ELSEVIER

Journal of Chromatography B, 732 (1999) 411–423

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Reversed-phase high-performance liquid chromatographic assay for the adenovirus type 5 proteome

Elisabeth Lehmberg[a,*], Joseph A. Traina[a], John A. Chakel[b], Ray-Jen Chang[a], Maria Parkman[a], Michael T. McCaman[a], Peter K. Murakami[a], Vafa Lahidji[a], Jeffrey W. Nelson[a], William S. Hancock[b], Eirik Nestaas[a], Erno Pungor Jr.[a]

[a]*Berlex Laboratories, Inc., 15049 San Pablo Avenue, P.O. Box 4099, Richmond, CA 94804-0099, USA*
[b]*HP Laboratories, Palo Alto, CA, USA*

Received 23 April 1999; received in revised form 25 June 1999; accepted 25 June 1999

Abstract

An RP-HPLC assay was developed for a recombinant adenovirus type 5. During chromatography, intact adenovirus dissociated into its structural components (DNA and proteins) and the viral proteome was separated yielding a characteristic fingerprint. The individual components were identified by matrix-assisted laser desorption ionization time-of-flight mass spectroscopy, N-terminal sequencing and amino acid composition. The assay was utilized to measure adenovirus particle concentration through quantification of structural proteins. Each structural protein provided independent measurement of virus concentration allowing verification of accuracy. The assay sensitivity is at or below $2 \cdot 10^8$ particles. Contrary to the benchmark spectrophotometric assay, the RP-HPLC assay was shown to be insensitive to contaminants common for partially purified adenovirus preparations.

Keywords: Protein fingerprinting; Adenovirus type 5; Proteome

1. Introduction

Recombinant adenovirus preparations are used for gene delivery in a variety of gene therapy settings. Adenoviral transfection offers an advantage if a short term (transient) gene expression is desired without integration into the host cell genome [1]. With the growing use of such viral vectors in clinical development, the need for their commercial production is anticipated in the near future. At that stage better analytical tools will be desired to define the virus product, i.e., to allow verification of product equivalency by chemical, biochemical and in vitro biological analysis following changes in the manufacturing process (including scale up, introducing new production sites, etc.). Without improved analytical tools the implementation of such manufacturing changes may require costly and time-consuming human clinical studies to prove product equivalency. Thus, similarly to the therapeutic proteins during the past decade, the recombinant adenovirus products will seek to ultimately reach the ''well characterized/understood biological product'' regulatory status.

The adenovirus particle consists of a double-stranded DNA molecule (36 kilobase pairs) and

*Corresponding author. Tel.: +1-510-6694-133; fax: +1-510-6694-920.
E-mail address: elisabeth_lehmberg@Berlex.com (E. Lehmberg)

0378-4347/99/$ – see front matter
PII: S0378-4347(99)00316-3

structural proteins. There are at least 11 distinct structural proteins (protein II–protein XII [2]) present in multiple copies. The total number of protein molecules in the virus is approximately 2500, and they are held together by non-covalent interactions [3,4]. The protein coat surrounding the viral DNA forms a symmetrical icosahedron which has been extensively studied by electron microscopy and crystallography [5,6]. The structural proteins forming the coat (coat proteins) are proteins II, III, IIIa, IV, VI and IX, their copy numbers in the virus are well established by electron microscopy. The remaining structural proteins are localized inside the virus (core proteins) together with the viral DNA. The copy numbers of the core proteins have been assessed by standard protein chemistry procedures [e.g., by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), radio-labeling] and are not supported by symmetry considerations [3,4]. Further structural information is available on the viral proteins (some three-dimensional structures based on crystallography, evidence of proteolytic processing of the structural proteins during virus assembly [7,8], evidence of glycosylation on some proteins [3,9], etc.). The current knowledge, however, does not meet the requirements of the analytical definition of the virus in comparison to the standards applied to the recombinant protein drugs (for example the evaluation of post-translational and manufacturing related protein modifications, including oxidation, deamidation and proteolysis).

The complexity of the adenovirus requires a systematic approach utilizing the available DNA and protein analysis techniques. The objective of the present study was to establish a reversed-phase high-performance liquid chromatography (RP-HPLC) assay to support such a systematic analysis of the viral structural proteins, the viral proteome. We aimed to completely dissociate the virus to allow for separation and further analysis of the individual structural proteins by N-terminal protein sequencing and by mass spectroscopy, thus leading to their identification. Quantification of the separated structural proteins offered a unique way to estimate virus concentration. We decided to compare the utility of this new protein based virus quantification assay to the current benchmark method: absorbance of the lysed virus at 260 nm [2] as well as to an assay based on viral DNA measurement.

2. Experimental

2.1. Virus

Adenovirus type 5 preparations with different transgenes were produced using HEK 293 cells adapted to serum free medium and suspension culture and purified by anion-exchange chromatography and ultrafiltration at Berlex Biosciences.

2.2. Protein and tRNA standards

Bovine serum albumin (BSA), 99% pure, and growth hormone releasing factor (GHRF) 1–29 amide were purchased from Sigma (St. Louis, MO, USA). Recombinant vampire bat salivary plasminogen activator (rDSPA α-1) was produced, purified and characterized at Berlex Biosciences [10]. The concentrations of the protein standards were determined by absorbances at 280 nm. The specific absorbances used were 0.69 ml $mg^{-1}cm^{-1}$ for BSA [11]; 1.71 ml $mg^{-1}cm^{-1}$ for DSPA (determined at Berlex Biosciences) and 0.89 ml $mg^{-1}cm^{-1}$ for GHRF, calculated according to Ref. [12]. The contribution of light scattering to the absorbance at 280 nm was estimated using a correction based on the absorbances at 340 and 360 nm [13]. tRNA was purchased from Sigma. The absorbance of 50 μg/ml tRNA is 1 AU cm^{-1} at 260 nm.

2.3. SDS–PAGE analysis

Samples (with or without prior drying in a Speed Vac concentrator) were mixed 1:1 with Laemmli reducing buffer (125 m*M* Tris–HCl, pH 6.8, 20% glycerol, 4% SDS, 0.005% bromophenol blue, 0.5% β-mercaptoethanol), boiled for at least 5 min and loaded onto a 1 mm×10-well 4–20% gradient Tris–glycine minigel (Novex, San Diego, CA, USA). Samples were electrophoresed at 120 V for 1.5 h. The gels were stained with Coomassie Brilliant Blue (0.1%, Sigma) or by silver stain (Novex).

2.4. Spectrophotometric measurement of virus concentration

We essentially followed the benchmark spectrophotometric method developed by Maizel, Jr., et al.

[2]. To each virus sample 10% SDS was added to a final concentration of 0.5%. After incubating the samples at 55°C for 5 min and a brief centrifugation in a microcentrifuge, the absorbance of the samples was measured at 260 nm. In this assay 1 AU at 260 nm corresponds to $1.1 \cdot 10^{12}$ viral particles/ml.

2.5. Measurement of viral DNA by fluorescent dye binding

Adenovirus samples were directly lysed by addition of SDS (final concentration of 0.1%) with or without pretreatment to degrade contaminating non-viral nucleic acids by incubation with RNase A (Sigma) and DNase I (Boehringer Mannheim, Indianapolis, IN, USA). A fluorescent reagent binding to double-stranded DNA, PicoGreen (Molecular Probes, Eugene, OR, USA) was diluted 380-fold in 10 m*M* Tris, pH 7.5, 1 m*M* EDTA and 190 μl of the diluted reagent was added to each well of a 96-well microplate (Serocluster, Corning Star, Cambridge, MA, USA) to measure DNA content. A 10-μl volume of standard DNA sample or virus lysate was added to the wells, the plate was incubated in the dark for 3 to 5 min and read on a fluorescent microplate reader (fmax plate reader from Molecular Devices, Sunnyvale, CA, USA), with settings of 485 nm (excitation) and 538 nm (emission). Bacteriophage λ DNA (provided with the PicoGreen reagent) was used to establish a standard curve from 70 to 5000 ng/ml DNA. For the estimation of virus concentration based on the DNA content we assumed that each virus particle contained a full set of the viral genome (36 000 base pairs, molecular mass= 26.7 MDa), therefore 1 ng DNA corresponds to $2.5 \cdot 10^7$ adenovirus particles. The limit of quantitation with this assay format is approximately 10^{10} viral particles/ml.

2.6. The optimized RP-HPLC procedure

The RP-HPLC experiments were performed on a Model 1090 HPLC unit equipped with diode-array detector, autosampler, micro flow cell and a temperature controlled column compartment (Hewlett-Packard, Palo Alto, CA, USA). A Jupiter column (150× 2 mm I.D.), packed with a 5 μm nominal diameter, 300 Å pore size C_4 resin (Phenomenex, Torrance, CA, USA) and a pre-column filter (0.5 μm pore size, Alltech, Deerfield, IL, USA) were used. The solvents were A: 0.1% trifluoroacetic acid (TFA; Pierce, Rockford, IL, USA) in water, and B: 0.1% TFA in acetonitrile (Baker, Phillipsburg, NJ, USA). Virus samples, without pretreatment, were injected into the RP-HPLC column. The column was equilibrated at 20% solvent B. The HPLC program started with injection and 10 min equilibration at 20% solvent B, followed by a linear gradient to 40% B in 5 min, and another linear gradient to 60% B in 40 min. The flow-rate was 200 μl/min, the column temperature was maintained at 40°C. Absorbance was monitored at 214 nm. The peaks were integrated: 1 area unit=1 milliabsorbance unit×s.

2.7. Matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (MALDI-TOF-MS)

MALDI-TOF-MS was performed on a HP 2030 system (linear, continuous extraction, Hewlett-Packard) or a Voyager-DE RP system operated in the linear, delayed extraction mode (Perkin-Elmer, Framingham, MA, USA). Sinapinic acid (Hewlett-Packard or Fluka, Ronkonkoma, NY, USA) was used as the matrix. The analyte:matrix ratio was variable, typically 1:5000. All spectra were taken in the positive ion mode. Aliquots of collected samples from the RP-HPLC system were either dried on the probe in a Model 2024A sample prep accessory (Hewlett-Packard) or, alternatively, allowed to air dry. Multiple laser shots were performed (normally 100) and the data were averaged to obtain the mass spectra. Mass calibration was performed by analyzing a set of external or internal standard proteins. The following proteins were used as calibration standards: BSA Calibration Standard (Hewlett-Packard), yeast enolase, apomyoglobin, bovine trypsinogen, carbonic anhydrase (Sigma), Sequazyme Calibration Mixture 3 (Perkin-Elmer). The deduced average molecular masses of the viral proteins were calculated using the Hewlett-Packard Peptide Tools software package.

2.8. N-Terminal sequencing, chloramine-T treatment

Automated Edman degradation was performed using a Model G1005A sequencer (Hewlett-Packard). For direct sequencing, samples without pre-

treatment were loaded to the sample cartridge and analyzed. In order to obtain sequences of proteins with blocked N-termini, samples were first incubated with 0.1 mg/ml (0.44 m*M*) Chloramine-T (Sigma) at room temperature for 30 min to achieve oxidative chain cleavage at the C-termini of tryptophanyl residues. The reaction was quenched by adding 0.2 mg/ml (1 m*M*) acetyl-methionine (Sigma). To minimize losses, the reaction mixture was loaded to the sequencer cartridge without separating the oxidized peptides. In most cases we were able to obtain sequences based on the similarity in the amino acid signal intensities in consecutive cycles. The initial sequencing signal intensities for the different peptides were different, as a result of variable chain cleavage efficacy by the Chloramine-T treatment or different initial yields for individual peptides in the sequence analysis. The repetitive yield of the sequencer was greater or equal to 90%, thus amino acids with less than 10% yield drop in consecutive cycles were assumed to originate from the same peptide. Otherwise, contiguous sequences were obtained by aligning the measured multiple amino acid signals in consecutive cycles with predicted sequences following tryptophan cleavage sites in the adenovirus type 5 proteins. The sequence based identifications were done by comparison to the SWISS-PROT database (University of Geneva) for protein VII and late L2 μ protein, for all others, the PIR database (Protein Information Resources, PIR, http://nbfro.georgetown.edu) was used. The deduced sequence of the late L2 μ protein is only available for adenovirus type 2 and we assumed homology.

2.9. Amino acid analysis

Samples were transferred into Pyrex culture tubes (50 mm×6 mm I.D.), and concentrated to dryness in a Speed Vac. The tubes were placed in glass reaction vials (Waters, Milford, MA, USA) for gas phase acid hydrolysis in 6 *M* HCl (Pierce) and 1% phenol. Norleucine (12 nmol, Pierce) was added as internal standard to each tube. The acid hydrolysis was performed under vacuum for 21 to 22 h at 114°C. The samples were subsequently derivatized with phenyl isothiocyanate (PITC, Pierce) for 20 min at room temperature. The PITC derivatives were separated by RP-HPLC on a ''Pico Tag'' C_{18} (300×3.9 mm I.D.) column from Waters operated at 48°C. The solvents were A: 130 m*M* sodium acetate, 0.047% triethylamine, 0.1 ppm EDTA and 6% acetonitrile in water, pH 6.4; and B: 60% acetonitrile in water. The absorption was monitored at 254 nm.

3. Results

3.1. Optimization of virus dissociation and the chromatographic conditions

Preliminary experiments were performed by injecting purified, intact adenovirus preparations without any pretreatment onto the RP column at 40°C, and the components were eluted using a linear acetonitrile gradient with 0.1% TFA. The appearance of a multitude of peaks in the chromatogram is indicative for an obviously occurring spontaneous dissociation of the virus under these conditions. The chromatographic procedure was optimized with respect to elution temperatures up to 50°C, acetonitrile concentrations for column equilibration between 0 and 20% and gradient slopes to improve peak resolution. A chromatogram obtained by the optimized procedure described in Experimental is shown in Fig. 1. During the first 15 min of the gradient no protein peaks were detected as verified by SDS–PAGE analysis of the corresponding eluate using silver staining. The viral DNA appeared not to be retained on the column as confirmed by spectral analysis of the flowthrough (system) peak which is

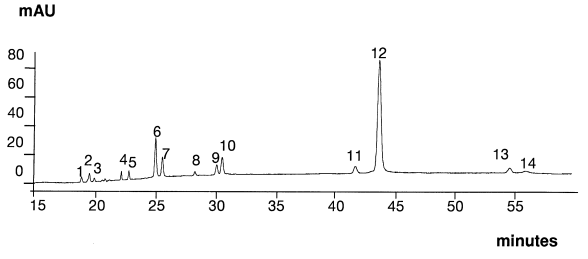

Fig. 1. A typical RP-HPLC chromatogram of the adenovirus type 5 proteome employing the optimized procedure. $5 \cdot 10^{10}$ viral particles were injected onto the column and chromatograms were obtained at 214 nm. No peaks appeared in the first 15 min of the profile with the exception of a flowthrough (system) peak that contained no detectable protein. Fractions were collected for further analysis of the 14 peaks detected.

also valid for purified viral DNA (data not shown). Viral preparations obtained by different purification methods (chromatography vs. density gradient centrifugation) and viruses carrying different transgenes gave very similar chromatograms (data not shown here).

In order to assure complete dissociation, attempts were made to dissociate the virus to its structural proteins prior to RP-HPLC analysis. These included incubations with acetonitrile (up to 50%), 6 *M* guanidine·HCl+20 m*M* dithiothreitol or 2% SDS, 10 min sonication, 10 min heating to 100°C, or five freeze–thaw cycles between −70°C and room temperature. None of these pre-treatments had any detectable effect compared with RP-HPLC of directly injected adenovirus preparations. The chromatographic profiles remained the same as judged by the number and areas of individual peaks. These observations strongly suggest that a complete dissociation of the virus occurs in the column under the conditions of the reversed-phase chromatography.

3.2. Identification of the chromatographic peaks

Chromatographic fractions containing individual peaks as shown in Fig. 1 were analyzed by N-terminal amino acid sequencing, MALDI-TOF-MS and by measuring amino acid composition. Approximately 10^{11} viral particles were used for each of the analyses. The sequencing and mass spectroscopy results, including the peak identifications are shown in Table 1.

Only peaks 6 and 10 gave detectable N-termini in the direct analysis, showing the authentic sequences of proteins VII and VI, respectively. The N-termini of all other peak samples appeared to be blocked. Chloramine-T treatment prior to sequencing resulted in internal sequences for all peaks except in peak 1, where no sequence could be identified. The N-termini of proteins VII and VI were detected in the Chloramine-T digests, too. These observations are supported by literature data on adenovirus type 2 reporting blocked N-termini for proteins II, III, V and IX but no blockage on proteins VI and VII (for review, see Ref. [4]). Based on the similarities between the different viral subtypes we expected blockages of the N-termini on the corresponding proteins from adenovirus type 5; to our knowledge this was not previously confirmed experimentally. The absence of a detectable free N terminus on protein IIIa was unexpected because the N-terminus of protein IIIa in adenovirus type 2 appears to have no blockage [14] and corresponds to residue G_{20} of its precursor protein.

All internal sequences found in the chromatographic peaks were consistent with a single component present in each peak, except in peak 1, where no sequence could be identified, and in peak 2, where two components were indicated. Nine out of 14 peaks contained a single detectable component by MALDI-TOF-MS. Peaks 3 and 5 appeared to be mixtures of two components and no signal could be detected in peaks 1, 2 and 4.

The molecular masses predicted for full length proteins were in excellent agreement with the measured masses (less than 0.1% difference to the predicted values) for four peaks; 7, 9, 12 and 13, corresponding to proteins V, III, II and IX, respectively. The mass spectroscopic results obtained with the remaining seven other peaks with detectable masses were consistent with proteolytic processing of the viral proteins as described below.

Both the N-terminal amino acid sequences and the observed molecular masses of proteins VI and VII suggest proteolytic processing of precursor proteins by a 23 kDa viral protease described by Anderson [8] for adenovirus type 2. The consensus recognition sequences of the protease are (I, L, M)XGG/X and (I, L, M)XGX/G with the cleavage sites marked with ''/''. Our data are in line with the deduced recognition sequence of MSGG/A starting at M_{30} in the protein VI precursor and MFGG/A starting at M_{21} in the protein VII precursor. The predicted mass for the mature (processed) protein VII (peak 6) in Table 1 was calculated accordingly. Interestingly, upon digestion of peak 3, we obtained a peptide sequence that lined up with a deduced internal sequence of the protein VII precursor following a tryptophan residue. The same sequence was also detected in peak 2. Although the error is somewhat large, approximately 3%, the MALDI-TOF-MS result with peak 3 may be consistent with the presence of the precursor segment spanning between the authentic N-terminus and G_{24}. The larger mass may also be a result of a modification of the peptide: for example acetylation of the N-terminal residue re-

Table 1
The results of N-terminal sequence analysis of the collected fractions from the RP-HPLC fingerprinting[a]

Peak	Sequences	Protein ID, position	W	Predicted mass (Da)	Measured mass (Da)
1	ND				ND
2	GLRFPSK	VII precursor, from W_{13}	1	2598	ND
	ALTXRLR	Late L2 μ, authentic N-terminal	0	8715	
3	GLRFPSK	VII precursor, from W_{13}	1	2598	2445*, 2515*
4	QDIGTSTN	VI precursor, from W_{22}	5	26 996	ND
5	SYQPQMG	VIII, from W_{11}	2	12 125	3541*
	PAALVYQ	VIII, from W_{73}			12 038
	FRHRVRS	VIII starting at F_{112}			
6	AKKRSDQ	VII, authentic N-terminal	2	19 412	19 410
	VRDSV	VII, from W_{155}			
7	KGRRVKR	V, from W_{66}	2	41 447	41 461
	MFRVSAP	V, from W_{270}			
8	GRSSFTP	C-terminal segment of VIII starting at G_{158}, except P_{160} predicted	0	7642	7642
9	GGDLKTI	III, from W_{119}	5	63 293	63 377
	VEFTLPE	III, from W_{165}			
	YLAYNYG	III, from W_{406}			
10	AFSWGSLW	VI, authentic N-terminal	4	22 100	22 104
	GSL	VI, from W_{4}			
	NSSTG	VI, from W_{26}			
	QSTLN	VI, from W_{196}			
11	LPPPGFY	IIIa, from W_{390}	6	63 502	63 736
	DDIDDSV	IIIa, from W_{412}			
	KTYAQEH	IIIa, from W_{523}			
12	SYXXISGQ	II, from W_{10}	13	108 008	107 950
	DEAXTXL	II, from W_{135}			
	NXAVDSY	II, from W_{387}			
	SLDYM	II, from W_{524}			
	NFRKDVN	II, from W_{583}			
	AFTRLKTK	II, from W_{679}			
	PGNXXLL	II, from W_{731}			
	RIPFSSNF	II, from W_{871}			
13	AGVRQNV	IX, from W_{22}	1	14 458	14 369
14	ILPLLIPLI	Late L2 μ, starting from I_{51}	0	2888	2889

[a] N-Terminal sequencing was performed on all chromatographic peaks directly and after chemical digestion with the exception of peak 8 which was only analyzed after chemical digestion. The chemical digestion and the subsequent sequencing procedure together with the assignment of the sequences are described in Experimental. The peak numbers correspond to Fig. 1. For protein IIIa, the predicted molecular mass is for the mature molecule as predicted by Boudin et al. [7], see also text. For proteins VI, VII, VIII and X (Late L2 μ) the predicted molecular masses are for the modified proteins, reflecting the proteolytic processing of the precursor molecules as predicted by Anderson [8] or reflecting additional processing(s) of the mature molecule by the same enzyme (see text). The number of tryptophan residues present in the molecule using the deduced sequence (i.e., the potential number of N-termini upon Chloramine T oxidation) of the molecule with the predicted mass is given in the table under the heading ''W''. The accuracy of the mass spectroscopic mass assignments is ±0.1% (except the data marked with * which have an accuracy of ±0.5%).

sulting in a blocked N-terminus. The MALDI-TOF-MS analysis data for protein VI in peak 10 are in agreement with a second proteolytic cleavage between L_{207} and G_{208} (positions in the protein after removing the precursor peptide) predicted and verified for adenovirus type 2 [8]. The recognition sequence for this case is IVGL/G starting from I_{204}. The peptide cleaved off by the protease from the C-terminus of protein VI has 11 residues and a predicted molecular mass of 1350 Da. Since the accuracy of the mass assignment for a 22 kDa protein (protein VI) is ±22 Da, the removal of this peptide can be detected by MALDI-TOF-MS. The predicted mass for protein VI in Table 1 was

calculated assuming this two-step processing. Similarly to the case of protein VII above, we obtained a peptide sequence in peak 4 that lined up with a deduced internal sequence of the protein VI precursor. However, no detectable mass signal was obtained with this peak.

The MALDI-TOF-MS results for peak 11 are in line with a proteolytic processing of protein IIIa at G_{571} as predicted for adenovirus type 2 [7,8]. The recognition sequence in this case matches the deduced sequence of LGGS/G starting with L_{567}. Since the accuracy of the mass assignment for the protein IIIa (having a molecular mass of 63 kDa) is ±63 Da, the removal of the peptide from L_{567} to the C-terminus (calculated mass of 1998 Da) can be detected by MALDI-TOF-MS. The predicted mass for protein IIIa in Table 1 is calculated for the molecule starting at the deduced authentic N-terminus and ending with S_{570} reflecting the above processing. Note, however, that the measured mass is 234 Da larger than the predicted mass which may indicate other modification(s).

N-Terminal sequencing of peak 8 yielded GRSSFTP, which is almost identical to a deduced internal sequence of protein VIII, GRPSFTP, starting from G_{158}. This observation is consistent with a processing by the viral protease since the detected sequence follows a deduced recognition site of IGGAG starting with I_{154}. The predicted molecular mass in Table 1 represents the fragment of protein VIII between G_{158} and the deduced authentic C-terminus, taking the observed P to S replacement into account. The MALDI-TOF-MS result is in excellent agreement with the predicted mass. We accept the data as tentative identification. The N-terminal sequences obtained for peak 5 line up with two positions in protein VIII following tryptophan residues and with a deduced sequence starting at F_{112}. The detection of F_{112} is consistent with a proteolytic processing at the deduced recognition site of LAGG/F starting at L_{108}. The MALDI-TOF-MS results seem to confirm this processing, one of the measured masses (12 038 Da) is in a good agreement with the calculated mass of protein VIII between the authentic N-terminus and G_{111} (12 125 Da). The presence of this peptide also explains the two detected internal sequences starting at S_{12} and P_{74}. At present we are unable to match the other detected mass in this peak to any known fragments of protein VIII, and were unable to obtain a mass spectroscopic identification of a peptide containing the detected sequence starting from F_{112}.

The MALDI-TOF-MS results for peak 14 are in line with a proteolytic processing of the late L2 μ core protein (protein X) by the viral protease. The recognition sequence in this case matches the deduced sequence of MRGG/I starting with M_{47}. The predicted mass in Table 1 is calculated for the molecule starting at I_{51} and ending with the deduced authentic C-terminus reflecting the above processing.

Amino acid composition analyses were performed on peaks 3 and 6–13. The percentages of 15 measured amino acids in the amino acid composition analysis (N+D, H, P, K, Q+E, R, Y, S, T, L, A, M and F) were calculated for the analyzed samples. A search of the on-line amino acid composition database [15] at the Heidelberg University (http://www.embl-Heidelberg.de/aaa.html) confirmed the predictions in Table 1 (data not shown).

3.3. Recovery of the viral proteins

One half of a virus sample ($9 \cdot 10^{10}$ particles) was injected onto the column and eluted with a steep gradient (20–80% B in 5 min). The unbound fraction and the effluent fractions were combined. The other half was directly analyzed as the control. The amino acid composition of both samples was measured and a recovery of 70.1±11.7% was obtained on the base of the so called ''well recovered amino acids'' (N+D, Q+E, A, L and K; [11]).

A recovery of 79% was measured by collecting, diluting 1:1 in 0.1% TFA and immediately re-injecting the largest chromatographic peak (peak 12, protein II) using the peak area at 214 nm.

The correlation between the areas of the three largest peaks in the chromatogram and increasing amounts of injected virus are shown in Fig. 2. Although the absolute peak areas differed more than ten fold, all three proteins showed a linear response in the range of $5.75 \cdot 10^{8}$ to $5 \cdot 10^{10}$ viral particles. At higher virus injections the response of the largest peak became non-linear while the two smaller peaks maintained linearity up to the largest injection tested ($2.3 \cdot 10^{11}$ particles).

The RP-HPLC assay is able to simultaneously

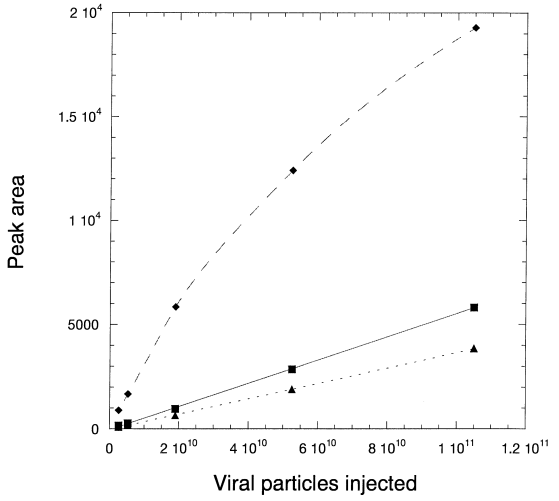

Fig. 2. Plots of the peak areas against the number of adenovirus type 5 particles in the range from $2 \cdot 10^8$ to 10^{11} particles injected onto the RP-HPLC column. The peak areas at 214 nm were integrated for peaks 6 (protein VII, squares), 10 (protein VI, diamonds) and 12 (protein II, triangles). The peak identification numbers are the same as in Fig. 1.

quantify several structural (coat and core) proteins in the adenovirus measuring the peak areas at 214 nm. The absorbance of a protein at 214 nm is primarily determined by the number of amide bonds [16]. As valid in most cases for high-molecular-mass species, the number of peptide bonds is approximately proportional to the injected mass of proteins. Although modest variations in specific absorbance at 214 nm may arise from differences in higher order structures [16], these are not anticipated to play significant role under the denaturing conditions applied in our assay (low pH, organic solvent and elevated temperature). Protein standard curves for quantification were generated by injecting different amounts of BSA, GHRF or rDSPA α-1 onto the RP-HPLC column, eluting the protein with the same gradient as used for the chromatography of the virus particles and integrating the protein peak at 214 nm. All three proteins eluted within the shallow part of the optimized gradient, i.e., between acetonitrile concentrations from 40 to 60%, where all detectable viral proteins elute. The protein standard curves were linear between 0.1 and 2.0 μg protein injected. The slopes were 2262±52 area units/μg BSA, 2205 area units/μg GHRF 1–29 and 2413±12 area units/μg rDSPA α-1. For the quantification of the viral proteins, we used BSA calibration.

The simultaneous measurement of the viral proteins also offers a unique way to verify the recovery of these proteins by demonstrating that no selective protein losses occur. Table 2 shows the results of the experiments conducted to determine the copy numbers of the structural proteins in the adenovirus by the RP-HPLC assay. Assuming that 720 copies of protein II are present in a virion [5], the copy number of each measured structural protein was calculated using the formula: copy number=720× mol protein/mol protein II. The copy numbers of the coat proteins obtained from crystallography and electron microscopy studies, called model copy numbers [6], and the copy numbers of core proteins determined by SDS–PAGE analysis, called bio-

Table 2
Calculation of the copy numbers of the main structural proteins in the virion[a]

Protein	Average mass (μg)	Measured molecular mass (Da)	Estimated copy number±SD	Model or biochemical copy number
II	0.681	107 950	720	720*
III	0.028	63 377	49±9	60*
IIIa	0.033	63 736	56±16	60*
V	0.063	41 461	170±15	157
VI	0.072	22 104	369±34	360*
VII	0.108	19 410	633±59	833
IX	0.032	14 369	256±41	240*

[a] Thirty reversed-phase chromatography runs were performed over a period of two months with a purified virus preparation using constant volume for injection. The average injected mass for each structural protein was calculated using measurement of peak areas at 214 nm and the calibration of 2262 area units/μg protein (obtained with BSA as the external standard). The molecular masses of the proteins were taken from Table 1. The model copy numbers for the coat proteins (see text, marked with *) and the biochemical copy numbers for the core proteins (see text) were obtained from Ref. [5].

chemical copy numbers [4] are also shown in Table 2. The calculated copy numbers for the coat proteins predicted by the RP-HPLC protein fingerprint are in very good agreement with the published values: all estimates are within less than one standard deviation (SD), with the exception of protein III. The peak area of protein III is the smallest among the identified coat proteins, which may explain the relatively lower accuracy. The calculated copy numbers for the two core proteins (V and VII) show somewhat weaker correlation with the corresponding biochemical copy numbers. The published copy numbers of the core proteins are considered less accurate and reliable than those of the coat proteins [3].

3.4. Estimation of virus concentration from the structural protein content

The concentration of an individual structural protein in a virus sample can be converted into equivalents of virus particle concentration. Using the example of protein II, having an apparent molecular mass of 107 950 Da (see Table 1) and a copy number of 720 molecules/virus [3–5]; 1 μg of protein II corresponds to $9.26 \cdot 10^{-12}$ (mol protein II/μg)$\times 6.022 \cdot 10^{23}$ (virus particles/mol of virus)/720 [mol protein II/mol virus)$=7.72 \cdot 10^{9}$ virus particles. Similar calculations can be made for all structural proteins of the virus with known copy numbers and molecular masses.

In order to evaluate the relative accuracy and precision of the virus concentration estimates based on the five coat proteins in the RP-HPLC fingerprint, increasing amounts of a purified adenovirus preparation were injected onto the RP-HPLC column. The number of injected viral particles was estimated using the signals of the five coat proteins (Table 3). The largest difference between the average estimates derived from the different proteins is 17%. Proteins II, IIIa, VI and IX showed a remarkably good correlation, giving estimates differing by less than 8%. As noted above, the peak of protein III is the smallest among the coat proteins which might explain the lower relative accuracy of its estimation. The results in Table 3 demonstrate linear responses for all five coat proteins within the interval of approximately $2.5 \cdot 10^{9}$–$5 \cdot 10^{10}$ injected virus particles.

3.5. Comparison of different methods to estimate virus concentration

The most reliable protein copy number in the adenovirus is for protein II: 720 copies/virus [3], therefore we chose protein II for further work. Protein II also represents approximately 60% of the protein content of adenovirus. We compared virus concentration estimates based on the (a) measurement of protein II content by the RP-HPLC assay to estimates obtained through (b) measurement of viral DNA and (c) the benchmark spectrophotometric (A260) method [2]. The results are shown in Fig. 3.

The data show excellent agreement between the virus concentration estimates based on protein II content and DNA content over a wide virus con-

Table 3
Independent virus concentration estimates using the five coat proteins identified in the reversed-phase chromatogram[a]

Injection volume (μl)	Virus concentration ($\cdot 10^{10}$ particles/ml)				
	Protein II	Protein III	Protein IIIa	Protein VI	Protein IX
50	5.94	5.25	2.16	5.49	5.30
100	5.69	5.36	6.07	6.01	6.51
370	5.26	4.86	6.06	5.65	5.46
1000	4.28	9.27	6.95	4.61	5.46
Mean	5.29	6.19	5.31	5.44	5.68
RSD (%)	13.85	33.42	40.30	10.94	9.78

[a] A purified virus preparation was injected into the RP-HPLC column using four injection volumes in triplicates for each volume. The virus concentrations for the original sample were calculated using the molecular masses and the model copy numbers from Table 2. The triplicates were averaged for each injection volume.

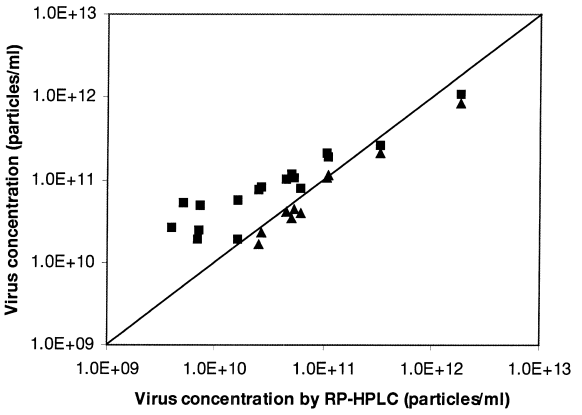

Fig. 3. Comparison of virus concentration estimates by the RP-HPLC assay (using the protein II peak), the benchmark spectrophotometric assay (squares) and the double-stranded DNA assay (triangles). Three purified adenovirus type 5 preparations were tested at various dilutions (all dilutions were handled as independent samples). For assay details, see Experimental.

centration range (approximately $3\cdot10^{10}$ to $3\cdot10^{12}$ viral particles/ml). Note that the protein based and the DNA based measurements had independent mass calibrations (BSA and bacteriophage λ DNA, respectively).

The spectrophotometric virus concentration estimates showed weaker correlation when compared to the two other methods. The differences were the largest in the lower virus concentrations (approximately $3\cdot10^{9}$ to 10^{11} viral particles/ml corresponding to the absorption range of 0.03–0.1 units at 260 nm). The increasing differences from the estimates by the RP-HPLC assay at lower virus concentrations led to an apparent slope of <1 and an apparent positive intercept for the spectrophotometric estimates in this plot. These characteristics are consistent with substantial light scattering in the SDS lysate samples (thus the need for centrifugation) reported by others (e.g., Refs. [17,18]).

3.6. Precision, linear range and accuracy of the RP-HPLC assay for quantification of adenovirus

The precision and the linear range of the RP-HPLC assay using the protein II signal for virus quantification was directly evaluated over the range from $2\cdot10^{8}$ to $3.2\cdot10^{10}$ viral particles injected. We obtained a linear response with a correlation coefficient of $r=0.983$. At the lowest point ($2\cdot10^{8}$ injected virus particles, corresponding to 2.6 ng protein II on the column) the SD of the measurement was 11.1%.

The specificity and accuracy of the assay was demonstrated by analyzing purified virus samples with constant virus concentration containing variable amounts of added tRNA by the RP-HPLC and the spectrophotometric assays. We identified transfer RNA as the major impurity co-purifying with the recombinant adenovirus type 5 in our first purification step (data not shown here). The results of this experiment are presented in Table 4. The data demonstrate that the tRNA interferes with the spectrophotometric assay, and as anticipated, the contribution of the tRNA appears to be additive leading to an overestimation of the virus particle concentration. At the same time, the reversed-phase chromatograms were not affected by the tRNA. The tRNA is not

Table 4
Comparison of the RP-HPLC and the spectrophotometric assays to estimate virus concentration[a]

	Spectrophotometric assay		RP-HPLC assay
	A260	Calculated virus concentration (particles/ml)	Calculated virus concentration (particles/ml)
Buffer+6.6 μg/ml tRNA	0.141	$1.6\cdot10^{11}$	0
Buffer+33 μg/ml tRNA	0.666	$7.3\cdot10^{11}$	0
Virus	0.088	$9.7\cdot10^{10}$	$7.0\cdot10^{10}$
Virus +6.6 μg/ml tRNA	0.232	$2.6\cdot10^{11}$	$7.0\cdot10^{10}$
Virus+33 μg/ml tRNA	0.761	$8.4\cdot10^{11}$	$6.6\cdot10^{10}$

[a] Buffer [phosphate-buffered saline (PBS)+2 m*M* $MgCl_2$] and a virus preparation in the same buffer were tested in the spectrophotometric assay and the RP-HPLC assay with and without addition of tRNA spikes (at 6.6 and 33 μg/ml). The protein II signal and the molecular mass of 107 950 Da, assuming 720 copies/virus were used to calculate virus concentration in the RP-HPLC assay.

retained under the chosen RP-HPLC conditions and therefore, the presence of tRNA did not change the protein II signal used to calculate virus concentration.

4. Discussion

Intact adenovirus type 5 was subjected to RP-HPLC in order to obtain a chromatographic fingerprint of the viral proteome. The results are consistent with a complete dissociation of the virus to individual structural proteins, excellent resolution and high recovery of the protein peaks under the chromatographic conditions. The characterization of the chromatographic peaks led to the identification of all structural proteins of the adenovirus type 5 with greater than 2% contribution to the protein content of the virus. Also, proteins II, III, IIIa, V, VI, VII and IX together represent over 95% of the protein mass of adenovirus type 5. This is calculated with our own MS data and employing model copy numbers for proteins II, III, IIIa, VI and IX [6] and biochemical copy numbers for proteins V and VII [4]. We have not yet been able to identify the fiber protein (protein IV), which is the next largest contributor with approximately 1.8% and which is implicated to have a primary role in the infection process [19,20].

We determined N-terminal and internal sequences of 10 viral proteins and assigned molecular masses with MS. To our knowledge this is the first study published with such data on adenovirus type 5. Combining these techniques, we were able to verify proteolytic processing on several structural proteins (IIIa, VI, VII, VIII, late L2 μ), most of which were anticipated using predictions from adenovirus type 2 data [8]. Our data also suggest a not yet reported processing event on protein VIII. We discovered that the precursor parts of proteins VI, VII, VIII and late L2_μ are present in the purified adenovirus type 5 preparations suggesting their incorporation into the virus. We are not aware of other reports with such observations. Our mass spectroscopy data suggest post-translational or purification related modification(s) on protein IIIa indicated by an increase in molecular mass relative to the prediction.

Besides providing an avenue for characterization of individual structural proteins of the viral proteome, the RP-HPLC assay can be used for the quantification of the virus particles. The method offers significant and unique advantages over the benchmark A260 assay [2] and other potential approaches (including direct microscopic or chromatographic counting, immunological detection of structural proteins or measurement of viral DNA). The RP-HPLC assay (a) measures structural proteins by an absolute calibration with BSA as an external standard using the 214 nm absorbance peak areas, (b) provides structural protein selectivity through efficient separation by means of RP-HPLC, (c) offers independent and parallel virus concentration measurement using each structural protein with known molecular mass and copy number in the virus and (d) offers high sensitivity.

Basing the virus particle concentration assay on proteins is important because empty virions with no DNA encapsulated may exist [21]. Thus for estimating the virus concentration based on DNA content one must verify (or assume) that the virus preparation does not have empty particles, otherwise the total particle concentration might be underestimated. Although the benchmark spectrophotometric assay is based on a DNA calibration [2,22], the published studies using this method (for direct measurement or calibrating other assays) generally only assume the absence of empty particles without verification. An important exception is the anion-exchange chromatographic assay developed by Shabram et al. [18] for counting intact viruses. There the available data suggest that the ion-exchange chromatography may resolve the empty virions from the ones with full DNA load, therefore the total virus particle count could be corrected for the empty viruses. The RP-HPLC assay measures structural protein(s) of the virus, and consequently the results are independent of the DNA content. Note that in our study the DNA and structural protein based virus concentration estimates were similar, suggesting that the contribution of empty virions (if any) to the protein based estimates was negligible for our viral preparations.

The direct spectrophotometric method [2] does not provide information on the presence of interfering contaminants. The purity of the sample has to be independently verified. Similarly, the absence of non-viral nucleic acids must be verified, or these nucleic acids must be removed from the preparation

before a direct DNA assay can be applied to measure virus concentration. The virus concentration estimation based on the anion-exchange chromatographic method [18] introduced a selectivity into the assay with respect to some impurities. Benzonase treatment of the samples was proposed to degrade the DNA fraction that would otherwise coelute with the virus on the column. The protein purity of the virus peak, however, needs to be independently verified for the samples of interest. In contrast, the RP-HPLC protein fingerprint can be applied to virus preparations of relatively low purity, because it allows simultaneous estimation of the virus concentration based on individual structural proteins in the virus sample. The feasibility of this approach was demonstrated in our study by (a) showing that the estimated copy numbers of the coat proteins in the RP-HPLC fingerprinting assay were in good agreement with the predictions by the crystallographic and electron microscopic studies, and by (b) testing the assay in the presence of a known contaminant (tRNA). Significant differences between the virus concentration estimates would only be anticipated if impurities co-eluted with coat proteins in the chromatogram. However, as long as there is one coat protein peak with acceptable purity (which can be confirmed by on-line MS), the fingerprinting assay can provide accurate estimates of virus concentration.

A further important selectivity related problem with the spectrophotometric virus concentration measurement is the precipitation and/or light scattering in the virus lysates in SDS reported by several other authors (e.g., Ref. [17]). Centrifugation of the SDS lysates is recommended to remove precipitates. While removal of compounds absorbing at 260 nm by centrifugation can lead to underestimation of the virus concentration, residual light scattering may contribute to overestimation. The RP-HPLC protein fingerprint bypasses these difficulties; no precipitation and light scattering could be observed.

In addition, the RP-HPLC assay is very sensitive. A linear response was demonstrated down to $2 \cdot 10^8$ injected virus particles. This sensitivity is about two-orders of magnitude better than the sensitivity of the spectrophotometric assay, in particular because the light scattering problem of the latter method increases toward lower particle concentrations.

The proposed RP-HPLC assay, based on external protein calibration, can also be used to calibrate reference standards for chromatographic [18] or for direct particle counting techniques (such as electron microscopy or microscopy with fluorescent labeled viruses). It is generally recognized [23] that the recovery of the viruses throughout the microscopic measurement is variable and therefore must be established with a well characterized sample with a known virus concentration.

Acknowledgements

The authors wish to thank Professor Csaba Horváth for his helpful suggestions.

References

[1] K.L. Berkner, in: N. Muzycka (Ed.), Current Topics in Microbiology and Immunology, Springer, Berlin, Heidelberg, New York, Tokyo, 1992, p. 39.
[2] J.V. Maizel Jr., D.O. White, M.D. Scharff, Virology 36 (1968) 115.
[3] R.M. Burnett, in: W. Chiu, R.M. Burnett, R.L. Garcea (Eds.), Structural Biology of Viruses, Oxford University Press, Oxford, New York, 1997, p. 209.
[4] J. van Oostrum, R.M. Burnett, J. Virol. 56 (1985) 439.
[5] P.L. Stewart, S.D. Fuller, R.M. Burnett, EMBO J. 12 (1993) 2589.
[6] P.L. Stewart, R.M. Burnett, Curr. Top. Microbiol. Immunol. 199 (1995) 25.
[7] M.-L. Boudin, J.-C. D'Halluin, C. Cousin, P. Boulanger, Virology 113 (1980) 781.
[8] C.W. Anderson, Virology 177 (1990) 259.
[9] K.G. Mullis, R.S. Haltiwanger, G.W. Hart, R.B. Marchase, J.A. Engler, J. Virol. 64 (1990) 5317.
[10] A. Appfel, J. Chakel, S. Udiavar, W.S. Hancock, C. Souders, E. Pungor Jr., in: A.P. Snyder (Ed.), Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry, American Chemical Society, Washington, DC, 1995, p. 432.
[11] Personal communication with B.F. Parten, based upon B.F. Parten, G.E. Petrie, R.L. Marlowe, J.C. Anders, poster presented at the 10th Symposium of the Protein Society, 3–7 August 1996, San Jose, CA.
[12] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, Prot. Sci. 4 (1995) 2411.
[13] G.H. Beaven, E.R. Holiday, Adv. Protein Chem. 7 (1952) 319.
[14] P. Lemay, M.-L. Boudin, M. Milleville, P. Boulanger, Virology 101 (1980) 131.

[15] U. Hobohm, T. Houthaeve, C. Sander, Anal. Biochem. 222 (1994) 202.
[16] R.R. Becklin, D.M. Desiderio, Anal. Lett. 12 (1995) 2175.
[17] N. Mittereder, K.L. March, B.C. Trapnell, J. Virol. 70 (1996) 7498.
[18] P.W. Shabram, D.D. Giroux, A.M. Gordeau, R.J. Gregory, M.T. Horn, B.G. Huyghe uygh, X. Liu, M.H. Nunnally, B.J. Sugarman, S. Sutjipto, Hum. Gene Ther. 8 (1997) 453.
[19] J.M. Bergelson, J.A. Cunningham, G. Drougett, E.A. Kurt-Jones, A. Krithivas, J.S. Hong, M.S. Horwitz, R.L. Crowell, R.W. Finberg, Science 275 (1997) 1320.
[20] D. von Seggern, C.Y. Chiu, S.K. Fleck, P.L. Stewart, G.R. Nemerow, J. Virol. 73 (1999) 1601.
[21] M. Toth, B. Tarodi, I. Beladi, Acta Virol. 26 (1982) 217.
[22] N. Green, N. Pina, Proc. Natl. Acad. Sci. 51 (1964) 1251.
[23] K.P. Hennes, C.A. Suttle, Limnol. Oceanogr. 4 (1995) 1050.