Mass-spectrometric analysis of ADP-ribosylation factors from bovine brain: identification and evidence for homogeneous acylation with the C_{14:0} fatty acid (myristate)

Scott J. BERGER,* Katheryn A. RESING,† Timothy C. TAYLOR*‡ and Paul MELANÇON*§

*Department of Chemistry and Biochemistry, and †Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado at Boulder, Boulder, Colorado 80309-0215, U.S.A.

The two proteins from bovine brain previously shown to be required for the guanosine 5'-[γ-thio]triphosphate-dependent inhibition of a well-characterized intra-Golgi transport assay, termed GGBF and GGBF*, have been definitively identified as members of the ADP-ribosylation factor (ARF) family by electrospray MS analysis of the intact proteins, and of their tryptic fragments. Extensive protein-sequence information obtained from this analysis identified GGBF and GGBF* as bovine ARF1 and ARF3 respectively. The sequence of bovine ARF3, which had not previously been determined, appears identical to

that predicted from the rat and human ARF3 cDNAs. Further analysis of the N-terminal tryptic fragments of both bovine ARFs demonstrates N-terminal acylation solely with the C_{14:0} fatty acid (myristate). This finding establishes that the previously reported specific-activity difference between ARF1 and ARF3 in the intra-Golgi transport assay is not due to lipid heterogeneity at the N-terminus. This finding also indicates that the heterogeneity of N-terminal fatty-acyl groups previously observed on other myristoylated proteins is not universal.

INTRODUCTION

ADP-ribosylation factors (ARFs) are a family of ras-related low-molecular-mass GTPases that were originally identified as necessary cofactors in the ADP-ribosylation of the α subunit of stimulatory G-protein (G_sα) by cholera toxin [1]. Recent studies carried out *in vivo* or using a variety of cell-free assays have revealed an important physiological role for ARFs in the regulation of multiple steps in intracellular protein transport (recently reviewed in [2,3]). Using a cell-free assay that reconstitutes transport between Golgi compartments [4], two soluble 21 kDa proteins from bovine brain were purified which mediate the guanosine 5'-[γ-thio]triphosphate (GTP[S])-dependent vesiculation of Golgi membranes and the resulting block of intra-Golgi transport [5,6]. These factors, termed GGBF (GTP-dependent Golgi binding factor) and GGBF*, differed 3–5-fold in specific activity, but were both found to be GTP-binding proteins immunologically related to the ARF family [5]. ARFs have been implicated in additional intracellular transport steps, including endoplasmic-reticulum–Golgi traffic [7,8], endosome–endosome fusion [9] and reassembly of the nuclear envelope [10].

ARFs are believed to function in intracellular transport by recruiting cytoplasmic coat complexes to the membrane before, or contemporaneously with, the budding of transport vesicles [2,3,6,11]. An ARF requirement has been demonstrated for the membrane association of two types of coat proteins, COP I (or Coatomer) [12,13] and the clathrin adaptor protein AP-1 [14,15], and suggested for a third, p200 [16]. It has been proposed that the multiplicity of ARFs is required for ordered assembly of multiple coats at different steps of intracellular transport [5]. The difference in specific activity between the two GGBFs observed by using the intra-Golgi assay is consistent with this hypothesis, and suggests

that members of the ARF family have related, but distinct, functions in regulating assembly of multiple protein coats. However, other roles for ARFs cannot be ruled out; for example, ARFs have recently been identified as potential upstream regulators of phospholipase D activity [17–19].

The assignment of GGBF and GGBF* as specific members of the ARF family is complicated by several factors. The mammalian ARF family is highly conserved at the protein-sequence level, both within a single species and between species, making production of antisera for a specific ARF protein difficult. Mammalian ARFs can be separated into three groups by sequence similarity, Group I (ARF1, ARF2, ARF3), Group II (ARF4, ARF5), and Group III (ARF6), that share at least 90 % sequence identity within each group, and 64–80 % sequence identity between groups [20]. Additional complications in identifying GGBF and GGBF* arise from the fact that much of the unique sequence in ARFs is present within the 14 N-terminal residues, blocked to traditional peptide sequencing by N-terminal fatty acylation [21,22]. Lastly, although Genbank data for human and rat ARF sequences are somewhat extensive, the sequences for bovine ARF family members are limited to two Group I ARF proteins, ARF1 and ARF2.

The ARF family also contains ARF-like proteins (ARLs), which show 40–60 % protein-sequence identity with ARF1, but are distinguishable from ARFs by their inability to serve as cofactors in the cholera-toxin-mediated ADP-ribosylation of G_sα subunits [23]. Our previous analysis of GGBF and GGBF* [5] demonstrated that they possess this cofactor activity, making their assignment as ARLs unlikely.

Recent developments in MS have greatly facilitated analysis of large biomolecules by the introduction of nebulization-assisted electrospray ionization to volatilize proteins and peptides for

Abbreviations used: ARF, ADP-ribosylation factor; ARL, ARF-like protein; GGBF, GTP-dependent Golgi binding factor; G_s, the stimulatory G-protein; G_sα, the transducing G-protein α subunit; GTP[S], guanosine 5'-[γ-thio]triphosphate; LC, liquid chromatography; LC/MS, LC coupled with MS; LC/MS/MS, LC coupled with tandem MS; MS/MS, tandem MS; NMT, peptide N-myristoyltransferase (EC 2.3.1.97); Tp, tryptic peptide.

‡ Present address: Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute–BCMM, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536, U.S.A.

§ To whom all correspondence should be addressed.

mass analysis (reviewed in [24]). Proteins or peptides are introduced into a triple-quadrupole electrospray mass spectrometer either by direct injection (MS) or as eluates from a liquid chromatograph (LC/MS). Masses of proteins or peptides can be determined on a single quadrupole with precision up to 1 part in 10000. Additionally, sequence information from a mixture of peptides can be obtained by tandem mass spectrometry (MS/MS) [25–27]. This process involves selecting a precursor ion of particular m/z ratio in the first quadrupole, collisionally activating the selected ion in the second quadrupole, and analysing the resulting product ions in the third quadrupole. The precision of the technique, coupled with its high sensitivity, makes electrospray MS a powerful technique for the analysis and identification of GGBF and GGBF*.

ARFs belong to a diverse group of proteins that are acylated with myristate or related fatty acids at their N-terminus. N-terminal protein myristoylation is an irreversible eukaryotic co-translational modification that occurs shortly after cleavage of the initiating methionine residue [28]. This modification has been shown to serve several functions that include increasing protein stability [29], modulating protein/protein interactions [30,31] and regulating protein/membrane association [32,33]. Previous electrospray MS studies have revealed unexpected heterogeneity of N-terminal fatty-acyl groups on recoverin and the α subunit of transducin (G_{α}) purified from bovine retina [34–36]. These two phototransduction proteins were modified primarily with the mono-unsaturated $C_{14:1}$ fatty acid, and to a lesser extent with equal amounts of the $C_{14:2}$, $C_{12:0}$ and $C_{14:0}$ fatty acids. It has been suggested that heterogeneity of acylation at the N-terminus is a mechanism for fine regulation of protein function by the fatty acid [35,36]. More recent evidence indicates that acyl-group heterogeneity is both species- and tissue-specific, but whether it is protein-specific remains to be determined [37].

We now report the assignment of GGBF as bovine ARF1 and GGBF* as bovine ARF3 through electrospray mass analysis of the intact proteins and extensive sequencing of tryptic fragments. Furthermore, LC/MS and LC/MS/MS analysis of the N-terminal tryptic fragments of these ARFs purified from bovine brain indicates N-terminal acylation by only the 14:0 fatty acid (myristate).

EXPERIMENTAL

Purification of ARFs/GGBFs

GGBFs were purified from a bovine brain post-nuclear supernatant as described previously [5]. Chromatographically purified GGBF and GGBF* were prepared for MS, LC/MS or proteolytic digestion by reverse-phase HPLC using a Vydac narrow-bore C_{18} column (2.1 mm $\times$ 250 mm). Buffer A (6 mM HCl) and Buffer B (90% acetonitrile/6 mM HCl) were filtered through 0.5 μ m-pore type FH filters (Millipore) before use. GGBFs were eluted with a 40–80% B gradient (1% B/min) at a flow rate of 0.3 ml/min. Protein in the eluate was measured at 215 nm on a Dynamax UV-D detector (Rainin), and the GGBF or GGBF* peaks were collected manually. Typical gradient elution conditions for GGBF and GGBF* were 70% B (63% apparent acetonitrile) and 72% B (65% apparent acetonitrile) respectively.

Preparation of GGBF and GGBF* proteolytic digests for peptide sequencing

All reactions were performed in Titer Tubes (Bio-Rad), and all stock solutions were stored in polypropylene tubes washed with formic acid to remove residual polymers and plasticizing agents. The HPLC-purified GGBF or GGBF* was prepared for proteo-

lytic digestion by evaporation of residual acetonitrile under a nitrogen stream. Initial difficulties in proteolytic digestion of GGBF and GGBF* by trypsin or a mixture of trypsin and endoproteinase LysC were alleviated by carrying out the digestion in the presence of 1.6 M urea (UltraPure, Fisher). Tris (pH 8) and $CaCl_2$ were also added to final concentrations of 200 mM and 2 mM respectively. Proteolytic digestion of GGBFs was initiated by addition of sequencing-grade trypsin (Gibco) (GGBF and GGBF*) and Endoproteinase LysC (Wako, Richmond, VA, U.S.A.) (GGBF) to 10% (w/w) and 5% (w/w) relative to the GGBF protein in the sample. After 2 h at 37 °C, the reaction was quenched by addition of formic acid (final concn. 5%, v/v). Samples were stored on ice for up to several hours before electrospray analysis.

MS analysis

Mass determinations were carried out with an API-III triple-quadrupole mass spectrometer (Sciex, Toronto, Canada) equipped with a nebulization-assisted electrospray source (ionspray). The ionspray needle was kept at a potential of 4.6 kV with a nitrogen curtain gas flow rate of 1.0 litre/min that served to desolvate ions entering the spectrometer. In some cases the sample was directly injected into the mass spectrometer via an infusion pump (model 5511; Harvard, S. Natick, MA, U.S.A.) at a flow rate of 3–5 μ l/min. For separation of peptides or proteins before MS or tandem MS LC/MS or LC/MS/MS, a liquid chromatograph (model 140A; Applied Biosystems/Bodman, Ashton, PA, U.S.A.) was directly coupled to the mass spectrometer.

LC/MS analysis of intact protein or proteolytic digests was carried out on a 500 μ m $\times$ 20 cm capillary column packed with POROS 120R resin (PerSeptive Systems, Cambridge, MA, U.S.A.) equilibrated with Buffer A (0.1% formic acid) and peptides or proteins were eluted with Buffer B (80% acetonitrile/0.1% formic acid) with a 1–2% B/min gradient at a flow rate of 20 μ l/min. Mass analysis was conducted on-line as the peptides were eluted from the capillary column. Hydrophilic peptides from the break-through volume of the POROS column were re-analysed on a Vydac C_{18} capillary column (500 μ m $\times$ 15 cm) under the same chromatography conditions as for the POROS. Capillary columns were packed by the method of Davis and Lee [38], except that Teflon filter paper was used as a frit.

Mass spectra (LC/MS) of HPLC-purified GGBF (2 pmol/ μ l, 40 pmol total) were obtained by using the chromatographic conditions described above. Spectra were acquired at unit resolution, scanning from m/z values of 750 to 1500 Da/e with a dwell time of 0.65 ms and a 0.1 Da/e step size (9.8 s/scan) using an orifice voltage of 90 V. Data represent the average of nine scans taken during elution of the protein.

Mass spectra (MS) of HPLC-purified GGBF* (10 pmol/ μ l) were obtained by direct injection via a Harvard infusion pump at 5 μ l/min. Spectra were acquired at unit resolution, scanning from m/z values of 600 to 1500 Da/e with a dwell time of 1.0 ms and a 0.1 Da/e step size, using an orifice voltage of 100 V. Data represent the sum of 12 scans and approx. 100 pmol of protein.

The trypsin/Endoproteinase LysC digest of approx. 200 pmol of GGBF was analysed by LC/MS as described above. All other settings were identical with those used for MS analysis. Spectra for the initial identification of peptides were acquired from m/z values of 380 to 1550 Da/e with a dwell time of 0.80 ms and a 0.2 Da/e step size (5.9 s/scan) using an orifice voltage of 70 V. Spectra used for subsequent sequence analysis were acquired from m/z values of 140 to 1570 Da/e with a dwell time of 0.65 ms and a 0.2 Da/e step size (6.0 s/scan), raising the orifice

Figure 1 Determination of the masses of GGBF and GGBF* by electrospray MS

The LC/MS spectrum of GGBF (a) and the MS spectrum of GGBF* (c) were processed with the hypermass module of the program MassSpec 3.2 to generate a deconvoluted mass spectrum in the ARF mass region (20–21 kDa) for both proteins. The deconvoluted mass spectrum for GGBF (b) contains a single symmetrical peak centred at 20774 Da. The deconvoluted mass spectrum for GGBF* (d) indicates a single prominent peak centred at 20678 Da, with a shoulder due to ion adducts trailing up to 20800 Da. Further examination of this data set (results not shown) did not identify any other components.

voltage from 75 V to 95 V to induce fragmentation upon collision with the nitrogen curtain gas. Analysis of the spectra collected during, before and after elution of each peptide verified that all fragment ions attributed to a given peptide were coincident with the parent ion.

Alternatively, for the analysis of GGBF* tryptic digests, a 2 mm-bore Hewlett–Packard C₁₈ HPLC column was used with post-column splitting, which directed approx. 10–15% of the effluent to the mass spectrometer. The remaining effluent was analysed by an ABI model 785A UV detector, and fractions were collected manually. The column was equilibrated in Buffer A (0.04% trifluoroacetic acid) and peptides were eluted with Buffer B (80% acetonitrile/0.04% trifluoroacetic acid) using a 2% B/min gradient at a flow rate of 200 μ l/min. The collected fractions were freeze-dried, and resuspended in 25 or 50% methanol containing 0.1% formic acid for subsequent analysis by MS/MS. The tryptic digest of approx. 250–500 pmol of GGBF* was initially analysed by LC/MS, fractionated, and subjected to re-analysis by MS/MS. MS/MS spectra were obtained by using either an API III high-pressure collision cell with an argon-gas thickness of 3.5×10^{14} atoms/cm³ and an ion kinetic energy of 28 V, or the older API III collision cell at 4×10^{14} atoms/cm³ gas thickness and an ion kinetic energy of 25 V. For MS/MS analysis, the resolution of the first quadrupole was set to allow a 2–3 Da window around the mass of the selected ion, and the resolution of the third quadrupole was

decreased below unit resolution to increase sensitivity. The data acquisition conditions used during MS/MS were optimized for each peptide fragment ion studied.

Data collection and analysis were accomplished by using programs provided with the Sciex instrument (Tune 2.4, MacSpec 3.2, MacBioSpec 1.0.1, and PeptideSeq), except that MS and MS/MS figures were prepared by importing data into Cricket Graph or Deneba Artworks for the Macintosh.

RESULTS

GGBF and GGBF* were purified to apparent homogeneity by standard chromatography as described previously [5], then prepared for electrospray MS by reversed-phase HPLC as described in the Experimental section. The HPLC purification indicated that purified GGBF* was eluted as a single A_{215} peak, whereas GGBF was eluted as two peaks, the earlier peak representing an 11% contamination with GGBF* (results not shown; see [5]). In all preparations examined to date ($n > 4$), GGBF and GGBF* were eluted as distinct symmetrical peaks whose elution in the gradient did not vary between preparations.

Mass determinations for GGBF and GGBF*

Samples of HPLC-purified GGBF were subjected to LC/MS

Table 1 Observed masses of GGBFs isolated from bovine brain, and predicted masses of the known mammalian ARFs and ARLs

Protein	Observed mass of the protein purified from bovine brain (Da) ^a		
GGBF	20774 ± 3		
GGBF*	20678 ± 3		
Mass predicted from cDNA sequence (Da) ^{b,c}			
Protein	Human	Rat	Bovine
ARF1	20777	20777	20777
ARF2	unk.	20826	20826
ARF3	20681	20681	unk.
ARF4	20591	20476	unk.
ARF5	20610	20610	unk.
ARF6	20162	20162	unk.
ARL1	unk.	20491	unk.
ARL2	20916	unk.	unk.
ARL3	20535	20534	unk.
ARL4	unk.	20668	unk.

^a Masses were calculated from the spectra presented in Figures 1(a) and 1(c) by the hypermass module of the MassSpec 3.1 program. Values were rounded off to the nearest integer.

^b Sequences from the following GENBANK files were used to calculate the predicted protein masses using the program MacBioSpec v1.0 (b = bovine, h = human, r = rat): bARF1 (J03275), bARF2 (J03794), hARF1 (M36340, M25203), hARF3 (M33384, M25204), hARF4 (M36341, M31890), hARF5 (M57567), hARF6 (M57763), hARL2 (L13687), hARL3 (U07151), rARF1 (L12380), rARF2 (L12381), rARF3 (L12382), rARF4 (L12383), rARF5 (L12384), rARF6 (L12385), rARL1 (X76920), rARL3 (X76921), rARL4 (X77235). Masses represent the average mass of the protein, and are rounded off to the nearest integer. Abbreviation: unk. indicates that sequence data are not currently available for this protein.

^c Predicted masses assume cleavage of the N-terminal methionine and modification of the N-terminus by myristate.

Figure 2 Tryptic peptide maps of mammalian ARF1 and ARF3

Amino acid sequences of bovine ARF1 and human ARF3 were obtained from GENBANK (accession nos. J03275 and M33384 respectively). Dots indicate identity between sequences. Boxes above the aligned sequences depict the peptides generated by tryptic digestion, with the darkened boxes representing peptides unique between the two ARF sequences. Abbreviation: Myr, myristoyl.

using a capillary column to concentrate the sample for mass analysis. Upon elution, a single peak of ion current was observed; nine scans were averaged across this peak to produce the spectrum in Figure 1(a). Deconvolution of the series of multiply charged protein ions in this spectrum indicated the presence of a single species with a mass of 20774 Da, with a S.D. of 3 Da (Figure 1b).

This mass is consistent with the mass of myristoylated ARF1 predicted from human, rat and bovine cDNA sequences (Table 1). Similar results (not shown) were obtained by using several samples obtained from four different ARF preparations, yielding a mean molecular mass of 20775 Da ($n = 8$), with a S.E.M. of 2 Da.

Samples of HPLC-purified GGBF* were sufficiently concentrated for mass analysis by direct infusion (Figure 1c). Deconvolution of this spectrum indicated the presence of a single species with an average compound mass of 20678 Da, with a S.D. of 3 Da (Figure 1d). This mass is consistent with the mass of myristoylated human and rat ARF3 (Table 1). Consistent results (not shown) were obtained with three different ARF preparations, yielding a mean molecular mass of 20680 Da ($n = 4$), with a S.E.M. of 2 Da. The decreased signal-to-noise ratio in the MS spectrum in Figure 1(c) most likely results from the presence of ion adducts, as suggested by the trailing shoulder on the deconvoluted peak in Figure 1(d).

Sequencing of GGBF and GGBF* by MS analysis of tryptic fragments

Observation of masses similar to those predicted from the cDNA could arise from mass-neutral amino acid substitutions in the sequence, and therefore does not provide definitive identification. In order to confirm that GGBF and GGBF* are ARF1 and ARF3 respectively, LC/MS and LC/MS/MS sequencing of tryptic peptides was undertaken. Although Group I ARFs share 95% identity, three of the predicted tryptic peptides out of 21 (ARF1) or 22 (ARF3) are unique between ARF1 and ARF3 (Figure 2) and thus provide a method of distinguishing these two closely related proteins. Two of these tryptic peptides (Tp1, Tp2) are unique among all known ARFs, and the third peptide (Tp20) can be used to distinguish ARF1 from all ARFs except ARF2.

Peptides of GGBF were resolved on a POROS (a C_4 equivalent) column. GGBF peptides were first identified by comparing observed masses with those predicted for proteolytic fragments of ARF1 (Table 2a). Sequence information on GGBF peptides was then obtained on a subsequent LC/MS run employing a higher orifice voltage to promote collisional activation of peptides with the nitrogen curtain gas (see footnote d, Table 2). Some of the ions that fragmented poorly under these conditions were further analysed by LC/MS/MS; in those cases, a sequence score is given in Table 2(a).

GGBF peptides identified by mass account for 93% of the total ARF1 sequence. The remaining 7% of the ARF sequence are peptides of 4 residues or less that were poorly retained on the capillary columns used during mass analysis. Peptides identified with extensive sequence information cover 72% of the total ARF1 sequence. More importantly, the two ARF1 unique peptides, Tp1 and Tp2, were identified by mass, and a full set of sequence information for the N-terminal peptide was obtained. The Tp20 peptide, shared only by ARF1 and ARF2 sequences, was also identified by mass, but its large size and low recovery precluded further analysis by MS/MS. This combination of electrospray mass analysis of the intact protein and sequence data gathered from tryptic fragments unambiguously confirms the assignment of GGBF as bovine ARF1.

Analysis of trypsin-treated GGBF* was completed by LC/MS and LC/MS/MS as described in the Experimental section. GGBF* peptides were identified by comparing the observed masses with those determined for proteolytic fragments of human ARF3 (Table 2b). The sequence predicted from the human ARF3 cDNA was used for this comparison, because no equivalent bovine cDNA has yet been obtained. All predicted

Table 2 Masses of tryptic peptides (a) of bovine ARF1 observed by LC/MS and LC/MS/MS, and (b) of bovine ARF3 observed by LC/MS and MS/MS, compared with calculated masses, and identity of the major peptide fragments that were observed

Tryptic peptide number ^a	Predicted mass (Da) ^b	Observed mass (Da) ^c	Peptide fragments observed ^d	Sequence score (%) ^e
(a) ARF1				
Tp1 ^{1*}	1233.6	1233.1	b1–8, y2–6	98.4
Tp2*	520.3	520.5	Mass only	
Tp3	146.1	n.s.		
Tp4	434.2	n.d.		
Tp5	1086.6	1086.8	b2–4, y4–11	
Tp6	737.4	737.6	b2–4, y1–6	
Tp7	259.0	n.s.		
Tp8	2323.7	2323.8	b4, 5, 7, 8 y3, 6, 7, 13, 14	
Tp9	1565.7	1565.3	b2, b6–9, y4–14	
Tp10	839.5	839.8	b2, 4, 5, y1–6	
Tp11	2153.4	2153.3	b5–10, y4–10	
Tp12	303.2	n.s.		
Tp13	587.3	586.9	Mass only	
Tp14	676.3	676.4	b1–3, b5, y1–5	
Tp15	975.5	975.8	b2, b4, y1–8	
Tp16	1088.6	1089.1	b2–7, y1–10	
Tp17	1630.8	1630.3	b2, 8, 13, y3–15	
Tp18	794.5	794.7	b6, b7, y1–7	
Tp19	311.2	n.s.		
Tp20*	3075.4	3075.2	Mass only	
Tp21	388.2	n.s.		85.1
(b) ARF3				
Tp1 ^{1*}	1184.8	1184.9	b0–b4, b7, y5–7	88.6
Tp2*	516.3	516.4 ^g	b2, b4, y1–5	98.9
Tp3	147.1	n.s.		
Tp4	434.2	434.2 ^g	Mass only	
Tp5	1086.6	1086.6	b2, y7–9	
Tp6	737.4	737.4	b2, b4, y1–5	
Tp7	259.0	n.s.		94.5
Tp8	2323.7	2323.2	b3–5, b7, b8, b11 y5, y10, y13–16	87.2
Tp9	1565.7	1564.9	b2, b3, b6, y4–12	97.0
Tp10	839.5	839.4	b2–5, y1–y4	94.3
Tp11	2153.4	2152.6	b1–3, y4, y17(2 +)	94.2
Tp12	303.2	n.s.		
Tp13	587.3	587.4 ^g	Mass only	
Tp14	676.3	676.1	y1–5, b2, b3	
Tp15	975.5	975.1	y1–6, b2, b5	
Tp16	1088.6	1088.2	Mass only	
Tp17	1630.8	1630.7	b2, b3, y8–10, y12–14(2 +)	
Tp18	794.5	794.4	y1, y3–6, b2, b4	
Tp19	311.2	n.s.		
Tp20*	3029.4	3029.6	Mass only	
Tp21	260.2	n.s.		
Tp22	147.1	n.s.		

^a Tryptic peptides are labelled as indicated in Figure 2 from the N-terminus (Tp1) to the C-terminus (Tp21 for ARF1 and Tp22 for ARF3) of each ARF protein. Peptides marked with a * contain sequence information that can be used to distinguish between ARF1 and ARF3.

^b Masses of peptides were calculated from the amino acid sequence using the program MacBioSpec v1.0. Masses shown are the monoisotopic mass for peptides smaller than 1200 Da and the average mass for peptides above 1200 Da.

^c Reported mass is taken from the LC/MS analysis of a peptide, and is the average of the calculated masses for each charged form of the parent peptide ion. Those masses obtained from the LC/MS/MS of a peptide are indicated by the presence of a sequence score for that peptide. Abbreviations: n.s., not scanned; n.d., not detected.

^d Fragment ions produced by collisional activation by the nitrogen curtain gas (LC/MS) or in the high-pressure collision cell (LC/MS/MS) are reported. Ions reported are the 'y' and 'b' ions (in the nomenclature system of Biemann [26]), and result from a single cleavage between the carboxyl carbon and nitrogen of the amide bond. 'a' ions, 'c' ions, and ions resulting from

ARF3 peptides larger than 3 residues, accounting for 94 % of the total human ARF3 sequence, were identified by mass. Of the three peptides unique to ARF3, Tp1 and Tp2 were identified by mass and sequenced, whereas Tp20 was observed solely by mass, for the reasons explained above. Overall, GGBF* peptides identified with sequence information cover nearly 70 % of ARF3 and show complete identity with those predicted for the human and rat proteins. This electrospray mass analysis not only confirms the assignment of GGBF* as bovine ARF3, but provides novel and extensive sequence data that suggest that the sequence of bovine ARF3 is identical with that previously determined for the mammalian ARF3s listed in Table 1.

Analysis of the N-terminus of bovine brain ARF1 and ARF3

The greatest sequence diversity among ARFs exists within the fatty-acylated N-terminal peptide, blocked to traditional gas-phase sequencing methods. Additional heterogeneity may be contributed by the lipid moiety, as described for two other myristoylated proteins [34–36]. Analysis by MS not only allows sequence determination of acylated peptides, but can also provide information about the nature of the blocking group itself.

The N-terminal tryptic peptide of ARF1 (Tp1, $MH^+ = 1234.6$ Da/e) was analysed by LC/MS/MS (Figure 3a). The blocked N-terminus necessitated working with a singly charged form of this peptide, resulting in a low level of fragment ions. However, collisional activation led to cleavage between all peptide bonds, resulting in a full set of b ions and a partial set of y ions (y2–y4 and y6). Additional ions resulting from secondary cleavages (internal fragments, neutral loss of H_2O or NH_3) were also observed, but are not labelled in Figure 3. The sequence score for this peptide indicates that 98.4 % of the total ion current was due to predicted mass ions (Table 2a).

MS/MS of Tp1 obtained from ARF3 ($MH^+ = 1185.8$ Da/e) by collision-induced dissociation with argon produced few fragment ions (results not shown). Cleavage of this peptide was achieved by raising the orifice voltage during LC/MS analysis (Figure 3b). The resulting spectrum showed overlapping sets of y ions (y5–y8) and b ions (b1–b4). Ions resulting from secondary cleavages yielding internal fragment ions and neutral loss of H_2O or NH_3 were also observed. The sequence score for this peptide indicates that the predicted sequence accounted for 88.6 % of the total fragment ion current (Table 2b). An ion consistent with the mass calculated for the N-terminal myristoyl group ($MH^+ = 212.2$ Da/e) was also observed.

In order to determine if the N-terminus was modified with acyl groups varying in length or degree of saturation, a computer program that searches the data for predicted ion masses was utilized. This program produces chromatograms of the ion current observed for each predicted ion mass as a function of scan number during the LC/MS analysis, a measure comparable with LC elution time. Ion chromatograms for the parent ions expected for the singly charged form of Tp1 for ARF1 (Figure 4a) and ARF3 (Figure 4b), modified with the $C_{14:0}$ (myristate), $C_{14:1}$, $C_{14:2}$ or $C_{12:0}$ fatty acid, were extracted from the LC/MS

secondary cleavages, including the loss of water or ammonia and internal fragment ions are not reported. Ions observed in the doubly charged form [referred to as (2+)] correspond to the $(y+1H)^{2+}$ or the $(b+1H)^{2+}$ in the Biemann nomenclature.

^e The sequence score represents the percentage of the total product ion current that is attributable to the fragment ions predicted for the fragmentation of a given peptide.

^f The mass calculated for the N-terminal peptide assumes cleavage of the N-terminal methionine and addition of myristate at the penultimate glycine.

^g Peptide was detected using C_{18} but not C_4 reversed-phase chromatography.

Figure 3 Sequence analysis of Tp1 (N-terminal peptide) of ARF1 and ARF3 from bovine brain

Fragmentation of a peptide at each amide linkage produces a sequence-specific set of ions incremented by the mass of the next amino acid. In the nomenclature of Biemann [26], 'b' ions represent the N-terminal fragment, and 'y' ions represent the C-terminal fragment produced by this type of cleavage. (a) LC/MS/MS of the singly charged ARF1 tryptic peptide 1 ($MH^+ = 1234.6$ Da/e) provides a complete series of singly charged b ions (b1–b8), and a partial set of y ions (y2–y4, y6), consistent with the published sequence of mammalian ARF1 and also consistent with N-terminal modification with myristate. (b) LC/MS analysis of the singly charged ARF3 tryptic peptide ($MH^+ = 1234.6$ Da/e) reveals an incomplete set of singly charged b ions (b1–b4), and a partial set of y ions (y5–y8). The identified ions are consistent with the published sequence of mammalian ARF3, and cleavage is observed at all positions, except between leucine and lysine at the C-terminus. Also labelled are the doubly charged parent ion and the ion representing the N-terminal modifying myristate (predicted $m/z = 212.2$ Da/e). In both panels, the fragment-ion intensity is normalized relative to the intensity of the singly charged parent ion. For clarity, ions including internal fragment ions, 'a' type ions, and ions representing the loss of ammonia or water were not labelled. Abbreviation: Myr, myristate.

data sets for tryptic digests of ARF1 and ARF3. The $C_{14:0}$ -modified Tp1 peptide from each ARF appears on these ion chromatograms as a single peak eluted late during the LC/MS run (peptides modified with fatty acids are typically eluted at acetonitrile concentrations greater than 50%). On the other hand, ion chromatograms resulting from a search for the m/z ions of Tp1 that would be expected from previously reported lipid heterogeneity show no signal above background. Further examination of the LC/MS data for Tp1 ions modified with acyl moieties other than those previously reported proved similarly negative. Finally, LC/MS analysis of a partially purified ARF preparation subjected only to anion-exchange and size-exclusion chromatography showed no evidence of ARF heterogeneity (results not shown).

DISCUSSION

The two proteins from bovine brain shown to be required for the GTP[S] inhibition of a well-characterized intra-Golgi transport assay, termed GGBF and GGBF*, have been definitively identified by electrospray MS as bovine ARF1 and ARF3, respectively. This identification is based not only on mass determination of

the intact protein, but also on extensive sequence information obtained from tryptic peptides. In particular, the unique N-terminal peptide of both proteins, inaccessible to traditional gas-phase sequencing methods, has been sequenced by electrospray MS. Previous purification of soluble bovine brain ARFs on the basis of their cofactor activity towards cholera toxin yielded two proteins that were also identified as Group I ARFs [39]. However, this identification was tentative, as it was largely based on partial sequence of a single 10-residue internal peptide of each protein that did not distinguish between ARF1 and ARF2. The extensive sequence information reported here provides definitive assignment of the two GGBF activities as ARF1 and ARF3. Taken together, these data indicate that ARF1 and ARF3 are the two major ARFs present in soluble brain extracts.

Previously, we have shown that GGBF/ARF1, the less abundant GGBF species isolated from bovine brain extracts, had 3–5-fold greater specific activity in the transport inhibition assay than GGBF*/ARF3 [5]. This observation is now particularly intriguing, because both proteins are identified as Group I ARFs that are 96% identical at the amino acid level. It is nevertheless consistent with the demonstration that, in the presence of GTP[S], ARF1 associates with Golgi-enriched membranes to a greater

Figure 4 Ion chromatograms for the N-terminus of bovine ARF1 and ARF3 as modified by myristate and other possible N-terminal fatty acid modifications

Trypsin/Endoproteinase LysC digests of ARF1 (a) and ARF3 (b) were analysed by LC/MS as described in the Experimental section. The expected m/z for the $C_{14:1}$ -modified peptide, the $C_{14:2}$ -modified peptide and the $C_{12:0}$ -modified peptide were calculated as the mass of the $C_{14:0}$ -modified peptide by subtracting 2, 4 or 28 Da respectively. The mass ion signal was extracted by using a 0.5 Da window on each side of the calculated m/z for each peptide mass ion, and plotted as ion signal versus scan number. The peak of ion current from both Tp1- $C_{14:0}$ ions was eluted over 10 LC/MS scans, or approx. 59 s (ARF1) or 68 s (ARF3) of LC elution time. The apparent difference in peak widths simply reflects compression of the chromatogram resulting from the additional data scans collected during the ARF3 LC/MS analysis.

extent than ARF3 [39]. These differences in activity cannot be due to acylation by different populations of lipid moieties, since we have shown here that the N-terminus of both proteins is modified solely with the $C_{14:0}$ fatty acid (myristate). Therefore, the differences in specific activity may be due to factors that potentiate or inhibit one of the ARFs, or could be attributable to differential affinities with putative receptors for ARFs on Golgi membranes.

Early attempts to obtain N-terminal sequence of purified ARFs revealed that these proteins were blocked to Edman degradation [21,22]. Analysis of bovine ARF hydrolysates to identify the blocking group revealed the presence of a peak that co-migrated on reverse-phase chromatography with a myristoylglycine standard [22]. Heterogeneity at the N-terminus of acylated proteins was only recently described [34–36], and therefore was not addressed in the earlier work. Analysis of the N-terminus of bovine ARF1 and ARF3 reported here demonstrates modification solely by the $C_{14:0}$ fatty acid. The absence of heterogeneity is not likely to result from selective purification of the $C_{14:0}$ -modified protein, since the chromatographic methods used for purification cannot even resolve myristoylated from non-myri-

stoylated recombinant ARF [40]. Furthermore, lack of heterogeneous fatty acylation was also confirmed by electrospray MS analysis of crude ARF preparations obtained at intermediate steps of purification (results not shown).

Our work with brain ARFs clearly demonstrates that heterogeneity of N-terminal acylation observed on several retinal proteins is not universal. This finding is in agreement with a recent report that cyclic-AMP-dependent protein kinase purified from bovine heart and bovine brain is homogeneously myristoylated, whereas the retinal form is heterogeneously modified [36,37]. Two possible mechanisms could yield tissue-specific heterogeneity in acylation. First, it could arise from tissue-specific expression of N-myristoyltransferase (NMT) isoforms; multiple isoforms of NMT have been previously identified in mammalian cell extracts, but the lipid substrate specificity of these isoforms has not been compared [41,42]. Second, heterogeneity could simply result from an unusual composition of the pool of activated fatty acids available to retinal NMT(s). An alternative explanation is that the N-terminal acyl heterogeneity is protein-specific, and could be restricted to a subset of myristoylated proteins. For example, analysis of four myristoylated proteins from bovine retina has revealed significant differences in the relative amounts of acyl groups attached to these heterogeneously acylated proteins [37]. Further insights into the extent and functional significance of fatty acyl heterogeneity may be provided by cross-tissue analysis of several myristoylated proteins by electrospray MS. The ARF protein family is ideal for such a study, since ARFs are soluble, and are abundantly expressed in all tissues of eukaryotic organisms studied to date [22,43,44]. Furthermore, the small size, ease of purification, and reverse-phase-HPLC behaviour of ARFs facilitate their analysis by LC/MS, even in complex mixtures.

Electrospray MS is fast becoming an established technique for the analysis of proteins and their modifications. As presented in this study, this method yields accurate mass determination of proteins as well as sequence information from complex mixtures of proteolytic peptides. Thus electrospray MS is ideally suited to identify proteins, even within closely related families. These capabilities, coupled with the explosive growth of nucleotide sequence information, is likely, in the next decade, to make electrospray MS a powerful tool for the characterization and identification of proteins and their modifications.

We thank Dr. Kenneth Walsh (University of Washington, Seattle), as well as Dr. Martin McDermott and John Hunt (Synergen Inc., Boulder, CO, U.S.A.) for providing access to their respective MS facilities. Additional thanks go out to Dr. Natalie Ahn for providing expert advice on the preparation of ARF tryptic digests, and to Dr. Richard Johnson for many helpful discussions. Dr. Robert Kutchtla and Jian-Ping Yan are also acknowledged for their critical reading of the manuscript. This work was supported in part by NIH grants GM43378 (P.M.) and AR39730 (K.A.R.), and an NIH instrument grant RR05543 provided to Dr. Ken Walsh (University of Washington).

REFERENCES

- Kahn, R. A. and Gilman, A. G. (1984) *J. Biol. Chem.* **259**, 6228–6234
- Donaldson, J. G. and Klausner, R. D. (1994) *Curr. Opin. Cell Biol.* **6**, 527–532
- Rothman, J. E. (1994) *Nature (London)* **372**, 55–63
- Balch, W. E. and Rothman, J. E. (1985) *Arch. Biochem. Biophys.* **240**, 413–425
- Taylor, T. C., Kahn, R. A. and Melancon, P. (1992) *Cell* **70**, 69–79
- Taylor, T. C., Kanstein, M., Weidman, P. and Melancon, P. (1994) *Mol. Biol. Cell* **5**, 237–252
- Dascher, C. and Balch, W. E. (1994) *J. Biol. Chem.* **269**, 1437–1448
- Stearns, T., Willingham, M. C., Botstein, D. and Kahn, R. A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1238–1242
- Lenhard, J. M., Kahn, R. A. and Stahl, P. D. (1992) *J. Biol. Chem.* **267**, 13047–13052
- Boman, A. L., Taylor, T. C., Melancon, P. and Wilson, K. L. (1992) *Nature (London)* **358**, 512–514

- 11 Ostermann, J., Orci, L., Tani, K., Amherdt, M., Ravazzola, M., Elazar, Z. and Rothman, J. E. (1993) *Cell* **75**, 1015–1025
- 12 Palmer, D. J., Helms, J. B., Beckers, C. J., Orci, L. and Rothman, J. E. (1993) *J. Biol. Chem.* **268**, 12083–12089
- 13 Donaldson, J. G., Cassel, D., Kahn, R. A. and Klausner, R. D. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6408–6412
- 14 Stamnes, M. A. and Rothman, J. E. (1993) *Cell* **73**, 999–1005
- 15 Traub, L. M., Ostrom, J. A. and Kornfeld, S. (1993) *J. Cell Biol.* **123**, 561–573
- 16 Narula, N., McMorrow, I., Plopper, G., Doherty, J., Matlin, K. S., Burke, B. and Stow, J. L. (1992) *J. Cell Biol.* **117**, 27–38
- 17 Cockcroft, S., Thomas, G. M. H., Fensome, A. et al. (1994) *Science* **263**, 523–526
- 18 Kahn, R. A., Yucel, J. K. and Malhotra, V. (1993) *Cell* **75**, 1045–1048
- 19 Brown, H. A., Gutowski, S., Moomaw, C. R., Slaughter, C. and Sternweis, P. C. (1993) *Cell* **75**, 1137–1144
- 20 Serventi, I. M., Moss, J. and Vaughan, M. (1992) *Curr. Top. Microbiol. Immunol.* **175**, 43–67
- 21 Price, S. R., Nightingale, M., Tsai, S. C. et al. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5488–5491
- 22 Kahn, R. A., Goddard, C. and Newkirk, M. (1988) *J. Biol. Chem.* **263**, 8282–8287
- 23 Tamkun, J. W., Kahn, R. A., Kissinger, M., Brizuela, B. J., Rulka, C., Scott, M. P. and Kennison, J. A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3120–3124
- 24 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989) *Science* **246**, 64–71
- 25 Biemann, K. (1990) *Methods Enzymol.* **193**, 455–479
- 26 Biemann, K. (1990) *Methods Enzymol.* **193**, 886–887
- 27 Hunt, D. F., Yates, J. R. III, Shabanowitz, J., Winston, S. and Hauer, C. R. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233–6237
- 28 Wilcox, C., Hu, J. and Olson, E. N. (1987) *Science* **238**, 1275–1278
- 29 Yonemoto, W., McGlone, M. L. and Taylor, S. S. (1993) *J. Biol. Chem.* **268**, 2348–2352
- 30 Kataoka, M., Mihara, K. and Tokunaga, F. (1993) *J. Biochem. (Tokyo)* **114**, 535–540
- 31 Linder, M. E., Pang, I., Duronio, R. J., Gordon, J. I., Sternweis, P. C. and Gilman, A. G. (1991) *J. Biol. Chem.* **266**, 4654–4659
- 32 Haun, R. S., Tsai, S.-C., Adamik, R., Moss, J. and Vaughn, M. (1993) *J. Biol. Chem.* **268**, 1–5
- 33 Jones, T. L. Z., Simonds, W. F., Merendino, J. J., Brann, M. R. and Spiegel, A. M. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 568–572
- 34 Dizhoor, A. M., Ericsson, L. H., Johnson, R. S. et al. (1992) *J. Biol. Chem.* **267**, 16033–16036
- 35 Kokame, K., Fukada, Y., Yoshizawa, T., Takao, T. and Shimonishi, Y. (1992) *Nature (London)* **359**, 749–752
- 36 Neubert, T. A., Johnson, R. S., Hurley, J. B. and Walsh, K. A. (1992) *J. Biol. Chem.* **267**, 18274–18277
- 37 Johnson, R. J., Ohguro, H., Palczewski, K., Hurley, J. B., Walsh, K. A. and Neubert, T. A. (1994) *J. Biol. Chem.* **269**, 21067–21071
- 38 Davis, M. T. and Lee, T. D. (1992) *Protein Sci.* **1**, 935–944
- 39 Tsai, S.-C., Adamik, R., Haun, R. S., Moss, J. and Vaughn, M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9272–9276
- 40 Randazzo, P. A., Weiss, O. and Kahn, R. A. (1992) *Methods Enzymol.* **219**, 362–369
- 41 King, M. J. and Sharma, R. K. (1992) *Mol. Cell. Biochem.* **113**, 77–81
- 42 Boutin, J. A., Gilles, F., Ernould, A., Maes, P., Remond, G. and Vincent, M. (1993) *Eur. J. Biochem.* **214**, 853–867
- 43 Tsai, S. C., Adamik, R., Tsuchiya, M., Chang, P. P., Moss, J. and Vaughan, M. (1991) *J. Biol. Chem.* **266**, 8213–8219
- 44 Tsuchiya, M., Price, S. R., Tsai, S. C., Moss, J. and Vaughan, M. (1991) *J. Biol. Chem.* **266**, 2772–2777