Journal of Chromatography B, 880 (2012) 19–26

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Extraction, purification and characterization of the plant-produced HPV16 subunit vaccine candidate E7 GGG

J.F. Buyel[a,1], J.A. Bautista[a], R. Fischer[b,c,*], V.M. Yusibov[a,2]

[a] Fraunhofer USA Center for Molecular Biotechnology, 9 Innovation Way, Suite 200, Newark, DE 19711, USA
[b] Institute for Molecular Biotechnology, Worringer Weg 1, RWTH Aachen University, 52074 Aachen, Germany
[c] Fraunhofer Institute for Molecular Biology and Applied Ecology, Forckenbeckstraße 6, 52074 Aachen, Germany

ARTICLE INFO

Article history:
Received 26 July 2011
Accepted 6 November 2011
Available online 16 November 2011

Keywords:
Plant-produced vaccine
Purification
Recombinant protein
Biopharmaceutical
HPV16
Transient expression

ABSTRACT

Several studies indicated that biopharmaceuticals based on the recombinant protein E7 of human papillomavirus (HPV) can serve as therapeutic vaccines preventing the development of cancer in women infected with high-risk types of HPV such as HPV16. Here, we report effective extraction and purification of a plant-produced E7GGG-lichenase fusion protein, an HPV16 subunit vaccine candidate, from *Nicotiana benthamiana* plants, to a high yield. The target contains the modified HPV16 E7 protein internally fused to the surface loop of a truncated, hexa-His- and KDEL-tagged variant of bacterial lichenase, and has been previously shown to possess anti-cancer activity in an animal model [18]. We purified the protein using a combination of immobilized metal-ion affinity chromatography and gel filtration. The achieved purity of the final product was 99% as confirmed by Coomassie or SYPRO Ruby staining after sodium dodecyl sulfate-polyacrylamide gel electrophoresis and by analytical size exclusion chromatography coupled with multi-angle laser light scattering. The overall yield was 50% corresponding to 0.1 g of protein per 1 kg plant biomass. Only slight changes in these parameters were observed during the process scale-up from 50 g to 1 kg of processed leaf biomass.

1. Introduction

Human papillomavirus (HPV) 16 is one of the principal causes of cervical cancer [1–5]. The development of HPV vaccines has been explored using a range of strategies, including DNA vaccines, viral and bacterial vectors, peptide-based vaccines, autologous dendritic cells (DCs) pulsed with a desired antigenic determinant, interfering RNA (RNAi), antisense RNA, ribozymes and combined techniques (such as prime-boost regimens), each with specific benefits and drawbacks [6–9]. Two protein-based vaccines against HPV have been developed, i.e. Cervarix by GSK and Gardasil by Merck [10–12]. These vaccines are based on the HPV capsid protein (L1) and only prevent the initial infection. They do not protect against cancer in women already carrying the virus.

The development of cervical cancer requires persistent HPV infection resulting from the integration of HPV DNA. In most cases, this is accompanied by disruption of the E2 open reading frame (ORF), which leads to disregulated expression of the viral genes E6 and E7. The E7 protein is highly conserved among different HPV types [1,3]. It interacts with a number of partners including pocket proteins such as pRb, cell cycle proteins such as E2F-Cyclin A complex, Cyclin E, p21 and p27, and transcription factors such as pCAF [13,14]. E7 reinitiates the cell cycle by binding to pRb, releasing E2F and thereby skipping the G_1-S checkpoint. This contributes to the accumulation of further mutations leading to so-called cervical intraepithelial neoplasia (CIN) and ultimately to cancer [1,2,15].

Because of its central role in malignant cell transformation, E7 is a promising therapeutic vaccine candidate. Initial studies have shown that E7-based vaccines can elicit protective immune responses in mice [6–8,16–19]. Furthermore, vaccination of animals already challenged with the virus significantly reduces the likelihood of cancer [18]. The vaccine candidate used in the above

Abbreviations: CD, circular dichroism; CIN, cervical intraepithelial neoplasia; CV, column volume; DCs, dendritic cells; DTT, dithiothreitol; ESI, electro spray ionization; HIC, hydrophobic interaction chromatography; HPV, human papillomavirus; IEC, ion exchange chromatography; IMAC, immobilized metal affinity chromatography; LB, lysogeny broth; Lic, lichenase; LOQ, limit of quantification; MALLS, multi-angle laser light scattering; MMC, multi-modal chromatography; MRME, mean residue molar ellipticity; MS, mass spectrometry; MWCO, molecular weight cut off; ORF, open reading frame; PAA, polyacrylamide; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate buffered saline; POI, protein of interest; RT, room temperature; SDS, sodium dodecyl sulfate; SEC, size exclusion chromatography.

* Corresponding author at: Institute for Molecular Biotechnology, Worringer Weg 1, RWTH Aachen University, 52074 Aachen, Germany. Tel.: +49 241 6085 11020; fax: +49 241 6085 10000.
E-mail addresses: johannes.buyel@rwth-aachen.de (J.F. Buyel), jab95005@yahoo.com (J.A. Bautista), fischer@molbiotech.rwth-aachen.de (R. Fischer), vyusibov@fraunhofer-cmb.org (V.M. Yusibov).
[1] Permanent address: Institute for Molecular Biotechnology, Worringer Weg 1, RWTH Aachen University, 52074 Aachen, Germany. Tel.: +49 241 6085 13162.
[2] Tel.: +1 302 369 8955.

1570-0232/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2011.11.010

studies is a fusion protein comprising a modified E7 protein internally fused in-frame into the surface loop of a truncated bacterial lichenase (LicKM-E7GGG) [18]. The latter fusion part contained hexa-His and KDEL tags at the C-terminus. In this report, we describe the procedures used to extract and purify this recombinant antigen from *Nicotiana benthamiana* plants, including immobilized metal affinity chromatography (IMAC) and gel filtration. For the 50 g scale protocol, a 50% recovery yield was achieved based on the soluble protein of interest after extraction.

2. Materials and methods

2.1. Expression vector

The plant expression vector that was used to express the mutated HPV16 E7 protein of interest (POI) as a fusion to the surface loop of circularly permuted modified β-1,3-1,4-glucanase (lichenase) from *Clostridium thermocellum* (LicKM) was pBID4-LicKM-E7GGG [18], a derivative of pBID4 [20].

The LicKM modifications included adding an N-terminal signal sequence derived from the PR1a protein of *Nicotiana tabacum* as well as a hexa-His tag followed by the endoplasmic reticulum (ER) retrieval signal KDEL at the C-terminus. The vector was introduced into *A. tumefaciens* strain GV 3101 by electroporation [21]. Cells were regenerated in lysogeny broth (LB) medium and positively transformed clones selected using LB agar plates containing 0.1 μg/mL kanamycin.

2.2. Plant growth, infiltration, harvest and storage

N. benthamiana seedlings were grown for 12 weeks at 25 °C, 70% humidity with a 12 h light period per day. For infiltration with *A. tumefaciens*, plants were submerged upside down in bacterial solution with OD_{600} of 0.5–0.6 for 1.5 min under a vacuum of 5 kPa. After a brief rinse in water, plants were maintained for 7 days at 25 °C with a 12 h light period per day and non-regulated humidity before harvest. Whole plants were cut, frozen on dry ice, crushed to homogeneity and stored at −80 °C until use.

2.3. Extraction

For 50 g scale: 50 g of frozen plant tissue was weighed and transferred to a cooled Oster Fusion blender, and 150 mL (3, v/w) of the phosphate-based extraction buffer was added. The suspension was ground in four steps, each 45 s, at the highest speed setting, interspersed with 1 min breaks. *For 1 kg scale*: 1 kg of frozen plant tissue was weighed and transferred into a cooled beaker, 3 (v/w) of extraction buffer was added, and the suspension was processed twice through a Comitrol homogenizer equipped with a 220 blade ring (Urschel). After the second passage, 0.25 (v/w) of the phosphate-based chaser buffer was used to rinse the Comitrol. This rinse fraction was collected together with the rest of the homogenized plant material suspension.

2.4. Clarification

For 50 g scale: The homogenate was split into 30 mL centrifugation tubes and large particulate matter was spun down in an Avanti J-26 XPI centrifuge with a JA 25.50 rotor for 30 min at 75,600 × *g* and 10 °C. The supernatant was then filtered through Miracloth (Calbiochem). *For 1 kg scale*: Particulate matter was removed from the homogenate using a semi-continuous disc-stack centrifuge (Clarifier FSD 1-06-107, GEA Westfalia) at 12,000 × *g* and regulated back-pressure between 0.10 and 0.15 MPa (15–20 psi) at 4 °C. The last 1 L of an extract was flushed out from the centrifuge bell with 1.75 L of the chaser buffer. The supernatant was applied to a CUNO BioLife PB capsule and a Sartopore 2 filter followed by rinsing with 200 mL of the chaser buffer. For both scales, the pH of the cleared extract was adjusted to 7.5 if necessary.

Table 1
Parameters of IMAC capture for different column sizes.

Parameter	Scale	
	50 g	1 kg
Resin material	Ni Sepharose 6 Fast Flow[a]	Chelating Sepharose Big Beads
Column volume [mL]	10	146
Bed height [cm]	5	27.5
Column diameter [cm]	1.6	2.6
Volumetric loading flow rate [mL/min]	3	20
Linear loading flow rate [cm/h]	90	226
Volumetric elution flow rate [mL/min]	1	10
Linear elution flow rate [cm/h]	30	113

[a] Pre-packed resin.

Table 2
Parameters of SEC polishing step.

Parameter	Scale	
	50 g	1 kg
Resin	HiLoad Superdex 200	
Column	XK 16/100	XK 50/100
Column volume [mL]	177	1728
Loaded sample volume [mL]	1.0	10.0
Sample volume/column volume ratio [%]	0.56	0.58
Protein concentration in sample [mg/mL]	11	21
Bed height [cm]	88	88
Column diameter [cm]	1.6	5
Volumetric flow rate [mL/min]	1	10
Linear flow rate [cm/h]	30	30

2.5. Chromatography

An ÄKTA purifier 100 was used in all chromatography steps. *Capture*: For both 50 g and 1 kg scales, IMAC with Ni^{2+} as an ion and Chelating Sepharose Big Beads (GE Healthcare) as a resin was used to capture the hexa-His-tagged target proteins. Columns were equilibrated with 5 column volumes (CV) of the loading buffer (50 mM sodium phosphate, pH 7.5; 250 mM sodium chloride; 50 mM imidazole; 20%, v/v, glycerol), and clarified extracts were loaded under conditions given in Table 1. Columns were washed with 10 CV of the loading buffer followed by a wash with 9.3% of the elution buffer (as Ni Buffer A but with 500 mM imidazole) for 5 CV. *Elution*: Elution of the target protein was achieved using 100% elution buffer. Then eluates were concentrated 15- or 30-fold using AMICON 15 mL or Centricon Plus-70 concentrators (Millipore) with 10,000 molecular weight cut off (MWCO) for 50 g or 1 kg scales, respectively. *Polishing*: Size exclusion chromatography (SEC) using HiLoad Superdex 200 resin (GE Healthcare) was used in different column formats for the two scales as a polishing step, conditions are given in Table 2. Running buffer for SEC was 50 mM sodium phosphate, pH 7.5 with 250 mM sodium chloride, 10 mM dithiothreitol (DTT) and 0.5% (v/v) EMPIGEN-BB. *Conditioning*: SEC eluates from 1 kg scale were dialyzed against phosphate buffered saline (PBS) and concentrated 4-fold using AMICON 15 mL or Centricon Plus-70 concentrators (Millipore) with 10,000 MWCO prior to analysis with SEC-multi-angle laser light scattering (MALLS) or circular dichroism (CD).

2.6. SDS-PAGE, Coomassie/SYPRO Ruby staining and Western Blotting

Samples were diluted 1:5 in 5× sample buffer (250 mM Tris–HCl, pH 6.8; 10% (w/v) sodium dodecyl sulfate (SDS); 0.5% (v/v) bromophenol blue; 50% (v/v) glycerol; 500 mM DTT), boiled for 5 min, mixed and spun down. Concentrated samples were diluted with 1× sample buffer to match the linear range of the according assay. Samples were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using discontinuous polyacrylamide (PAA) gels consisting of a 4% stacking gel and a 10% separation gel [22]. Gels were either stained with GelCode Blue (Pierce) at room temperature (RT) as described elsewhere [23] and analyzed using the GeneTools software (Syngene Bio Imaging), or stained with 50 mL of SYPRO Ruby gel staining solution [24,25]. In the latter case, fluorescence was detected using an UV-transilluminator through a 590 nm filter and the same software was used for analysis. For immunodetection, proteins were transferred from gels onto a PVDF membrane at a constant voltage of 100 V for 1 h at 4 °C [26] and detected using either mouse anti-tetra-His or rabbit anti-LicKM as a primary antibody and peroxidase-labeled goat anti-mouse F_c or goat anti-rabbit F_c, respectively as a secondary antibody. The luminescence signal was developed using SuperSignal West Pico Stable Peroxidase Solution and SuperSignal West Pico Luminol Enhancer Solution.

2.7. SEC-MALLS

PBS with 0.5% EMPIGEN-BB was used as the running buffer and samples with a concentration of 1 mg/mL and 2 mg/mL were analyzed using a GL 10/300 Superdex 200 analytical column from GE healthcare and a subsequent Astra MALLS detector connected to an ÄKTA purifier 10. Collected UV and MALLS data were analyzed using the ASTRA 5.3.4.14 software.

2.8. Circular dichroism

A J-810 spectropolarimeter (Jasco) was used for all measurements. For sample concentrations of 0.105, 0.210 and 0.420 mg/mL in PBS, CD spectra were measured from 280 to 195 nm in a 1 mm cuvette at 4 °C in triplicate, each consisting of 4-fold measurement of each replicate. After blank subtraction, measured values were transformed from ellipticity to mean residue molar ellipticity (MRME) using Eq. (1) and plotted against the measured wavelength. For a thermal melt, a sample's CD was measured at 218 nm in a 10 mm path cuvette at temperatures from 5 to 95 °C, with every 5 °C increment at a rate of 1.5 °C/min. Furthermore, CD spectra from 280 nm to 195 nm were measured at temperatures of 40, 55, 60, 65, 70, 75 and 80 °C as well as after 10 min boiling at 90 °C. Data were analyzed using different deconvolution software packages: k2d, k2d2, SOMCD, CDNN 2.1, ContinLL, SELCON3 and CDSSTR.

$$\theta_M = \frac{\theta M_r}{\gamma l N_{aa} 10} \quad (1)$$

where θ_M is the mean residual molar ellipticity (MRME) in [deg cm^2 dmol^{-1}], θ the ellipticity in [deg], M_r the molar mass [g mol^{-1}], γ the mass concentration in [mg/mL], l the cuvette path length in [cm] and N_{aa} the number of amino acid residues of the POI as well as the transformation factor 10 to suit the desired unit.

2.9. Model building

The structure of circularly permuted lichenase from *Paenibacillus macerans* (UniProtKB entry P23904 GUB_PAEMA; pbd code 1cpn) was used as a scaffold to model LicKM and LicKM-E7GGG, circular permutations of the *licB* gene from *C. thermocellum* strain ATCC 27405/DSM 1237 (UniProtKB entry A3DBX3 GUB_CLOTH). The modeling was performed either at Fraunhofer USA Center for Molecular Biotechnology using the Swiss-PDB Viewer 4.0.1 by threading LicKM and LicKM-E7GGG through the structure of LicB or by an external service [27].

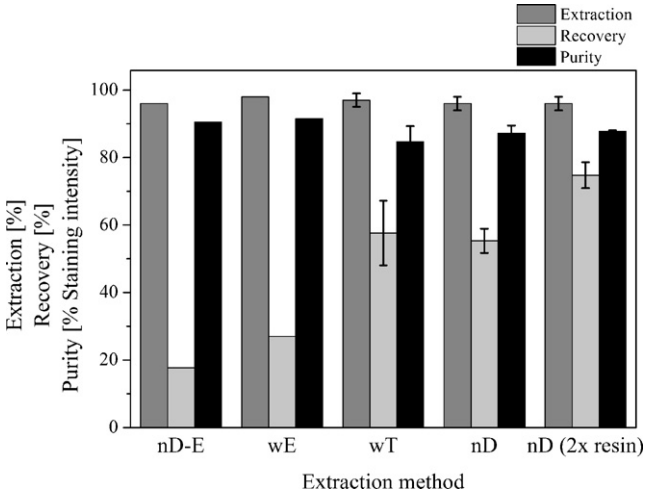

Fig. 1. Extraction, recovery and purity levels of LicKM-E7GGG during extraction and IMAC capture for different buffers and IMAC column volumes. nD-E, no detergent during extraction, 0.5% EMPIGEN-BB during IMAC; wE, 0.5% EMPIGEN-BB during extraction and IMAC; wT. 0.5% Triton X 100 during extraction and IMAC; nD, no detergent; and nD (2× resin), 10 mL resin instead of 5 mL. Error bars denote the standard deviation.

2.10. N-terminal sequencing

N-terminal sequencing of the final product was performed by Proteos, Kalamazoo, MI. After separation by SDS-PAGE, the 38 kDa band was isolated and treated with pyroglutamate aminopeptidase to remove the first N-terminal amino acid which was identified as pyroglutamate. Automated protein sequencing was then performed using a Perkin Elmer Applied Biosystems Procise 494 sequencer equipped with an on-line Perkin Elmer Applied Biosystems Model 140C PTH Amino Acid Analyzer. Sequencing was achieved by Edman degradation [28], which detects the N-terminal residue previously coupled to phenylthiocarbamyl at each degradation cycle. The model 610A version 2.1 software was employed for data acquisition and processing.

2.11. Mass spectrometry

Electro spray ionization (ESI) mass spectrometry (MS) analysis of the final product was performed by M-scan (West Chester, PA) [29].

3. Results and discussion

3.1. Extraction

The average expression level of LicKM-E7GGG was 233 ± 5 mg (n = 3) per 1 kg of plant biomass. This expression level is in the range of what has been commonly reported for recombinant protein expression in *N. benthamiana*: 100–600 mg/kg [20,30–33]. From the expressed 233 mg/kg biomass, 225 ± 14 mg (n = 3) were found in the final extract, which indicates an average extraction efficiency of almost 97%. Given the accuracy of the quantitation method [34], we assumed this to be a complete extraction. We found that the extraction efficiency of the target protein was not affected by the use of either 0.5% (v/v) Triton X-100 or EMPIGEN-BB (Fig. 1). The main

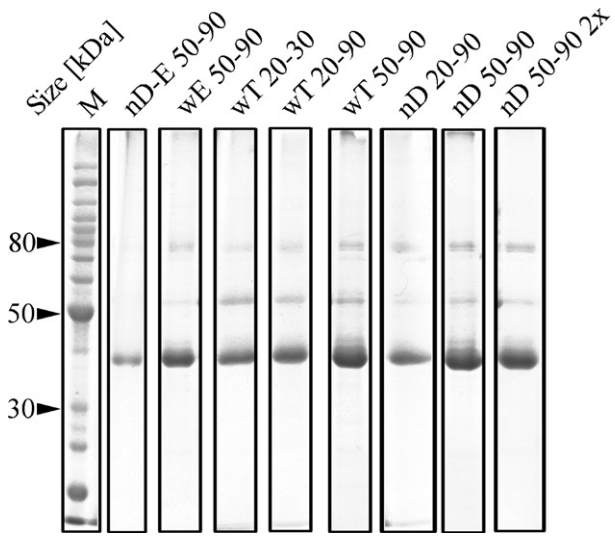

Fig. 2. Impact of varying imidazole concentrations, detergents and sample/CV ratio on purity of IMAC elution fractions containing LicKM-E7GGG. nD-E, no detergent during extraction, 0.5% EMPIGEN-BB during IMAC; wE, 0.5% EMPIGEN-BB during extraction and IMAC; wT, 0.5% Triton X 100 during extraction and IMAC; and nD, no detergent; dash-separated numbers indicate imidazole concentration in mM during sample loading (first number) and wash step (second number); 2×, 10 mL resin instead of 5 mL were used.

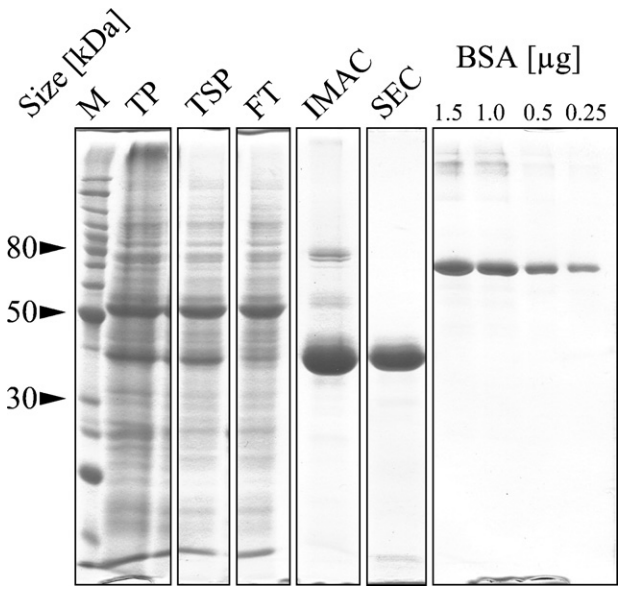

Fig. 3. Coomassie-stained process samples separated using 4–10% SDS-PAGE. M, pre-stained protein ladder; TP, total protein from non-clarified homogenate; TSP, total soluble protein (after solids removal); FT, IMAC capture flow-through fraction; IMAC, IMAC elution fraction; SEC, SEC elution fraction; and BSA, bovine serum albumin quantification standard. *Note*: lanes M through IMAC and the lane BSA are from the same gel, whereas the lane SEC is from a different gel. For clarity, the lanes were compiled into a single figure.

effect of the detergents, especially EMPIGEN-BB, during extraction was solubilization of host proteins notably, of a 25 kDa protein, which is likely to be the tonoplast intrinsic protein (TIP) according to the size (data not shown) [35].

3.2. Clarification

Any losses of target protein that might have occurred during the centrifugation/filtration steps were below the limit of quantification (LOQ; here ~50 ng/band) in the colorimetric quantification assays (compare the 37 kDa POI bands of the TP and TSP fractions in Fig. 3). Thus, we considered these losses to be negligible for the given process.

3.3. Capture

We defined a target recovery of >60% and a target purity of >70% as the objectives of the capture step. For varying imidazole concentrations in the loading buffer and a wash step ranging from 20 mM to 50 mM and from 20 mM to 90 mM, respectively, the recovery remained at 59 ± 9% (n = 11) (Fig. 1). However, using 50 mM instead of 20 mM imidazole in the loading buffer in combination with a 90 mM imidazole wash step increased the purity achieved by the capture step from 76 ± 3% (n = 3) to about 87 ± 2% (n = 8) (Fig. 2). By decreasing the loaded sample volume/CV ratio from 40 to 20 (increasing the CV two-fold) we were able to increase the recovery from 59% to 75 ± 4% (n = 3). The detergents tested during extraction were found to have no effect on the purity after the IMAC capture. Although they had no effect on the extraction efficiency, they had none or a negative impact on the recovery of the LicKM-E7GGG protein after IMAC, also when added after extraction (Fig. 1). In case 0.5% (v/v) Triton X-100 was added no significant effect on recovery was observed. Using 0.5% (v/v) EMPIGEN-BB during extraction or after clarification, the recovery dropped to 46% or 31% of the detergent-free setup, respectively. Thus, the use of detergents was rejected for this process.

Others have reported that processing of plant cell extracts can lead to the accumulation of polyphenols on the capture resin which causes non-specific protein binding in the initial chromatographic step [36–38]. We assume that in the presence of the detergents more proteins (and secondary metabolites such as polyphenols) are solubilized, as observed for the 25 kDa protein, leading to a higher protein concentration and greater non-specific binding to the IMAC resin, reducing the effective column capacity. In our study, the use of the IMAC capture resulted in 87% purity of the LicKM-E7GGG protein with 75% recovery, corresponding to 169 ± 9 mg (n = 3) of target protein per 1 kg of plant biomass. Two major contaminants that remained included the protein with an apparent size of about 53 kDa which may correspond to ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) large subunit and the protein of unknown identity with an apparent size of 80 kDa (Figs. 2 and 3).

3.4. Polishing

For the polishing step, our objectives were defined as the recovery of >50% of the IMAC eluate and >97% target purity. Despite extensive screening, we did not find conditions for ion exchange (IEX) chromatography, hydrophobic interaction chromatography (HIC) or multi-modal chromatography (MMC) that allowed for efficient separation of LicKM-E7GGG and the two abovementioned impurities. The purities we achieved never exceeded 97%. For SEC, we tested a series of additives alone or in combination, including the uncharged detergent Triton X-100, the anionic detergents sodium dodecyl sulfate and deoxycholate, the carbohydrates sucrose and cellobiose, as well as EDTA and guanidine HCl (Table 3). None of the tested additives, alone or in combination, did increase the purity above 97% and exhibited a yield >50% at the same time. Instead, under slightly reducing conditions with 10 mM DTT and in the presence of 0.5% (v/v) of the zwitterionic detergent EMPIGEN-BB, we were able to purify the target protein on a Superdex 200 column with 88 cm bed height to a purity of more than 99% (Fig. 4). The yield of this polishing step was 67%, leading to an overall process yield of about 50% corresponding to 118 mg of LicKM-E7GGG per 1 kg of plant biomass. Adding 20% (v/v) glycerol to the SEC running buffer for protein stabilization purposes had no effect on the purity but reduced the yield to only 57 ± 2% (n = 2). When DTT

Table 3
Buffer variants tested for SEC polishing with Superdex 200 resin. Variants 8 through 13 were used for optimization.

Variant	Buffer	Glycerol [%v/v]	Detergent [%v/v]	DTT [mM]	Other additives/modifications	Bed height [cm]	Purity [%]	Recovery [%]
1	PBS	0	0.1% TX-100	0	None	88	92	36
2	PBS	0	0.1% TX-100	5	None	30	88	18
3	PBS	0	0.16% DOC	5	None	30	96	41
4	SB	0	0.5% TX-100	100	500 mM sucrose	30	92	66
5	SB	20	None	10	125 mM cellobiose	30	92	21
6	SB	20	0.5% EMPIGEN-BB; 0.1% SDS	20	50 mM guanidine HCl; 10 mM EDTA	30	93	85
7	SB	20	0.5% EMPIGEN-BB	10	None	30	95	46
8	SB	20	0.5% EMPIGEN-BB	10	None	88	98	56
9	SB	0	0.5% EMPIGEN-BB	10	None	88	98	67
10	SB	20	0.5% EMPIGEN-BB	0	None	88	96	39
11	SB	0	None	10	None	88	95	35
12	SB	20	0.5% EMPIGEN-BB	10	2-Fold sample concentration	88	98	48
13[a]	SB	0	0.5% EMPIGEN-BB	10	2-Fold sample concentration	88	99	51

DTT, dithiothreitol; PB, phosphate buffer; TX-100, Triton X-100; and DOC, deoxycholate
[a] This setup was selected for scale-up.

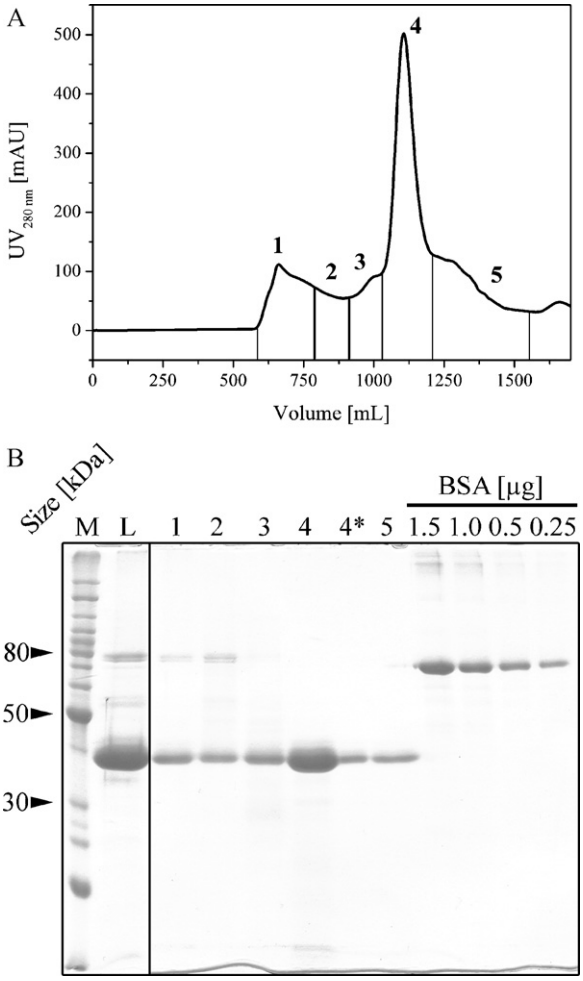

Fig. 4. Preparative SEC polishing step for the isolation of highly pure target protein. (A) UV_{280nm} elution profile of LicKM-E7GGG on Superdex 200, see Table 2 for column specifications. Numbers 1 through 5 indicate the collected fractions. (B) Coomassie-stained fractions from preparative SEC after separation by SDS-PAGE. The POI was the dominating protein in all fractions but purest in fraction 4. The 80 kDa impurity eluted in fractions 1 and 2, corresponding to high molecular masses. Fraction numbers match to those in (A); 4*, 1:10 dilution of fraction 4; M, pre-stained protein ladder; and L, load; this corresponds to IMAC eluate. A BSA dilution series was used for initial quantification.

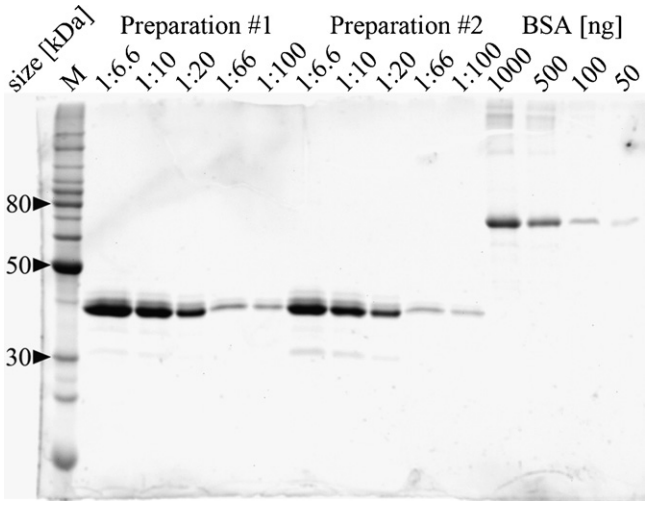

Fig. 5. Dilutions of LicKM-E7GGG from the final bulk of 1 kg preparations #1 + 2 after separation on a discontinuous, SYPRO Ruby-stained 4–10% SDS-PAA gel. M, protein ladder; BSA, bovine serum albumin standard.

was removed from the running buffer in the presence of glycerol, the yield dropped further to 38% and the purity was 96%. When EMPIGEN-BB and glycerol were removed, the yield was only 35% and the purity was 95%. Hence, all components of the running buffer described under Section 2.5 were considered necessary for the described purification step.

3.5. Scale-up

During scale-up of the IMAC capture step, the achieved purity remained the same, but the yield dropped from 75 to 60 ± 5% (n = 2). This might have been caused by changes in the process parameters during scale-up (Table 1). These changes were introduced to balance parameter constancy and process time. In the future, an increase in the IMAC CV could potentially prevent the losses of POI and compensate for the parameter changes. For SEC, we faced similar problems. Even though we were able to keep most process parameters constant (Table 2), the protein concentration had to be increased 4-fold to keep the loaded sample volume/CV ratio constant. The higher protein concentration did not affect the efficiency of chromatography; however, it further reduced reproducibility and the yield during sample handling. Therefore, the yield was only 50 ± 12% (n = 2) instead of 67%, corresponding to about 85 mg of LicKM-E7GGG per 1 kg of plant biomass. The purity achieved with SEC was not affected by the scale-up. In summary, an overall process yield of 38% was achieved after scale-up, with a product purity of more than 99%. We assume that with optimized sample handling and column geometry an overall process yield of 50% can easily be restored for the 1 kg scale.

Table 4
Predicted and observed results for N-terminal sequencing of the POI.

Residue number	Expected amino acid	Observed amino acid
1	Glutamine (Gln, Q)	Pyroglutamate[a]
2	Asparagine (Asn, N)	Asparagine (Asn, N)
3	Glycine (Gly, G)	Glycine (Gly, G)
4	Glycine (Gly, G)	Glycine (Gly, G)
5	Serine (Ser, S)	Serine (Ser, S)

[a] This amino acid residue had been enzymatically removed prior to Edman degradation.

Table 5
Potential modifications responsible for the observed mass shifts of the POI during MS.

Observed masses [Da]	Δm to predicted mass [Da]	Possible modification
37,438	−487	Lost KDEL tag
37,949	23	Incorporated sodium
38,061	136	As before + all seven cysteines oxidized
38,083	157	As before + sodium
38,158	233	As before + glycerol ester
38,360	435	Dioctyl phthalate sodium adduct + sodium

3.6. Product analysis

We confirmed the product purity of more than 99% previously determined by colorimetric analysis of Coomassie-stained SDS-PAA gels with a more sensitive SYPRO Ruby staining (Fig. 5) [24,25,39]. The target protein identity was assured by immunostaining using either lichenase-or tetra-His tag-specific antibodies (Fig. 6). The integrity of the LicKM-E7GGG protein within the bulk formulation was confirmed by the N-terminal sequencing and MS analysis of the single protein band of an approximate size of 38 kDa isolated from SDS-PAA gels. When the isolated protein was pre-treated with pyroglutamate aminopeptidase, the amino acid sequence listed in Table 4 was observed. Since a N-terminal glutamine can spontaneously transform into pyroglutamate [40], the observed N-terminal sequence was considered to be as expected, providing strong evidence that the 38 kDa protein was indeed LicKM-E7GGG. According to the nucleotide sequence of the gene the predicted mass of the LicKM-E7GGG protein after cleavage of the 35-residue signal peptide was 37,942 Da. However, with respect to the pyroglutamate formation found in the N-terminal sequencing a mass of 37,925 Da was expected. Observed masses were conclusively matched to the expected protein modifications typically occurring during protein MS, as listed in Table 5, confirming the intact state of the target protein after purification [41–44].

A combination of SEC and MALLS was used to investigate the oligomeric state of the target protein in PBS with 0.5% EMPIGEN-BB.

Table 6
Observed UV peaks during SEC-MALLS analysis of samples from the final bulk product and calculated molar masses of the according species.

Buffer	Sample concentration [mg/mL]	Peak	Retention volume [mL]	Maximum peak height [mAU]	Average molar mass [kDa]	Likely oligomeric state
PBS with 0.5% EMPIGEN-BB	1.0	1	14.6	195	72 ± 4	Dimer
	2.0	1	14.6	406	72 ± 3	Dimer

Table 7
Comparison of the POI's secondary structure composition calculated by different methods. k2d through CDSSTR were freely distributable CD deconvolution software packages. k2d through CDNN 2.1 use neural network algorithms.

Fitting method	1 kg preparation	Data range [nm]	Ratio of structural feature [%][a]						Sum helix [%]	Sum sheet [%]	Sum random/ unassigned [%]	Total [%][b]
			α-Helix		β-Sheet		β-Turn	Random coil				
			r	d	r	d						
k2d	#1	200–240	4		48		–	48	4	48	48	100
	#2		5		48		–	48	5	48	48	101
k2d2	#1		4		44		–	–	4	44	52	100
	#2		4		44		–	–	4	44	52	100
SOMCD	#1		17		39		10	34	17	39	43	100
	#2		17		39		10	34	17	39	43	99
CDNN 2.1[c]	#1	210–260	15		17	16	22	50	15	33	72	119
	#2		15		17	16	21	50	15	32	71	118
ContinLL	#1	200–280	2	5	28	12	22	32	7	40	54	100
	#2		1	3	19	10	21	31	4	28	51	84
SELCON3	#1		0	2	19	9	21	34	2	28	55	85
	#2		1	3	19	10	21	31	4	28	51	84
CDSSTR	#1		1	4	26	13	23	32	5	39	56	99
	#2		1	4	27	13	25	30	6	39	55	99
Average[d]	#1	n/a	8 ± 6		40 ± 6		22 ± 6	38 ± 8	8 ± 6	40 ± 7	60 ± 9	108 ± 10
	#2		8 ± 6		39 ± 7		22 ± 6	37 ± 9	8 ± 6	39 ± 8	59 ± 9	105 ± 12
Calculated from E7 and LicB diffraction data	n/a	n/a	14		40		48		14	40	48	102
Calculated from modeled POI[e]	n/a	n/a	11		40		49		11	40	49	100

[a] : r, regular; d, distorted.
[b] Values larger than 100% arise from algorithms that do not force the sum of all secondary structures to equal unity.
[c] In this case r is anti-parallel sheet and d is parallel sheet.
[d] Value behind plus-minus indicates standard deviation.
[e] For details of the modeling see Section 2.9.

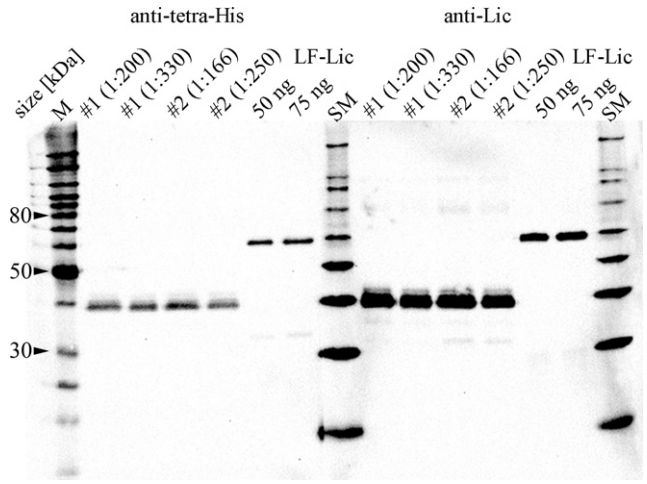

Fig. 6. Western blot of LicKM-E7GGG from the final bulk of 1 kg preparations #1 + 2 after separation on discontinuous SDS-PAGE. For the His tag detection, mouse anti-tetra-His (primary) and goat anti-mouse F_c (secondary) antibodies were used. For lichenase detection, rabbit anti-LicKM (primary) and goat anti-rabbit F_c (secondary) antibodies were used. LF-Lic, a fusion of lethal factor protein from *Bacillus anthracis* and lichenase, was used as a quantification standard. M, protein ladder; SM, 1:1 mixture of SeeBlue and MagicMark protein ladder.

We observed single peak UV chromatograms for injections of 1 mg/mL and 2 mg/mL of LicKM-E7GGG (Table 6). For these peaks, we calculated the molecular weight and hence the oligomeric state based on the measured UV signals and Rayleigh ratios. According to these calculations, the target protein was present as a dimer in PBS with 0.5% EMPIGEN-BB. This observation is in a good agreement with the findings of other groups reporting dimerization of the E7 C-terminal domains [45–47]. To gain initial insights into the structural integrity and functionality of the target protein, we compared the secondary structure elements of a model based on diffraction data of homologue proteins with recorded CD spectra of protein samples after purification. The obtained CD data were consistent for all analyzed samples and matched well with model-based predictions (Table 7), indicating that the LicKM-E7GGG protein was correctly folded. Slight deviations between the calculation and the model were found: the model overestimated helices and underestimated the number of random structures. We attribute this observation to the origin of the diffraction data where the proteins are present in a rigid, crystallized form, suggesting a more ordered state of the secondary structure compared to the measurement conditions of CD where the molecules are soluble. In general, calculation of the ratios of the different secondary structure elements using several programs as suggested in current literature led to reasonable estimates [48–51]. For information about different deconvolution software packages used in this study, see the abovementioned literature.

4. Conclusions

In the current study, we describe a fast and efficient two-step process for extraction and purification of the plant-produced HPV16 subunit vaccine candidate LicKM-E7GGG. After homogenization and extraction, the target protein was purified using an IMAC capture step based on the hexa-His tag binding and then polished using SEC. Thereby, the target protein was purified to more than 99% with an overall process yield of 50% (118 mg/kg). The process was scaled up from a 50 g to a 1 kg batch size. Despite of the limited availability of large-scale equipment, the average overall yield was still around 40% (85 mg/kg) and purity remained at 99%. The purified POI was characterized using different techniques including N-terminal sequencing, MS, CD and SEC-MALLS. The results were used to build a reference data set for subsequent production batches and confirmed the identity and integrity of the LicKM-E7GGG protein after purification. The results also provided a good indication that the protein was present in a folded state and will thus contain linear and conformational epitopes. The purification process described here will facilitate the reproducible production of highly pure LicKM-E7GGG needed for forthcoming animal studies. With proper evaluation of endotoxin, host cell protein and residual DNA levels, it can also be used to provide material for phase I clinical trials.

Acknowledgements

The authors are grateful to Drs. Mark Jones and Chad Blamey for sharing their expertise and Robert Eleuterio for growing the plants used in this study. We wish to thank Drs. Vadim Mett and Birgit Orthen for their scientific comments on the manuscript. We also would like to thank Drs. Richard Twyman and Natasha Kushnir for their assistance with editing the manuscript.

References

[1] M.A. Stanley, M.R. Pett, N. Coleman, Biochem. Soc. Trans. 35 (2007) 1456.
[2] C.M. Hebner, L.A. Laimins, Rev. Med. Virol. 16 (2006) 83.
[3] S. Motoyama, C.A. Ladines-Llave, S. Luis Villanueva, T. Maruo, Kobe J. Med. Sci. 50 (2004) 9.
[4] A. Gonzalez Martin, Clin. Transl. Oncol. 9 (2007) 347.
[5] T. Hoory, A. Monie, P. Gravitt, T.C. Wu, J. Formos. Med. Assoc. 107 (2008) 198.
[6] V.A. Govan, Ann. N.Y. Acad. Sci. 1056 (2005) 328.
[7] A. Mahdavi, B.J. Monk, Oncologist 10 (2005) 528.
[8] A. Mahdavi, B.J. Monk, Curr. Oncol. Rep. 8 (2006) 465.
[9] M. Smahel, P. Sima, V. Ludvikova, V. Vonka, Virology 281 (2001) 231.
[10] R. Kirnbauer, F. Booy, N. Cheng, D.R. Lowy, J.T. Schiller, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 12180.
[11] European Medicines Agency, Gardasil: EPAR -Summary for the public, (2010) EMA/429427/2010.
[12] European Medicines Agency, Cervarix: EPAR -Summary for the public, (2010) EMA/423507/2010.
[13] I.J. Yu, D.L. Spector, Y.S. Bae, D.R. Marshak, J. Cell Biol. 114 (1991) 1217.
[14] P. Massimi, L. Banks, Virology 276 (2000) 388.
[15] M. Stanley, Gynecol. Oncol. 109 (2008) S15.
[16] A. Venuti, S. Massa, V. Mett, L.D. Vedova, F. Paolini, R. Franconi, V. Yusibov, Vaccine 27 (2009) 3395.
[17] P. Di Bonito, F. Grasso, G. Mangino, S. Massa, E. Illiano, R. Franconi, E. Fanales-Belasio, M. Falchi, E. Affabris, C. Giorgi, Int. J. Immunopathol. Pharmacol. 22 (2009) 967.
[18] S. Massa, R. Franconi, R. Brandi, A. Muller, V. Mett, V. Yusibov, A. Venuti, Vaccine 25 (2007) 3018.
[19] R. Franconi, S. Massa, E. Illiano, A. Mullar, A. Cirilli, L. Accardi, P. Di Bonito, C. Giorgi, A. Venuti, Int. J. Immunopathol. Pharmacol. 19 (2006) 187.
[20] K. Musiychuk, N. Stephenson, H. Bi, C.E. Farrance, G. Orozovic, M. Brodelius, P. Brodelius, A. Horsey, N. Ugulava, A.M. Shamloul, V. Mett, S. Rabindran, S.J. Streatfield, V. Yusibov, Influenza Other Respi. Viruses 1 (2007) 19.
[21] J.J. Lin, Methods Mol. Biol. 47 (1995) 171.
[22] B.J. Davis, Ann. N.Y. Acad. Sci. 121 (1964) 404.
[23] S. Fazekas de St Groth, R.G. Webster, A. Datyner, Biochim. Biophys. Acta 71 (1963) 377.
[24] K. Berggren, E. Chernokalskaya, T.H. Steinberg, C. Kemper, M.F. Lopez, Z. Diwu, R.P. Haugland, W.F. Patton, Electrophoresis 21 (2000) 2509.
[25] M.F. Lopez, K. Berggren, E. Chernokalskaya, A. Lazarev, M. Robinson, W.F. Patton, Electrophoresis 21 (2000) 3673.
[26] W.N. Burnette, Anal. Biochem. 112 (1981) 195.
[27] P.A. Bates, L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM, Proteins: Struct., Funct. Genet., Suppl. 5 (2001) 39–46.
[28] P. Edman, Arch. Biochem. 22 (1949) 475.
[29] M. Wilm, Principles of electrospray ionization, Mol. Cell. Proteomics (2011) 10, M111 009407.
[30] D.A. Goldstein, J.A. Thomas, QJM 97 (2004) 705.
[31] T.O. Kohl, I.I. Hitzeroth, N.D. Christensen, E.P. Rybicki, BMC Biotechnol. 7 (2007) 56.
[32] P.M. Doran, Trends Biotechnol. 24 (2006) 426.
[33] M.E. Nichols, T. Stanislaus, E. Keshavarz-Moore, H.A. Young, J. Biotechnol. 92 (2002) 229.
[34] W. Weiss, F. Weiland, A. Gorg, Methods Mol. Biol. 564 (2009) 59.
[35] C. Maurel, J. Reizer, J.I. Schroeder, M.J. Chrispeels, EMBO J. 12 (1993) 2241.
[36] R. Fischer, E. Stoger, S. Schillberg, P. Christou, R.M. Twyman, Curr. Opin. Plant Biol. 7 (2004) 152.
[37] T.J. Menkhaus, Y. Bai, C. Zhang, Z.L. Nikolov, C.E. Glatz, Biotechnol. Prog. 20 (2004) 1001.

[38] W. Wang, F. Tai, S. Chen, J. Sep. Sci. 31 (2008) 2032.
[39] N. Yamada, S. Ozawa, N. Kageyama, H. Miyano, J. Agric. Food Chem. 52 (2004) 5329.
[40] L.E. Hanford, J.J. Enghild, Z. Valnickova, S.V. Petersen, L.M. Schaefer, T.M. Schaefer, T.A. Reinhart, T.D. Oury, J. Biol. Chem. 279 (2004) 50019.
[41] H. Barsnes, S.O. Mikalsen, I. Eidhammer, BMC Res. Notes 1 (2008) 130.
[42] E.S. Witze, W.M. Old, K.A. Resing, N.G. Ahn, Nat. Methods 4 (2007) 798.
[43] C.F. Rodriquez, X. Guo, T. Shoeib, A.C. Hopkinson, K.W. Siu, J. Am. Soc. Mass Spectrom. 11 (2000) 967.
[44] B.N. Pramanik, Applied Electrospray Mass Spectrometry: Practical Spectroscopy Series, Marcel Dekker Inc., 2002.
[45] C.G. Ullman, P.I. Haris, D.A. Galloway, V.C. Emery, S.J. Perkins, Biochem. J. 319 (Pt 1) (1996) 229.
[46] K.E. Clemens, R. Brent, J. Gyuris, K. Munger, Virology 214 (1995) 289.
[47] A. Clements, K. Johnston, J.M. Mazzarelli, R.P. Ricciardi, R. Marmorstein, Biochemistry 39 (2000) 16033.
[48] N.J. Greenfield, Nat. Protoc. 1 (2006) 2876.
[49] N. Sreerama, R.W. Woody, Protein Sci. 13 (2004) 100.
[50] C. Perez-Iratxeta, M.A. Andrade-Navarro, BMC Struct. Biol. 8 (2008) 25.
[51] N.J. Greenfield, Anal. Biochem. 235 (1996) 1.