) equilibrated with buffer A (chromatography profiles are summarized in Fig. 1). After washing the column with five column volumes of buffer A, the column was developed with a linear gradient of buffer A to buffer A containing 500 m*M* NaCl (total of fifteen column volumes). The flow-rate was maintained at a superficial linear velocity of 300 cm/h. The 62 kDa protein eluted early in the gradient (Fig. 1A). The presence of ORF-2 related 62 kDa protein was confirmed by Western blotting of

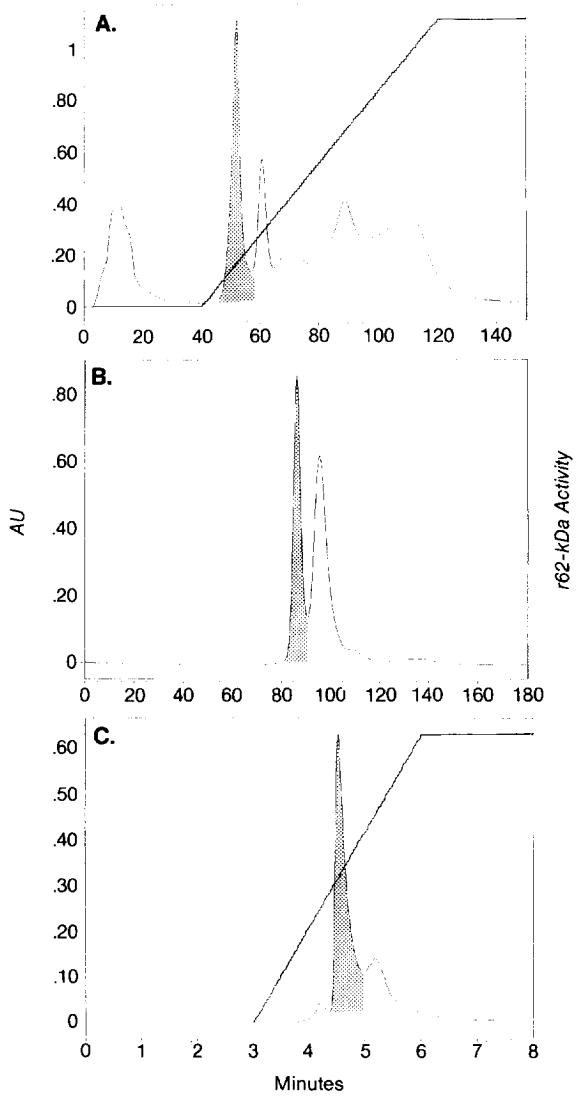

Fig. 1. Chromatographic steps in r62-kDa protein purification. (A) DEAE EMD 650 (S) column chromatogram. The gradient ranges from 50 m*M* to 0.5 *M* NaCl. (B) Sephacryl S-100 column chromatogram. (C) Poros HQ/F column chromatogram. The gradient ranges from 0 to 1 *M* NaCl. Shaded areas represent fractions that were determined to be immunoreactive to 1L6 antibody by Western blot analysis.

column fractions with rabbit polyclonal antibody, 1L6 [7], an antibody raised against the carboxyl terminal region of the parent ORF-2 protein (Fig. 2A).

3.2.3. Sephacryl S-100 chromatography

The r62-kDa-containing fractions from the DEAE column were pooled, concentrated by centrifugal ultrafiltration (Amicon) and chromatographed on a Sephacryl S-100 column that had been equilibrated in buffer B (5×100 cm) at a superficial linear velocity of 30 cm/h (Fig. 1B). Fractions containing the 62 kDa protein were pooled and concentrated by centrifugal ultrafiltration. This step in the procedure was used primarily as a buffer-exchange step.

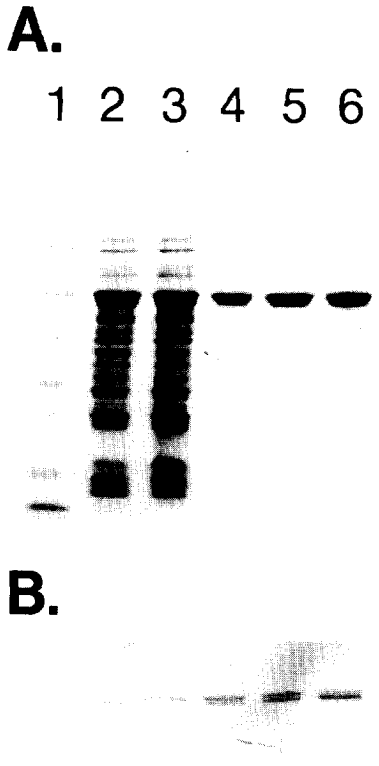

Fig. 2. r62-kDa Purification process. 4–20% SDS–polyacrylamide gel electrophoresis and corresponding Western blot. (A) Lane 1, molecular mass markers; lane 2, cell lysate; lane 3, cell lysate supernatant; lane 4, DEAE peak pool; lane 5, Sephacryl S-100 peak pool; lane 6, Poros HQ/F peak pool. Molecular mass markers are Novex SeeBlue pre-stained standards and their molecular mass range is as follows (from top to bottom): myosin, 250 kDa; BSA, 98 kDa; glutamic dehydrogenase, 64 kDa; alcohol dehydrogenase, 50 kDa; carbonic anhydrase, 36 kDa; myoglobin, 30 kDa; lysozyme, 16 kDa; aprotinin, 6 kDa; insulin B chain, 4 kDa. (B) Western blot of lanes 1–6 from above. Samples were diluted fifteen-fold prior to SDS–PAGE and 1L6 antibody was used as the primary antibody.

3.2.4. Poros HQ/F chromatography

The 62 kDa protein-containing pool from the S-100 column was applied to a Poros Q/F column (10×4.6 mm) that had been equilibrated with buffer B at a superficial linear velocity of 3000 cm/h. The column was washed with five column volumes of buffer B, and the 62 kDa protein was eluted with a linear gradient of buffer B to buffer B containing 1 *M* NaCl (total of ten column volumes). The 62 kDa protein eluted at approximately 300 m*M* NaCl (Fig. 1C). At this stage, the 62 kDa protein was nearly homogeneous, based upon SDS–PAGE and Western blotting (Fig. 2A, lane 6).

3.3. Characterization of the recombinant 62 kDa protein

3.3.1. SDS–PAGE and Western blotting

Process fractions of the 62 kDa protein were denatured and analyzed by SDS–PAGE, as described in Section 2.11. We used the tricine buffer system [15] in order to obtain higher resolution and minimize interference with Edman reagents in solid phase sequencing from PVDF membranes. Analysis of the initial cell lysate (Fig. 2A, lane 2) indicated that the 62 kDa protein was adequately expressed in the baculovirus expression system and was suitable for further processing. The cell lysis supernatant (Fig. 2A, lane 3) appeared to yield essentially quantitative recovery of the 62 kDa protein as a soluble product. The DEAE pool (Fig. 2A, lane 4) yielded a significant purification of the 62 kDa protein from the previous lysate supernatant pool with the Sephacryl S-100 and Poros HQ/F chromatography steps, resulting in further purification of the 62 kDa protein (Fig. 2A, lanes 5 and 6). One curious aspect of the gel electrophoresis profiles was the purification of the 62 kDa band as a doublet. The Western blot illustrated in Fig. 2B contains a fifteen-fold dilution of the samples run on SDS–PAGE in parallel. In the Western blot, the protein is visualized as a doublet with the lower band appearing to accumulate over the course of the purification procedure. However,

proteolysis appears to be minimal. This suggests that the protein may exist as a heterogenous species through potential side chain modifications, such as glycosylation, N-terminal or C-terminal modification.

3.3.2. Amino terminal sequence and amino acid composition

The 62 kDa protein was resolved by SDS–polyacrylamide gels, transferred to PVDF (Bio-Rad) membranes and sequenced by amino terminal sequencing. The levels of amino acids present in the first five cycles were approximately ten-fold lower than expected (data not shown) suggesting that a major portion of the 62 kDa molecule was blocked at the N-terminus. The sequence obtained appeared to be identical to the processed form of the 62 kDa protein that was originally produced in Sf9 cells expressing the full length ORF-2 [9]. Presumably a methionine aminopeptidase present in bacuolovirus-infected Sf9 cells cleaved the N-terminal methionine introduced into the coding sequence of the recombinant protein to ensure correct initiation of translation.

3.3.3. Tryptic peptide analysis

A 2-pmol quantity of the 62 kDa protein was digested in situ with trypsin in an excised polyacrylamide gel slice. A blank gel slice and a gel slice containing 50 pmoles of transferrin were also digested and analyzed in parallel, as controls. The resulting peptides were resolved by reversed-phase HPLC (Fig. 3). Out of a possible total of 38 tryptic peptides and 143 potential peaks detected by re-

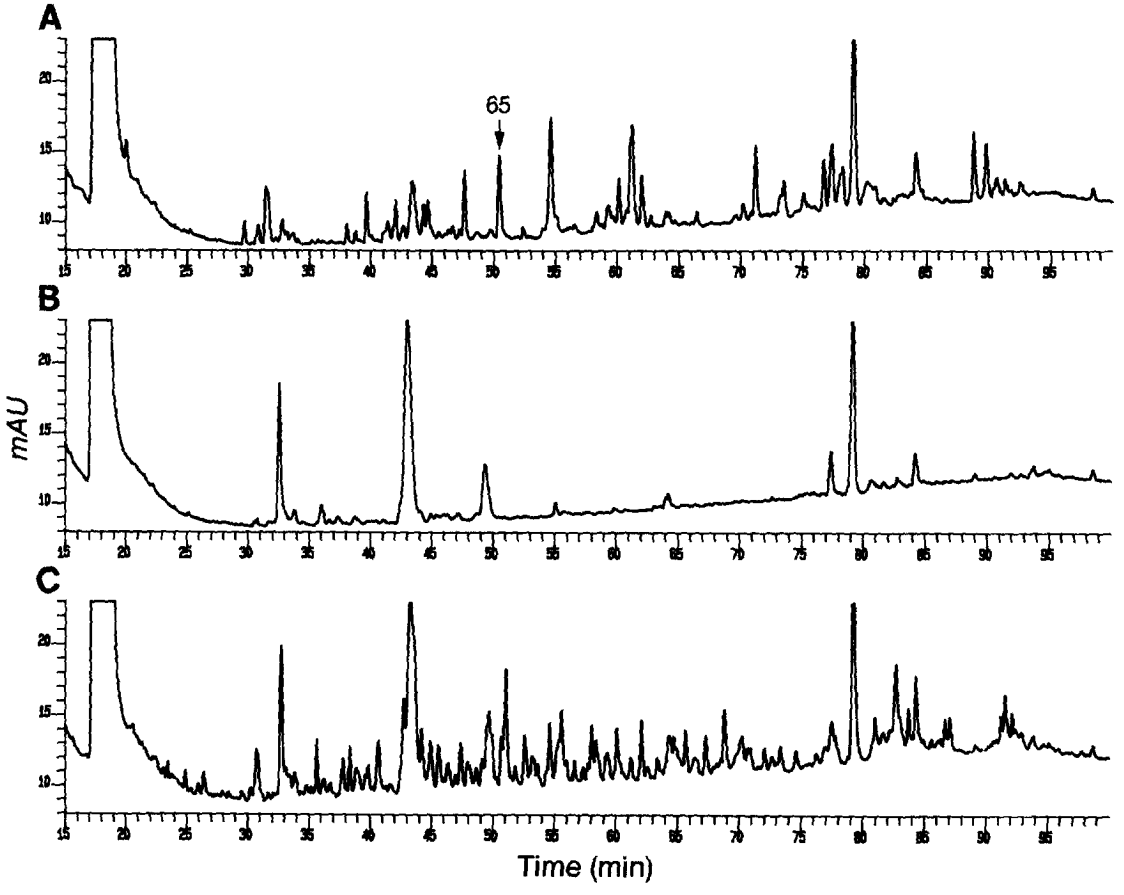

Fig. 3. Reversed-phase HPLC of tryptic peptides from r62-kDa. Numbered peaks correspond to LD–MS profiles and sequenced peptides from r62-kDa in Figs. 4 and 5, respectively. HPLC conditions were as described in Section 2. Full scale corresponds to 0.1 absorbance unit at 214 nm. (A) Trypsin-digested r62-kDa chromatogram. (B) Blank chromatogram. (C) Trypsin-digested transferrin chromatogram.

versed-phase HPLC analysis of the digested 62 kDa protein, eight peaks were selected for laser desorption mass spectroscopy, since the HPLC profile suggested that these peaks appeared to contain only one major species (Fig. 4).

3.3.4. Laser desorption mass spectroscopy (LD–MS)

Peaks 45, 50, 62, 65, 73, 82, 101 and 116 were further evaluated by LD–MS. LD–MS was used to resolve several issues; (1) whether the peak was an artifact peak or was in fact a peptide and (2) whether the peak contained more than one peptide and (3) whether the LD–MS calculated mass was comparable to the predicted mass that was derived from the predicted amino acid sequence encoded by the viral RNA. Through this approach, post-translational modifications would be readily determined. Peaks 65, 73, 101 and 116 appeared to be suitable for direct sequencing, as the LD–MS results (data not shown) indicated the presence of only one major species. Peaks 45, 50, 62 and 82 were also sequenced. Although the peaks appeared to be mixtures that would optimally require HPLC repurification prior to

Fig. 4. Sequence analysis of r62-kDa tryptic peptides. Tryptic peptide sequences are indicated in the figure, above the predicted sequence. A lower case a, b, or c following the peptide designation indicates that the peptide was obtained as a mixture of peptides. IUPAC nomenclature is used for amino acid abbreviations.

individual sequencing, we chose to directly sequence these mixtures as we had the mass information and predicted amino acid sequence.

3.3.5. Sequence analysis

Peptide 65 did not yield an interpretable sequence. Upon further examination, the mass observed by LD–MS was consistent with the N-terminal residues of the 62 kDa protein with the addition of an N-terminal acetyl group (predicted 1786.9 Da versus 1785.5 Da observed; corresponding to a 0.03% error; Fig. 5A). Evaluation of a tryptic peptide data base using the predicted sequence of the 62 kDa protein revealed only one other peptide that was similar in molecular mass (residues 408–423). Post source decay analysis revealed that Peak 65 was indeed the predicted amino terminal tryptic peptide (Fig. 4B). All other peptide peaks matched various internal sequences of the 62 kDa protein (data not shown).

3.3.6. LC–MS and carboxyl terminal sequence analysis

In order to evaluate the nature of the 62 kDa protein doublet observed by SDS–PAGE, the purified 62 kDa protein was chromatographed on a Vydac C_{18} reversed-phase capillary column with the eluting peak being evaluated by electrospray mass spectrometry (ES–MS) (Fig. 6). The 62 kDa protein resolved into two primary peaks by ES–MS corresponding to 56.5 and 58.1 kDa, respectively (Fig. 7).

The predicted molecular mass of the 62 kDa protein using the coding sequence of residues 112 to 660 of the ORF-2 region is 59.1 kDa. These data suggested that a deletion occurred in the molecule, most likely at the amino or carboxyl terminus. The protein was not found to be glycosylated (data not shown) either by periodate oxidation or by GC–MS analysis. Molecular mass determination by ES–MS is typically 0.01% [16]. With the confirmation of the amino terminus, the ES–MS data suggests that the carboxyl terminus may be clipped between residues 551–552 and residues 536–537. Automated carboxyl terminal sequencing was performed using intact 62 kDa protein to confirm the putative carboxyl terminal processing. The initial sequencing cycle gave rise to two very strong peaks corresponding to glutamine and lysine, neither of which are located at the predicted carboxyl terminus of the 62 kDa protein (Fig. 8). The second cycle revealed a very strong (>200 pmol) leucine peak, indicating the presence of more than one leucine in the polypeptide mixture. The third cycle was somewhat ambiguous due to increasing background. However, arginine was clearly present in the third cycle along with either a glutamic acid or a glycine residue. Thus, the carboxyl sequencing data supports the existence of a heterogenous, truncated protein.

4. Discussion

We have constructed a baculovirus vector that directs the efficient expression of a recombinant 62 kDa protein encoded by the Hepatitis E virus ORF-2 region [8]. This protein was purified by weak anion-exchange chromatography followed by gel filtration and strong anion-exchange chromatography. The 62 kDa protein was determined to be greater than 95% pure by this procedure (Table 1).

Biochemical characterization of the purified 62 kDa protein was undertaken using a variety of methods. Transfer of the recombinant 62 kDa sequence to PVDF membrane and subsequent amino acid sequence analysis indicated that the amino terminus of the 62 kDa protein was intact, with the exception that the N-terminal methionine that had been introduced to ensure correct initiation of translation had been removed. The overall yields in each of the first five cycles during sequencing were low, indicating that a significant portion of the 62 kDa protein was probably blocked at the amino terminus.

Tryptic peptide analysis revealed as many as 143 peaks by reversed-phase HPLC. Eight peaks were selected for LD–MS to determine the structural integrity and the potential post-translational modifications. Of these eight peaks, four gave rise to single species peptides. Three of these peptides matched various internal regions of the 62 kDa protein, as determined by Edman degradation sequencing. One peak, peak 65, did not yield a sequence by Edman degradation. However, the molecular mass agreed very well with the predicted mass for the amino terminal tryptic peptide, taking into account the removal of the N-terminal methionine by a cellular aminopeptidase followed by the acylation of the adjacent alanine residue. Post

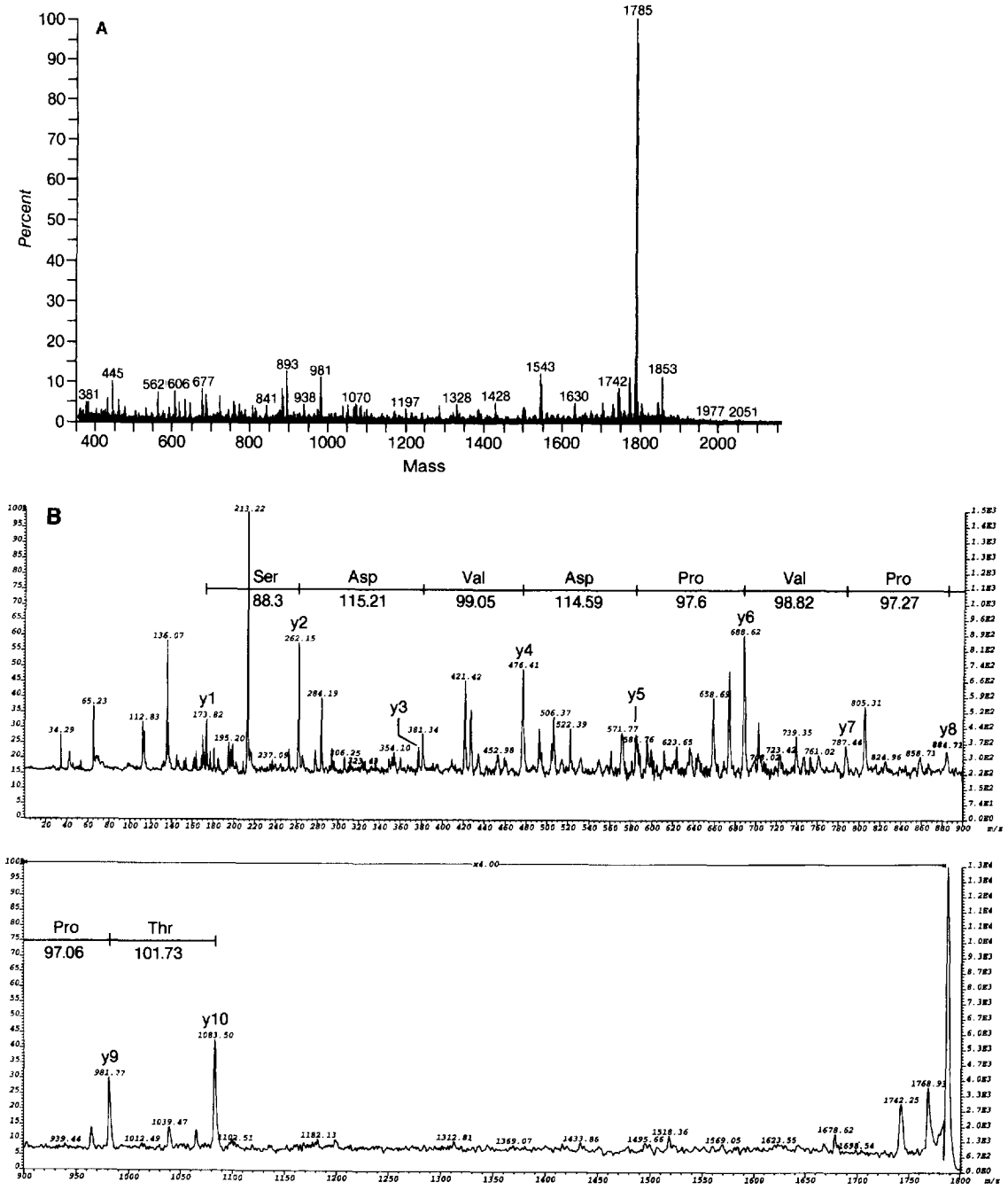

Fig. 5. Mass spectrometry analysis of tryptic peptide 65. (A) LD–MS of tryptic peptide 65. LD–MS was carried out on a 3-μl volume of tryptic peptide 65 from RP-HPLC using a VG/Fisons TofSpec mass spectrometer that was operated in the positive linear mode, at an accelerating voltage of 25 kV. The instrument was equipped with a nitrogen laser (337 nm) and a 0.65-m linear flight tube. The data for peak 65 indicates a molecular mass of 1786.81 Da. The predicted mass for this peptide (residues 1–17) is 1743.9 Da. (B) Post Source Decay on blocked peptide 65. A sequencing ladder consisting of residues 8–16 was generated by post source decay of the blocked peptide using a VG TofSpec SE LD–MS.

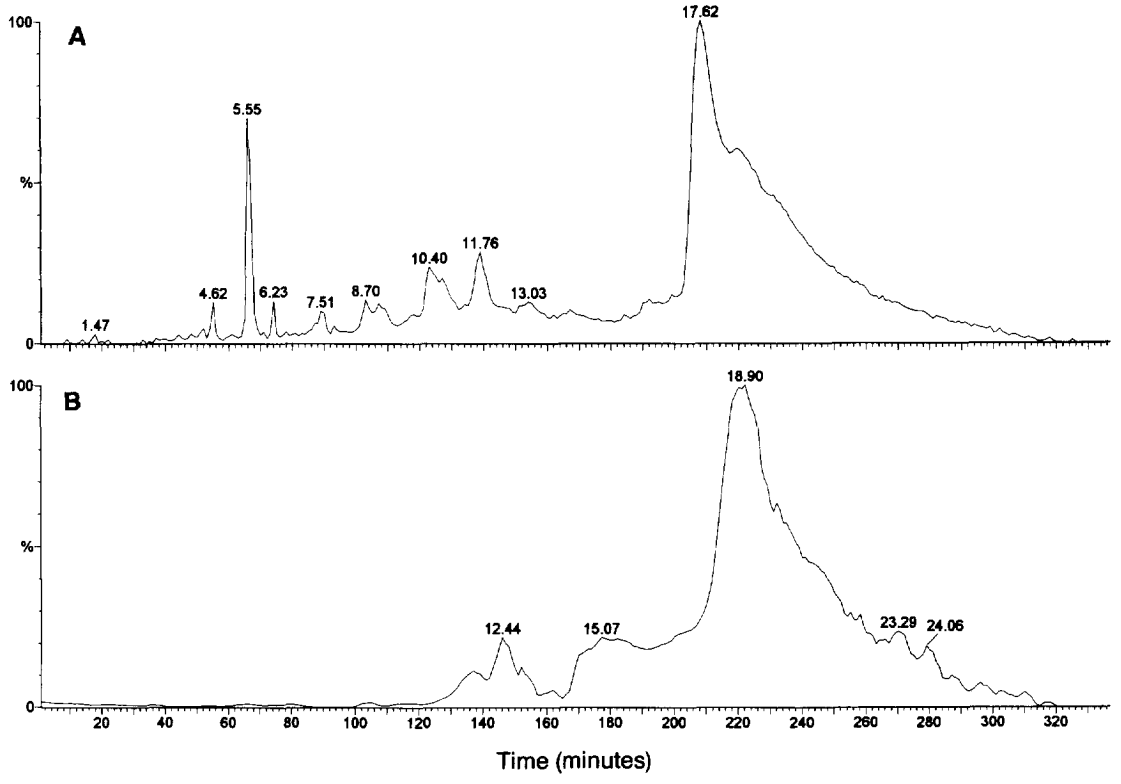

Fig. 6. Microbore reversed-phase HPLC analysis of r62-kDa. Microbore LC was performed on the recombinant 62 kDa protein using an ABI syringe pump model 410 connected to a Vydac C_{18} microbore 50×1 mm reversed-phase column at a flow-rate of 50 μl/min. The starting buffer (buffer A) was 0.1% TFA in water and the gradient buffer (buffer B) consisted of 0.1% TFA in acetonitrile–water (9:1, v/v). Running conditions consisted of 0% B for 5 min followed by a linear gradient to 100% B over 55 min. (A) UV trace of column eluent. (B) Total ion content of column eluent.

source decay analysis by laser desorption mass spectrometry indicated that peak 65 was the amino terminal tryptic peptide. The other four tryptic digest HPLC peaks that gave rise to multiple peptide species were also sequenced and provided further confirmation that the authentic HEV sequence was preserved. LC–MS data established the true molecular masses of the 62 kDa doublet that was observed by SDS–PAGE. With the elucidation of the amino terminus in previous experiments, it was possible from ES–MS data to predict the putative carboxyl terminal processing steps that gave rise to the bimodally distributed '62 kDa' species. Automated carboxyl terminal sequencing validated the carboxyl terminal processing of the protein and firmly established residues 551–552 and 536–537 as the carboxyl termini of the 58.1 kDa and the 56.5 kDa proteins, respectively. Thus, the intended protein was expressed at high levels and purified to greater than 95% purity, although the apparent carboxyl terminal processing was not anticipated. This processing did not appear to interfere with the ability of the protein to serve as an effective immunogen however, as the purified doublet material elicited protective immunity in a group of macaques [10].

The baculovirus expression system (BEVS) has proven to be an excellent system for the expression and production of foreign gene products [18]. In particular, the generation of viral-like particles (VLPS) has been well documented [19]. These VLPS have been found in several studies to represent superior antigens for diagnostic or vaccine development purposes. In previous studies, we found that a 62 kDa HEV ORF-2-derived protein produced in

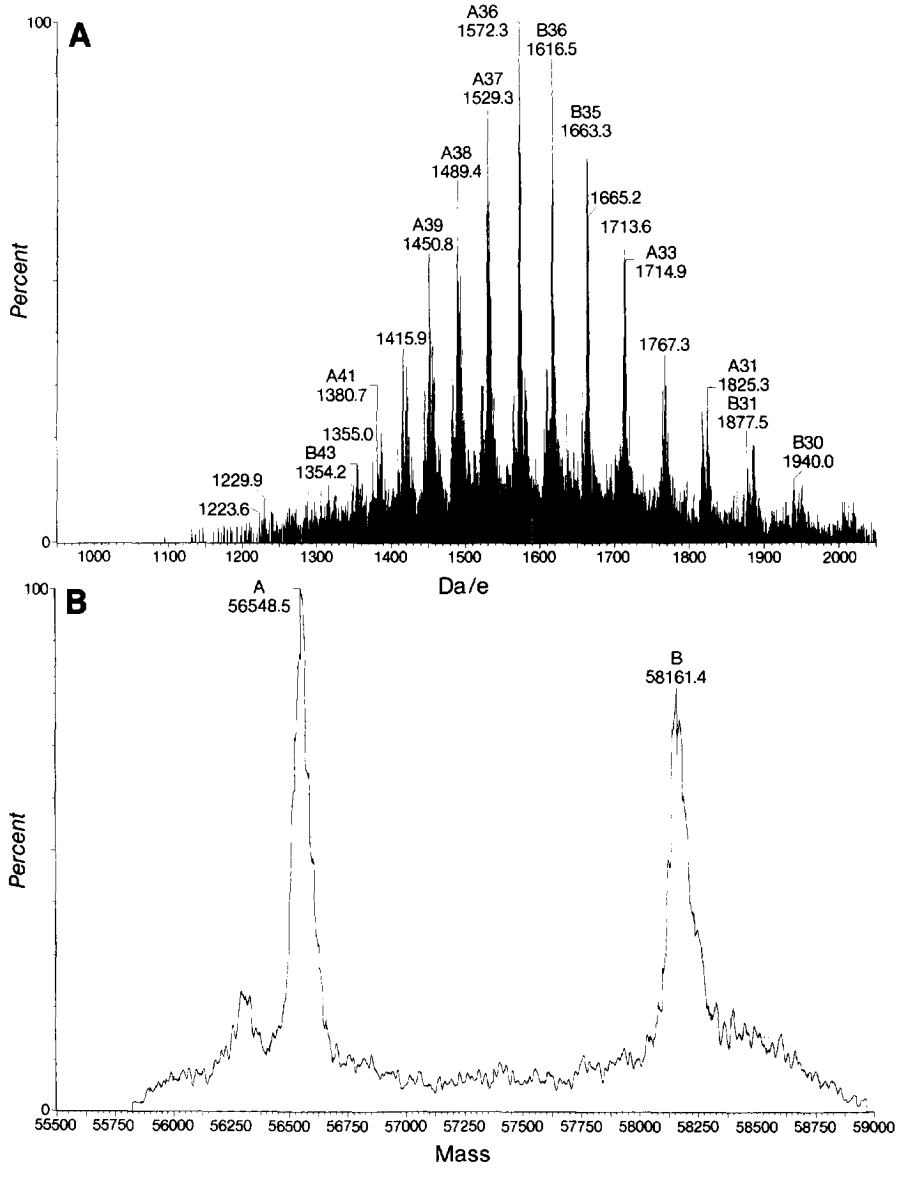

Fig. 7. LC–MS electrospray mass spectrometry analysis of r62-kDa. MS detection of capillary eluent was achieved using a VG BioQ triple quadrupole mass spectrometer operating in the positive ion electrospray ionization mode. The coupled LC–MS was facilitated by a Carlo Erba Phoenix 20CV pump and the use of a flow splitter achieving an on-line flow-rate of 10 μl/min. (A) Positive ion ES–MS multiple charged spectra. (B) Deconvoluted spectra.

baculovirus represented an improved antigen in comparison to bacterial-expressed proteins in HEV diagnostic assays [8,9]. The excellent immunogenic properties of this antigen were also apparent as we were able to elicit protective immune responses in primates after heterologous challenge with HEV [10]. These observations suggest that the baculovirus-expressed protein may contain an immunological structure that closely resembles the native virus capsid protein. Further evaluation of the recombinant

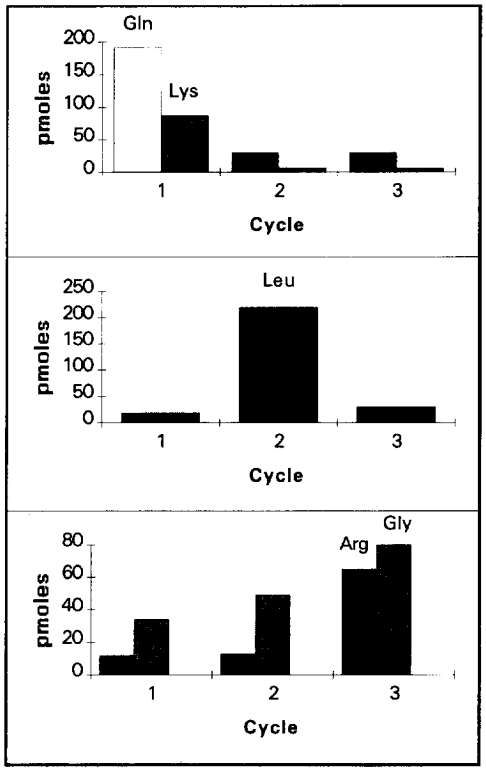

Fig. 8. Carboxy terminal sequencing analysis of r62-kDa. Recombinant 62 kDa protein samples were spotted in amounts greater than 200 pmoles onto isopropanol-treated Zitex membranes that were inserted into Hewlett Packard sequencing columns and sequenced by a Hewlett Packard G1009A automated carboxy-terminal sequencing system. The sequences were determined to be:
Gln(1) – Leu(2) – Gly(3)

Lys(1) – Leu(2) – Arg(3)
(A) Histogram plot of cycle 1 analysis indicating the presence of two C-terminal amino acids. (B) Histogram plot of cycle 2 analysis indicating the presence of more than one leucine in the heterogenous mixture at position 2. (C) Histogram plot of cyle 3 indicating the presence of two possible amino acids in the third position of the C-terminal sequence.

Table 1
Purification Table

	Protein (mg)	r62-kDa (mg)	Yield (%)	Purification
Cell lysate	200.92	8.20	–	1
Supernatant	190.03	8.14	99	1.05
DEAE EMD 650(S)	7.79	7.01	85	20.91
Sephacryl S-100	5.40	5.08	62	21.87
Poros HQ/F	4.58	4.50	55	22.85

62 kDa protein is in progress to address the issue of particle formation as well as more in-depth biochemical analysis of the protein itself.

Acknowledgments

This work was supported by USPHS Grant 1 R43 AI35400-01 from the National Institutes of Allergy and Infectious Diseases. We thank each of our collaborators, Drs. Bradley, Ticehurst and Skidmore, for kindly providing serum samples. We would also like to thank Dr. Jeff Lifson for his critical review of this manuscript.

References

[1] M.S. Balayan, A.G. Andjaparidze, S.S. Sarinskaya, E.S. Ketiladze, D.M. Braginsky, A.P. Savinov and V.F. Poleschak, Intervirology, 20 (1983) 23.
[2] D.W. Bradley, K. Krawczynski, E.H. Cook, Jr., K.A. McCaustland, C.D. Humpherey, J.E. Spelbring, H. Myint and J.E. Maynard, Proc. Natl. Acad. Sci. U.S.A., 84 (1987) 6277.
[3] K. Krawczynski, Hepatology, 17 (1993) 932.
[4] V.A. Arankillo, J. Ticehurst and M.A. Sreenivason, Lancet, 1 (1988) 550.
[5] A.W. Tam, M.M. Smith, M.E. Guerra, C.C. Huang, D.W. Bradley, K.E. Fry and G.R. Reyes, Virology, 185 (1991) 120.
[6] G. Reyes and B.M. Baroudy, Adv. Virus Res., 40 (1991) 57.
[7] P.O. Yarbough, A.W. Tam, K.E. Fry, K. Krawczynski, K.A. McCaustland, D.W. Bradley and G.R. Reyes, J. Virol., 65 (1991) 5790.
[8] Y. Zhang, P. McAtee, P.O. Yarbough, M. Qiu, T. Fuerst and A. Tam, Virology, submitted for publication.
[9] C.P. McAtee, Y. Zhang, P.O. Yarbough, T.R. Bird T.R. and Fuerst, Protein Expr. Purif., in press.
[10] P.O. Yarbough, A. Tam, Y. Zhang, P. McAtee, J. Lifson, K. McCaustland, J. Spellbring, J.-P. Prieels, N. Garcon, M. Slaoui, F. Myriam, K. Krawczynski and T. Fuerst, J. Infect. Disease, submitted for publication.
[11] J. Rosenfeld, J. Capdevelle, J. Guillemot and P. Ferrara, Anal. Biochem., 203 (1992) 173.
[12] K.R. Williams and K.L. Stone, in J. Crabb (Editor), Techniques in Protein Chemistry VI, Academic Press, New York, 1995, pp. 143–153.
[13] K.R. Williams, S. Samandar, K.L. Stone, M. Saylor and J. Rush, in J.M. Walker (Editor), Protein Protocols Handbook, Humana Press, Toledo, 1996, in press.
[14] J.I. Elliott, K.L. Stone and K.R. Williams, Anal. Biochem., 211 (1993) 94.
[15] E. Schagger and O. von Jagow, Anal. Biochem., 166 (1987) 368.

[16] H.A. Scoble, J.E. Vath, W. Yu and S.A. Martin, in P. Matsudaira (Editor), A Practical Guide to Protein and Peptide Purification, Academic Press, New York, 1993, pp. 125–153.

[17] P.O. Yarbough, A.W. Tam, K. Gabor, E. Garza, R.A. Moeckli, I. Palings, C. Simonsen and G.R. Reyes, in K. Nishioka, H. Suzuki, S. Mishiro and T. Oda (Editors), Viral Hepatitis and Liver Disease, Springer-Verlag, Tokyo, 1994, pp. 367–370.

[18] D.L. Jarvis, in M.F.A. Goosen, A.J. Daugulis and P. Faulkner (Editors), Insect Cell Culture Engineering, Marcel Dekker, New York, 1993, pp. 195–220.

[19] M.V. Neermut, D.J. Hockley, J.B.M. Jowett, I.M. Jones, M. Garreau and D. Thomas, Virology, 198 (1994) 288.