

## Analysis of the Adenovirus Type 5 Proteome by Liquid Chromatography and Tandem Mass Spectrometry Methods

Dirk Chelius,<sup>\*,†</sup> Andreas F. R. Hühmer,<sup>†</sup> Chia H. Shieh,<sup>†</sup> Elisabeth Lehmberg,<sup>‡</sup>  
 Joseph A. Traina,<sup>‡</sup> Timothy K. Slattery,<sup>‡</sup> and Erno Pungor Jr.<sup>‡</sup>

*Thermo Finnigan, 355 River Oaks Parkway, San Jose, California 95128, and  
 Berlex Biosciences, 15049 San Pablo Avenue, Richmond, California 94804*

Received April 24, 2002

We compared detection sensitivity and protein sequence coverage of the adenovirus type 5 proteome achievable by liquid chromatography and tandem mass spectroscopy (LC/MS/MS) using three sample preparation and clean up methods. Tryptic digestion was performed on either purified viral proteins or whole virus, and followed by shotgun sequencing using tandem mass spectrometry for peptide identification. We used a recombinant adenovirus type 5 as a test system. The methods included separation of adenoviral proteins by reversed-phase high-performance liquid chromatography followed by tryptic digestion and analysis by LC/MS/MS. Alternatively, the purified whole virus was digested with trypsin and the peptides separated either by one-dimensional (reversed-phase) or by two-dimensional (cation exchange and reversed-phase) chromatography and analyzed by tandem mass spectrometry. A total of 11 protein species were identified from 154 peptides. All of the major viral proteins were found. In addition, two minor proteins, the 23 kDa viral protease and the late L1 protein, were identified for the first time by chromatography based assays. The 23 kDa viral protease, present at only 10 copies per virus, and representing 0.2% of the protein content of the virus, was detected by the 2D LC/MS/MS analysis of the whole virus digest from a sample containing only 70 fmols of the protein. This demonstrates the high sensitivity and selectivity of the method. The 2D LC/MS/MS analysis of the whole virus digest was also able to detect all viral proteins with copy numbers at or above 10/virus particle, with broad coverage of the amino acid sequences. Coverage ranged from 2 to 54%, a majority between 20 and 35%, suggesting the possibility of using this analysis to assess the purity of the virus preparations. This broad coverage may also provide a useful approach to identify posttranslational modifications on the structural proteins of the adenovirus.

**Keywords:** tandem mass spectroscopy • adenovirus type 5 • protein identification • shotgun sequencing

### Introduction

Mass spectrometry methods, in particular tandem MS (MS/MS), provide a fundamental set of tools for proteome analysis. To illustrate the power of such methods, Washburn et al. identified 1484 proteins of the yeast proteome in a single assay by performing a separation of the tryptic peptides generated from whole yeast cell lysate and "shotgun sequencing" by LC/MS/MS.<sup>1</sup> The LC portion of the analysis is normally based on reversed-phase separation. This large-scale protein identification technology is only limited by the ability to adequately separate peptides of suitable size for MS/MS fragmentation<sup>2</sup> and the detection limits of the mass spectrometer.

It is only a matter of a short time before these powerful analytical tools, used primarily in research today, will also be

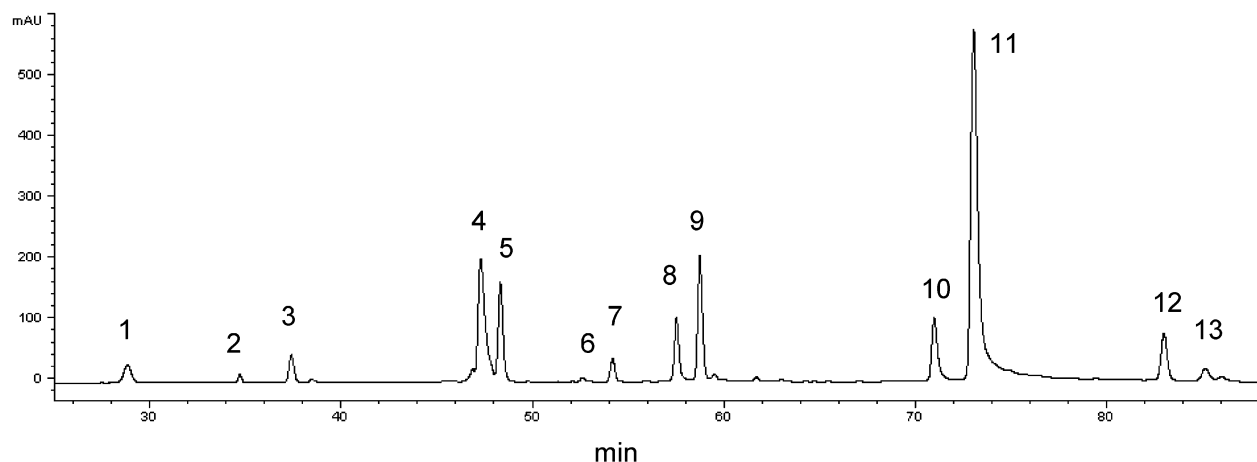
applied to the routine characterization of complex biological drugs in support of manufacturing. The drug targets could include protein preparations with high degrees of heterogeneity or more complex biological systems such as therapeutic viruses and cells. The goals of such characterization could include purity analysis, identification of posttranslational and/or post-production modifications, assessment of the effects of different manufacturing schemes on complex drugs, and so forth.

The objective of our study was to evaluate the performances (sensitivity and attainable protein sequence coverage) of one-dimensional (1D) and two-dimensional (2D) LC/MS/MS analysis of the adenovirus proteome. The adenoviral proteome is a very appropriate test system for such comparisons: it is fairly complex but well understood and characterized. It consists of 11 protein species, with a total of only about 2500 molecules comprising each virus particle. Quantitative information is available on copy numbers based both on electron microscopy results with symmetry considerations and direct measurements with various other methods.<sup>3-7</sup> As there are large differences in the copy numbers of the structural proteins (covering almost

\* To whom correspondence should be addressed. Dirk Chelius, Thermo Finnigan, Proteomics Division, 355 River Oaks Parkway, San Jose, CA 95128. Phone: (408) 965-6326. Fax: (408) 9656138. E-mail: dchelius@thermofinnigan.com.

<sup>†</sup> Thermo Finnigan.

<sup>‡</sup> Berlex Biosciences.



**Figure 1.** Reversed-phase HPLC chromatogram of the adenovirus type 5 proteome as described in the Experimental Section.  $1.63 \times 10^{12}$  viral particles were injected into the column and chromatograms were obtained at 214 nm. No peaks appeared in the first 15 min of the profile, with the exception of the flow-through (system) peak that contained no detectable protein. Fractions were collected manually for further analysis of the 13 peaks detected.

2 orders of magnitude), the selectivity, sensitivity, and dynamic range of the different 1D and 2D LC/MS/MS approaches can be assessed using the adenovirus as a test system.

We decided to evaluate three analytical strategies involving different sample preparation techniques and levels of sample clean up before or during the LC/MS/MS analysis. The first of these used viral protein fractions (separated by a reversed-phase HPLC method<sup>7</sup>), that were reduced, alkylated, digested with trypsin, and introduced into a standard 1D LC/MS/MS analysis where the analytical 1D LC was a reversed-phase separation. The two other approaches used whole virus preparations. These were reduced and alkylated, digested with trypsin, and introduced into a 1D (reversed-phase separation) or a 2D (cation exchange followed by reversed-phase separation) LC/MS/MS analysis system. In all cases, proteins were identified by “shotgun sequencing” using tandem mass spectrometry data of tryptic peptides and database searching by the SEQUEST algorithm.<sup>8</sup>

### Experimental Section

**Materials.** Adenovirus type 5 samples were prepared using HEK 293 cells adapted to serum free medium and suspension culture. The samples were purified by anion-exchange chromatography and ultrafiltration at Berlex Biosciences. Unless otherwise noted, all materials were obtained from Sigma (St Louis, MO).

**HPLC Separation of Viral Proteins.** Viral proteins were separated by reversed-phase HPLC as described earlier.<sup>7</sup> In short, the RP-HPLC experiments were performed on an Agilent 1100 HPLC unit equipped with a diode-array detector, auto-sampler, micro flow cell and temperature controlled column compartment (Agilent, Palo Alto, CA). A Jupiter column (250  $\times$  4.6 mm i.d.) packed with a 5  $\mu$ m nominal diameter, 300 Å pore size C<sub>4</sub> resin (Phenomenex, Torrance, CA) and a precolumn filter (0.5  $\mu$ m pore size, Alltech, Deerfield, IL) were used. The solvents were (A) 0.1% trifluoroacetic acid (TFA; Pierce, Rockford, IL) in water, and (B) 0.1% TFA in acetonitrile (Baker, Phillipsburg, NJ). Virus samples, without pretreatment, were injected into the RP-HPLC column. The column was equilibrated at 20% solvent B. A linear gradient from 20 to 65% B was run over 95 min. The column temperature was maintained at 40 °C. A total of  $1.63 \times 10^{12}$  virus particles were injected onto the column for fraction collection. The chromatogram is shown

in Figure 1. The fractions were collected manually as indicated on the Figure 1.

**Reduction, Alkylation, and Digestion.** Each HPLC fraction was concentrated to 10  $\mu$ L with a speed vac. The samples were reduced by adding 90  $\mu$ L of ammonium bicarbonate (100 mM pH 8.5) and DTT (3  $\mu$ L of a 100 mM solution in H<sub>2</sub>O). The mixture was incubated for 30 min at 37 °C. To alkylate the protein, iodoacetic acid (7  $\mu$ L of 100 mM solution in 100 mM KOH) was added, and the mixture was incubated for an additional 30 min at room temperature in the dark. An additional 13  $\mu$ L of DTT (100 mM solution in H<sub>2</sub>O) was added to react with the excess iodoacetic acid. The reduced and alkylated proteins were digested by adding 2  $\mu$ L of trypsin (Promega, Madison, WI, 0.5 mg/mL). The mixture was incubated for 6 h at 37 °C, then an additional 2  $\mu$ L of trypsin was added (final sample volume: 127  $\mu$ L) and incubation was continued for 16 h at 37 °C.

Purified virus at a concentration of  $2.49 \times 10^{11}$  particles/ml was used for whole virus digestion experiments. Aliquots of 150  $\mu$ L ( $3.73 \times 10^{10}$  particles, corresponding to a protein content of approximately 10  $\mu$ g) were reduced in the presence of 15  $\mu$ L of acetonitrile with DTT (100 mM, 3  $\mu$ L, final concentration at 1 mM). The sample was incubated overnight at 37 °C. Iodoacetic acid (100 mM, 7  $\mu$ L) was added. The mixture was incubated for an additional 30 min at room temperature in the dark to alkylate the protein. An additional 13  $\mu$ L DTT (100 mM) was added to react with the excess iodoacetic acid. The reduced and alkylated whole virus preparation was digested by adding 4  $\mu$ L of trypsin. The mixture was incubated for 6 h at 37 °C, then an additional 4  $\mu$ L of trypsin was added (final sample volume: 196  $\mu$ L) and incubation was continued for 16 h at 37 °C.

**One-Dimensional LC/MS/MS.** The digested samples were analyzed using a fully automated nanoflow LC/MS/MS system (Figure 2). Aliquots of 10  $\mu$ L (whole virus digest) and 20  $\mu$ L (digested and collected HPLC fractions) were placed in wells of a 96-well plate (Nalge Nunc International, Rochester, NY). The plate was sealed with plastic film to minimize evaporation and inserted into the auto-sampler of a Surveyor HPLC system (Thermo Finnigan, San Jose, CA), where it was kept at 4 °C while awaiting analysis (maximum waiting time not more than 8 h). The auto-sampler was equipped with no-waste injection capability, which enables injection volumes as low as 1  $\mu$ L. The

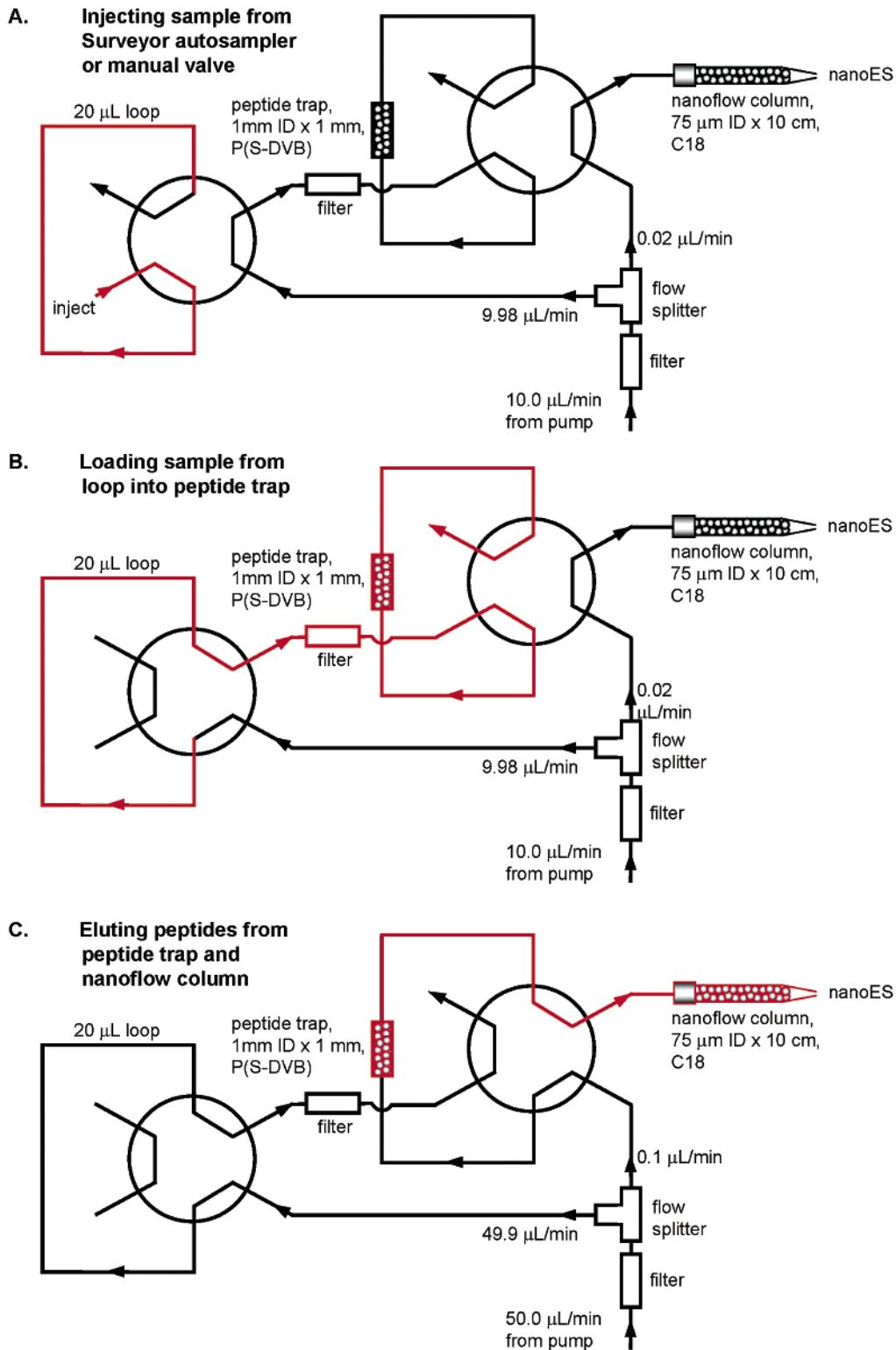


Figure 2. Nanoflow system used in this study. Location of sample is shown in red.

injected peptides were first loaded onto a reversed-phase poly(styrene-divinylbenzene) peptide trap (Michrom Bioresources, Auburn, CA) with a flow rate of 10  $\mu\text{L}/\text{min}$  for 3 min (Figure 2). The peptides were eluted from the trap and separated on a reversed-phase capillary column (PicoFrit; 5  $\mu\text{m}$  BioBasic C18, 300  $\text{\AA}$  pore size; 75  $\mu\text{m}$  x 10 cm; tip 15  $\mu\text{m}$ , New Objective, Woburn, MA) with a 30-min linear gradient of 0 to 60%

acetonitrile in 0.1% formic acid/water at a flow rate of approximately 0.1  $\mu\text{L}/\text{min}$  after split. The HPLC was directly coupled to a Thermo Finnigan LCQ DecaXP ion trap mass spectrometer equipped with a nano-spray ionization source. The spray voltage was 2.0 kV, and the capillary temperature was 150  $^{\circ}\text{C}$ . The ion-trap collisional fragmentation spectra were obtained using collision energies of 35%. Each full scan mass

spectrum was followed by three data dependent MS/MS spectra of the three most intense peaks. The Dynamic Exclusion feature was enabled (repeat counts, 2; repeat duration, 0.2 min; exclusion duration, 5 min; and exclusion mass width, 2 Da).

**Two-Dimensional LC/MS/MS.** Whole virus digests (aliquots of 17  $\mu$ L) were also analyzed using a fully automated two-dimensional LC/MS/MS system (ProteomeX, Thermo Finnigan, San Jose, CA). The flow-rate of the mobile phase was maintained at 200  $\mu$ L/min before splitting and at 1.2  $\mu$ L/min after the splitting. The first dimension of the separation was performed on a strong cation exchange column (BioBasic-SCX; 300 Å pore size; 0.32 mm  $\times$  100 mm, Thermo Hypersil-Keystone, Bellefonte, PA). The column was equilibrated in 5% ACN and 0.1% formic acid in water. After loading the samples, the peptides were eluted in a sequential step gradient with 0, 50, 100, 200, 300, 400, and 500 mM ammonium chloride (in 5% ACN and 0.1% formic acid in water). Each salt gradient step was maintained for 50 min. Eluted samples from the cation exchange were directly loaded onto a reversed-phase column (BioBasic-C18, 300 Å pore size; 0.18 mm  $\times$  100 mm, Thermo Hypersil-Keystone, Bellefonte, PA) to perform separation in the second dimension. The reversed-phase column was equilibrated with 5% ACN and 0.1% formic acid in water, loaded and reequilibrated with the same solvent for 3 min. The peptides were eluted in a 30 min linear gradient from 0 to 65% ACN and 0.1% formic acid in water, followed by a 5 min ramp to 80% ACN and 0.1% formic acid in water and a hold at this condition for another 5 min. The 2D-LC system consisted of two Surveyor HPLC pumps and the Surveyor Autosampler and was directly coupled to the orthogonal API source of a Thermo Finnigan LCQ DecaXP ion trap mass spectrometer.

Samples eluting from the reversed-phase separation column were introduced to MS via the ion source chamber of the ESI source with a 30  $\mu$ m ID metal spray needle. The temperature of the ion transfer tube was set at 160 °C. The spray voltage was set at 2.6 kV, sheath gas flow at 6 units and the normalized collision energies were set at 35% for MS/MS. Each full mass spectrum was followed by three MS/MS spectra of the three most intense peaks. Dynamic exclusion was enabled using the following conditions: Repeat count, 2; repeat duration, 0.2 min; exclusion duration, 5 min; and exclusion mass width, 2Da.

**Data Analysis.** Peptides and proteins were identified automatically by the computer program BioWorks version 3.1 (Thermo Finnigan, San Jose, CA) which correlated the experimental tandem mass spectra against theoretical tandem mass spectra from human and adenovirus proteins using the SEQUEST algorithm.<sup>8</sup> The database for the human and adenovirus proteins were obtained from the National Center for Biotechnology Information (NCBI) and the adenovirus protein database was appended to the database containing the human proteins. No enzyme parameter was chosen, and the alkylated cysteines were examined as static modifications. Peptide identification was further evaluated using the "Unified Score" filter of the BioWorks software, combining all three correlation coefficients generated by the SEQUEST algorithm. The score was calculated according to the following formula: Score = (10 000  $\times$  DelCn<sup>2</sup> + Sp)  $\times$  Xcorr.<sup>8</sup> The equation was developed empirically in house and tested on several thousands of analyses. Only peptides with a score of more than 2400 were accepted.

## Results

**Analysis of Collected RP-HPLC Fractions.** The results of the 1D LC/MS/MS analysis of the collected HPLC fractions after

**Table 1.** Identification of Viral Proteins from the Collected Reversed-Phased HPLC Fractions in Figure 1

fraction	protein ID	no. of peptides identified	sequence coverage (%)	model or biochemical copy numbers <sup>g</sup>
1	ND			
2	IX	1	9	240
3	N-terminal fragment of VIII after processing by the 23 kD protease <sup>a,b</sup>	2	17	
4	mature VII <sup>a,c</sup>	8	30	833
5	V <sup>a,c</sup>	9	30	157
6	VIII	3	20	
7	C-terminal fragment of VIII after processing by the 23 kD protease <sup>a,d</sup>	1	20	
8	III <sup>a</sup> , 23 kD protease	17, 8	29, 47	60, 10
9	mature VI <sup>a</sup> , 23 kD protease	7, 8	36, 47	360, 10
10	mature III <sup>a</sup>	10	26	60
11	II <sup>a</sup>	22	37	720
12	mature IX <sup>a</sup> , II <sup>f</sup>	8	54	240
13	L1, II <sup>f</sup>	4	10	

The numbering of the fractions is the same as in Figure 1. For experimental details, see the Experimental Section. <sup>a</sup> The protein identified is the same as assigned to the corresponding RP-HPLC peak in the study reported by Lehmeberg et al.<sup>7</sup> <sup>b</sup> The peptides ILLEQAAITTPR and NNLNPR were detected from the N-terminal fragment of the full length protein VIII, which is processed by the 23 kD viral protease at G<sub>111</sub>.<sup>7</sup> This processing site was taken into account for the sequence coverage calculation. <sup>c</sup> The collected samples showed traces of proteins in both fractions. The sequence coverage is given for the major protein detected in the peak. <sup>d</sup> The peptide QAIL-TLQTSSEPR was detected from the C-terminal fragment of the full length protein VIII, starting from the site processed by the 23 kD viral protease at N-terminus of G<sub>157</sub>.<sup>7</sup> This processing site was taken into account for the sequence coverage calculation. <sup>e</sup> Sequences for the 23 kD viral protease were detected in both fractions 8 and 9; 5 peptides in fraction 8 (LACAIVNTAGR, FPGFVSPHK, NQEQLYSFLER, DLGCGPYFLGTGDK, QVYQFEYESLLR) and with 6 peptides in fraction 9 (LACAIVNTAGR, FPGFVSPHK, NQEQLYSFLER, TCYLFEPFGFSDQR, SAIASSPDRCTLEK and SATSFCHLK). The sequence coverage (47%) was calculated for all detected peptides in the two fractions. <sup>f</sup> Traces of protein II were detected with several identifiable peptides in fractions 12 and 13, likely due to tailing. Sequence coverage for protein II was not reported in these fractions. <sup>g</sup> Copy numbers were obtained from Stewart et al.<sup>5</sup>

reduction, alkylation, and tryptic digestion are summarized in Table 1. The peptides detected from each protein are shown in Tables 2–12.

As anticipated, the protein identification data in the 1D LC/MS/MS assay are in very good agreement with the results of our previous investigation.<sup>7</sup> In that study, collected RP-HPLC fractions were analyzed by MALDI-TOF and N-terminal sequencing following treatment with Chloramine T to perform oxidative cleavage at the C-terminal side of tryptophan residues to obtain internal sequences of proteins with blocked N-termini. There were, however, some important differences between the results of the two studies. Although some viral proteins identified in the RP-HPLC chromatogram in our previous work were not detectable (e.g., the precursor part of proteins VI and VII, and the L2  $\mu$  protein), the current study led to some new peak assignments in the reversed-phase chromatographic profile of adenovirus.<sup>7</sup>

Fraction 2 contained one peptide (ELNVVSQQLDLR, identified based on MS/MS spectra) attributable to protein IX. Protein IX could also be identified in fraction 12 with 7 peptides (see also Table 10), consistent with the full-length protein as anticipated. We speculate that fraction 2 may contain a fragment or a minor isoform of protein IX (214 nm peak area is much smaller in fraction 2 than 12 in Figure 1).

Table 2. Amino Acid Sequence of Viral Protein II<sup>a</sup>

	ATPSMMPQWSYMHISGQDASEYLSPLVQFAR <b>ATETYFSLN</b> <b>FR</b> <b>NPTVAPTHDVT</b> <b>TDR</b> SQR LTLR
HPLC	***** ( ** ***** )
1D	
2D	
	<b>FIPVDR</b> <b>EDTAYS</b> <b>YK</b> <b>AR</b> <b>FTLAVGDNR</b> <b>VLDMASTYFDIR</b> GVLDR GPTFK PYSGTAYNALAPK
HPLC	*****
1D	*****
2D	
	GAPNPCEWDEAATALEINLEEDDDNEDEVDEQAEQQK <b>THVFGQAPYSGINITK</b> <b>EGIQIGVEGQTPK</b>
HPLC	*****
1D	*****
2D	*****
	YADK <b>TFQPEPQIGESQWYETEINHAAGR</b> VLK K <b>TTPMK</b> <b>PCYGSYAK</b> <b>PTNENGGQILVK</b>
HPLC	*****
1D	*****
2D	
	QQNGK LESQVEMQFFSTTEATAGNGDNLTPK <b>VVLYSEDVDIETPDTHISYMP</b> <b>TIK</b> EGNSR
HPLC	*****
1D	
2D	
	<b>ELMQQSM</b> <b>PNR</b> <b>PNYIAFR</b> DNFIGLMYYNSTGNMGVLAGQASQLNAVVDLQDR <b>NTELSYQLLLDSIGDR</b>
HPLC	*****
1D	*****
2D	
	TR YFSMWNQAVDSYDPDVR IENHGTEDELPNYCFPLGGVINTETLTK VK PK <b>TGQENGWEK</b>
HPLC	*****
1D	
2D	
	<b>DATEFSDK</b> NEIR VGNNFAMEINLNANLWR <b>NFLYSNIALYLPDK</b> LK YSPSNVK <b>ISDNPNTYDYMNK</b>
HPLC	*****
1D	*****
2D	
	<b>R</b> <b>VVAPGLVDCYINLGAR</b> WSLDYMDNVNPFNHHR NAGLR YR SMLLGNGR <b>YVPFHIQVPQK</b> FFAIK
HPLC	*)
1D	
2D	*****
	NLLLLPGSYTYEWNFR K <b>DVNMVLQSSLGNDLR</b> VDGASIK FDSICLYATFFPMAHNTASTLEAMLR
HPLC	*****
1D	
2D	
	NDTNDQSFNDYLSAANMLYPIPANATNPISIPSR <b>NWAAFR</b> GWAFTR LK TK
HPLC	*****
1D	
2D	
	ETPSLGSYDPYYTYSGSIPYLDGTFYLNHTFK K <b>VAITFDSSVSWPGNDR</b> <b>LLTPNEFEIK</b> <b>R</b>
HPLC	*****
1D	
2D	
	<b>SVDGEGYNVAQC</b> <b>NMTK</b> DWFLVQMLANYNIGYQGFYIPESYK DR MYSSFR <b>NFQPM</b> <b>SR</b> QVVDLTK YK
HPLC	*****
1D	*****
2D	
	DYQQVGILQHNNSGFVGYLAPTMR <b>EGQAYPANFPYPLIGK</b> TAVDSITQK K FLCDR TLWR
HPLC	*****
1D	
2D	
	IPFSSNFMSMGALTDLGQNLLYANSAHALDMTFEVDPMDEPTLLYVLFVFDVVR VHR PHR
HPLC	
1D	
2D	
	<b>GVIETVYL</b> <b>TPFSAGNATT</b>
HPLC	*****
1D	
2D	

<sup>a</sup> Peptides detected from either the analysis of the HPLC fractions or the whole virus digestion and analysis by 1D or 2D chromatography are labeled (\*). The predicted tryptic digestion sites (C-termini of R and K residues) are marked with a space. All amino acids of the detected peptides are typed in **bold**, the first amino acid of the confirmed mature protein is marked with **bold underline**. All other consensus cleavage sites for the 23 kD protease are marked with **bold italic underline**. Parenthesis "(" and ") indicate that both peptides, with and without the missed tryptic cleavage were identified.

The analysis of fraction 3 resulted in two peptides from the N-terminal region of protein VIII, whereas fraction 7 showed one peptide in the C-terminal region, consistent with the

N-terminal and C-terminal segments of the processed protein, as anticipated.<sup>7</sup> Interestingly, protein VIII was also identified with all three of these peptides in fraction 6, possibly indicating

Table 3. Amino Acid Sequence of Viral Protein III<sup>a</sup>

	MR R AAMYEEGPPPSYESVVSAPVAAALGSPFDAPLDPPFVPPR <b>YLR PTGGR</b> NSIR
HPLC	
1D	*****
2D	
	<b>YSELAPLFDTR VYLVDNK</b> STDVASLNYQNDHSNFLTTVIQNNDYSPGEASTQTINLDDR
HPLC	*****
1D	*****
2D	*****
	<b>SHWGGDLK TILHTNMPNVNEFMFTNK</b> FK AR VMVSR <b>LPTK DNQVELK</b>
HPLC	*****
1D	*****
2D	
	YEWVEFTLPEGNYSETMTIDLMNNAIVEHYLK VGR QNGVLES DIGVK FDTRNFR
HPLC	
1D	
2D	
	LGFDVPTGLVMPGVYTNEAFHPDIILLPGCGVDFTHSR <b>LSNLLGIR K R QPFQEGFR</b>
HPLC	***** (*** ) *****
1D	( * *****
2D	***** *****
	<b>ITYDDLEGGNIPALLDVDAYQASLK</b> DDTEQGGGAGGSNSGSGAEENSNAAAAAMQPVEDMNDHAIR
HPLC	
1D	*****
2D	
	GDTFATR AEEK <b>R AEAEAAAAAQAQPEVEK POK K PVIK PLTEDSK K R</b>
HPLC	***** ( *
1D	*****
2D	*****
	<b>SYNLISNDSTFTQYR SWYLAYNYGDPQTGIR</b> SWTLTCTPDVTCGSEQVYWSLPDMMQDPVTFR
HPLC	*****
1D	
2D	*****
	STR <b>QISNFPVVGAE LLPVHSK SFYNDQAVYSQLIR QFTSLTHVFNR</b> F PENQILAR
HPLC	*****
1D	
2D	*****
	PPAPTITTVSENVPALTDHGTLPLR NSIGGVQR VTITDAR R R <b>TCPYVYK ALGIVSPR</b> VLSSR TF
HPLC	*****
1D	
2D	*****

<sup>a</sup> Same as footnote a of Table 2.

an unprocessed form of protein VIII detected in empty virions.<sup>9</sup>

As shown in Table 1, we were able to make two additional new peak assignments: we found the late L1 protein in fraction 13 and the 23 kDa viral protease in fractions 8 and 9. The MS/MS spectra of one of the identified peptides in the 23 kDa protease (FPGFVSPHK) is shown in Figure 3. Fractions 8 and 9 seem to be well separated, and it is not yet understood why the protease is present in both fractions. The fact that the 23 kDa protease coelutes with protein VI in the RP-HPLC analysis of the viral proteins is very interesting. It has been reported earlier that the adenovirus protease requires activation by an 11-residue peptide, GVQSLKRRRCF, to achieve maximum proteolytic activity. Derived from the C-terminus of the viral protein VI, the activating peptide forms a disulfide bond with C<sub>104</sub> of the protease and causes a conformational change that accompanies the development of proteolytic activity.<sup>10</sup> Our results could indicate the 23 kDa protease binding to the intact protein VI.

The successful protein identifications in the 1D LC/MS/MS analysis of the collected HPLC fractions indicate that the methods used for reduction, alkylation, and tryptic digestion were appropriate. No peptides with free SH groups could be identified (data not shown), all cysteine residues were alkylated. All identified peptides were tryptic fragments and only a few peptides with missed cleavage sites were found (Tables 2–12).

They were seen mostly in K/RP sequences and in cases where a second consensus cleavage site was within 3 or less amino acid residues. This digestion pattern has been observed in many applications and does not cause any problems in our protein identification technique.

The sequence coverage of the identified proteins ranges from 54% for protein IX to 10% for the late L1 protein with no clear correlation to copy number. Note that the 10 μL sample injected to the 1D LC/MS/MS system theoretically contained the proteins of 2.56 × 10<sup>11</sup> particles, see the Experimental Section. This corresponds to 425 fmols of protein injected if the protein is present as a single copy in each virus particle. This calculation assumes full recovery of the peptides during the entire sample preparation. It is, therefore, likely that factors in sample preparation and/or peptide properties (e.g., recovery of proteins after the lyophilization, solubility of the proteins, digestibility with trypsin, ability to quantitatively elute the peptides from the peptide trap, etc.) have dominant effects on the sequence coverage.

**Analysis of Whole Virus Digest by 1D LC/MS/MS.** The analysis of the whole virus tryptic digest by 1D/MS/MS resulted in the identification of 10 viral proteins. The results are summarized in Table 13. The peptides detected for each protein are shown in Tables 2–12.

All major structural proteins (II, III, IIIa, IV, V, VI, VII, VIII, and IX) and a minor virus component, the late L1 protein, could

Table 4. Amino Acid Sequence of the Viral Protein IIIa<sup>a</sup>

	MMQDATDPAVR	<b>AA</b> LQSQP <b>S</b> GLNSTDDWR	QVMDR	<b>IMS</b> LTAR	NPDAFR	QQFQANR	LSAILEAVV <b>F</b> AR				
HPLC		*****									
1D		*****		*****							
2D		*****		*****							
	ANP <b>T</b> HEK	<b>V</b> LAI <b>V</b> NALAE <b>N</b> R	<b>A</b> IR	<b>P</b> DEAG <b>L</b> V <b>D</b> ALL <b>Q</b> R	<b>V</b> AR	<b>Y</b> NSGN <b>V</b> Q <b>T</b> N <b>L</b> DR	<b>L</b> V <b>G</b> D <b>V</b> R	E <b>A</b> V <b>A</b> Q <b>R</b>			
HPLC		*****	*****	*****	*****	*****	*****				
1D		*****	*****	*****	*****	*****	*****				
2D		*****	*****	*****	*****	*****	*****				
	ER	<b>A</b> QQ <b>Q</b> GN <b>L</b> GS <b>M</b> VAL <b>N</b> AF <b>L</b> ST <b>Q</b> PAN <b>V</b> PR	<b>G</b> Q <b>E</b> D <b>Y</b> T <b>N</b> F <b>V</b> S <b>A</b> L <b>R</b>			<b>L</b> M <b>V</b> T <b>E</b> T <b>P</b> Q <b>S</b> E <b>V</b> Y <b>Q</b> S <b>G</b> P <b>D</b> Y <b>F</b> F <b>Q</b> T <b>S</b> R					
HPLC		*****									
1D		*****		*****							
2D		*****		*****							
	Q <b>G</b> L <b>Q</b> T <b>V</b> N <b>L</b> S <b>Q</b> A <b>F</b> K	<b>N</b> L <b>Q</b> L <b>W</b> G <b>V</b> R	A <b>P</b> T <b>G</b> D <b>R</b>	A <b>T</b> V <b>S</b> S <b>L</b> L <b>T</b> P <b>N</b> S <b>R</b>	<b>L</b> L <b>L</b> L <b>L</b> I <b>A</b> P <b>F</b> T <b>D</b> S <b>G</b> S <b>V</b> S <b>R</b>						
HPLC					*****						
1D					*****						
2D		*****			*****						
	<b>D</b> T <b>Y</b> L <b>G</b> H <b>L</b> L <b>T</b> L <b>Y</b> R	E <b>A</b> I <b>G</b> Q <b>A</b> H <b>V</b> D <b>E</b> H <b>T</b> F <b>Q</b> E <b>I</b> T <b>S</b> V <b>S</b> R	<b>A</b> L <b>G</b> Q <b>E</b> D <b>T</b> G <b>S</b> L <b>E</b> A <b>T</b> L <b>N</b> Y <b>L</b> L <b>T</b> N <b>R</b>	<b>R</b>	<b>Q</b> K						
HPLC					*****						
1D		*****		*****	*)						
2D		*****		*****	*)						
	<b>I</b> P <b>S</b> L <b>H</b> S <b>L</b> N <b>S</b> E <b>E</b> E <b>R</b>	<b>I</b> L <b>R</b>	<b>Y</b> V <b>Q</b> Q <b>S</b> V <b>S</b> L <b>N</b> L <b>M</b> R	<b>D</b> G <b>V</b> T <b>P</b> S <b>V</b> A <b>L</b> D <b>M</b> T <b>A</b> R	<b>N</b> M <b>E</b> P <b>G</b> M <b>Y</b> A <b>S</b> N <b>R</b>	<b>P</b> F <b>I</b> N <b>R</b>	<b>L</b> M <b>D</b> Y <b>L</b> H <b>R</b>				
HPLC											
1D		*****		*****	*****	*****					
2D		*****	*****	*****	*****	*****	*****				
	<b>A</b> A <b>V</b> N <b>P</b> E <b>Y</b> F <b>T</b> N <b>A</b> I <b>L</b> N <b>P</b> H <b>W</b> L <b>P</b> P <b>P</b> G <b>F</b> Y <b>T</b> G <b>G</b> F <b>E</b> V <b>P</b> E <b>G</b> N <b>D</b> G <b>F</b> L <b>W</b> D <b>D</b> I <b>D</b> D <b>S</b> V <b>F</b> S <b>P</b> Q <b>P</b> Q <b>T</b> L <b>L</b> E <b>L</b> Q <b>Q</b> R	<b>E</b> Q <b>A</b> E <b>A</b> A <b>L</b> R									
HPLC					*****						
1D					*****						
2D					*****						
	<b>K</b>	<b>E</b> S <b>F</b> R	<b>R</b>	<b>P</b> S <b>S</b> L <b>S</b> D <b>L</b> G <b>A</b> A <b>A</b> P <b>R</b>	<b>S</b> D <b>A</b> S <b>S</b> P <b>F</b> P <b>S</b> L <b>I</b> G <b>S</b> L <b>T</b> S <b>T</b> R	<b>T</b> T <b>R</b>	<b>P</b> R	<b>L</b> L <b>G</b> E <b>E</b> E <b>Y</b> L <b>N</b> N <b>S</b> L <b>L</b> Q <b>P</b> Q <b>R</b>			
HPLC	*	*****		*****	*****	*****	*****				
1D		*****		*****	*****	*****	*****				
2D		*****		*****	*****	*****	*****				
	<b>E</b> K	<b>N</b> L <b>P</b> P <b>A</b> F <b>P</b> N <b>G</b> I <b>E</b> S <b>L</b> V <b>D</b> K	<b>M</b> S <b>R</b>	<b>W</b> K	<b>T</b> Y <b>A</b> Q <b>E</b> H <b>R</b>	<b>D</b> V <b>P</b> G <b>P</b> R	<b>P</b> P <b>T</b> R	<b>R</b>	<b>Q</b> R	<b>H</b> D <b>R</b>	<b>Q</b> R
HPLC	**	*****									
1D		*****			*****						
2D		*****			*****						
	<b>G</b> L <b>V</b> W <b>E</b> D <b>D</b> S <b>A</b> D <b>D</b> S <b>V</b> L <b>D</b> L <b>G</b> G <b>S</b> G <b>N</b> P <b>F</b> A <b>H</b> L <b>R</b>	<b>P</b> R	<b>L</b> G <b>R</b> M <b>F</b>								
HPLC											
1D											
2D											

<sup>a</sup> Same as footnote a of Table 2.

be identified by this approach. Very importantly, two peptides could be assigned to protein IV (this protein could not be detected in the RP-HPLC assay).<sup>7</sup> The tandem mass spectrum of one of the two peptides (GLPLAISLLK) is shown in Figure 4. The 23 kDa viral protease, identified in the analysis of the collected HPLC fractions, was not detectable in the whole virus digest with the 1D LC/MS/MS approach. We first speculated that this was due to the sample size. A total of  $3.73 \times 10^{10}$  particles were used for the sample preparation, and an aliquot of 10  $\mu$ L of the final sample volume of 196  $\mu$ L (see the Experimental Section) was used for the analysis. This represents  $1.90 \times 10^9$  virus particles and, consequently, 32 fmols of the 23 kDa protease (with 10 copies/virus) assuming that no loss occurs during sample preparation and analysis. However, we were unable to detect the 23 kDa protease even with increasing the amounts of virus digest introduced into the 1D LC/MS/MS system. This suggests that additional factors (e.g., the limited loading capacity of the system and the inability to adequately separate coeluting peptides so that minor components are selected for MS/MS) might play an important role in determining the relative sensitivity of the analysis.

The data obtained in this analysis suggest that the reduction, alkylation, and tryptic digestion conditions selected were appropriate. Similar to the analysis of the HPLC fractions, no peptides with free SH groups could be identified (data not shown), again indicating that all cysteines were alkylated under

the employed conditions. All identified peptides are tryptic peptides and the digestion pattern was similar to the pattern of the analysis of the HPLC fractions. Missed cleavages were observed only in K/RP sequences or if a second cleavage site was within 3 or less amino acid residues.

The sequence coverage of the identified proteins was similar to the coverage of the analysis of the HPLC fractions, again showing no clear correlation to copy numbers (concentrations). The sequence coverage of the hexon (protein II) was unexpectedly low compared to the coverage in the analysis of the HPLC fraction (see Table 1). This may indicate that this protein is more protected from the tryptic digestion in a partially or fully assembled virus (as under the reduction/alkylation conditions the virus was not anticipated to completely disassemble) than in denatured form in the collected HPLC fraction.

**Analysis of Whole Virus Digest by Two-Dimensional LC/MS/MS.** Analysis of the same whole virus digest by 2D LC-MS/MS resulted in the identification of 11 proteins. The results are summarized in Table 13, and the peptides detected from each protein are shown in Tables 2–12.

All proteins identified by the analysis of the HPLC fractions (Table 1) or by the 1D LC-MS/MS analysis of the digested whole virus (Table 13) could be identified using the 2D analysis. This included the 23 kDa viral protease. In these experiments, 17  $\mu$ L of digest was analyzed (almost twice the amount used for the 1D analysis, see the Experimental Section) corresponding

Table 5. Amino Acid Sequence of the Viral Protein IV<sup>a</sup>

	METR GR	<b>RPAALQHQQDQPQAHFGQR</b>	AAR	SAPLHR	DPDYADEDPAVER	HDPGPSGR	APTAVQR
HPLC							
1D		*****					
2D							
	K	PPQPAK R	GDMLDR	DAVEQVTEWDR	LELLGQTLK	SMPTADGLK	PLK <b>NFASLQELLSLGGER</b>
HPLC							
1D							*****
2D							
	LLADLVR	ENMR VR	DMLNEVAPLLR	DDGSCSSLNYPVIGVIYGTGCGK	SQLLR	NLLSSQLI	
HPLC							
1D							
2D							
	SPTPETVFFIAPQVDMIPPSELK	AWEMQICEGNYAPGPDGTIIPQSGTLR	PR	FKV	MAYDD		
HPLC							
1D							
2D							
	LILEHNYDVSDPR	NIFAQAAAR	GPIAIIMDECENLGGHK	GVSK	<b>FFHAFPSK</b>	LHDK	FPK
HPLC							
1D							
2D					*****		
	CTGYTVLVVLHNMNPR	R	DMAGNIANLK	IQSK	<b>MHLISPR</b>	MHPSQLNR	<b>FVNTYTK GLPLAISLLK</b>
HPLC							
1D							*****
2D					*****	*****	
	DIFR	HHAQR	SCYDWIINYNTTPQHEALQWCYLHPR	DGLMPMYLNIQSHLYHVLEK	IHR	TLNDR	DR
HPLC							
1D							
2D							
	WSR	AYR	AR	K	TPK		
HPLC							
1D							
2D							

<sup>a</sup> Same as footnote a of Table 2.

Table 6. Amino Acid Sequence of the Viral Protein V<sup>a</sup>

	MSK	R	K	<b>IK</b>	<b>EEMLQVIAPFIYGP</b>	K	EEQDYK	PR	K	LK	R	VK	K	K	<b>K K</b>
HPLC															****
1D				(**	*****										
2D					*****										
	<b>DDDDELDEVELLHATAPR</b>	R	R	VQWK	GR	R	VK	R	<b>VLR</b>	<b>PGTTVFTPGER</b>	STR	TYK	R		
HPLC															
1D															
2D															
	VYDEVYGEDLLEQANER	<b>LGEFAYGK</b>	R	HK	DMLALPLDEGNTPSLK	PVTLQQVLPALAPSEEK	R								
HPLC															
1D															
2D															
	GLK	R	<b>ESGDLAPTQMLVFK</b>	R	<b>QR</b>	<b>LEDVLEK</b>	<b>MTVEPGLPEVR</b>	VR	PIK						
HPLC															
1D															
2D															
	QVAPGLGVQTVDVQIPTTSSTSIATATEGMETQTS	SPVASAVADA	AAVQAVAAAASK	<b>TSTEVQTDPMFMR</b>											
HPLC															
1D															
2D															
	VSAPR	R	PR	GSR	K	<b>YGAASALLPEYALHPSIAPTPGYR</b>	<b>GYTYR</b>	<b>PR</b>	R	R	ATTR	R	R	TTTGTR	
HPLC															
1D															
2D															
	R	R	R	R	R	<b>R</b>	<b>QPV LAPISVR</b>	R	VAR	EGGR	<b>TLVLP TAR</b>	<b>YHPSIV</b>			
HPLC															
1D															
2D															

<sup>a</sup> Same as footnote a of Table 2.

to  $3.23 \times 10^9$  particles and 70 fmols of the 23 kDa protease injected to the 2D system (using 10 copies/virus) assuming no loss during sample preparation and analysis. Note that, as mentioned above, the 1D analysis of the same digest was not able to detect the 23 kDa protease even at large injection amounts.

Although the sequence coverage of the identified proteins was fairly similar to the coverage of the 1D analysis, the detected peptides were not always the same. Several peptides could be identified by the 1D analysis, but not by the 2D analysis and vice versa, which shows, as anticipated, that the

Table 7. Amino Acid Sequence of the Viral Protein VI<sup>a</sup>

	MEDINFASLAPR HGSR PFMGNWQDIGTSNMSSGGA <sup>a</sup> FSWGLSLWSGIK <b>NFGSTVK</b> NYGSK
HPLC	*****
1D	*****
2D	
	<b>AWNSSTGQMLR</b> DK <b>LK EQNFQOK</b> VVDGLASGISGVVDLANQAVQNK INSK
HPLC	*****
1D	*****
2D	*****
	<b>LDPR PPVEEPPPAVETVSPEGR</b> GEK <b>R PR PDR EETLVTQIDEPPSYEALK</b> QGLPTR
HPLC	***** (* *****)
1D	(* *****)
2D	
	PIAPMATGVLGQHTPVTLDLPPADTQQK <b>FVLPGP</b> TAVVTR <b>PSR</b> ASLR R AASGPR
HPLC	*****
1D	
2D	***** (**)
	<b>SLRPV</b> ASGNWQSTLNSIVGLG <sup>a</sup> QSLK R R R CF
HPLC	
1D	*****
2D	

<sup>a</sup> Same as footnote a of Table 2.

Table 8. Amino Acid Sequence of the Viral Protein VII<sup>a</sup>

	MSILISPSNNTGWGLR FPSK MFGGAK K R SDQHPVR VR GHYR APWGAHKR GR TGR
HPLC	
1D	
2D	
	<b>TTVDDAIDAVVEEAR</b> NYTPTPPPVSTVDAAIQTVVR GAR R YAK MK R R R R VAR R HR R
HPLC	*****
1D	*****
2D	*****
	R PGTAAQR <b>AAAALLNR</b> ARR TGR R AAMR AA R <b>R LAAGIVTVPPR</b> SR R R
HPLC	***** (* ***** (**
1D	***** *****
2D	***** *****
	<b>AAAAAAAISAMTQGR</b> R <b>GNVYWVR</b> DSVSGLR VPVR TR PPR N
HPLC	***** (* *****
1D	***** (* *****
2D	***** *****

<sup>a</sup> Same as footnote a of Table 2.

Table 9. Amino Acid Sequence of the Viral Protein VIII<sup>a</sup>

	MSK EIPTPYMWSYQPQMGLAAGAAQDYSTR INYMSAGPHMISR VNGIR AHR NR <b>ILLEQAAITTTPR</b>
HPLC	*****
1D	
2D	*****
	<b>NNLNPR</b> SWPAALVYQESPAPTTVVLPR DAQAEVQMTNSGAQLAGGFR HR VR <b>SPGQGITHLTIR</b> GR
HPLC	*****
1D	*****
2D	*****
	<b>GIQLN</b> DES <sup>a</sup> VSSSLGLR PDGTFQIGGAGR PSFTPR <b>QAILTLQTSSEPR</b>
HPLC	*****
1D	*****
2D	
	SGGIGTLQFIEEFVPSVYFNPFSGPPGHYPDQFIPNFDAVK DSADGYD
HPLC	
1D	
2D	

<sup>a</sup> Same as footnote a of Table 2.

type(s) of separation(s) performed on the digest would have an important impact on which peptides will become detectable.

Discussion

The objective of this study was to evaluate the performances of 1D and 2D LC/MS/MS (shotgun sequencing) analyses on the adenovirus proteome. The adenoviral proteome is a

convenient test system for such comparison: it is fairly complex but well understood and characterized. Structural proteins are present in the virus particles with well established copy numbers. The copy numbers of these proteins and their contribution to the virus mass are highly variable covering differences over 2 orders of magnitude (Table 13).

One-dimensional (reversed-phase separation) LC/MS/MS analysis was performed on both fractions of the viral proteome

Table 10. Amino Acid Sequence of the Viral Protein IX<sup>a</sup>

	MSTNSFDGSIVSSYLTR	<b>MPPWAGVR</b>	QNVMGSSIDGR
HPLC		*****	
1D			
2D			
	PVLPANSTTLTYETVSGTPLETAASAAAASAAAATAR	<b>GIVTDFAFLSPLASSAASR</b>	<b>SSAR DDK</b>
HPLC		***** (**** (**	
1D		*****	
2D		*****	
	<b>LTALLAQLDLTLR</b>	<b>ELNVVSQQLDLR</b>	<b>QQVSALK ASSPPNAV</b>
HPLC	*****	*****	*****
1D	*****	*****	*****
2D	*****	*****	

<sup>a</sup> Same as footnote a of Table 2.

Table 11. Amino Acid Sequence of 23 kD Viral Endopeptidase<sup>a</sup>

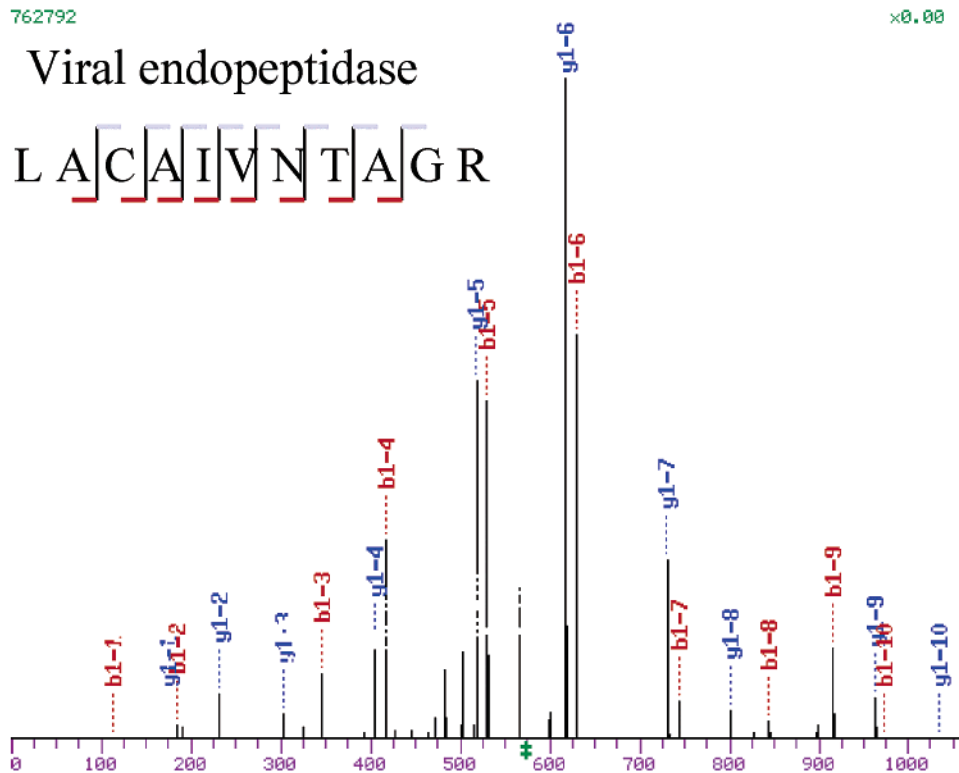
	MGSSEQELK	AIVK	<b>DLGCGPYFLGTYDK</b>	<b>R</b>	<b>FPGFVSPHK</b>	<b>LACAIVNTAGR</b>	ETGGVHWMFAWNPFSK
HPLC			*****		*****	*****	
1D							
2D					*****		
	<b>TCYLFEPFGFSDQR</b>	LK	<b>QVYQFEYESLLR</b>	<b>R</b>	<b>SATASSPDR</b>	<b>CITLEK</b>	
HPLC	*****		*****		*****	*****	
1D							
2D							
	STQSVQGPNSAACGLFCCMFLHAFANWPQTPMDHNPTMNLITGVPNSMLNSPQVQPTLR						R
HPLC							
1D							
2D							
	<b>NQEQLYSFLER</b>	HSPYFR	SHSAQIR	<b>SATSFCHLK</b>			NM
HPLC	*****			*****			
1D							
2D							

<sup>a</sup> Same as footnote a of Table 2.

Table 12. Amino Acid Sequence of the L1 Protein<sup>a</sup>

	MHPVLR	QMR	PPPQQR	QEQEQR	QTCR	APSPPTASGGATS	SAVDAAADGDYEPP	R	R	R	AR
HPLC											
1D											
2D											
	<b>HYLDLEEGELAR</b>	LGAPSPER	YPR	VQLK	R	DTR	EAYVPR	<b>QNLFR</b>	DR	EGEEPEEMR	
HPLC	*****							*****			
1D	*****										
2D	*****										
	DR	K	FHAGR	ELR	HGLNR	ER	<b>LLR EEDFEPDAR</b>	TGISPAR	<b>AHVAAADLV</b>	<b>TAYEQTVNQEINFQK</b>	
HPLC							*****			*****	
1D										*****	
2D							*****				
	SEFNNHVR	TLVAR	<b>EEVAIGLMHLWDFVSALEQNPNSK</b>				PLMAQLFLIVQHSR			DNEAFR	
HPLC			*****								
1D			*****								
2D											
	<b>DALLNIVEPEGR</b>	WLLDLINILQ	SIVVQER	<b>SLSLADK</b>	<b>VAAINYSMLSLGK</b>	FYAR	K	<b>IYHTPYVPIDK</b>			
HPLC	*****			*****	*****			*****			
1D	*****							*****			
2D				*****	*****			*****			
	EVK	<b>IEGFYMR</b>	MALK	<b>VLTLSDDLGVYR</b>	NER	IHK	AVSVSR	R	R	ELSDR	<b>ELMHSLQR</b>
HPLC											*****
1D											*****
2D		*****		*****							
	ALAGTGS	<b>DR</b>	<b>EAESYFDAGADLR</b>	WAPSR	R	<b>ALEAAGAGPGLAVAPAR</b>					
HPLC						*****					
1D						*****					
2D			*****								
	AGNVGGVEEYDEDEYEPEDGEY										
HPLC											
1D											
2D											

<sup>a</sup> Same as footnote a of Table 2.



**Figure 3.** Tandem mass spectrum derived by collision induced dissociation of the  $(M + 2H)^{2+}$  precursor ion,  $m/z$  574.5. Fragment ions in the spectrum represent mainly single-event preferential cleavage of the peptide bonds resulting in the sequence information recorded from both N (b-ions) and C (y-ions) termini of the peptides simultaneously. SEQUEST searches (Eng et al., 1994) of the MS/MS data against homo-sapiens and adenovirus database downloaded from National Center for Biotechnology (NCBI) revealed the peptide identification LACAIVNTAGR, which is a tryptic peptide of viral endopeptidase.

**Table 13.** Identification of Viral Proteins from Whole Virus Digests Using One-dimensional Chromatographic (1D) or Two-dimensional Chromatographic (2D) Separation of the Peptides and Analysis by Mass Spectrometry. For Experimental Details, See Experimental Section

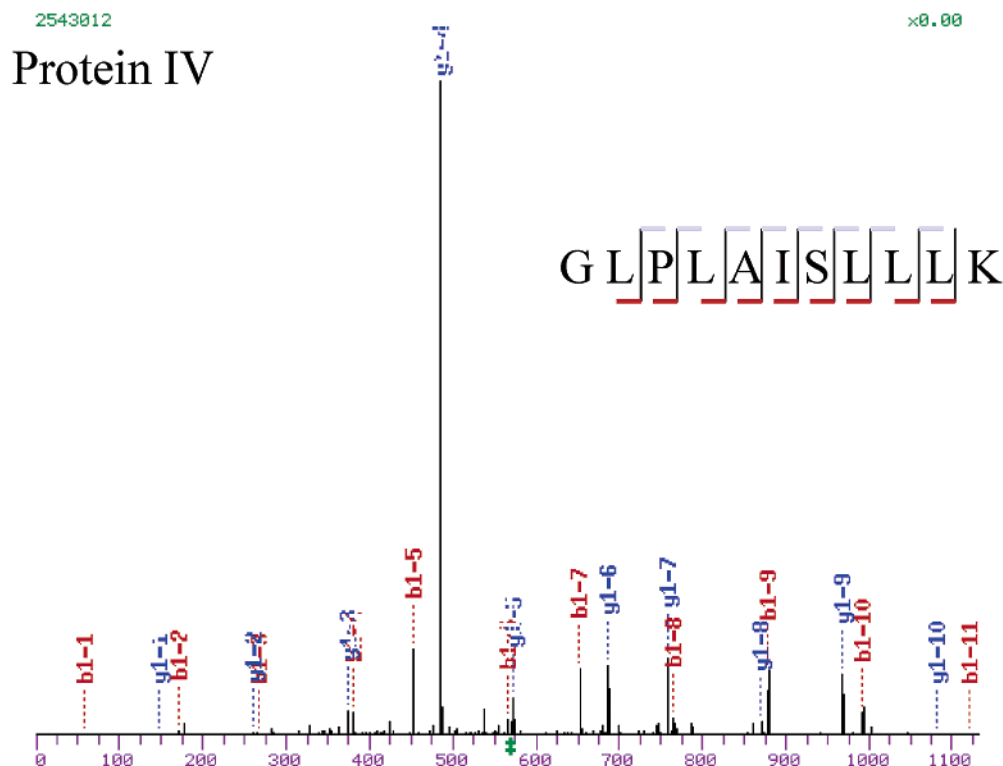
protein ID	1D analysis		2D analysis		copy no.
	no. of peptides identified	sequence coverage [%]	no. of peptides identified	sequence coverage [%]	
II	8	11	2	2	720
III	8	19	10	22	60
IIIa	15	34	17	33	60
IV	3	10	3	3	36
V	12	34	9	28	157
VI	6	36	3	11	360
VII	7	40	5	29	833
VIII	2	11	2	11	
IX	4	38	3	32	240
L1	7	23	8	19	
23 kD protease	—	—	1	4	10

(following separation of the viral proteins by RP-HPLC, reduction/alkylation and tryptic digestion of the fractions) and on the tryptic digest of the whole virus (following reduction/alkylation). Two-dimensional (cation exchange and reversed-phase separation) LC/MS/MS was performed on the tryptic digests of the whole virus only.

We were able to identify most or all major viral proteins in the different analysis approaches. In addition, we were able to refine the peak assignments in the RP-HPLC chromatography profile of the adenoviral proteome.<sup>7</sup> We identified two minor components, the L1 protein and the 23 kDa viral protease. Their

presence in the sample was also confirmed by the analysis of the whole virus digests. In our hands, these proteins had not been detected previously by chromatography and mass spectrometry based methods. A further tentative identification has been made of the full-length protein VIII.

The ability to detect the 23 kDa viral protease, which is present at a low copy number (10/virion), gives a measure of the sensitivity of the different analytical methods. The 1D LC/MS/MS was unable to detect it, whereas the method using fractionated viral proteins and 2D LC/MS/MS analysis could. These results fit neatly with expectations based on sample complexity. The 1D LC/MS/MS method loads a large number of peptides on the capillary column all at once and elutes them into the mass spectrometer in a single separation. This presents two challenges. The first challenge is sample loading. The large number of peptides may exceed the capacity of the capillary column, losing a portion of the sample. Also, it is possible that selective binding may bias the analysis for the peptides that are able to bind. The second challenge is to deal with multiple peptides entering the mass spectrometer simultaneously. To maximize the number of peptides analyzed, each scan-cycle collects MS/MS data for the three most abundant masses detected in the full scan. To further increase the number of different peptides analyzed Dynamic Exclusion was used to exclude masses that had been analyzed in previous scans from being included in subsequent rounds of MS/MS. Although this scheme will maximize the number of peptides analyzed, many peptides could still be missed. The low abundance peptides will be the ones most likely to be missed because they may never reach the threshold for inclusion in the MS/MS scan,



**Figure 4.** Tandem mass spectrum derived by collision induced dissociation of the  $(M + 2H)^{2+}$  precursor ion,  $m/z$  569.9. Fragment ions in the spectrum represent mainly single-event preferential cleavage of the peptide bonds resulting in the sequence information recorded from both N (b-ions) and C (y-ions) termini of the peptides simultaneously. SEQUEST searches (Eng et al., 1994) of the MS/MS data against homo-sapiens and adenovirus database downloaded from the National Center for Biotechnology (NCBI) revealed the peptide identification GLPLAISLLK, which is a tryptic peptide of viral protein IV.

before they finish eluting from the column. When the complexity of the sample is reduced, the probability of acquiring spectra all of the components increases.<sup>2</sup>

The analysis of fractionated viral proteins and the 2D LC/MS/MS both simplify the analysis by presenting only a subset of the total number of peptides for analysis. The fractionated proteins are analyzed as a single chromatographic peak at a time (though this peak may contain multiple components). The 2D LC/MS/MS loads only those peptides eluting at a specific salt concentration from the cation exchange column. In both cases, the sample complexity is considerably reduced over the 1D LC/MS/MS method. This provides higher sensitivity because the simpler chromatogram should contain fewer coeluting peaks, so the mass spectrometer has a greater chance of analyzing all the peptides injected. The result should be greater coverage of all peptides, including those from minor components.

The results of this comparison also demonstrated potential shortcomings of protein fractionation prior to LC/MS/MS analysis. Protein IV (fiber) was well detected in the analysis of the whole virus digest (Table 13). But, at the same time, we were unable to detect protein IV in any of the collected RP-HPLC peak fractions consistent with earlier observations,<sup>7</sup> suggesting that protein IV does not form a peak in this separation. It elutes on the RP-HPLC column on the baseline (data not shown). Thus, protein IV can be lost for further analysis, if only peak collection is used in the initial fractionation of the sample.

When comparing the potential sensitivities of the 1D vs the 2D LC/MS/MS methods for peptide identification in an identical whole virus digest sample (Table 13), the advantage of the

2D method is quite obvious. This is due to the initial sample clean up by the cation exchange separation and the resulting relatively higher loading capacity on the analysis (reversed-phase separation) column. This sample clean up allowed us to detect 70 fmols of the 23 kDa viral protease (the quantity of the protease in the sample injected to the 2D LC/MS/MS system was calculated ignoring any loss during sample preparation). Although the step elution from the cation exchange cleanup is very convenient, attention must be paid to ensure a complete elution of peptides, to avoid splitting peptides between subsequent elution fractions. This is particularly important if the goal is to achieve high-sequence coverage. Alternatively, one might consider using two or more different step elution conditions.

The sequence coverage observed in the different shotgun sequencing approaches ranged from 2 to 54%, with the majority between 20 and 35% (see Tables 1 and 13). On the basis of such sequence coverage (and the established low detection limits), the shotgun sequencing should provide a valuable approach to identify modifications (posttranslational or post-production) on the structural proteins of the adenovirus. The initial results generated in our laboratories already show that the shotgun sequencing combined with selected ion monitoring was able to identify at least one glycosylated protein from the adenovirus proteome.

A well understood complex system, like the adenovirus, where the identity of the protein constituents (even the minor ones) and the copy numbers of the different proteins in a single virus are well-known, may allow the use of shotgun sequencing methods to detect (and to a limited extent) quantitate contaminants. As shown in Table 13, the 2D LC/MS/MS analysis

of the whole virus digest was able to detect all known viral proteins with copy numbers at or above 10/virus particle. These protein species are quite different from each other in their sequences. Consequently, one should anticipate that a contaminant protein with an arbitrary sequence and with a concentration higher than the equivalent of 10 copies/virus particle in the sample would be detectable. Such analysis would nicely complement other (chromatographic, immunological, etc.) tests to establish purity of the virus preparations.

## References

- (1) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–47.
- (2) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **1997**, *69*, 767–776.
- (3) van Oostrum, J.; Burnett, R. M. Molecular composition of the adenovirus type 2 virion. *J. Virol.* **1985**, *56*, 439–448.
- (4) Stewart, P. L.; Fuller, S. D.; Burnett, R. M. Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy. *EMBO J.* **1993**, *12*, 2589–2599.
- (5) Stewart, P. L.; Burnett, R. M. Adenovirus structure by X-ray crystallography and electron microscopy. *Curr. Top. Microbiol. Immunol.* **1995**, *199*, 25–38.
- (6) Burnett, R. M. *Structural Biology of Viruses*; Oxford University Press: Oxford, New York, 1997; pp 209–238.
- (7) Lehmborg, E.; Traina, J. A.; Chakel, J. A.; Chang, R. J.; Parkman, M.; McCaman, M. T.; Murakami, P. K.; Lahidji, V.; Nelson, J. W.; Hancock, W. S.; Nestaas, E.; Pungor, E., Jr. Reversed-phase high-performance liquid chromatographic assay for the adenovirus type 5 proteome. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, *732*, 411–423.
- (8) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- (9) Vellekamp, G.; Porter, F. W.; Sutjipto, S.; Cutler, C.; Bondoc, L.; Liu, Y.-H.; Wylie, D.; Cannon-Carlson, S.; Tang, J. T.; Frei, A.; Voloch, M.; Zhuang, S. Empty capsids in column-purified recombinant adenovirus preparations. *Human Gene Therapy* **2001**, *12*, 1923–1936.
- (10) Cabrita, G.; Iqbal, M.; Reddy, H.; Kemp, G. Activation of the adenovirus protease requires sequence elements from both ends of the activating peptide. *J. Biol. Chem.* **1997**, *272*, 5635–5639.

PR025528C