FEBS Letters 552 (2003) 170–176 FEBS 27604

A transient tobacco expression system coupled to MALDI-TOF-MS allows validation of the impact of differential targeting on structure and activity of a recombinant therapeutic glycoprotein produced in plants

Nathalie Mokrzycki-Issartel[a], Bernadette Bouchon[b], Sibille Farrer[a], Patricia Berland[a], Hélène Laparra[a], Jean-Claude Madelmont[b], Manfred Theisen[a,*]

[a] *MERISTEM Therapeutics, 8 Rue des Frères Lumière, 63100 Clermont-Ferrand, France*
[b] *UMR 484 Inserm–Université d'Auvergne, Rue Montalembert, 63005 Clermont-Ferrand, France*

Received 13 June 2003; revised 9 August 2003; accepted 11 August 2003

First published online 22 August 2003

Edited by Ulf-Ingo Flügge

Abstract Tobacco-based transient expression was employed to elucidate the impact of differential targeting to subcellular compartments on activity and quality of gastric lipase as a model for the production of recombinant glycoproteins in plants. Overall *N*-linked glycan structures of recombinant lipase were analyzed and for the first time sugar structures of its four individual *N*-glycosylation sites were determined in situ by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) on a trypsin digest without isolation or deglycosylation of the peptides. Three glycosylation sites contain both complex-type *N*-glycans and high-mannose-type structures, the fourth is exclusively linked to high-mannose glycans. Although the overall pattern of glycan structures is influenced by the targeting, our results show that the type of glycans found linked to a given Asn residue is largely influenced by the physico-chemical environment of the site. The transient tobacco system combined with MALDI-TOF-MS appears to be a useful tool for the evaluation of glycoprotein production in plants.

Key words: Differential targeting; Transient expression; Gastric lipase; *N*-linked glycosylation; Recombinant therapeutic proteins from plants; MALDI-TOF-MS

1. Introduction

The growing demand for safe protein drugs for human or animal health has led to the development of several production systems that are based on recombinant technologies for gene expression. In addition to the more conventional fermentation-based systems such as bacteria, yeast, insect- and mammalian cells, transgenic organisms have been employed for the production of high-value recombinant proteins [1]. Transgenic animal-based expression of recombinant proteins has been demonstrated and some products have entered clinical development [2,3]. Production of recombinant proteins in various plant species has also been developed within the last decade, and a first series of potential products has reached the clinic [4].

Plant-based expression systems accumulate several advantages as compared to other production systems, such as increased biological safety, capacity to produce complex molecules and last but not least the potential for highly economical large-scale production of biopharmaceuticals [5]. This large-scale production is based on the generation of stable transformed plants that can be grown in greenhouses or open fields.

The quality of a given recombinant product is largely influenced by the processing of the primary polypeptide chain, such as the addition of necessary post-translational modifications, among them *N*-linked glycosylation, by the host cell, the capacity of the heterologous system to correctly fold the peptide and, if required, to assemble multiple subunits into functional hetero- or homo-multimers. In plants it is possible to target proteins specifically to the extracellular space (apoplast) or to subcellular organelles such as the vacuole, the endoplasmic reticulum (ER) or to chloroplasts [6]. This will influence the level of expression, the stability but also the different post-translational modifications of a given molecule. The choice of subcellular targeting is therefore an important decision during the strategy design for expression of a given protein. Our aim was to develop a rapid tobacco-based transient expression system that allows the validation of the best expression targeting hypothesis for a given candidate protein within the shortest possible time and without generation of stable transgenic plants. We have previously shown that an adaptation of the agrobacterium infiltration technique [7] allowed us to demonstrate the introduction of a post-translational modification pathway into tobacco cells by transient co-expression of animal-derived prolyl-hydroxylase and its substrate, collagen, to increase hydroxyprolin content and thereby thermal stability of the collagen triple helix [8]. Vaquero et al. have demonstrated with a similar system that different expression approaches can be tested for recombinant antibodies [9].

To further validate this transient expression system, notably to produce sufficient amounts of the recombinant protein for a detailed functional and structural analysis of the candidate protein, we have studied recombinant dog gastric lipase (rGL) as a model.

Gastric lipase (GL) is a 50 kDa glycoprotein with a potential interest as treatment of pancreatic insufficiency linked to cystic fibrosis [10]. We have chosen this molecule as a model for our analyses, since preliminary data in stable tobacco on expression of gastric lipase that was either secreted or targeted

*Corresponding author. Fax: (33)-4-73986819.
E-mail address: manfredtheisen@free.fr (M. Theisen).

0014-5793/03/$22.00 © 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00916-5

to the vacuole were previously reported [11] and were available for comparison with results that we obtained by transient expression. In addition, we show the influence of retention within the ER on the structure and activity of rGL and present for the first time a detailed analysis of *N*-linked glycan structures on the differentially targeted molecules of this recombinant enzyme by a combination of Fluorophore-assisted carbohydrate (FACE) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analyses.

2. Materials and methods

2.1. Plasmid constructs

cDNA encoding mature dog GL was fused to the signal peptide of gastric lipase from rabbit (PS) for secretion targeting (s-rGL) [11]. For targeting to the ER, a DNA fragment encoding the KDEL tag was fused to the C-terminal end of the coding region (r-rGL) by polymerase chain reaction (PCR). For vacuolar targeting the C-terminal propeptide of barley lectin [12] was amplified by PCR and fused to the coding region of lipase (v-rGL). The coding regions were placed under the control of the enhanced 35S promoter (ep35S) [13,14] and 35S terminator (t35S) of cauliflower mosaic virus [15] in the plant expression plasmid pGAZE derived from pGA492 [16]. Constructs were first cloned in *Escherichia coli* strain DH5α, using standard techniques [17]. After control by DNA sequencing the plasmids were transferred into *Agrobacterium tumefaciens* strain LBA4404 [18] to obtain strains A746, A711 and A732 containing T-DNA encoding respectively s-rGL, r-rGL and v-rGL.

2.2. Agroinfiltration of tobacco leaves

Tobacco plants were agroinfiltrated according to the protocol of Kapila et al. [7]. Plant tissue was harvested for extraction 4 days after transformation as described in Merle et al. [8].

2.3. Extraction and purification of rGL

Frozen agroinfiltrated tobacco leaves were ground in liquid N_2 and macerated at 4°C in 250 mM NaCl, 50 mM glycine, 1 mM Triton X100 pH 2.5 (1 g of leaf material per 2 ml of extraction solution) for 1 h. Crude extract was clarified by centrifugation at 42 000×*g*. The sample was loaded onto a cation exchange sulfopropyl ceramic hyperD column (Cyphergen Biosystems, Biosepra, Fremont, CA, USA) buffered with 50 mM glycine, 100 mM NaCl, pH 2.5. The protein was eluted with 500 mM NaCl, 50 mM sodium acetate, pH 4.0 and loaded on 500 μl Fractogel EMD chelate charged with Cu^{2+} (Merck, Whitehouse Station, NJ, USA) in a Microbio Spin column (Bio-Rad, Hercules, CA, USA) equilibrated with 500 mM NaCl, 50 mM sodium acetate, pH 4.0. Elution was performed with 1.2 M ammonium acetate, 10 mM glycine pH 2.8. Fractions containing rGL were dialyzed against 20 mM citric acid pH 4.0. The pure product was obtained by collecting the rGL peak in RP-HPLC fractionation on a C4 column (Vydac, Hesperia, CA, USA) (150 mm×4.6 mm).

2.4. Determination of specific lipase activity

GL activity was measured using a pH Stat (718 STAT titrino; Methrom, Switzerland) at pH 5.5. Short-chain triacylglycerol (tributyrin, Fluka, St. Louis, MO, USA) was used as substrate. Lipolytic activity was measured at 37°C in the conditions described by Gargouri et al. [19]. One lipase unit is the amount of enzyme which catalyzes the release of 1 μmol of fatty acid per min.

2.5. Protein analysis by SDS–PAGE and N-terminal sequencing

Different forms of GL were analyzed on 12% SDS–PAGE gel according to Laemmli et al. [20]. The gel was stained with Coomassie blue. N-terminal sequencing was performed by automated Edman degradation on a Procise 491 sequencer (Applied Biosystems, Framingham, MA, USA).

2.6. FACE analysis of rGL N-glycans

100 μg of each of the different purified rGLs was used for *N*-glycan analysis. *N*-glycans from horseradish peroxidase and ribonuclease B (Sigma, St. Louis, MO, USA) were used as reference structures. *N*-glycans were released with PNGase A (Roche, Basel, Switzerland) after digestion of glycoproteins by trypsin (Sigma). *N*-glycans were purified by successive elution with water on a Sep-Pack Vac C18 cartridge and Fractogel EMD $SO3^-$ and evaporated. *N*-glycans were labeled with 8-aminonaphtalene-1,3,6-trisulfonate (ANTS; Molecular Probes, Eugene, OR, USA) according to Jackson et al. [21]. A standard ladder of glucose oligomers was obtained from wheat starch digest that was labeled with ANTS [22]. ANTS-oligosaccharides were loaded on a 30% polyacrylamide gel without SDS and subjected to separation at 4°C. They were visualized under UV light. Each oligosaccharide band was assigned by calculating its degree of polymerization value (DP) using the standard ladder of glucose oligomers [23] and confirmed using oligosaccharides released from ribonuclease B and horseradish peroxidase. The assignment was confirmed by MS after elution of each oligosaccharide from the gel [24].

2.7. Mass spectrometry

MALDI mass spectra were acquired on a Voyager DE-PRO TOF mass spectrometer (Applied Biosystems) equipped with a delayed extraction MALDI source and a pulsed nitrogen laser (337 nm).

Carbohydrate analysis of the ANTS-labeled oligosaccharides was performed according to the published procedure [24] in a negative linear delayed extraction mode.

Analysis of the intact molecules was done in a positive linear delayed extraction mode. All spectra were acquired using an accelerating voltage of 25 kV, a pulse delay time of 600 ns and a grid voltage of 93%. About 200 scans were averaged for each spectrum to improve signal-to-noise level. Protein samples (10–30 μM) in 0.1% aqueous TFA solution were mixed with a matrix solution and deposited onto a stainless-steel target according to the dried droplet method. The matrix solutions were prepared by dissolving 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid, SIN; Acros Organics, Van Overbeek, Belgium) in CH_3CN/water (30:70 or 50:50, v/v) containing 0.1% TFA at a concentration of 15 μg/μl, or 2,5-dihydroxybenzoic acid (DHB; Acros Organics) in CH_3CN/water (30:70 or 50:50, v/v) containing 0.1% TFA at a concentration of 10 μg/μl. Calibration was performed in a close external mode using bovine serum albumin (BSA; singly charged M^+ and doubly charged M^{2+}) ions.

For analysis of the glycopeptides, the purified protein was dissolved in CH_3CN (10% v/v, final concentration) and diluted with 25 mM ammonium hydrogenocarbonate, pH 8.2, containing 1 mM calcium chloride. Reductively alkylated porcine trypsin (Promega, Madison, WI, USA), dissolved in the same buffer, was added to the protein at a molar enzyme/substrate ratio of 1/20 and the solution was incubated at 37°C for 16 h. The tryptic digests (20 pmol/μl) in the digest solutions were diluted to a final concentration of 10 pmol/μl in the matrix solution. Sample preparation was carried out according to the dried droplet method, using SIN and DHB as matrices. The matrix solutions were prepared as for the analysis of the intact molecule.

Spectra were obtained in positive reflector mode with an accelerating voltage of 25 kV, a pulse delay time of 250 ns and a grid voltage of 75%. Spectra were averaged from about 200 laser shots to improve signal-to-noise level. Mass assignment was done using a peptide mixture (insulin B chain, bovine insulin, aprotinin and ubiquitin; Laser-Bio Labs, Sophia-Antipolis, France) as external standards. A trypsin autodigestion peptide (m/z = 5561.2) was also used as an internal calibration standard.

3. Results

3.1. Production of s-rGL, r-rGL and v-rGL in tobacco leaves

Three different expression cassettes for differential targeting of rGL in transient expression in tobacco leaves were generated (Fig. 1A). The coding region of mature dog GL [25] was fused to the signal peptide of GL from rabbit to assure secretion out of the cell. This construct was termed s-rGL. A fusion of this construct with the C-terminal propeptide of barley lectin was generated for targeting to the vacuole (v-rGL). A third construct was generated by fusion of the C-terminal KDEL (Lys-Asp-Glu-Leu) sequence to the coding region of GL to assure retention in the ER [26]. Transcription of all constructs was driven from the enhanced 35S promoter of cauliflower mosaic virus and transcription was terminated

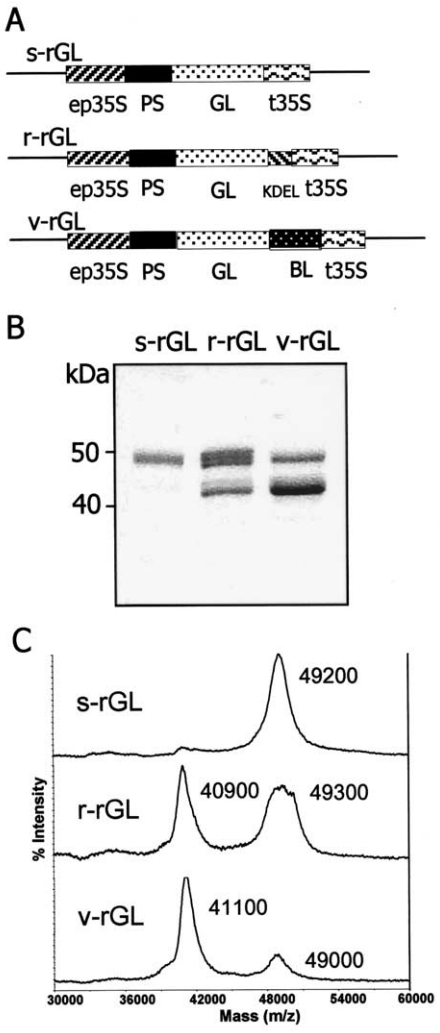

Fig. 1. Analysis of purified s-rGL, r-rGL and v-rGL. A: Schematic representation of expression cassettes. The coding region of mature dog GL is indicated as GL, the signal peptide from rabbit GL as PS. KDEL represents the ER retention signal of r-rGL and BL the vacuolar targeting peptide of v-rGL. ep35S is the enhanced 35S promoter and t35S the 35S terminator of cauliflower mosaic virus. B: SDS–PAGE analyses. The purified proteins were analyzed on a 12% polyacrylamide gel, and stained with Coomassie blue. Size markers for 50 and 40 kDa are indicated in the left margin. C: MALDI-TOF mass spectra of the purified proteins using DHB as matrix. The mass range 30 000–60 000 corresponds to rGL spectra having singly charged ions. Indicated masses correspond to peak centroids and were determined with an external calibration using BSA (singly and doubly charged ions).

using the t35S termination sequence. An *Agrobacterium*-mediated transient system [7] was used to express s-rGL, r-rGL and v-rGL after vacuum infiltration of tobacco leaves. Extraction and purification of the rGLs were carried out 4 days after agroinfiltration. Between 0.5 and 1 mg of recombinant protein per 100 g of fresh leaves were purified. A detailed structural and functional analysis of the three different lipase forms was performed.

3.2. Comparison of catalytic activities of s-rGL, r-rGL and v-rGL in tobacco leaves

The catalytic activities of recombinant rGLs after transient expression were measured on short-chain triacylglycerols [19]. s-rGL is the most active recombinant protein with a specific activity of 360 U/mg. This activity is lower than but in the same range as that observed for the native protein, which reached 570 U/mg [27]. v-rGl exhibits two-fold less specific activity (140 U/mg) than s-rGL. Comparable values were observed when the proteins were produced in stable tobacco [11], thus showing a good equivalence between the transient and stable tobacco expression systems. Specific activity of r-rGL was 310 U/mg, which is close to the activity of the secreted form.

3.3. Primary structure of recombinant rGLs transiently expressed in tobacco leaves

Purified rGLs were analyzed by SDS–PAGE (Fig. 1B) and Western blot (data not shown). The recombinant proteins were analyzed by MALDI-TOF-MS (Fig. 1C) and submitted to N-terminal sequencing. Different patterns were obtained according to the targeting of the protein. For s-rGL one major molecular species was detected on SDS–PAGE and by mass spectrometry, with a molecular weight around 49 kDa, which is very close to the molecular weight of native dog GL

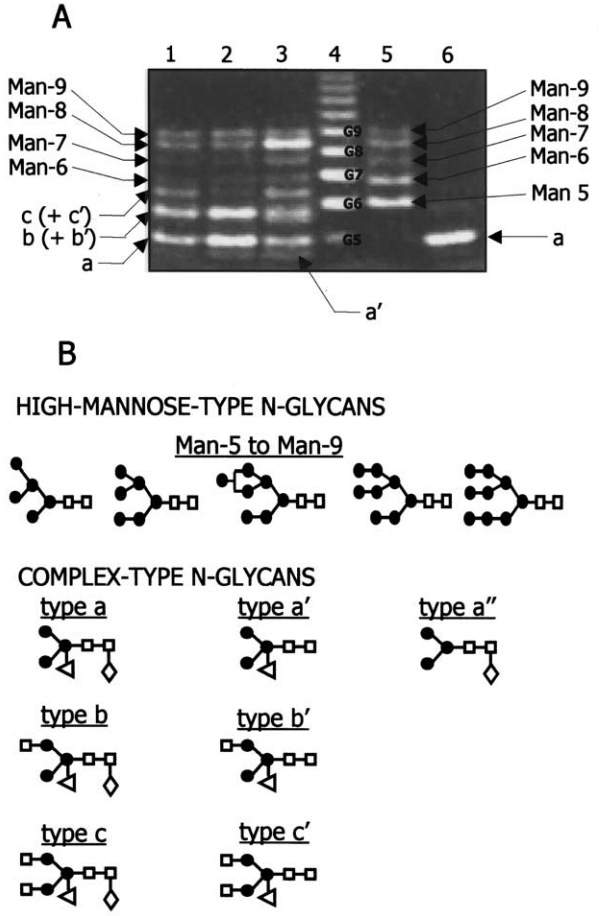

Fig. 2. FACE analysis. A: Comparison of FACE profiles of ANTS-labeled *N*-glycans isolated from purified proteins. Lanes 1–3: ANTS-labeled *N*-glycans isolated from respectively s-rGL, v-rGL and r-rGL. Lane 4: standard ladder of glucose oligomers (G5–G9). Each band was assigned according to its DP value and according to standard oligosaccharides from ribonuclease (Man-5 to Man-9, lane 5) and peroxidase (lane 6). Band identifications are indicated in the margins of the figure. B: Theoretical structures of *N*-glycans linked to rGL. ● Man; □ GlcNAc; ◇ Fuc; ◁ Xyl.

Table 1
Theoretical masses of tryptic glycosylated peptides of recombinant GL

	UN[a]	Man-5	Man-6	Man-7	Man-8	Man-9	a	b	c	a′	b′	c′	a″
Asn 15	**4647.2**	**5864.3**	**6026.5**	**6188.6**	**6350.8**	**6512.9**	5818.3	6021.5	6224.7	**5672.1**	**5875.3**	**6078.5**	5686.2
Asn 80	4685.3	5902.4	6064.6	**6226.7**	**6388.9**	6551.0	**5856.4**	**6059.6**	**6262.8**	5710.3	5913.4	6116.6	**5724.3**
Asn 252	2922.4	4139.6	4301.7	**4463.8**	**4626.0**	**4788.1**	4093.5	4296.7	4499.9	3947.4	4150.6	4353.8	3961.4
Asn 308	**6329.2**	7546.3	7708.4	7870.6	**8032.7**	8194.9	**7500.3**	**7703.5**	**7906.6**	7354.1	7557.3	7760.5	7368.1

Each glycopeptide is defined by its glycosylation position along the peptidic chain (Asn 15, Asn 80, Asn 252, Asn 308) which corresponds to a specific tryptic peptide. Expected glycosylation pattern (high-mannose or complex-type, see Fig. 2B for detailed structures) are supposed for each tryptic peptide, and theoretical masses are calculated from these expected tryptic peptides. Average masses are expressed as $[M+H]^+$.
Numbers in bold correspond to identified structures.
[a]UN: unglycosylated peptide.

(50 kDa) [27]. The recombinant secreted proteins have an N-terminal sequence, which corresponds to correct N-terminal processing of the protein. For v-rGL two molecular species were detected, a major one around 41 kDa and a minor one at about 49 kDa. N-terminal sequencing revealed the presence of two protein sequences, a minor one corresponding to the full-length protein, whereas the dominant sequence corresponded to a truncated form of GL beginning at amino acid Ile 55. A more complex pattern was observed in the case of r-rGL. By MALDI-TOF-MS, a very broad peak around 49 kDa was detected together with a form around 41 kDa. This pattern is consistent with the SDS–PAGE profile that shows at least three individual bands around 49 kDa. Western blotting with a lipase-specific antibody shows the same pattern of bands. Incubation of the blot with an anti-KDEL antibody does not show a correlation of these bands to a potential proteolytic cleavage of the KDEL sequence (data not shown). A truncated form of the protein starting at Ile 55 was detected by N-terminal sequencing together with the full-length protein. These results indicate the presence of heterogeneous forms of rGL after targeting to the ER. The relative broadness of the peaks in MALDI-TOF-MS spectra is a first indication of a potential heterogeneity of the different forms of rGL that could eventually be attributed to differences in glycosylation.

3.4. *Analysis of* N-*glycosylation of rGLs by FACE*

GL has four potential *N*-glycosylation sites located on Asn 15, Asn 80, Asn 252 and Asn 308. To study the extent of *N*-glycosylation on the total protein, the different rGLs were submitted to FACE analysis [21]. A large number of different but typical plant *N*-glycans were found. High-mannose-type *N*-glycans from Man-6 to Man-9 and complex-type *N*-glycans, type a, b and c containing fucose and xylose residues attached to the core structure, were identified (Fig. 2A,B). Intermediate complex structures lacking fucose residue were also detected. These structures were named type a′, b′ and c′. Type a′ was clearly resolved on the gel. Due to lower resolution, types b′ and c′ migrate at the same positions as type b and type c, respectively, and their presence could only be confirmed by MALDI-TOF-MS analysis of the isolated glycan bands (data not shown).

Although each rGL contains the two classes of *N*-glycans, the relative proportion of high-mannose and complex structures differs from one targeting to the other. Whereas s-rGL contains more complex-type *N*-glycan (c+c′) than v-rGL, 'truncated' structures types a and (b+b′) (lacking one or two GlcNAc terminal residues) are the dominant complex structures in v-rGL (Fig. 2, lanes 1 and 2). When rGL is directed to the ER, Man-8 becomes the major high-mannose-type structure (Fig. 2, lane 3). These data correlate with the theoretical *N*-glycosylation processing in plants [28]. Nevertheless, a clear-cut difference of glycosylation patterns for the three proteins is not found since r-rGL displays complex structures in relatively high proportion and s-and v-rGL high-mannose-type structures.

3.5. *Site-specific analysis of* N-*glycosylation by MALDI-TOF-MS*

To further elucidate the glycan structures present on each *N*-linked glycosylation site of the rGLs obtained after differential targeting, tryptic digests of the proteins were analyzed by MALDI-TOF-MS. Trypsin produced glycosylated peptides which could be analyzed by mass spectrometry on a relatively narrow mass range (*m*/*z* 4000–8500). Glycosylation on the four potential glycosylation sites could thus be compared easily in situ without peptide isolation and deglycosylation.

Table 2
Relative amounts of *N*-glycans detected on s-rGL, r-rGL and v-rGL *N*-glycosylation sites

		UN	Man-5	Man-6	Man-7	Man-8	Man-9	a	b	c	a′	b′	c′	a″
Asn 15	s-rGL	<5	15	15	10	20	15	–	–	–	10	10	<5	–
	r-rGL	5	<5	5	10	50	10	–	–	–	<5	10	5	–
	v-rGL	5	–	10	10	30	20	–	–	–	5	15	5	–
Asn 80	s-rGL	–	–	–	<5	<5	–	30	45	15	–	–	–	5
	r-rGL	–	–	–	–	30	–	30	35	–	–	–	–	5
	v-rGL	–	–	–	<5	<5	–	45	40	–	–	–	–	15
Asn 252	s-rGL	–	–	–	20	45	35	–	–	–	–	–	–	–
	r-rGL	–	–	–	20	40	40	–	–	–	–	–	–	–
	v-rGL	–	–	–	15	40	45	–	–	–	–	–	–	–
Asn 308	s-rGL	10	–	–	–	<5	–	40	30	15	–	–	–	–
	r-rGL	20	–	–	–	25	–	30	25	<5	–	–	–	–
	v-rGL	15	–	–	–	–	–	50	30	<5	–	–	–	–

Glycosylated peptides were estimated for each glycosylation site. Relative amounts of oligosaccharide structures are expressed as percent (rounded to the nearest 5%). Data are representative of three experiments. See Fig. 2B for detailed structures.
–: not detected.

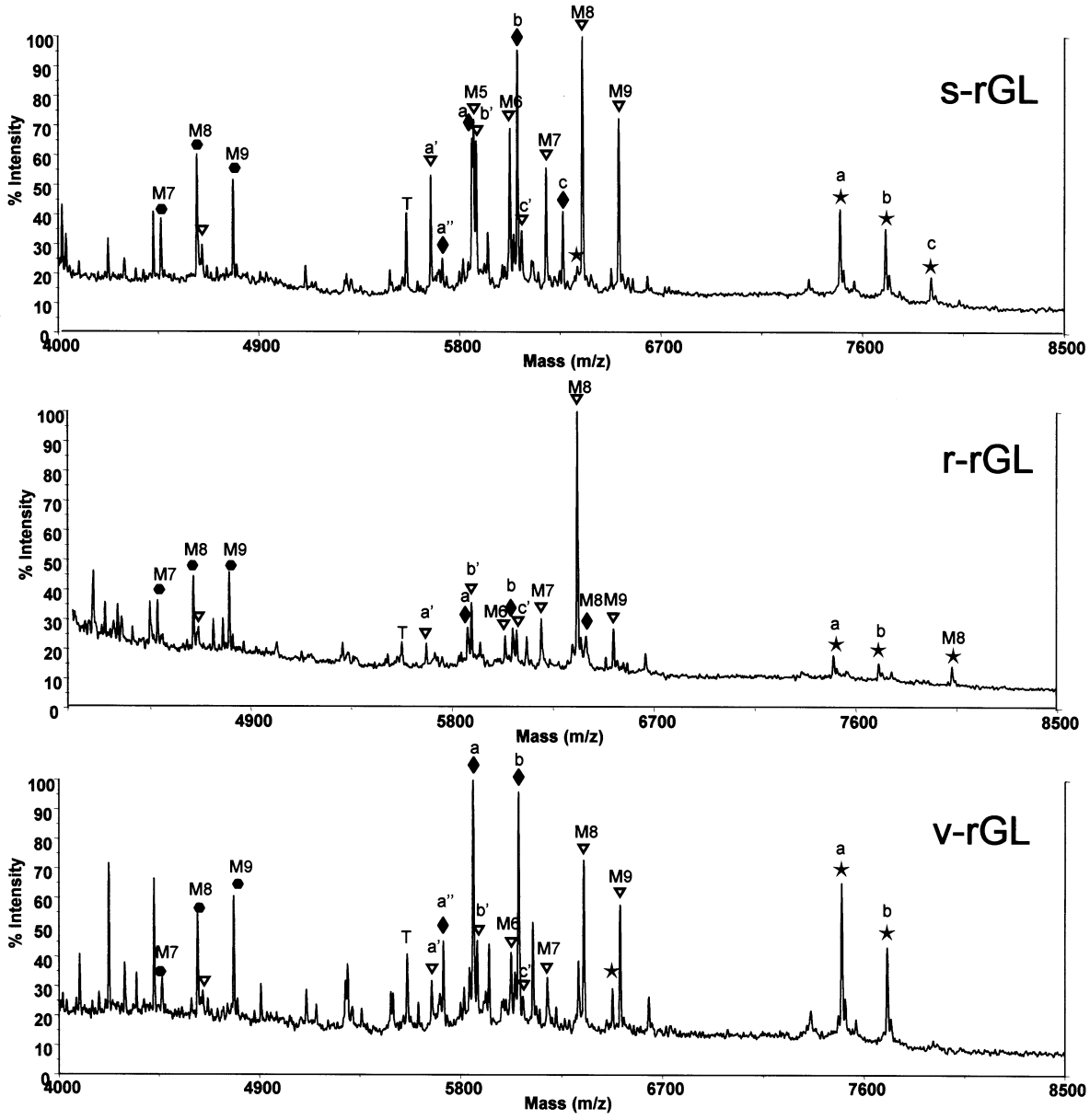

Fig. 3. MALDI-TOF analysis of rGL tryptic digests. Tryptic peptide mixtures were analyzed directly using SIN as matrix. Theoretical glycosylation was calculated for each potential glycosylation site (Table 1). ▽ Tryptic peptide corresponding to Asn 15 glycosylation site; ♦ tryptic peptide corresponding to Asn 80 glycosylation site; ● tryptic peptide corresponding to Asn 252 glycosylation site; ★ tryptic peptide corresponding to Asn 308 glycosylation site; T, trypsin autolytic peptide. Each glycosylated form was identified as high-mannose-type (Man-5 to Man-9) or as complex-type (a, b, c, a′, b′, c′, a″). See Fig. 2B and Table 1 for structure assignment and Table 2 for relative amounts of *N*-glycans. Non-assigned peaks correspond to other tryptic peptides of the protein.

N-glycan assignment was realized by comparison between calculated and measured masses for each theoretical tryptic peptide. The interpretation is based on a good correlation between a series of calculated masses and experimental values for a given tryptic peptide (Table 1). The four sites were found glycosylated. The *N*-glycopeptides containing the potential *N*-glycosylation sites Asn 15, Asn 80, Asn 252 and Asn 308 were clearly identified (Fig. 3) and only a very small fraction of peptides corresponding to Asn 15 and Asn 308 were non-glycosylated. A minor fraction of oxidized peptides was detected (M+16, not indicated in Fig. 3).

Independent of the subcellular targeting, each site displays a specific glycosylation pattern (Fig. 3 and Table 2). Peak intensities in mass spectra are calibrated to the most intense ion. In the case of r-rGL, with Man-8 representing a very intense ion, the peak intensities of complex glycan structures appear weaker than for other targetings (Fig. 3). For each of the three different targetings Table 2 gives the relative amount of a given glycan structure in % of the total glycan structures for a specific glycosylation site. Each line represents 100% of a given peptide for a specific targeting. Asn 80 and Asn 308 were mainly glycosylated with complex-type *N*-glycans including types a, b and c. In addition, Asn 80 is the only site that contains complex-type a″. Asn 252 was only glycosylated with high-mannose-type *N*-glycans, Man-7, Man-8 and Man-9. Asn 15 displayed a more complex glycosylation pattern; it

was linked to high-mannose-type *N*-glycans (Man-5 to Man-9) but also to non-fucosylated complex-type *N*-glycans a′, b′ and c′. Whereas each of the sites displays its specific pattern of glycosylation, some appear also partly influenced by the targeting. Asn 80 and 308 of s-rGL contain higher amounts of complex-type c than for vacuolar or ER targeting (Table 2 and Fig. 3). An increase in types a and b for v-rGL and r-rGL compensates this decrease at least in part, which correlates to the data observed by FACE analysis. v-rGL has a higher amount of type a″ on Asn 80. When directed to the ER, a clear increase of high-mannose-type *N*-glycan Man-8 is observed on Asn 15, Asn 80 and Asn 308. Glycosylation of Asn 252 is not influenced by the targeting; whatever the targeting, high-mannose-type structures with roughly identical percentages were detected.

The relative proportions of certain glycopeptides differ according to the targeting of the proteins (i.e. r-rGL having a higher amount of high-mannose-type *N*-glycans) but the type of glycans found appears to be more site-specific than targeting-specific (Table 2).

4. Discussion

The choice of a recombinant expression system for the production of a therapeutic protein is influenced by several technical, economical and regulatory parameters. The initial question to be answered is whether a given expression system can produce the molecule in an active form that can be administered to the patient. Characteristics such as integrity, conformity, activity, immunogenicity and stability of the protein have to be elucidated. Our aim was to test a transient tobacco-based expression system that allows validation of the production of a given recombinant product in plants in a minimum of time and with a maximum level of fidelity as compared to results obtained in stable plants. This system has previously been used to validate the expression of antibody molecules [9,29], or to study the possibility to metabolically engineer plant cells to improve the thermal stability of recombinant human collagen [8]. In the latter case, results obtained by transient expression were highly reproducible and predictive for the biochemical quality of protein obtained in stable plants.

In this study we performed a detailed analysis of differentially targeted recombinant lipases and their *N*-glycosylation patterns after purification from transiently transformed tobacco without going through the time-consuming creation of stable tobacco lines. GL is a highly glycosylated globular protein [30,31] of approximately 50 kDa containing four *N*-glycosylation sites localized on Asn 15, Asn 80, Asn 252 and Asn 308. We have previously shown that targeting of GL to two different subcellular locations in stable tobacco had a significant influence not only on the structure of the molecule, but also on its catalytic activity [11]. In [11] it was also shown by immuno-gold labeling that the secreted form of rGL is found in the apoplast, while v-rGL accumulates in the vacuole. The results obtained for secretion and vacuolar targeting in transient expression confirm these results. As for stable tobacco a large proportion of v-rGL is cleaved between amino acids Asn 54 and Ile 55, thus confirming its vacuolar targeting. This cleavage can be attributed to proteases that are located in the plant vacuole [32,33] and appears to be related to a proteolytic sensitivity of this peptide bond in pre-duodenal lipases [34]. Although it appears that the N-terminal peptide remains associated to the rest of the molecule in the non-denatured globular structure [11] it is readily detached under denaturing conditions in SDS–PAGE and MALDI-TOF-MS. This proteolytic instability of v-rGL has an influence on the catalytic activity of the molecule, which loses more than 50% of its activity as compared to the secreted s-rGL. It therefore appears to be obvious that secretion, but not vacuolar targeting, would be the better choice for production of GL in plants.

Interestingly, a large proportion of r-rGL which is supposedly blocked in the ER through its KDEL C-terminal sequence also shows a cleavage between amino acids Asn 54 and Ile 55. This could be an indication for the leakiness of this targeting and might indicate that r-rGL is partly shuttled to the vacuole, a result which has already been described by Gomord et al. [35]. Although we did not express r-rGL in stable tobacco, it again appears obvious that targeting to the ER is not a viable expression hypothesis for this enzyme.

GLs were already expressed in heterologous organisms such as *E. coli* [10], insect cells [36] or stable tobacco plants [11], but the exact structures of *N*-glycan chains of the native or the recombinant enzymes were, to our knowledge, not investigated.

The purified s-, v- and r-rGL from transient expression in tobacco allowed us to perform this analysis. Our aim was thereby to compare the glycan structures on recombinant proteins produced by differential targeting in plants and not to elucidate the glycan structure of recombinant versus native protein. FACE data on the overall glycan composition of the different forms of lipase confirmed the presence of a large number of different glycan structures. This important heterogeneity of plant glycan structures has already been described for several plant proteins. See for example Lerouge et al. [28] and references cited therein. Our FACE analysis does not show a clear-cut distinction of glycan structures in relation to the targeting of the protein. In particular, complex and truncated structures were found in the ER-directed protein, suggesting a leakage through other compartments [35]. The presence of truncated complex glycan structures on s-rGL can be explained by the presence of exoglycosidases in the apoplast [28,37]. It does not necessarily indicate a transport of s-rGL to the vacuole, which is also confirmed by the lack of the Ile 55 truncation in this secreted form of rGL. In general the FACE technique does not allow a detailed site-specific comparison of structures, but gives a first appreciation of the overall glycosylation pattern in relation to differential targeting.

A more precise identification of the glycan structures on GL could be obtained through a site-by-site in situ analysis by MALDI-TOF-MS on tryptic digests of the differentially targeted molecules. In the case of this study this was a very fast and easy procedure which did not require isolation of the individual glycopeptides or deglycosylation of the protein. For other proteins additional isolation of glycopeptides or additional mass ranges might be required for MS analysis of the structures. Each of the four glycosylation sites has its characteristic pattern of glycan structures. Data from MALDI-TOF-MS and FACE analysis are complementary but a direct comparison of relative values from MALDI-TOF-MS and band intensities from the FACE gel is not possible. Complex form c for r-rGL, for example, represents only a low relative intensity in the MALDI-TOF-MS analysis of

Fig. 3, while a band representing total (c+c′) forms of r-rGL is clearly visible in the FACE gel (Fig. 2).

Asn 15 contains high-mannose glycans as well as complex-type glycans a′, b′ and c′ and a minor percentage of non-glycosylated residues. Asn 80 and Asn 308 contain predominantly complex structures and only few high-mannose forms. Asn 252 is only linked to high-mannose-type glycans. Apart from some quantitative differences, the glycan structures on a given site are independent of the targeting of the molecule. Asn 15, for example, in all cases contains a mixture of high-mannose and complex-glycans while Asn 252 always contains only high-mannose-type glycans. The processing of the glycan side-chains is closely related to their accessibility to the glycosyltransferases in the Golgi [28,38,39]. It therefore appears obvious that Asn 252, after having been glycosylated with high-mannose-type glycans, is buried within the tertiary structure of the molecule in a way that prohibits further modification of the structures.

The site-by-site analysis of glycan structures after differential targeting of lipase as a model enzyme indicates that, at least for this protein, the biochemical and/or physical environment of the Asn residue has a significantly greater impact on the sugar structures that are added to the polypeptide chain than the differential routing of the protein through the compartments of the cell.

In conclusion, our transient expression system allowed us to purify sufficient amounts of recombinant protein to perform a functional and structural analysis of the differentially targeted molecules and to select the best targeting approach for a given protein without the necessity to produce stable transformed plants.

Acknowledgements: We thank our colleagues at Meristem Therapeutics and specifically Nathalie Duchateau, Marianne Mure and Véronique Gruber for their assistance. We also thank Gregg Wallis and Christine Merle for critical reading of the manuscript. This work was partly funded by a grant (Saut Technologique) from the French Research Ministry.

References

[1] Andersen, D.C. and Krummen, L. (2002) Curr. Opin. Biotechnol. 13, 117–123.
[2] Young, M.W., Meade, H., Curling, J.M., Ziomek, C.A. and Harvey, M. (1998) Res. Immunol. 149, 609–610.
[3] Houdebine, L.M. (2000) Transgenic Res. 9, 305–320.
[4] Giddings, G., Allison, G., Brooks, D. and Carter, A. (2000) Nat. Biotechnol. 18, 1151–1155.
[5] Mison, D. and Curling, J. (2000) Biopharmacology 13, 48–54.
[6] Daniell, H., Streatfield, S.J. and Wycoff, K. (2001) Trends Plant Sci. 6, 219–226.
[7] Kapila, J., De Rycke, R., Van Montagu, M. and Angenon, G. (1997) Plant Sci. 122, 101–108.
[8] Merle, C., Perret, S., Lacour, T., Jonval, V., Hudaverdian, S., Garrone, R., Ruggiero, F. and Theisen, M. (2002) FEBS Lett. 515, 114–118.
[9] Vaquero, C., Sack, M., Schuster, F., Finnern, R., Drossard, J., Schumann, D., Reimann, A. and Fischer, R. (2002) FASEB J. 16, 408–410.
[10] Bénicourt, C., Blanchard, C., Carrière, F., Verger, R. and Junien, J.L. (1993) in: Clinical Ecology of Cystic Fibrosis (Escobar, H., Baquero, C.F. and Suarez, L., Eds.), pp. 291–295, Elsevier, Amsterdam.
[11] Gruber, V., Berna, P.P., Arnaud, T., Bournat, P., Clément, C., Mison, D., Olagnier, B., Philippe, L., Theisen, M., Baudino, S., Bénicourt, C., Cudrey, C., Bloës, C., Duchateau, N., Dufour, S., Gueguen, C., Jacquet, S., Ollivo, C., Poncetta, C., Zorn, N., Ludevid, D., Van Dorsselaer, A., Verger, R., Doherty, A., Mérot, B. and Danzin, C. (2001) Mol. Breed. 7, 329–340.
[12] Bednarek, S.Y. and Raikhel, N.V. (1991) Plant Cell 3, 1195–1206.
[13] Guilley, H., Dudley, R.K., Jonard, G., Balazs, E. and Richards, K.E. (1982) Cell 30, 763–773.
[14] Kay, R., Chan, A., Daly, M. and McPherson, J. (1987) Science 236, 1299–1302.
[15] Guerineau, F. and Mullineaux, P. (1993) in: Plant Molecular Biology (Croy, R.R.D., Ed.), pp. 121–47, BIOS Scientific Publishers, Oxford.
[16] An, G. (1986) Plant Physiol. 81, 86–91.
[17] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Plainview, NY.
[18] Walden, R., Koncz, C. and Schell, J. (1990) Methods Mol. Cell. Biol. 1, 175–194.
[19] Gargouri, Y., Pieroni, G., Riviere, C., Sauniere, J.F., Lowe, P.A., Sarda, L. and Verger, R. (1986) Gastroenterology 91, 919–925.
[20] Laemmli, U.K. (1970) Nature 227, 680–685.
[21] Jackson, P. (1990) Biochem. J. 270, 705–713.
[22] Jackson, P. (1997) BioMethods 9, 122.
[23] Frado, L.Y. and Strickler, J.E. (2000) Electrophoresis 21, 2296–2308.
[24] Lemoine, J., Cabanes-Macheteau, M., Bardor, M., Michalski, J.-C., Faye, L. and Lerouge, P. (2000) Rapid Commun. Mass Spectrom. 14, 100–104.
[25] Blanchard, C., Bénicourt, C. and Junien, J.L. (1994) PCT Patent Appl. WO 94/13816.
[26] Andres, D.A., Rhodes, J.D., Meisel, R.L. and Dixon, J.E. (1991) J. Biol. Chem. 266, 14277–14282.
[27] Carriere, F., Moreau, H., Raphel, V., Laugier, R., Benicourt, C., Junien, J.L. and Verger, R. (1991) Eur. J. Biochem. 202, 75–83.
[28] Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Laine, A.C., Gomord, V. and Faye, L. (1998) Plant. Mol. Biol. 38, 31–48.
[29] Vaquero, C., Sack, M., Chandler, J., Drossard, J., Schuster, F., Monecke, M., Schillberg, S. and Fischer, R. (1999) Proc. Natl. Acad. Sci. USA 96, 11128–11133.
[30] Bodmer, M.W., Angal, S., Yarranton, G.T., Harris, T.J., Lyons, A., King, D.J., Pieroni, G., Riviere, C., Verger, R. and Lowe, P.A. (1987) Biochim. Biophys. Acta 909, 237–244.
[31] Vaganay, S., Joliff, G., Bertaux, O., Toselli, E., Devignes, M.D. and Benicourt, C. (1998) DNA Seq. 8, 257–262.
[32] Boller, T. and Kende, H. (1979) Plant Physiol. 63, 1123–1132.
[33] Hara-Nishimura, I., Inoue, K. and Nishimura, M. (1991) Eur. J. Biochem. 294, 89–93.
[34] De Caro, J., Verger, R. and De Caro, A. (1998) Biochim. Biophys. Acta 1386, 39–49.
[35] Gomord, V., Denmat, L.A., Fitchette-Laine, A.C., Satiat-Jeunemaitre, B., Hawes, C. and Faye, L. (1997) Plant J. 11, 313–325.
[36] Wicker-Planquart, C., Canaan, S., Riviere, M., Dupuis, L. and Verger, R. (1996) Protein Eng. 9, 1225–1232.
[37] Tezuka, K., Hayashi, M., Isshihara, H., Nishimura, M., Onozaki, K. and Takahashi, N. (1993) Anal. Biochem. 211, 205–209.
[38] Faye, L., Sturm, A., Bollini, R., Vitale, A. and Chrispeels, M.J. (1986) Eur. J. Biochem. 158, 655–661.
[39] Bardor, M., Loutelier-Bourhis, C., Marvin, L., Cabanes-Macheteau, M., Lange, C., Lerouge, P. and Faye, L. (1999) Plant Physiol. Biochem. 37, 319–325.