ELSEVIER

Journal of Chromatography A, 921 (2001) 39–48

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Polypeptide composition of an adenovirus type 5 used in cancer gene therapy

Francis Blanche*, Bertrand Monegier, Didier Faucher, Marc Duchesne, François Audhuy, Anne Barbot, Sophie Bouvier, Gaëlle Daude, Hervé Dubois, Thierry Guillemin, Laurent Maton

Aventis Pharma, Centre de Recherche de Vitry-Alfortville, 13 Quai Jules Guesde, BP 14, 94403 Vitry/Seine Cedex, France

Abstract

For cancer gene therapy, a recombinant adenovirus serotype 5 named RPR/INGN201 has been constructed by susbtitution of the E1 region with human tumor suppressor gene *p53*. The protein components of RPR/INGN201 virions were separated by reversed-phase HPLC and were individually identified by electrospray time-of-flight mass spectrometry and N-terminal sequencing, both on intact proteins and on their proteolytic fragments after trypsin digestion. Twenty-five peptide components of the proteome (including fiber) with greater than 0.25–0.5% contribution to the protein content of the virus were identified and characterized. Fiber was confirmed to be partially glycosylated (both the non-glycosylated and the monoglycosylated states were identified), and two proteins were isolated and identified as phosphorylation derivatives, namely protein V (non-phosphorylated and monophosphorylated) and protein IIIa (mono- and diphosphorylated). This new analytical tool proved to be very useful not only for refining our current knowledge of the polypeptide repertoire of purified infectious virions but also for monitoring and very rapidly identifying structural modifications resulting from changes in the manufacturing process. It was also used successfully for the characterization of various adenoviral constructs.

Keywords: Adenovirus; Peptides; Polypeptides

1. Introduction

Recombinant adenoviruses are widely used for transferring genes in a variety of current gene therapy protocols. An adenovirus named RPR/INGN201 containing a *p53* expression cassette is used at Aventis Pharma and Introgen Therapeutics in Phase I and II clinical trials to treat patients for head-and-neck cancers. To satisfy the growing use of these vectors in clinical development, new methodologies for production and purification of adenoviruses have been developed. Particularly the classical purification using density gradient centrifugation has been replaced by more sophisticated processes involving one [1] or several [2] chromatographic steps. In the same time, rapid and sensitive assays have been developed to monitor virus production and processing and to control the quality of final preparations to be used in clinical trials. Of particular relevance is the development of anion-exchange

*Corresponding author. Tel.: +33-1-5571-8093; fax: +1-33-1-5571-2135.

E-mail address: francis.blanche@aventis.com (F. Blanche).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)00896-2

high-performance liquid chromatography (HPLC) methods which allowed for rapid quantitation of adenovirus particles in crude as well as purified samples [3,4].

The adenovirus particle is a very large M_r $15 \cdot 10^6$ macromolecular assembly composed of a linear double-stranded 35 kilobase pair (kbp) DNA genome and at least 11 distinct structural proteins (polypeptides II–IX, IIIa, μ, and terminal) present in multiple copies and held together by non-covalent interactions [5–7]. Polypeptides II (hexon), III (penton base), IIIa, and IV (fiber) are the major constituents of the outer icosahedral capsid of the virion. The trimeric fiber protein is protruding outward from the 12 vertices of the icosahedron. Hexon-associated polypeptides VI and VIII are located on the inner surface of the capsid, the former also interacting with the viral DNA molecule, while polypeptide IX stabilizes the capsid structure through interactions with hexons. Polypeptides V, VII and μ are tightly associated with viral DNA to form a nuclear core and one terminal protein molecule is covalently bound to the 5′-end of the viral DNA. Six of these polypeptides (IIIa, VI, VII, VIII, μ, and terminal protein) are in fact cleavage products of their respective precursors (pIIIa, pVI, pVII, pVIII, and X) proteolytically processed by a specific viral protease (L3/p23) during viral maturation [8,9]. Other post-translational protein modifications have also been reported, including glycosylation of fiber [10] and phosphorylation of proteins IIIa [11], V, and VI [8].

Despite the very large body of structural information gained on adenovirus, we still have at the molecular level a sketchy knowledge of the protein components that make up an adenovirus particles. Therefore, there is a need for methods that could provide us with a refined analytical definition of the adenovirus particle within a short time frame.

To this end, and in order to get a deeper knowledge of the molecular composition of highly purified adenoviral preparations that are used in gene therapy protocols, we have used reversed-phase (RP) HPLC for the separation of the proteins and peptide which make up an adenoviral particle and we have identified all of them by a combination of N-terminal sequencing, peptide mapping with trypsin and mass spectrometry (MS).

2. Experimental

2.1. Preparation of adenovirus

Experiments were performed using a replication-deficient recombinant adenoviral construct derived from an adenovirus type 5 backbone [12] named RPR/INGN201. This construct contained the human *p53* gene included in the expression cassette designed for suppressing the growth of tumors through the expression of *p53*. The entire genome (35.3 kbp) was entirely sequenced on both strands at Lark Technologies (Houston, TX, USA). Adenovirus production and purification has been described previously [1]. Briefly, ATCC 293 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) medium supplemented with 10% fetal bovine serum (Hyclone) and infected at a multiplicity of infection of 50 viral particles (vp) per cell. Six days post infection, the culture liquid was collected, clarified and concentrated by ultrafiltration on an M_r 300 000 Biomax membrane (Millipore, Bedford, MA, USA). The viral suspension was purified on a column of Source 15Q (Amersham Pharmacia Biotech, Uppsala, Sweden) equilibrated with 20 m*M* Tris–HCl, 250 m*M* NaCl (pH 8.0) and eluted at a flow-rate of 150 cm/h with a linear gradient of 250 m*M* to 1 *M* NaCl over 30 column volumes. The fraction containing infectious adenovirus particles (eluted with 0.5 *M* NaCl) was collected and adjusted to 10% (v/v) glycerol, concentrated by ultrafiltration, and stored at −20°C until use.

2.2. Purification of empty particles

The peak corresponding to adenovirus particles (FI; $1.5 \cdot 10^{14}$ vp) was followed by a small peak (FII; approx. 10 mg of viral protein) containing empty capsids. This peak was collected, adjusted to 10% of glycerol, diluted to 150 m*M* NaCl and re-chromatographed on a FineLine 35 column packed with 100 ml of Source 15Q equilibrated with 50 m*M* Tris–HCl, 1 m*M* $MgCl_2$, 0.25 *M* NaCl, 10% glycerol. Empty capsids were eluted with a linear gradient of 0.25 to 1.0 *M* NaCl over 50 column volumes at a linear flow-rate of 60 cm/h and chromatographed again on a 10-ml HR10/10 Source 15Q column

operated under similar conditions. Fractions containing purified empty capsids were pooled and stored at −20°C until use.

2.3. Analysis of empty capsids by centrifugation

Purified empty capsids were mixed with 1.34 g/ml CsCl in 50 m*M* Tris–HCl (pH 8.0) and centrifuged overnight at 35 000 rpm, 18°C in a Beckman SW41Ti rotor. The band was collected and buffer-exchanged into 50 m*M* Tris–HCl, 1 m*M* $MgCl_2$, 10% glycerol (pH 8.0). A reference sample of purified adenovirus was banded similarly.

2.4. Purification of fiber

A preparation of purified adenovirus ($5 \cdot 10^{12}$ vp/ml) was dialyzed overnight against 100 volumes of water and centrifuged at 40 000 *g* for 3 h at +4°C. The supernatant was loaded onto a 1-ml column of Source 15Q equilibrated with 20 m*M* Tris–HCl (pH 7.2) and fiber was eluted with a linear gradient of 0 to 0.3 *M* NaCl over 30 column volumes. Fiber-containing fractions eluted with 50 m*M* NaCl were pooled and dialyzed against acetonitrile–0.1% aqueous trifluoroacetic acid (TFA) (25:75). Fiber detection was performed by Western blot analysis, using monoclonal anti-fiber antibody Ad2/2F7 [13] and electrochemiluminescence (ECL) detection.

2.5. Particle quantitation by anion-exchange HPLC

Quantitation was performed with a Waters quaternary LC 626 system equipped with a 996 photodiode array detector (set at 220 to 350 nm) as described previously [4]. Briefly, a 1-ml HR 5/5 column packed with Q Sepharose XL (Amersham Pharmacia Biotech) was used. The mobile phase was a gradient of 0 to 1 *M* NaCl in 20 m*M* Tris–HCl (pH 7.5) at 1.5 ml/min.

2.6. Electron microscopy

A 20-μl drop of an adenoviral preparation was deposited onto a carbon-coated grid and treated with 1.5% uranyl acetate. After air drying, the sample was examined at magnification of 20 000 to 50 000 with a Jeol 1010 electron microscope operated at 80 kV.

2.7. RP-HPLC

Preparative RP-HPLC was performed by using a modification of a method described recently [14]. HPLC experiments were carried out with a Waters LC 626 system equipped with a 996 photodiode array detector (set at 220 to 350 nm) using a C_4 Jupiter column (250×4.6 mm I.D., 5 μm; Phenomenex, Torrance, CA, USA) eluted at 50°C and 1.0 ml/min with a linear gradient of 20% to 65% acetonitrile in 0.1% aqueous TFA over 125 min. Virus samples ($5 \cdot 10^{12}$ vp) were injected onto the column after dilution with four volumes of acetonitrile–0.1% aqueous TFA (25:75). Each chromatographic peak was collected manually for further analyses, including N-terminal amino acid sequencing, proteolytic cleavage with trypsin, and MS. Analytical RP-HPLC analyses were carried out similarly using a 150×2.1 mm I.D. column eluted at 0.2 ml/min and 50°C with a linear gradient of 20% to 65% acetonitrile in aqueous TFA over 66 min. LC–electrospray ionization (ESI) MS (see below) was performed similarly with a slightly steeper gradient of acetonitrile (20–70%) in aqueous 0.05% TFA.

2.8. Mass spectrometry

RP-HPLC coupled with ESI-MS experiments on both intact material and fragment peptides obtained by proteolytic cleavage were carried out on an LCT mass spectrometer (Micromass) equipped with the Z-Spray ion source and a HP1100 HPLC accessory (Hewlett-Packard) including a photodiode array detector. LC–ESI-MS of peptides obtained after proteolytic cleavage was performed using a C_{18} Vydac column (250×4.6 mm I.D., 5 μm) eluted at 30°C and 1 ml/min with a gradient of 0 to 50% of acetonitrile in 0.05% aqueous TFA over 80 min. When flow-rates of 1 ml/min were used, a split at the outlet of the UV cell was added in order to direct 10% of the eluate towards the mass spectrometer and 90% to a fraction collector for microsequencing. For lower flow-rates (200 μl/min), the whole mobile

phase was injected into the mass spectrometer. UV spectra (190 to 600 nm) and mass spectra were recorded together in the same data file. The following operating conditions were used: ionization voltage: 3500 V, cone voltage: 45 V, source temperature: 110°C, nebulizer gas: 90 l/h, desolvation gas: 550 l/h, desolvation gas temperature: 130°C. The time-of-flight analyzer was calibrated with cluster ions of cesium iodide. Acquisition of one spectrum per second was done in the profile mode from 200 to 1800 u. LC–ESI-MS chromatograms were displayed as total ion current (TIC) and UV signal versus time. Mass spectra of peptides and proteins were reconstructed by averaging the spectra of the different peaks eluted with background subtracted from both sides. Masses of intact proteins obtained by LC–ESI-MS were all confirmed by introducing an aliquot of each of the collected fractions at a flow-rate of 10 μl/min in the spectrometer with a syringe pump (Harvard Apparatus) after dilution in 0.05% aqueous TFA–acetonitrile (50:50).

2.9. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analyses

SDS–PAGE separations were run under standard conditions using precast 4–20% gels (Novex), and proteins were transferred onto poly(vinylidene difluoride) (PVDF) membranes for sequencing using an Applied Biosystems 494 microsequensor. N-Terminally blocked proteins were submitted to CNBr cleavage prior to sequencing [15]. In this later case peptide sequences comparisons using FAST and SwissProt softwares were used for protein identifications.

3. Results

3.1. Purification of adenovirus and separation of empty capsids

The profile obtained from a crude cell lysate chromatographed on Source 15Q (Fig. 1) displayed a major peak (FI) corresponding to fully infectious particles immediately followed by a minor peak (FII). Contaminants were eluted in the flow-through or between 120 and 175 min before the adenovirus

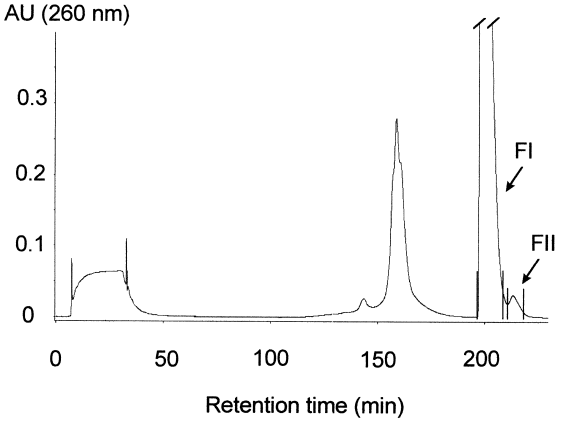

Fig. 1. Preparative-scale purification on Source 15Q of adenovirus type 5 from a cell culture ultrafiltrate. FI, Adenovirus peak; FII, empty capsids; $1.5 \cdot 10^{15}$ particles loaded onto a 2.3-l column.

peak. Peaks FI and FII were not completely baseline separated even under optimized conditions. However, the minor species FII could be obtained essentially free of adenovirus particles after two additional chromatography steps on Source 15Q. This species FII was identified as empty capsids (see below).

3.2. Identification of the protein components of purified virions

Adenovirus preparations obtained with the process described in the Experimental section were of high purity. They contained exquisitely low levels of bovine serum albumin (BSA) (<10 ng per $1.0 \cdot 10^{12}$ vp, i.e., less than 1 ng BSA per 25 μg viral protein) which is a good indicator of very low protein contamination. In addition, the level of contamination by empty capsid was also very low (see below). Therefore, these preparations could be confidently used to identify the protein composition of native particles.

A typical analytical RP-HPLC chromatogram obtained from purified virus preparations (Fig. 2) contained 21 peaks (named A to U) with greater than 0.25–0.5% contribution to the total protein content of the virus. Peak C was actually comprised of two small peaks which were not separated under preparative conditions (not shown). On-line LC–ESI-MS experiments (Fig. 3) gave a similar pattern, with a single molecular mass (ranging from 1000 up to

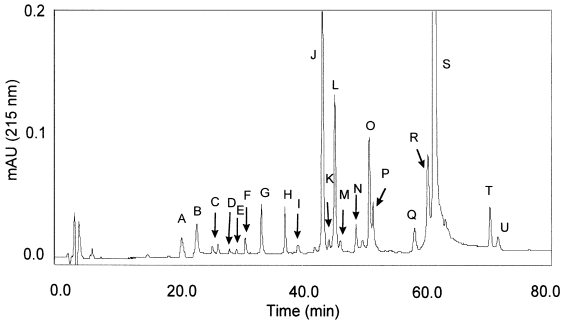

Fig. 2. Typical analytical RP-HPLC chromatogram of an adenovirus type 5 preparation purified by chromatography on Source 15Q. $3 \cdot 10^{10}$ viral particles were injected onto the column. Twenty one fractions (A to U) were collected at preparative scale for protein identification. Fraction C contains two peaks which were not separated at preparative scale.

110 000) for most of the peaks (Table 1), which is consistent with a single component in each fraction. Two molecular masses were obtained for peaks A, D, Q, and R, revealing a molecular heterogeneity which was characterized upon further analysis (see below). In the TIC profile, two peaks of M_r 432 and 460 eluted between peaks H and I and between M and N, respectively, had no UV counterpart (Fig. 3)

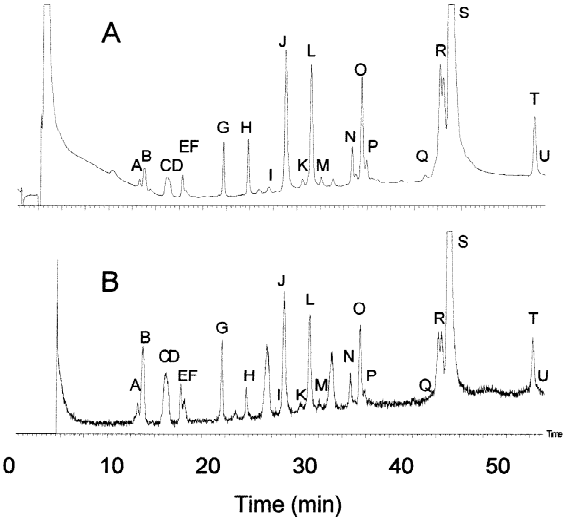

Fig. 3. Typical analytical chromatogram obtained by LC–ESI-MS. (A) Sum of the UV response from 190 to 600 nm versus time. (B) TIC obtained by electrospray mass spectrometry. In the particular experiment shown, peaks C and D, E and F, and T and U are co-eluted.

and were likely to correspond to small polypeptides. All of the 21 UV peaks were collected from a preparative RP-HPLC run and were analyzed by N-terminal amino acid sequencing. The mass data were combined with amino-terminal sequence information obtained from each chromatographic peak to identify some of the peptides and their primary sequence, particularly in peaks A, C, D, E, F, I, J, N, O, and U. For all these proteins, the difference between theoretical and predicted molecular masses was less than 2 u. In spite of the identification of two sequences for peak M, both belonging to protein VI (VI, 34–45 and VI, 109–120), no primary sequence could be assigned to the observed molecular mass of 22 105. On the other hand, no amino-terminal sequence could be detected for fractions B, G, H, K, L, P, Q, and S indicating that the corresponding proteins were blocked. A sequence of low abundance (less than 10-fold lower than expected) was identified from fractions R and T, indicating that only a minor component of the fraction was detected in these cases (see below).

Then, most of the chromatographic fractions containing individual peaks (see Table 1) were submitted to proteolytic cleavage with trypsin and the digests were analyzed by LC–ESI-MS. Owing to the post-column split, 10% of the eluate could be used to get mass data while the remaining 90% was fraction collected for sequencing of selected fractions. The masses of the fragments confirmed the initial identification of peaks D, E, F, J, N, O and U and allowed the identification of the proteins of peaks B, G, H, L, P, Q, R, S and T.

As a result of this study, all of the 25 individual peptides constituting the 21 RP-HPLC peaks could be unambiguously identified, including those which were N-terminally blocked. For all peptides (excluding fraction F), 60 to 100% of the primary sequence could be confirmed from mass analyses of the peptide digests. Moreover, by combining these results with the molecular masses obtained on intact proteins, we could identify both the amino- and the carboxy-terminal amino acids of some intact proteins. All these results are summarized in Table 1.

The two components of peak A were identified as fragments 14–24 and 15–24 of pVII precursor. Although not described previously, the cleavage at position 13/14 within the sequence NTGW/GL is

Table 1
Molecular composition of the adenovirus type 5 particle

Peak name	%	N-Terminal sequence (peptide ID)	Measured M_r	Predicted M_r	Protein identification	Protein coverage[a] (LC–MS of tryptic fragments)	Comments
A	1.5	LRFPSKMFGG (VII, 15–24)	1138	1138	VII 15–24	ND	Cleavage at NTGW/G
		GLRFPSKMFGG (VII, 14–24)	1197	1196	VII 14–24	ND	(13/14) not reported
B	1.7	SILISPSNNTG (VII, 2–12)[b]	1329	1329	VII Ac2–13	One single peak (M_r 1329); no cleavage site in the sequence	Cleavage at NTGW/G (13/14) not reported
C	0.90	ALTXRLRFPVPGFRG (X 2–16)	3035	3035	X 2–27	ND	
D	0.23	GIQLNDESVSSSLGL (VIII, 132–146)	2547	2547	VIII 132–156	Two peaks (M_r 2547 and 2618); no cleavage site in the sequence	Cleavage at QIGG/A (156/157) not reported
			2618	2618	VIII 132–157		
E	0.16	GNPFAHLRPRLGRMF (IIIa, 571–585)	1768	1768	IIIa 571–585	68% (N ter)	
F	0.64	ALTXRLRFPVPG (X, 2–13)	6071	6071	X 2–53	15% (position 7 to 15)	Cleavage site at 53/54 (GIL/P) unexpected
G	2.0	No sequence	3624	3624	VI Ac1–33	88% (N and C ter)	
H	1.3	No sequence	12 036	12 036	VIII Ac2–111	92% (C ter)	
I	0.81	XXXXVEEPPPAV (VI, 109–120)	13 967	13 968	VI 109–239	79% (C ter)	D/P(108–109) bond hydrolysis
J	11.8	AKKRSDQHPVRVRGH (VII, 25–39)	19 412	19 412	VII 25–198	91% (C ter)	
K	0.13	No sequence	41 536	41 537	V Ac2–368 (P)	ND	
L	4.9	No sequence	41 456	41 457	V Ac2–368	95% (N and C ter)	
M	0.27	AFSWGSLWSGIK (VI, 34–45) PRPPVEEPPPAV (VI, 109–120)	22 105	?	?	V and VI identified	D/P(108–109) bond hydrolysis
N	0.81	GRSXFXPRQAILTLQ (VIII, 158–172)	7642	7642	VIII 158–227	98% (C ter)	
O	4.26	AFSWGSLWSGIKNFG (VI, 34–48)	22 100	22 100	VI 34–239	76% (N ter)	
P	0.80	No sequence	63 340	63 335	III Ac1–571	66% (C ter)	
Q	0.89	No sequence	61 631	61 627	IV Ac1–581	No fragment	
			61 833	61 830	IV Ac1–581(GlcNAc)		
R	4.5		63 631	63 624	IIIa Ac1–570 (P)	62% (N ter)	One phosphorylation on threonine at position 8
			63 711	63 704	IIIa Ac1–570 (P2)		
		XMDEPTLLYVLFEVF, weak (II, 909–923)	5017	5017	II 909–952		D-P(908–909) bond hydrolysis
S	61.1	No sequence	107 868	107 860	II Ac2–952	85% (N and C ter)	
T	0.91		14 369	14 369	IX Ac2–140	64% (N and C ter)	
		XTNSFDGSIVSSYL, weak (IX, 2–15) (5%)	–				
U	0.53	ILPLLIPLIAAA; X 52–63	2888	2888	X 52–80	100% (N and C ter)	

ND, Not determined.

[a] Fractional number of amino acids identified in tryptic fragments relative to total amino acids in uncleaved peptide (%). Fragments comprising the N-terminal or the C-terminal residue are reported as N ter and C ter, respectively, where identified.

[b] Sequence determined from a peptide released during fiber purification experiments (see text).

fairly close to one of the two consensus recognition sequence of the M_r 23 000 viral protease [(M, L, I)XGX/G], the other one being [(M, L, I)XGG/X], (where X represent any amino acid and the cleavage site is designated by «/») [8]. In contrast, the cleavage between amino acids 14 and 15 is quite unexpected and the corresponding peptide 15–24 is likely to result from a secondary cleavage of the amino terminal glycine residue of peptide 14–24.

Peptide B was N-terminally blocked. Interestingly, during the first dialysis step of fiber purification (see below) a peptide was released from intact virions

which co-eluted with peak B by RP-HPLC and whose N-terminal sequence (SILISPSNNTG) was identical to that of fragment 2–12 of pVII precursor.

The two peptides in fraction D are fragments 132–156 and 132–157 of pVIII precursor. Both cleavage sites at positions 131/132 and 157/158 have been previously described [11]. In contrast, the cleavage site QIGG/A identified at position 156/157 is original and not quite consistent with the consensus recognition sequences of the protease. Overall, these results indicate that the consensus recognition sequences of the protease should probably have to be broaden to (M, L, I, Q)XGG/X and (M, L, I, N)XGX/G.

The protein component of fraction L was unambiguously identified as protein V (initiator methionine removed upon acetylation) and the small peak K eluted right before L was identified as its monophosphorylated derivative.

Molecular masses of 61 631 and 61 833 were obtained on fraction Q by LC–ESI-MS. In contrast, no mass was obtained either when fraction Q was re-injected into the mass spectrometer after collection or when the tryptic digest was analyzed by LC–ESI-MS, possibly because the tiny amount of material present in fraction Q. The difference of 202 u observed between the two components of this fraction being indicative of a difference in glycosylation (probably *N*-acetylglucosamine; see Ref. [10]) these two peptides were tentatively identified as fiber (protein IV, predicted mass 61 627) and its *N*-acetylglucosamine derivative (predicted mass 61 830), both of them being acetylated. This result was further unambiguously confirmed when authentic purified trimeric fiber was analyzed by LC–ESI-MS under identical conditions (not shown), giving the same two masses obtained from intact virions.

Fraction R was even more complex with three masses determined by LC–ESI-MS on intact proteins: 63 631, 63 711 and 5017. After trypsin digestion, LC–ESI-MS unambiguously identified protein IIIa as the major component with more than 60% of the primary sequence covered. The two large peptides of M_r 63 631 and 63 711 were identified as mono- and diphosphoester derivatives of protein IIIa (fragment 1–570, acetylated) with a difference of 7 u between predicted and calculated masses. Moreover, one of the two phosphorylation sites could be identified by carefully examining by LC–ESI-MS the peptide map of the tryptic digest. Two masses of 1355 and 1276 were measured at 16.2 and 16.4 min, respectively, which were interpreted as mono- and non-phosphorylated derivatives of fragment Ac1–12 of IIIa. Since one amino acid only (threonine at position 8) can be phosphorylated within this short sequence, we propose this amino acid as one of the phosphorylation sites in IIIa. The small peptide (5017) was identified as the carboxy-terminal fragment 909–952 of protein II (hexon). This cleavage product is likely to be formed by chemical hydrolysis of the fragile Asp–Pro peptide bond at position 908–909 in the hexon protein. Interestingly similar cleavage products of an Asp–Pro bond in pVI precursor (at position 108/109) were identified in fraction I (as fragment 109–239) and in fraction M. Fraction M, which was of exquisitely low abundance (0.27%), was obviously a complex mixture of peptides derived from pVI precursor.

Fraction T was identified as protein IX (acetylated, N-terminal methionine removed). Interestingly, a fraction of this protein (5%) was not acetylated, and could be sequenced at its N terminus.

3.3. Protein composition of empty capsids

The minor peak FII eluted right after the main adenovirus peak on Source 15Q (Fig. 1) was identified as empty capsids on the following interlocking grounds.

Cesium chloride banding of FII revealed one main light band (d, 1.266) and a minor heavy band (d, 1.30) in proportions of 98% and 2%, respectively, by anion-exchange HPLC. The ratio of absorbance ($A_{260\ nm}/A_{280\ nm}$) was substantially lower for FII (1.12) than for FI (1.28). The minor band was identified as complete native particles by anion-exchange HPLC, electron microscopy, and by comparison with a reference sample ultracentrifuged in parallel. DNA was extracted from the viral material from the light and heavy bands, digested with *Hind*III, and analyzed by agarose gel electrophoresis. No DNA could be detected from the light fraction in contrast with the sample of FI processed in parallel (not shown). In line with this, the infectivity of empty capsids was 100-fold lower than the infectivity of native virions.

Electron micrographs obtained from the light band after centrifugation (not shown) revealed the presence of irregular round-shaped pseudo-capsids structures of 60–70 nm in diameter. Some hexon capsomers were missing, leaving the interior of the pseudo-capsid and the outside medium in contact. A few hexon capsomers free or arranged in typical groups of nine (GON) [5,7] were also present in the preparations. Most of the pseudo-capsids were aggregated in small clumps.

RP-HPLC analysis (Fig. 4) indicated that empty capsids consist of hexon (more than 90%) and protein IX (2%), mainly. Traces of protein VI (Ac1–33), protein VII (25–198), protein VIII (Ac2–111 and 158–227) were detected in the preparation, which might result from a slight contamination of the preparation (<5%) by complete particles. None of the other capsid and core proteins could be detected. An additional peak (named X in Fig. 4; less than 1% of total peak area) not detected in purified infectious particles was present in the chromatogram. It was not studied any further.

In agreement with RP-HPLC results, SDS–PAGE analysis of FII (Fig. 5) revealed one major protein identified as protein II (hexon), one minor band at M_r 14 000 (protein IX), and numerous minor proteins forming a smear ranging from M_r 10 000 to 100 000.

No band corresponding to light empty capsids could be detected when a sample of FI was analyzed by centrifugation in CsCl which indicated that the actual contamination of purified adenovirus (FI) by empty capsids was below 0.5%.

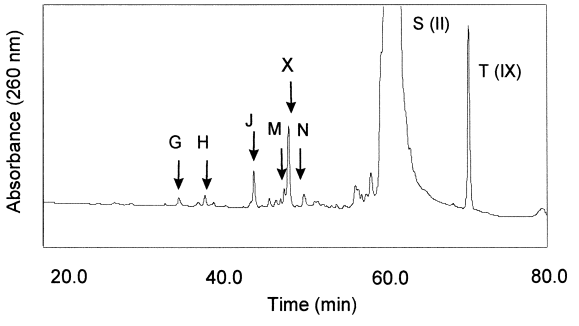

Fig. 4. RP-HPLC analysis of empty capsids (zoomed-in view). Empty capsids were purified by anion-exchange HPLC on Source 15Q followed by ultracentrifugation in CsCl gradient. Capital letters indicate the retention time of the corresponding component identified from infectious virions (see Fig. 2 and Table 1 for definitions). An unidentified polypeptide labeled X not present in infectious virions is found in empty capsids.

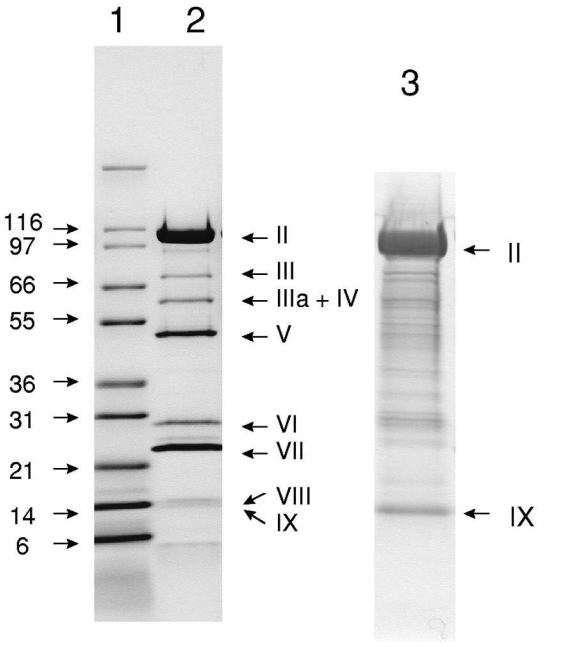

Fig. 5. SDS–PAGE of adenovirus. Lanes: 1=molecular mass markers (values in thousand, on left), 2=purified native virions, 4 μg of protein (viral proteins identified, on right), 3=purified empty capsids, 6 μg of protein (viral proteins II and IX identified, on right). Gels (Novex, 4–20%) were stained with Coomassie Blue.

4. Discussion

Adenovirus particles were efficiently separated from residual contaminants present in ultrafiltrates of production supernatants by a single anion-exchange HPLC step using Source 15Q [1], leading to preparations in which cellular or other contaminants were undetectable. In addition, this purification process also removed the empty capsids that were present in the crude harvest. To our knowledge this is the first report demonstrating that anion-exchange chromatography can efficiently remove empty capsids from intact adenoviral particles, and reporting on the protein composition of empty capsids by RP-HPLC. Empty capsids, i.e., viral particle lacking DNA, are well documented in the literature [16].

Intact particles were dissociated into their indi-

vidual protein components which were subjected RP-HPLC–ESI-MS analysis. In addition, each one of the 21 individual RP-HPLC peaks with greater than 0.2–0.5% contribution to the protein content of adenovirus was collected and the corresponding protein was identified. Overall, 25 individual peptides incorporated in native virions were identified. Importantly, this extensive characterization work could not have been achieved without prior knowledge of the entire adenovirus genome sequence without ambiguity. Post-translational modifications reported previously for adenovirus type 2 were confirmed for most of them and removal of the N-terminal methionine upon acetylation was confirmed for proteins II, and IX and was demonstrated for proteins V and VIII. In addition, the proteolytic processing and cleavage positions for proteins pIIIa, pVI, pVII, pVIII, and pX into IIIa, VI, VII, VIII, and X were verified.

The complex processing of polypeptide VIII at positions 111/112, 131/132, and 157/158 reported previously [11] was confirmed and refined with the discovery of an additional cleavage position at 156/157 which does not fit in strictly with the consensus recognition sequences of the M_r 23 000 viral protease [17].

Three proteolytic fragments of pX were identified (2–27, 2–53, and 52–80). The removal of the N-terminal methionine and the cleavage at positions 27/28 and 51/52 conforms to the proteinase cleavage consensus sequence. However, we could not find fragment 33–51, originally identified from virus preparations obtained from KB cells and named μ [18]. The occurrence of fragment 2–53 revealed a cleavage at position GGIL/PL which was quite unexpected owing to the known specificity of the viral protease. The significance of this observation has not been investigated any further. However, this peptide is likely to be tightly associated with DNA through its extremely basic segment 33–51, thereby facilitating DNA packaging into the virion.

The N-terminal part of protein VII (Ac2–24) was found to contain an internal cleavage site at position 13/14 not described previously [11].

In contrast to previous results obtained with adenovirus type 2 [5,19], protein IIIa was not processed at its N-terminus, and its N-terminal methionine was acetylated.

The presence of the peptides cleaved off by the proteolytic cleavage of precursor pIIIa, pVI, pVII, pVIII, and pX was demonstrated in purified particles. The fragments identified for pIIIa, pVII, and pX covered the entire polypeptide, while the fragments covering positions 240–250 and 112–131 were missing for polypeptides pVI and pVIII, respectively. The presence of the precursor parts of structural proteins within the viral capsid was firmly established during this study. This result is in agreement with previous studies which demonstrated the incorporation of numerous peptide fragments at low abundance in the virions, but these fragments were not identified by that time [8]. Interestingly one aspartic acid–proline peptide bond underwent partial cleavage at positions 909/910 and 108/109 in proteins II and VI, respectively. These chemical cleavages are likely to occur during chromatography when the proteins are exposed to low pH conditions [20].

Other post-translational modifications readily identified during this study involved: (i) the phosphorylation of protein V, giving a small peak (monophosphorylated form) eluted right before the major (non-phosphorylated) form of protein V, and (ii) two forms (mono- and diphosphorylated) of protein IIIa, both eluted under the same peak. This is in line with previously reported observations showing that protein IIIa can exist in several phosphorylation states in the virus [11] and that protein V is prominently labeled in virus labeled with [^{32}P]orthophosphate after infection [8]. In addition, fiber which is of low abundance in virions (about 1% of total protein content) was co-eluted with its *N*-acetylglucosamine derivative from the RP-HPLC column. A reference sample of fiber behaved similarly. Indeed, the exact structure of the hexosamine was not determined unambiguously. However, glycosylation of fiber with *N*-acetylglucosamine is abundantly documented in the literature [10,21,22].

A recent study by Lehmberg et al. reported on the development of an LC–matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS method to separate and identify the structural proteins of adenovirus [14]. In essence, the results reported were in line with the general conclusions drawn during our study. However Lehmberg et al.'s study identified 14 major polypeptides only and

reported sketchy data on the post-translational modifications which occur during protein maturation. No protein phosphorylation was reported and fiber was not detected.

With all viral peaks characterized, the established RP-HPLC method proved to be extremely powerful for the monitoring and identification of protein modifications. The fingerprint obtained being characteristic of all the polypeptides contained in a viral particle, any change in the purity of the preparation or in the structure of any of these viral components could be detected by a single RP-HPLC run and characterized by further LC–ESI-MS experiments. Of particular interest to formulation studies, changes in the chromatographic fingerprint were systematically correlated with drops of viral infectivity. In contrast, stable formulations gave unchanged protein fingerprints (not shown). In addition, this method proved also very valuable to support changes in the manufacturing process. Some of the changes tested led to virus preparations which had different RP-HPLC profiles but which were otherwise indistinguishable by any other biological (titration) or analytical (SDS–PAGE, anion-exchange HPLC, size distribution analysis, and electron microscopy) methods (not shown).

Acknowledgements

We are indebted to Dr. P. Roingeard (Tours, France) for EM analyses, to J. Chroboczeck for mAb 2F7, and to Fabrice Bacchetta for gene sequence interpretations. We also thank M. Laborie, J.-M. Guillaume, and J.-F. Chaubard for adenovirus production, M. Pisano for his contribution to the preliminary characterization work, and M. Vuilhorgne, J.F. Mayaux, and E. Hochuli for their constant support and interest in this work. The development of RPR/INGN201 is a joined development project between Aventis Pharma and Introgen Therapeutics (Houston, TX, USA)

References

[1] Pat. Appl. WO 98/00524. F. Blanche and J.-M. Guillaume, 1996.
[2] B.G. Huygue, X. Liu, S. Sutjipto, B.J. Sugarman, M.T. Horn, H.M. Shepard, C.J. Scandella, P. Shabram, Hum. Gene Ther. 6 (1995) 1403.
[3] P.W. Shabram, D. D Giroux, A.M. Goudreau, R.J. Gregory, M.T. Horn, B.G. Huygue, X. Liu, M.H. Nunnally, B.J. Sugarman, S. Sutjipto, Hum. Gene Ther. 8 (1997) 453.
[4] F. Blanche, B. Cameron, A. Barbot, L. Ferrero, T. Guillemin, S. Guyot, S. Somarriba, D. Bisch, Gene Ther. 7 (2000) 1055.
[5] J. van Oostrum, R.M. Burnett, J. Virol. 56 (1985) 439.
[6] P.L. Stewart, R.M. Burnett, in: W. Doerfler, P. Böhm (Eds.), The Molecular Repertoire of Adenoviruses, Vol. 1, Springer-Verlag, Berlin, 1995, p. 25.
[7] P.L. Stewart, S.D. Fuller, R.M. Burnett, EMBO J. 12 (1993) 2589.
[8] C.W. Anderson, Virology 177 (1990) 259.
[9] J.C. D'Halluin, in: W. Doerfler, P. Böhm (Eds.), The Molecular Repertoire of Adenoviruses, Vol. 1, Springer-Verlag, Berlin, 1995, p. 47.
[10] M. Ishibashi, J.V. Maizel, Virology 58 (1974) 345.
[11] W.C. Russell, G.D. Kemp, in: W. Doerfler, P. Böhm (Eds.), The Molecular Repertoire of Adenoviruses, Vol. 1, Springer-Verlag, Berlin, 1995, p. 81.
[12] J. Chroboczeck, F. Bieder, B. Jacrot, Virology 186 (1992) 280.
[13] P. Fender, A.H. Kidd, R. Brebant, M. Öberg, E. Drouet, J. Chroboczeck, Virology 214 (1995) 110.
[14] E. Lehmberg, J.A. Traina, J.A. Chakel, R.-J. Chang, M. Parkman, M.T. McCaman, P.K. Murakami, V. Lahidji, J.W. Nelson, W.S. Hancock, E. Nestaas Jr., E. Pungor, J. Chromatogr. B 732 (1999) 411.
[15] G. Alms, P. Sanz, M. Carlson, T. Haystead, EMBO J. 18 (1999) 4157.
[16] B. Edvardsson, E. Everitt, H. Jörnvall, L. Prage, L. Philipson, J. Virol. 19 (1976) 533.
[17] J.M. Weber, in: W. Doerfler, P. Böhm (Eds.), The Molecular Repertoire of Adenoviruses, Vol. 1, Springer-Verlag, Berlin, 1995, p. 227.
[18] K. Hosokawa, M. Sung, J. Virol. 17 (1976) 924.
[19] P. Lemay, M.-L. Boudin, M. Milleville, P. Boulanger, Virology 101 (1980) 131.
[20] D. Piszkiewicz, M. Landon, E.L. Smith, Biochem. Biophys. Res. Commun. 40 (1970) 1173.
[21] M.-L. Caillet-Boudin, G. Strecker, J.-C. Michalski, Eur. J. Biochem. 184 (1989) 205.
[22] K.G. Mullis, R.S. Haltinwanger, G.W. Hart, R.B. Marchase, J.A. Engler, J. Virol. 64 (1990) 5317.