EBULIN-RP, A NOVEL MEMBER OF THE EBULIN GENE FAMILY WITH LOW CYTOTOXICITY AS A RESULT OF DEFICIENT SUGAR BINDING DOMAINS

Rosario Iglesias[a], J. Miguel Ferreras[a], Antimo Di Maro[b], Lucía Citores [a,*]

[a] Department of Biochemistry and Molecular Biology and Physiology, Faculty of Sciences, University of Valladolid, E-47011 Valladolid, Spain

[b] Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania "Luigi Vanvitelli", I-81100 Caserta, Italy

*Corresponding author. Phone: +34 983 185856; e-mail: luciac@bio.uva.es

Abstract

Background

Sambucus ebulus is a rich source of ribosome-inactivating proteins (RIPs) and RIP-related lectins generated from multiple genes. These proteins differ in their structure, enzymatic activity and sugar binding specificity.

Methods

We have purified and characterized ebulin-RP from *S. ebulus* leaves and determined the amino acid sequence by cDNA cloning. Cytotoxicity was studied in a variety of cancer cells and a comparative study of the ability of ebulin-RP to bind sugars using "in vitro" and "in silico" approaches was performed.

Results

Ebulin-RP is a novel heterodimeric type 2 RIP present in *S. ebulus* leaves together with the type 2 RIP ebulin l, which displayed rRNA N-glycosidase activity but unlike ebulin l, lacked functional sugar binding domains. As a consequence of changes in its B-chain, ebulin-RP displayed lower cytotoxicity than ebulin l towards cancer cells and induced apoptosis as the predominant pattern of cell death.

Conclusions

Ebulin-RP is a novel member of the ebulin gene family with low cytotoxicity as a result of deficient sugar binding domains. Type 2 RIP genes from *Sambucus* have evolved to render proteins with different sugar affinities that may be related to different biological activities and could result in an advantage for the plant.

General Significance

The ebulin family of RIPs and lectins can serve as a good model for studying the evolutionary process which may have occurred in RIPs. The lack of cytotoxicity of ebulin-RP makes it a good candidate as a toxic moiety in the construction of immunotoxins and conjugates directed against specific targets.

Keywords: *Sambucus ebulus* L.; Lectin; Ribosome-inactivating protein (RIP); Ricin; rRNA N-glycosidase; apoptosis

1. Introduction

Ribosome inactivating proteins (RIPs) are a family of proteins widely distributed among angiosperms although they have also been found in fungi, algae and bacteria [1, 2]. RIPs belong to a class of enzymes (EC 3.2.2.22) which exhibit rRNA N-glycosidase activity, which leads to the release of a specific adenine residue in the conserved sarcin-ricin loop (SRL) of the large rRNA responsible for the interaction of elongation factor with the ribosome, resulting in the irreversible inhibition of protein synthesis [3]. Moreover, some RIPs do not exclusively act on ribosomes but display polynucleotide:adenine glycosidase (PNAG) activity on different nucleic acid substrates [4]. Other enzymatic activities such as chitinase [5], DNase [6] and lipase [7] have also been attributed to RIPs.

From a structural point of view, RIPs have been classified into two types depending on the presence (type 2 RIPs) or the absence (type 1 RIPs) of a lectin chain. Thus type 1 RIPs consist of a single enzymatic active (A) chain, whereas type 2 RIPs are composed of an A chain similar to type 1 RIPs linked by a disulphide bond to a binding (B) chain with lectin activity. The carbohydrate-binding domains (1-alpha and 2-gamma) of the B chain recognize glycosylated receptors on the cell surface facilitating the entry of the A chain into the cell. Some type 2 RIPs such as ricin, abrin or volkensin are potent toxins to mammalian cells and animals whereas despite the presence of a B chain, other type 2 RIPs (non-toxic type 2 RIPs) display much less toxicity (e.g. nigrin, ebulin, cinnamonin). On the other hand, four-chain type 2 RIPs [1] and type 1 RIPs with an inner removable peptide [8, 9] have also been reported.

The exact biological role played by RIPs is yet unknown but it is thought to represent a defense mechanism of the plant against predators, fungi and viruses. RIPs also show toxicity towards animal cells targeting the host protein synthesis machinery. In mammalian cells, both type 1 and type 2 RIPs have been related to apoptosis. The mechanisms by which apoptosis is activated by a particular RIP may differ and may be independent of protein synthesis inhibition. Besides, some RIPs exhibit strong toxicity towards cancer cells and low toxicity towards normal cells and they impede or inhibit tumor growth mostly via apoptosis [10]. Therefore, RIPs either alone or as part of conjugates are good candidates for developing selective antiviral and anticancer agents [11-13]. Conjugates consist of a targeting portion such as an antibody, a lectin or a

growth factor linked to a toxic portion. RIPs have been used as the toxic portion in several conjugates that have been tested in experimental therapies against various malignancies [13-15]. In agriculture, RIPs have been shown to increase resistance against virus and other parasites in transgenic plants [16].

Most RIPs from plants have been found in a small number of families [2] and most of them are encoded by small multigene families. That is the case of saporin, ebulin, PAP or ricin gene family. Several species of the genus *Sambucus* (*S. nigra*, *S. ebulus*, *S. sieboldiana* and *S. racemosa*) have been extensively studied for the presence of RIPs and more than 40 RIPs and structurally related lectins have been isolated from them in the last few years. Non-toxic type 2 RIPs with specificity for galactose [17, 18], homodimeric and monomeric galactose specific lectins structurally related to type 2 RIPs [19-21], dimeric and tetrameric type 2 RIPs with specificity for sialic acid [22, 23], non-toxic type 2 RIPs lacking lectin activity [24] and non-toxic type 2 RIPs with affinity for N-acetyl-glucosamine oligomers [25] have been reported for the first time in the genus *Sambucus*. These proteins have been found in bark, leaves, fruits, seeds, rhizomes and blossoms and their presence in the different tissues is subject to seasonal and developmental variations [21, 26, 27]. All the type 2 RIPs found in *Sambucus*, from which ebulin l from *S. ebulus* and nigrin b from *S. nigra* are the most representative and studied members, are considered as non-toxic type 2 RIPs since they have the striking feature that despite being as toxic as ricin at the ribosomal level, they are several times less toxic to cultured cells and *in vivo* than ricin [28]. The reason for this difference seems to be the altered ability of the B chain to bind to cells that would affect the uptake and the intracellular destination. The structure of ebulin l has been resolved by X-ray diffraction analysis and the tertiary structure closely resembles that of ricin [29], however ebulin l has a lower affinity for galactose than ricin due to a change in the structure of the 2-gamma-subdomain of the ebulin B chain [29].

Sambucus ebulus is an herbaceous plant with a perennial underground stem rhizome that has been found to contain RIPs and lectins in several tissues. Leaves contain the type 2 RIP ebulin l (A-B type) [17]; two lectins, a monomeric lectin SELlm (B chain) and a dimeric lectin SELld (B-B type) [21] as well as type 1 RIPs (A chain) [30]. Fruits contain ebulin f (A-B type) and the lectin SELfd (B-B type) [31] and the rhizome, two type 2 RIPs, ebulin r1 and r2 (A-B type) , a tetrameric type 2 RIP SEA (A-B-B-A type)

and the monomeric lectin SEAII (B chain) [32, 33]. All the hololectins and B chains of ebulins are galactose specific except SEA that is specific for sialic acid.

While working with the leaves of *S. ebulus*, we detected that they contain in addition to the well characterized Gal/GalNAc-specific type 2 RIP ebulin l, another type 2 RIP named ebulin-RP which failed to bind on the affinity matrix AT (acid treated)-Sepharose 6B and to agglutinate human erythrocytes. In this work, we report the biochemical characterization and the enzymatic and cytotoxic activities of ebulin-RP. Although the new protein inhibited protein synthesis in a cell free system, it displayed even lower cytotoxicity than ebulin l and SEA, suggesting that the differential toxicity was related to changes in its B-chain. In this work we employed molecular docking to explore detailed interactions between sugars and carbohydrate-binding sites of the B-chains of type 2 RIPs from *S. ebulus*. Moreover, the analysis of the phylogenetic tree reveals that ebulin family can serve as a good model for the study of the evolutionary process which may have occurred in type 2 RIPs.

2. Materials and methods

2.1. Materials

Immobilon membranes were purchased from Millipore Ibérica (Madrid, Spain). Leaves from dwarf elder were harvested at Cobos de Cerrato (Palencia, Spain) in early summer. AmpliTaq DNA polymerase was from Applied Biosystems. CM-Sepharose, Sepharose 6B and Superdex-75 HiLoad 26/60 columns were purchased from GE Healthcare (Barcelona, Spain). Ricin was from Sigma (St Louis, MO, USA) and SEA was purified as described [32].

2.2. Cell lines and culture

COLO 320 (human colon carcinoma), HCT15 (human colon adenocarcinoma), HeLa (human cervix epitheloid carcinoma) and B16 Mel 4A5 (mouse melanoma) cells, were obtained from the European Culture Collection (ECACC) and grown in RPMI 1640 medium (GIBCO BRL, Barcelona, Spain) supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin and 0.1 mg/mL streptomycin under 5% CO_2 at 37^{o}C.

Human mesenchymal stem cells from human adipose tissue, (hMSCs) were obtained from GIBCO and cultured in DMEM supplemented with 10% FBS, 100 U/mL penicillin and 0.1 mg/mL streptomycin under 5% CO_2 at 37 ºC. Raji (Burkitt's lymphoma) cells, kindly provided by Dr A. Bolognesi (University of Bologna, Italy) were grown in RPMI 1640 supplemented with 10% FBS, 100 U/mL penicillin and 0.1 mg/mL streptomycin under 5% CO_2 at 37 ^{o}C.

2.3. Purification of ebulin-RP from Sambucus ebulus leaves

140 g of *S. ebulus* leaves were ground and extracted overnight with seven volumes of PBS (140 mM NaCl, containing 5 mM sodium phosphate, pH 7.5). The extract was filtered and then centrifuged at 14,300g for 30 min at 2 ºC in a Beckman JA 10 rotor, and the supernatant was decanted and filtered again. Glacial acetic acid was added to the crude extract until pH 4.05 was reached, after which it was centrifuged again at 14,300g. The acidified crude extract was subjected to ion exchange chromatograhy on a CM-Sepharose column (2.5 × 11 cm). The column was equilibrated with 10 mM sodium acetate (pH 4.5) at a flow rate of 0.42 L/h, the sample was applied, and the column was washed with 10 mM sodium acetate (pH 4.5). Then, the column was eluted first with 5 mM sodium phosphate (pH 7.5) and finally with 0.5 M NaCl in the same buffer. The protein eluted with 0.5 M NaCl was concentrated by ultrafiltration (using an Amicon YM-10 membrane) and subjected to molecular exclusion chromatography in a Superdex-75 HiLoad 26/60 equilibrated and eluted with PBS obtaining mainly three peaks. The fractions corresponding to the second peak which contained ebulin l and ebulin-RP were pooled, concentrated with Amicon YM-10 membrane and chromatographed again through Superdex-75 HiLoad 26/60. The fractions containing ebulin l and ebulin-RP were pooled and chromatographed through AT-Sepharose 6B at 0 ºC to separate the D-galactose-binding protein ebulin l from ebulin-RP. Ebulin l was eluted with 0.2 M lactose in PBS and ebulin-RP was obtained in the non-retained fraction. Both proteins were extensively dialyzed against water, and finally freeze-dried.

2.4. Amino acid sequence of peptide and protein samples

Native ebulin-RP, its A or B chains, and its CNBr peptides were sequenced by automated Edman degradation on a Procise sequencer, Model 491C (Applied Biosystems, Foster City, CA). Native, untreated ebulin-RP was subjected directly to automatic Edman degradation, following the manufacturer's procedures. For A and B chain N-terminal sequencing, chains were separated by SDS-PAGE in the presence of 5% 2-mercaptoethanol and then transferred to PVDF membranes (Applera) by electroblotting with the Mini Trans-Blot cell (Bio-Rad) in 10 mM CAPS, pH 11.0, containing 10% methanol. PVDF membranes were then stained for 1 min with Coomassie Blue R-250, destained with the washing solution (50% methanol), dried and directly analysed by Edman degradation [34]. Chemical digestion with cyanogen bromide was performed on the native protein dissolved in 75% formic acid, as reported [34]. CNBr-cleavage fragments were subjected to S-pyridylethylation. Separation of fragments were obtained using a Alltech C4 column (0.46 × 15 cm; 5 μm particle size) eluted with a linear gradient of solvent A (TFA 0.1%) and solvent B (acetonitrile + TFA 0.1%), from 5 to 70% of solvent B over 150 min at a flow rate of 1 mL/min [34]. The major peaks were subjected to automatic Edman degradation.

2.5. Cloning of ebulin-RP from cDNA.

The full length sequence of the RNA transcript coding for ebulin-RP was obtained by the RACE technique as described previously for ebulin l [29]. All the primers were synthesized from the sequences obtained as indicated in the section 2.4 and based on the cDNA sequence of ebulin l (GenBank accession no. AJ400822.1).

3'RACE-PCR: cDNA was synthesized using total RNA from leaves of *S. ebulus* (RNeasy Plant Mini kit. Qiagen GmbH, Hilden, Germany) as a template, and the MuLV reverse-transcriptase (Applied Biosystems. Roche) in the presence of an oligo-dT-adaptor primer (T1: 5'CGTCTAGAGTCGACTAGTGCTT(19) 3'). PCR was then used to amplify 3' cDNA using a sense primer corresponding to 5' end of the gen (A1f: 5' TTCYTIAATTTGGCGGGTGCC3') and an anti-sense primer complementary to the adaptor sequence (T2: 5'CGTCTAGAGTCGACTAGTGCT3').

The PCR product was cloned into the pCRII Dual Promoter vector using the TA Cloning Kit (Invitrogen, Barcelona, Spain) and then transformed into the *E. coli* strain Invα F' (Invitrogen). DNA sequencing in both directions was performed automatically.

5’RACE-PCR: The 5’end sequence was obtained using the 5’RACE System for Rapid Amplification of cDNA Ends, version 2.0 (Invitrogen).

cDNA was synthesized by reverse transcription, using an anti-sense primer R1r: 5’GCTTGTCCAGATTCGTAACTC3’, that recognizes a known sequence of the gene previously obtained. Then, a homopolymeric tail of cytosine was added to the 3’ end of the cDNA and the 5’ end fragment was amplified by nested PCR using the antisense gene specific primer R2r:5’GTATCTGAACCTTGCTGCTTC3’ and the sense universal primer AAP (5’RACE Abridged Anchor Primer): 5’GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG3’ that binds the homopolymeric tail added to the 3' ends of the cDNAs. A dilution of the original PCR (0.1%) was re-amplified using AUAP (Abridged Universal Amplification Primer): 5’GGCCACGCGTCGACTAGTAC3’ and R2r primers. All PCR products were cloned into the pCRII Dual Promoter vector and sequenced as indicated above. The cDNA sequence for ebulin RP was submitted to GenBank (accession number: MF170617).

2.6. rRNA N-glycosidase assays on rabbit reticulocytes lysates, S-30 lysates from yeast and COLO 320 cells.

The depurination assay was conducted as described by [35]. Rabbit reticulocytes lysates (80 µl), were incubated with 3 µg of RIP at 30 °C for 1 h. N-glycosidase activity on yeast ribosomes was assayed in 50 µl samples of S-30 lysates from yeast in 10 mM Tris–HCl buffer (pH 7.6) containing 10 mM KCl, 10 mM magnesium acetate and 6 mM 2-mercaptoethanol, which were incubated with 2 µg of RIPs at 30 °C for 1 h. After treatment, the RNA was extracted by phenolization, treated with 1 M aniline acetate (pH 4.5) and precipitated with ethanol. COLO 320 cells ($1 \cdot 10^6$/plate) were incubated for 48 h in presence of 1 µM ebulin-RP. After treatment, cells were harvested by centrifugation at 1,000g for 5 min. The pellets were lysed and the RNA was isolated following the instruction of the RNeasy Mini Kit (Qiagen GmbH, Hilden, Germany). RNA was treated with 1 M aniline acetate (pH 4.5) for 10 min at 0 °C and precipitated with ethanol. The RNAs were subjected to electrophoresis at 15 mA for 2 h (rabbit and COLO 320 cells) or 1 h 30 min (yeast) in a 7 M urea/5% (w/v) polyacrylamide gel and stained with ethidium bromide.

2.7. Polynucleotide:adenosine glycosidase activity on salmon sperm DNA and Tobacco mosaic virus (TMV) RNA.

The adenine release was measured according to the method reported by [36] with a few modifications. 10 μg of salmon sperm DNA or 10 μg of Tobacco mosaic virus RNA were incubated with 3 μg of RIP in 300 μL of a reaction mixture which contained 100 mM KCl, 50 mM magnesium acetate (pH 4), at 30 °C for 60 min. After incubation, the DNA was precipitated with ethanol at -80 °C for 3 h and centrifugated at 10,000g for 15 min. Adenine released from RIP-treated DNA or RNA was determined in the supernatants spectrophotometrically at 260 nm. The N-glycosidase activity of RIPs on TMV RNA was also assayed in 25 μL samples containing 5 μg of TMV RNA, which were incubated with 3 μg of the corresponding protein. After treatment, the RNA was analyzed by phenolization, treatment with 1M aniline acetate (pH 4.5), ethanol precipitation, RNA electrophoresis (15 mA for 1 h 15 min) and ethidium bromide staining as described elsewhere [35].

2.8. DNA cleavage experiments

DNA cleavage experiments were performed as previously reported [37]. Each reaction contained 1 μg of RIP and 200 ng of the plasmid pCR2.1 (Invitrogen) in a final volume of 10 μL of 10 mM Tris–HCl, 5 mM $MgCl_2$, 50 mM NaCl and 50 mM KCl, pH 7.8. Samples were incubated for 1 h at 37 °C, run on agarose gel (0.8%) in TAE buffer (0.04 M Tris, 0.04 M acetate, 1 mM EDTA, pH 8.0) and visualized by Gel Red nucleic acid staining (Biotium Inc., Hayward, CA). EcoRI linearization was achieved by incubating 250 ng of pCR2.1 with 1.5 units of EcoRI enzyme according to manufacturer instructions (Roche Diagnostics S.L., Barcelona, Spain).

2.9. DNA fragmentation analysis

COLO 320 cells ($1 \cdot 10^6$/plate) were incubated for 48 and 72 h in presence of 1 μM ebulin-RP. After treatment, cells were harvested by centrifugation at 1,000 × g for 5 min. The pellets were lysed and the DNA was isolated following the instruction of the Genomic Prep Cells and Tissue DNA Isolation Kit (GE Healthcare). DNA

electrophoresis was carried out in 1.8% agarose gels using TBE buffer (0.089 M Tris, 0.089 M boric acid, 2 mM EDTA, pH 8.0). DNA was stained with Gel Red (Biotium, Inc., Hayward, CA, USA) and visualized with an ultraviolet lamp.

2.10. Caspase-3/7 activity

The caspase-3/7 activity was assessed by the luminescent assay Caspase-GloTM 3/7 (Promega). COLO 320 cells ($4 \cdot 10^3$/well) were seeded in 96-well microtiter plates in 80 µL RPMI complete medium and incubated at 37 °C under 5% CO_2 in the absence or the presence of 1 µM of ebulin-RP. After 48 h of incubation, 70 µl/well of Caspase-Glo™ 3/7 was added. Plates were shaken for 1 min and then incubated for 1 h at room temperature in the dark. The luminescence was measured by SpectroMax L (integration time 10 s) and the values were normalized for the viability.

2.11. Cell viability analyses

Cell viability was determined with a colorimetric assay based on cleavage of the tetrazolium salt WST-1 to formazan by mitochondrial dehydrogenases in viable cells. $3 \cdot 10^3$ cells in 0.1 mL of medium were seeded in 96-well plates and incubated for 48 h at 37 °C under 5% CO_2 in the absence or presence of ricin, ebulin l, ebulin-RP or SEA, as described in the Figure legends. Then, the cells were incubated for further 2 h with 10 µL/well of the cell proliferation reagent WST-1 (Roche Diagnostics S.L., Barcelona, Spain) at 37 °C under 5% CO_2. After plate shaking, the sample absorbance was measured using a microtiter plate reader set at 450 nm, with 620 nm as reference. A background of the culture without cells was subtracted. To assess the effects of sugars (D-glucose, D-lactose, D-galactose, *N*-acetyl-D-galactosamine and *N*-acetyl-D-glucosamine), Brefeldin A and NH_4Cl, cells were preincubated for 1 h with these substances at a final concentration of 50 mM for D-glucose, D-galactose and *N*-acetyl-D-glucosamine; 40 mM for *N*-acetyl-D-galactosamine and D-lactose; 5 µg/ml for Brefeldin A and 20 mM for NH_4Cl, and then RIPs were added and after 20 or 24 h of further incubation, viability was determined as described above. The concentration of RIPs causing 50% reduction of viability (IC_{50}) was calculated by linear regression analysis.

2.12. Annexin V/Propidium Iodide (PI) analysis

Surface exposure of phosphatidylserine in apoptotic cells was detected using Annexin V-FITC (Propidium Iodide) apoptosis detection kit (Clontech). Cells were seeded in 24-well microtiter plates in 400 μL of medium and incubated with 1 μM ebulin-RP. After 48 h of incubation, cells were treated according to the manufacturer's instructions. Apoptotic cells (Annexin V+/PI−), necrotic cells (Annexin V−/PI+) and apoptotic cells at late stage (Annexin V+/PI+) were observed under a Nikon Eclipse Ti-E fluorescence microscope (Nikon, Melville, NY).

2.13. Fluorescence microscopy

Ebulin l and ebulin-RP were labelled with Cy3-maleimide according to the manufacturer's protocol (GE Healthcare, Barcelona, Spain). HeLa cells grown on coverslips at 37 ^{o}C were incubated with Cy3-ebulin l or Cy3-ebulin-RP for 1 h in HEPES medium at 4 ^{o}C. The cells were then washed three times in PBS and fixed or incubated for 30 min in HEPES medium at 37 ^{o}C. Cells were then fixed in 3% paraformaldehyde in PBS for 15 min, washed three times in PBS, mounted in Mowiol, and examined with a a Nikon Eclipse Ti-E fluorescence microscope.

2.14. Sequence retrieval and data treatment

All the amino acid sequences of RIP and lectins used in this study are available in the National Center for Biotechnology Information (NCBI) sequence database (http://www.ncbi.nlm.nih.gov/protein/) except those from sieboldin and SSA from *Sambucus sieboldiana* (Miq.) Blume ex Graebn. which were obtained from [38, 39] respectively. For the phylogenetic analysis, the signal peptide, A-chain and connecting peptide were removed using the following criteria by order of preference: information in the data bank entry, information in the literature from N-terminal sequencing, and comparison with other close related sequences.

2.15. Sequence alignment

Sequence alignment was performed using the ClustalW tool included in the Mega 6 suite (http://www.megasoftware.net) [40] with default parameters and edited manually to align all conserved Cys. Then, the sequences included between each pair of conserved Cys were aligned automatically, and finally the complete sequences as well. Multiple sequence alignments from the sugar-binding sites were graphically represented by sequence logos [41] created with WebLogo 3 (http://weblogo.threeplusone.com/) [42].

2.16. Protein structure studies and graphical representation

The structure of ebulin l complexed with galactose (accession code 1HWN) is available in the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). Three-dimensional structural modelling of SEA and ebulin-RP was carried out on the I-TASSER server [43]; http://zhanglab.ccmb.med.umich.edu/I-TASSER). Several structures of RIPs and lectins were chosen by I-TASSER as the templates in the modelling of ebulin-RP and SEA (Table S2). Study and graph representations of protein structures were performed with the aid of the Discovery Studio Visualizer suite (v16.1.0) (http://accelrys.com/).

2.17. Molecular docking

Docking was carried out using Autodock 4.2 [44]; (http://autodock.scripps.edu/) as has been previously described [45]. Docking was performed on a grid of 120×120×120 points, with the addition of a central grid point. The grid was centred on the mean of the coordinates of the sugar-binding site. Grid spacing was 0.125 Angstroms, leading to a grid of 15×15×15 Angstroms. For each molecule, 100 docking runs were performed. The generated 100 docking poses were clustered by root mean square (RMS) difference with a cutoff value 0.5 nm for each case. The top-ranked pose of the most populated cluster was retained and further analyzed with the Discovery Studio Visualizer suite (v16.1.0).

2.18. Phylogenetic analysis

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [46] as has been reported for other RIPs [2]. Briefly, the bootstrap consensus tree inferred from 1,000 replicates [47] is taken to represent the evolutionary history of the taxa analysed [47]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches in the figures [47]. Initial tree for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.1218)). The analysis involved 30 amino acid sequences. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA6 [40].

2.19. Other procedures

Analysis of proteins by SDS-PAGE was carried out using 10 and 12% acrylamide gels. The detection of glycan chains in proteins was performed with the Glycan Detection Kit from Roche, following their procedure. Protein synthesis was performed with a coupled transcription–translation *in vitro* assay using a rabbit reticulocytes lysate system as described elsewhere [27]. Red blood cell agglutination was determined at room temperature on microtitre plates containing 100 μL of 5 mM sodium phosphate (pH 7.5), 0.14 M NaCl, and 1% of human red blood cells.

3. Results and Discussion

3.1. Purification and characterization of ebulin-RP

S. ebulus leaves contained the type 2 RIP ebulin l and two D-galactose-binding lectins, one dimeric called SELld and one monomeric called SELlm [20, 21]. Using a new method for the isolation of ebulin l based on ion-exchange and gel filtration chromatography instead of affinity chromatography, we found a new RIP that co-purified with ebulin l in the leaves of *S. ebulus*. The acidified crude extract from leaves was subjected to ion-exchange CM-Sepharose chromatography. As shown in Fig. 1a,

only a small part of the material bound to the column was eluted with 5 mM sodium phosphate pH 7.5 while a large protein peak was eluted with NaCl. The protein fraction eluted with NaCl was concentrated and subjected to molecular exclusion chromatography on Superdex 75. As shown in Fig. 1b several peaks were obtained. SDS-PAGE and protein synthesis inhibition assays revealed that the second peak (shaded area) contained ebulin l together with another protein (Fig 1b, inset). In order to separate these two proteins, the second peak was chromatographed again through Superdex 75 and then subjected to affinity chromatography through AT-Sepharose 6B yielding two peaks (Fig. 1c).The first peak contained protein that was not retained by the column while the second peak corresponded to ebulin l that was bound and eluted from the column with 200 mM D-lactose. Both peaks were analysed by SDS-PAGE. In the absence of 2-mercaptoethanol, the analysis of the first peak revealed the presence of a homogeneous protein with an apparent *Mr* of 62,500 (Fig. 1d, lane 2). However, in the presence of 2-mercaptoethanol, 2 protein bands with a molecular weight of 32,000 and 30,500 were visible in the gel (Fig. 1d, lane 4), This new protein was named ebulin Related Protein (ebulin-RP) and the estimated yield was 1.3 mg/100 g of total starting plant tissue. The bound protein ebulin l is a heterodimeric protein with an apparent *Mr* of 57,000 Da with two subunits of 30,900 and 26,600 Da (Fig. 1d, lanes 1, 3). SDS-PAGE analysis of ebulin-RP and ebulin l followed by blotting onto Immobilon membrane and staining for carbohydrate using a glycan detection kit, indicated that ebulin-RP and ebulin l B-chain stained for carbohydrate, while ebulin A chain did not show any staining (Fig. 1e). This indicates that ebulin-RP contains sugar chains in both subunits.

Ebulin l consists of two subunits one with lectin activity and the other with N-glycosidase activity [17]. Therefore, ebulin l can be purified using affinity chromatography through AT-Sepharose 6B which exposed galactose residues. On the other hand, due to the lectin activity, ebulin l promoted human red blood cell agglutination. Since ebulin-RP was not retained on AT-Sepharose, we studied the agglutination capacity of this protein and found that even a concentration of 400 μg/mL ebulin-RP did not have any effect on human erythrocytes (data not shown). Under the same conditions, ebulin l agglutinated human red blood cells at concentrations as low as 12.5 μg/mL [17].

3.2. Molecular cloning

To determine the amino acid sequence of the new protein, we cloned the gene. Primers derived from ebulin-RP peptide sequences (Table S1) were designed and total RNA from *S. ebulus* leaves was used to synthesize the cDNA that served as a template in PCR amplifications. Amplicons obtained by the RACE technique were cloned into the vector pCRII-dual promoter and sequenced. Sequences overlapped in a wide region and the analysis allowed us to obtain the complete sequence of the gene encoding ebulin-RP. The complete sequence of cDNA was 1,917 bp long (GenBank accession number MF170617) and consisted of an open reading frame (ORF) of 1,716 bp, and included a 5'-UTR located 41 bp upstream of the start codon (ATG) and a 3'-UTR of 149 nucleotides that ended in a poly (A) tail. The open reading frame of 1716 bp encoded a precursor polypeptide of 572 amino acids (Fig. 2). The proteolytic processing of the signal peptide involves the removal of 26 amino acid residues resulting in a polypeptide of 546 amino acids. Additional processing removes 21 amino acids residues of the connecting peptide (Fig. 2, red) and generates the two disulfide-linked ebulin-RP chains, an A chain of 261 amino acids with a calculated molecular weight of 28,962 Da and a B chain of 264 amino acids with a calculated molecular weight of 28,823 Da. Such sequences were identified as the sequences of ebulin-RP since the N-terminal amino acid sequences obtained from native ebulin-RP, its A and B chains, and its CNBr fragments (Table S1) were found in the reading frame (Fig. 2).

From the deduced amino acid sequence, 9 potential N-glycosylation sites (5 for the A chain and 4 for the B chain) were identified which could link polysaccharide chains as revealed with the glycan detection procedure (Fig. 1e). Therefore, glycosylation could account for the molecular weight difference (about 4,715 Da) between the Mr assessed by SDS-PAGE (62,500 Da, Fig. 1d) and that deduced from the cDNA (57,785 Da).

Analysis of the deduced amino acid sequence indicated a high sequence identity with RIPs and lectins isolated from *S. ebulus* (ebulin l, SEA, SELlm and SELld). As shown in Fig. 2, ebulin-RP shares approximately 52.72% amino acid identity with the type 2 RIP ebulin l and 67.35% with the tetrameric RIP SEA. When comparing the B chain of ebulin-RP with those of SELlm and SELld we found identities of 51.08% and 48.84% respectively. The precursors of these lectins are truncated and contain the signal peptide,

a piece of the A chain, the connecting peptide and the B chain. After the processing the lectin protein contains only the B chain [20, 21].

The invariant residues within the active site of ebulin-RP A chain which have been reported to be highly conserved among RIPs were all well conserved (Fig. 2). Ebulin-RP B chain consists of two structurally equivalent domains and each domain comprises three subdomains (1-alpha, 1-beta and 1-gamma for domain 1 and 2-alpha, 2-beta and 2-gamma for domain 2). The Cys5 in the B chain of ebulin-RP is responsible for the formation of the inter-chain disulphide bridge with the A chain. On the other hand, the B chain contains eight conserved cysteinyl residues which could form four intra-chain disulphide bridges. In the following section the structures of the sugar-binding sites will be described in detail.

3.3. Structure and binding to sugars

To ascertain the main structural characteristics that may be involved in the deficient sugar-binding of ebulin-RP, a three-dimensional structure was predicted by comparative modelling using several type 2 RIP crystal structures as templates (Fig. S1). The selected best model was found to have a confidence score (C-score) of 1.49 for the A chain and 1.04 for the B chain, template modelling (Tm) score of 0.92 ± 0.06 for the A chain and 0.86 ± 0.07 for the B chain, and root-mean-square deviation (RMSD) of 2.9 ± 2.1 Å for the A chain and 3.8 ± 2.6 Å for the B chain, which satisfied the range of parameters for molecular modelling. To compare the structural characteristics that may be involved in the different sugar affinities of *Sambucus* type 2 RIPs, a new model for SEA was also predicted (Fig. S1). Ebulin l, ebulin-RP and SEA display similar three-dimensional structures. One of the most important differences is that SEA B-chain possesses an additional cysteine that allows dimerization to render a tetrameric type 2 RIP [32]. This is the consequence of mutations that change the codon AGA to TGT resulting in Cys (SEA) instead of Arg (ebulin l).

Ebulin l is a galactose-binding lectin [29] and SEA specifically binds to sialic acid [32]. SNLRPs are type 2 RIPs isolated from *S. nigra* bark that were first considered as carbohydrate-binding-defective lectins [48], but have recently been found to strongly interact with N-acetyl-D-glucosamine oligomers as well as with many glycan structures containing N-acetyl-D-glucosamine in glycan microarray analyses [25]. The B-chain of

ebulin-RP shares a strong homology (83.14%) with the B-chains of SNLRPs. We therefore selected D-galactose, N-acetylneuraminic acid (Neu5Ac) and N-acetyl-D-glucosamine as the sugars to assay the binding to the 1-alpha and 2-gamma binding sites of ebulin l, SEA and ebulin-RP. For this purpose we perform a comparative molecular docking study using Autodock 4.2. For each molecule, 100 docking runs were performed and clustered by root mean square (RMS) difference with a cutoff value 0.5 Å for each case. The top-ranked pose of the most populated cluster was considered because it fitted better to the experimental data reported by [29] for both ebulin l B-chain galactose-binding sites (Fig. S2). In both cases, experimental X-ray diffraction and in silico docking, the binding of D-galactopyranose is the result of the C-H-π interaction between an aromatic amino acid (W39 in the 1-alpha site and F249 in the 2-gamma site) and the apolar face of the pyranose ring (Fig. S2). The polar groups of the other face can establish hydrogen bonds with donor o acceptor atoms of the amino acids located in the opposite side of the binding pocket (D24, V25, N27 and Q37 in the 1-alpha site or D235, K337 and N256 in the 2-gamma site). Figure 3a shows the surface appearance of the 1-alpha site of ebulin l, SEA and ebulin-RP. Ebulin l displayed a well-defined pocket in which galactose can fit. Such a pocket was defined in the non-polar side, by two amino acids with large side chains (W39 and R115), allowing the stacking of the pyranose ring with W39. By contrast such a pocket has fully disappeared in SEA and almost fully disappeared in ebulin-RP because arginine has changed to glycine (without side chain) and tryptophan has change to serine (small side chain) in the case of SEA (Fig. 3a). Figure 3b shows the interactions of D-galactopyranose with key amino acids in the 1-alpha site of ebulin l. As indicated above, binding of galactose implies a C-H-π interaction between W39 and the apolar face of the pyranose ring and hydrogen bonds between the polar face and amino acids located in the opposite side of the binding pocket (D24, V25, N27 and Q37). Two of these amino acids (D24 and N27) are not present in ebulin-RP. In ebulin l, such a type of interaction was not possible with Neu5Ac because glycerol and N-acetyl groups avoided the entry of the pyranose ring in the pocket and its stacking with the tryptophan ring (Fig. 3b). By contrast N-acetyl-D-glucosamine bound to the 1-alpha site in almost the same way as galactose, however the orientation of the C4OH group avoids interaction with the tryptophan ring. Consequently, while 1-alpha site of ebulin l can bind galactose it cannot bind either sialic acid or N-acetylglucosamine.

As shown in Fig. 3c, ebulin l, SEA and ebulin-RP displayed a well-defined pocket in the 2-gamma sugar-binding site: a tight and deep cavity in ebulin l, a wide and deep cavity in SEA and a wide and more superficial cavity in ebulin-RP. Binding of galactose to the ebulin l 2-gamma site involved a C-H-π interaction between F249 and the apolar face of the pyranose ring and hydrogen bonds between the polar face and amino acids located in the opposite side of the binding pocket (D235, K237 and N256). This kind of interaction is not possible with sialic acid because of the presence of the N-acetyl group (Fig. 3d). N-acetyl-D-glucosamine bound to the 2-gamma site in almost the same way as galactose, however the orientation of the C4OH group avoids the interaction with the tryptophan ring. Consequently, the 2-gamma site can bind galactose but neither sialic acid nor N-acetyl-D-glucosamine can be bound to this site. Galactose and sialic acid bound to the 2-gamma site from SEA in a similar way whereas it was different for N-acetyl-D-glucosamine (Fig. 3d). The pocket was wide enough to accommodate sialic acid and numerous hydrogen bond interactions were established between several sugar groups and the amino acids D228, A230, Q231, Y242, N245 and E247. This is consistent with the fact that the substitutions A233R and Q234A in the tetrameric type 2 RIP SSA from *Sambucus sieboldiana* completely abolished the binding to the sialoglycoprotein fetuin [49]. In addition, the 2-gamma site from SEA allowed the binding of D-galactose because the non-polar face of the pyranose ring pointed towards the Y242 ring allowing C-H-π interaction (Fig. S3) and the groups of the polar face could bind to the opposite face of the pocket by hydrogen bonds with D228, Q231 and E247 (Fig. S3). By contrast, the pyranose ring of *N*-acetyl-D-glucosamine remained out of the binding pocket and only the *N*-acetyl group could form hydrogen bonds with A230 and Q231. Therefore, even if the 2-gamma site from SEA can bind sialic acid and galactose most probably it cannot bind *N*-acetyl-D-glucosamine.

Ebulin-RP 2-gamma site did not allow the stacking of pyranose rings from the tested sugars (Fig. 3d) and interactions might only occur by hydrogen bonds with the amino acids E233 (galactose, sialic acid and N-acetyl-D-glucosamine), Q236 and N254 (galactose and N-acetyl-D-glucosamine), Y247 (sialic acid), or N252 (N-acetyl-D-glucosamine), which could not be enough for efficient binding.

Based on these data, it could be proposed that the loss of lectin activity of ebulin-RP may be due to the presence of inactive 1-alpha and 2-gamma sites as has been reported for some abrin isoforms (abrin-b and abrin-c) that lack the ability to bind galactose [50].

Considering that ebulin-RP is a glycoprotein, we have performed a tryptic peptide mass fingerprinting by using MALDI-TOF mass spectrometry (see supplementary material) to check if amino acid residues of the sugar-binding pockets [1-alpha site (i.e. B-chain sequence position: 23-27, 36-39, 43-46 and 113-144) and 2-gamma site (i.e. B-chain sequence position: S199, 233-238 and 245-255); Fig. 2] were glycosylated.

The accurate Mr of found peptides obtained from in-gel tryptic digestion of ebulin-RP B-chain is reported in Table S3, while the same peptides mapped on the deduced B-chain amino acid sequence from ebulin-RP gene used as reference sequence, are displayed in Figure S4. With this set of data, the tryptic peptides provided about 81% of the B-chain amino acid sequence (214 out of 264 amino acid residues). Furthermore, tryptic T-3, T-4, T-6, T-8, T-9, T-10 and T-11 peptides allowed the mapping of the amino acid residues involved in the formation of the alpha or gamma pocket indicating the absence of glycosylation in these amino acid regions. In particular, the asparaginyl residue (N238 in the T-10) present in the 2-gamma site of the B chain is not N-glycosylated.

3.4. Enzymatic activities

RIPs are rRNA N-glycosidases that split an adenine from the Sarcin Ricin Loop (SRL) from the mammalian 28S rRNA disabling the ribosomes to bind the elongation factor 2 and arresting protein synthesis [3]. To ascertain whether ebulin-RP was able to inactivate ribosomes, we tested its activity in a cell-free translation system and we found that ebulin-RP inhibited protein synthesis with an IC_{50} of $9.5 \cdot 10^{-11}$ M in rabbit reticulcytes lysate (Table 1). This value is comparable to those of type 2 RIPs ebulin l ($8 \cdot 10^{-11}$ M) and ricin ($2 \cdot 10^{-11}$ M). On the other hand, SEA inhibited protein synthesis with an IC_{50} of 10^{-9} M, which is 10-50-fold less potent than those of type 2 RIPs ebulin l, ebulin-RP and ricin. The reason for this low toxicity is not clear but its tetrameric structure may play a role. The active site or the ribosome binding site of SEA may be blocked by the presence of the B subunits as have been reported for ricin and Shiga toxin holotoxins. Ricin could not interact with the ribosome or depurinate the ribosome due to the blockage of the ribosome binding site by the B subunit of ricin. [51].

We next studied whether ebulin-RP also exhibited rRNA N-glycosidase activity. As shown in Fig. 4a ebulin-RP depurinated the rRNA from rabbit reticulocyte ribosomes, which, upon acid aniline treatment of the isolated rRNA, released the diagnostic RNA fragment. We further tested the RNA N-glycosidase activity of ebulin-RP on ribosomes of different origin. Treatment with ebulin-RP of ribosomes from the enterobacterium *Escherichia coli* and from the parasitic bacterium *Agrobacterium tumefaciens* did not release the Endo's fragment indicating that bacteria ribosomes are not sensitive to this RIP (data not shown). On the other hand, ebulin-RP displayed rRNA N-glycosidase activity on yeast ribosomes as indicated by the release of the diagnostic fragment upon treatment with aniline (Fig. 4b). The N-glycosidase activity was comparable with other RIPs from *S. ebulus*, such as ebulin l and SEA that also depurinate yeast ribosomes (Fig. 4b).

RIPs do not exclusively act on ribosomes but display polynucleotide:adenine glycosidase (PNAG) activity on different nucleic acid substrates. Most of the RIPs release adenine from DNA, RNA, and poly(A) although the amount released vary widely [4]. We examined the ability of the three RIPs of *S. ebulus*; ebulin l, ebulin-RP and SEA to release adenine from substrates other than ribosomes. Thus we tested whether they exhibit PNAG activity against salmon sperm DNA and TMV RNA and compared this activity with that of ricin, which possesses a moderate activity [4]. As shown in Figure 5a these type 2 RIPs displayed similar PNAG activity on DNA and RNA than ricin and much lower than that of type 1 RIPs [52]. Comparing the ability of the three *S. ebulus* proteins, slight differences can be observed depending on the protein and the substrate. Ebulin-RP displayed similar PNAG activity on both substrates while ebulin l was more active on salmon sperm DNA than on TMV RNA. On the other hand, the ability of SEA to release adenine is very low on DNA and was inactive on TMV RNA.

The potential effects of these type 2 RIPs on TMV genomic RNA was investigated further. As shown in Figure 5b ebulin l and ebulin-RP promoted an extensive depurination of TMV RNA which, upon treatment with acid aniline, led to a large degradation of the polyphosphate RNA backbone. By contrast, SEA displayed very low activity against the TMV RNA. These results are in agreement with those obtained when the released adenines were determined spectrophotometrically (Fig. 5a).

Another enzymatic activity associated with some RIPs is endonuclease activity exerted on supercoiled plasmid DNA producing relaxed or even linear plasmids [53, 54]. Therefore we assayed the endonuclease activity of ebulin l, ebulin-RP and SEA on the plasmid pCR2.1 in the presence or absence of magnesium ions since it has been reported that this activity was enhanced by divalent metal ions such as Mg^{2+} [52, 53]. As shown in Figure 5c, only in the presence of magnesium ions, ebulin-RP and SEA cleaved supercoiled pCR2.1 dsDNA generating mainly relaxed forms. The highest activity was observed for SEA that promoted the conversion of all the supercoiled plasmid into the relaxed and, in less extension, linear forms, while ebulin-RP only converted around half of supercoiled plasmid into relaxed forms. Linear forms were observed for both proteins indicating that they possess nuclease activity, acting directly on DNA by introducing a cleavage into the DNA strands. By contrast, the activity of ebulin l on pCR2.1 was not dependent of magnesium ions and the mobility of the plasmid in the presence of ebulin l was similar to the control (Fig. 5c).

It has been reported that some RIPs play a role in plant defense [52] due to biological activities display for them that could have a defensive role against pathogens. In *S. ebulus*, type 2 RIPs exhibit similarity in their rRNA N-glycosidase activities against mammalian and fungi, but they differed in their activities against viral RNA and plasmid DNA. This suggests that the presence of different type 2 RIPs might optimize the response of the plant against several types of pathogens.

3.5. Cytotoxic effect in cell cultures

Cytotoxicity of type 2 RIPs is influenced by many variables. Among them, it depends on the efficiency of uptake into target cells which also depends on the characteristics both of the cells and of the carbohydrate-binding activity of the B-chain. Ebulin l and ricin are both galactose-binding proteins, but the reduced cytotoxicity of ebulin l compared with ricin is in part due to some defect that limits its ability to bind galactosides on cell surfaces [29]. Since glycans are abundant on the surface of tumor cells, we study the effect of the lack of sugar-binding of ebulin-RP on cytotoxicity in a variety of cancer cells compared with ebulin l and SEA. In all cases the cytotoxicity of ebulin l, ebulin-RP and SEA was much less than that exerted by ricin, which affects viability with IC_{50} several orders of magnitude lower (Table 1). When comparing ebulin

l, SEA and ebulin-RP, ebulin l was the most active toxin, especially on COLO320 cells. The cytotoxicity of ebulin-RP for cancer cells was the lowest with IC_{50} values greater than 10^{-6} M except for HeLa cells that were the most sensitive cells to this toxin. On the other hand SEA, a sialic-binding tretrameric type 2 RIP displayed higher toxicity than ebulin-RP but lower than ebulin l (Table 1) despite the high amount of terminal sialic acid present on the cells.

Although ebulin l had a cytotoxic effect on the tumor cells tested, it did not affect significantly the viability of non-tumorigenic mesenquimal stem cells (hMSC). Both ebulin l and ebulin-RP affected viability with similar IC_{50} value of 8 10^{-6} M in these cells suggesting that ebulin l may preferentially target cancer cells having a minimum effect on normal cells. Ebulin l can bind glycoproteins and glycolipids on the plasma membrane through the B chain. However, the low toxicity of ebulin-RP lacking sugar-binding activity, compared to ebulin l, may indicate that the cytotoxicity is induced in part by the carbohydrate binding activity of the B chain.

To further study RIPs carbohydrate-binding specificity, inhibition of cell binding with sugars was performed followed by cell viability measurements. The competitive inhibition by certain sugars will reduce the amount of toxin uptake by the cells. The availability of both, ebulin l and SEA to cells was greatly reduced by galactose, lactose or N-acetyl-D-galactosamine, thus improving cell viability whereas glucose and N-acetyl-D-glucosamine had no effect on the cytotoxicity (Fig.6a). Interestingly, none of the sugars affected viability of ebulin-RP treated cells (Fig. 6a). Therefore the binding and the uptake of ebulin-RP into cells might not be dependent on receptors containing those sugars. The lack of sugar interaction found for ebulin-RP is in agreement with the molecular docking studies carried out and with the failure of the lectin to agglutinate erythrocytes or to bind AT-Sepharose 6B. To visualize the binding and transport of ebulin-RP in cells, ebulin l and ebulin-RP were labelled with the fluorophore Cy3 and then probed in HeLa cells. Fig. 6b shows that when the cells were incubated with Cy3-ebulin l at 4 ^{o}C, it was bound to the cell surface in a uniform manner. When cells were incubated for further 30 min at 37 ^{o}C, the amount of ebulin l at the surface appeared reduced and the fluorescent ebulin l appeared as intracellular dots, indicating its accumulation in intracellular vesicles. However, the amount of bound CY3-ebulin-RP to the HeLa cell surface was almost undetectable and the amount of internalized protein

was extremely low compared to ebulin l (Fig. 6b). Therefore, a higher concentration of ebulin-RP is required to achieve the uptake of a cytotoxic dose.

Although the low cytotoxicity of ebulin-RP compared to ebulin l seems to be related to deficient sugar-binding domains, which is the major difference with ebulin l, it has been reported that articulatin-D, a type 2 RIP with a B-chain lacking sugar-binding activity displays similar cytotoxicity as classical toxic-type 2 RIPs, indicating that recognition of sugar receptors on the cell surface by B-chain may not be vital for internalization and subsequent cytotoxicity of all type 2 RIPs [55].

3.6. Apoptosis induction

Based on the above studies, ebulin-RP displays low toxicity against cells. The B chain allows binding of the toxins to the cell surface followed by internalization. A defective binding may result in low internalization (Fig. 6b) and therefore reduced translocation to the cytosol. To see if ebulin-RP was able to reach the cytosol and inactivate the ribosomes after being endocytosed we analyzed the ribosomal RNA from COLO 320 cells treated with ebulin-RP for 48 h. Figure 7a showed that the ribosomes were depurinated releasing the diagnostic fragment after treatment of the RNA with acid aniline indicating that ebulin-RP was able to reach the ribosomes to inhibit protein synthesis. Despite this, the low toxicity of ebulin-RP and ebulin l to whole cells compared to ricin suggests that, once internalized, ebulins likely follow a different and less productive intracellular route than ricin. After binding to N-glycosylated molecules, ricin enter mammalian cells by endocytosis and undergo retrograde transport via the Golgi complex to reach the endoplasmic reticulum (ER) where ricin A-chain exploits the ER-associated degradation (ERAD) pathway to reach and inactivate its target ribosomes [56]. We therefore investigated the intracellular pathway followed by ebulins studying their toxicity on COLO 320 cells in the presence of substances interfering with intracellular routing such as the fungal inhibitor Brefeldin A that causes Golgi complex disassembly and ammonium chloride. As shown in Figure 6c, Brefeldin A markedly reduced the cytotoxicity of ricin as expected but had no significant effect on that of ebulin l or ebulin-RP, indicating that these toxins follow a Golgi-independent pathway to the cytosol. On the other hand, preincubation of COLO cells with ammonium chloride enhanced the cytotoxicity of ricin as well as that of ebulin l and ebulin-RP (Fig.

6c) indicating that ebulins, like ricin, do not require a low pH for translocation to the cytosol. Moreover, ammonium chloride may stimulate the cytotoxicity by preventing the lysosomal degradation of ebulins.

Cell death might be a consequence of protein synthesis inhibition that can itself result in apoptosis. Apoptosis might also be induced by lectin binding to specific glycosylated proteins on the cell surface leading to activation of cell death factor receptors. We investigated whether the observed cytotoxic effects of ebulin-RP at concentrations close to its IC_{50} were mediated via apoptosis, since treated cells exhibited the morphological features characteristic of apoptosis such as cell rounding and blebbing (data not shown). Cleavage of chromosomal DNA into oligonucleosomal fragments is a biochemical hallmark of apoptosis. When COLO cells were treated with 10^{-6} M ebulin-RP for 48 and 72 h the breakdown of the nuclear DNA into oligonucleosomal fragments was clearly observed specially after 72 h treatment (Fig. 7b). To demonstrate the involvement of caspase-dependent apoptosis, caspase 3/7 activation was measured in cells exposed to 10^{-6} M ebulin-RP for 48 h. This experiment showed activation of effector caspases in both COLO 320 and HeLa cells (Fig. 7c). To further determine the role of caspase-dependent apoptosis, the pan-caspase inhibitor Z-VAD, which irreversibly binds to the catalytic site of caspases, was used to selectively inhibit the apoptotic pathway. HeLa cells were pretreated and maintained in 100 μM Z-VAD, and the cell viability was determined for different ebulin-RP concentrations. As shown in Fig. 7d, caspase inhibition by Z-VAD considerably reduced the cytotoxicity of $5 \cdot 10^{-6}$ M of ebulin-RP after 48 h from a viability of 35% to 70%. By contrast, in the presence of the necroptosis inhibitor Nec-1, cell death induced by ebulin-RP was only slightly reduced, with a viability of approximately 42% for that ebulin-RP concentration (Figure 7d). Figure 7e demonstrated that significant apoptosis occurred in HeLa cells treated with ebulin-RP for 48 h as shown by reaction of these cells with Annexin V-FITC (green). Some PI staining (red) was also seen in the cells, indicating late stage apoptosis or necrosis. However, very little PI staining (red) was observed in cells pretreated with Z-VAD. In contrast, double fluorescence showing the typical features of late stage apoptosis was observed for cells pretreated with Nec-1. These findings confirm that apoptosis is the predominant pattern of cell death induced by ebulin-RP and suggest that apoptosis might be a consequence of protein synthesis inhibition. However, we cannot rule out the possibility that induction of apoptosis occurs before protein synthesis

inhibition takes place. Non-toxic type 2 RIPs from *Sambucus* (nigrin b and ebulin l) have been used as a moiety of conjugates and immunotoxins targeting tumor cells with high selectivity [14]. The main advantage over ricin and derivatives is their differential cytotoxicity. Antibodies or ligands are internalized and promoted the productive translocation of the toxins to the cytosol, thus allowing the intracellular actions of non-toxic type 2 RIPs. The even lower cytotoxicity displayed by ebulin-RP as compared with that of ebulin l against tumor cells, together with the high efficiency in the inhibition of protein synthesis at the ribosomal level suggests that ebulin-RP would be a good candidate as toxic moiety in the construction of immunotoxins and conjugates directed against specific targets. Since the cellular response to targeted toxins is a complex mechanism, toxins like ebulin-RP that induce apoptosis are a good choice to kill the cells under a strictly controlled process.

3.7. Phylogenetic analysis

Sambucus species possess a complex mixture of diverse types of RIPs and related lectins [14] and they have been shown to contain type 1 RIPs, heterodimeric type 2 RIPs (one A chain and one B chain), tetrameric type 2 RIPs (two A chains and two B chains), and monomeric and homodimeric pure lectins (one or two B chains respectively). To better understand the relationship among the different type 2 RIPs and related lectins from *Sambucus* we performed a phylogenetic analysis of the B chains of type 2 RIPs and related lectins using the Maximum Likelihood method. The tree was rooted using representative sequences of B chains of type 2 RIPs from angiosperms like outgroup. The resulting tree (Fig. 8) gives a general idea of the relationships among these proteins and showed very high bootstrapping values for most of the branches. All these proteins cluster in the same clade that is divided into two branches, one containing all the D-galactose/N-acetyl-D-galactosamine specific proteins and the other containing the sialic acid specific proteins together with proteins with specificity for N-acetyl-D-glucosamine oligomers or unkown specificity, thus indicating that the first duplication and divergence event was related with a change in the specificity for sugars. Because type 2 RIPs from other angiosperm species are specific for galactose, it is likely that all proteins from *Sambucus* originated from a type 2 RIP specific for galactose. Galactose/N-acetyl-D-galactosamine specific binding type 2 RIPs from *Sambucus* have

undergone deletions that have rendered monomeric and dimeric lectins. Such deletions have occurred at least five times. The first deletion produced the dimeric lectins upon dimerization when a new cysteine that allowed the formation of a disulphide bridge between two B chains appeared. The last deletion produced the monomeric lectin SNAII from nigrin b taken into account that both B chains share an almost identical amino acid sequence (only an 8 amino acids deletion in the N-terminus and one more change in the sequence differentiate SNAII from nigrin b B-chain). The other branch only included type 2 RIPs specific for sialic acid, specific for N-acetyl-D-glucosamine oligomers or with an unknown sugar-specific binding (ebulin-RP). Because ebulin l B-chain shares a higher degree of homology with that of ebulin-RP (54.44%) than with that of SEA (53.28%) most probably sialic acid specific binding tetrameric type 2 RIPs derived from type 2 RIPs such as SNLRPs and ebulin-RP by dimerization due to the appearance of a new cysteine that allowed the formation of a disulphide bridge between two B chains. The key alteration required is the mutation that changes the codon AGA (arginine) from ebulin l to AGT (serine) in ebulin-RP and the mutation that changes AGT to TGT (cysteine) in SEA. In this sense ebulin-RP can be considered an intermediate state between a galactose specific dimeric type 2 RIP and a sialic acid specific tetrameric type 2 RIP. It is worthy of mention that the dimeric SNAI' from *Sambucus nigra* also possesses the codon AGT.

Figure 3e compares the sequence logos of sugar-binding sites from type 2 RIP B-chains and lectins from *Sambucus* specific for Gal/GalNAc, sialic acid (Neu5Ac) and those of ebulin-RP and SNLRPs. In the D-galactose/N-acetyl-D-galactosamine specific binding proteins all key amino acids of the 1-alpha site involved in the interaction with D-galactose (D24, V25, N27, Q37 and W39) were conserved. The exceptions were the homodimeric lectins SELld and SNAld which did not conserve amino acids V25, N27, and W39 (SELld), probably because these proteins lack a functional 1-alpha site as has been suggested before [20]. Additionally, N27 changed to aspartic acid in sieboldin b and LECnig f. However only V25 and Q37 seem to be conserved in Neu5Ac specific binding proteins (V24 and Q36) and SNLRPs and ebulin-RP (V24 and Q36) supporting that 1-alpha site is not functional in such proteins. Key amino acids of the 2-gamma site involved in the interaction with D-galactose are also conserved (D235, K237, F249 and N256) allowing, however, the conserved substitutions of lysine by arginine and phenylalanine by tyrosine. Most of them are also conserved in the 2-gamma site from

sialic acid specific binding proteins probably allowing also galactose binding. Unlike galactose specific proteins, sialic acid specific binding proteins showed a lower degree of conservation in the key amino acids of the 2-gamma site and only Q231 was found in all the sequences, thus suggesting that several combinations of amino acids could allow the interaction with sialic acid. Finally is worthy of mention that all the key amino acids that could bind sugars in the group of ebulin-RP and SNLRPs seem to be conserved although this might be due to the limited number of sequences available.

4. Concluding remarks

RIPs have been implicated in a variety of processes, from antiviral, antifungal or plant defence to storage, programmed senescence, antifeedant, stress protection or development regulation. *Sambucus ebulus* is a rich source of RIPs and RIP-related lectins generated from multiple genes. The existence and maintenance of multigene RIP families may have a functional significance. The occurrence of multigene families has also been reported for *Ricinus communis*, *Saponaria officinalis*, *Viscum album*, *Iris x hollandica*, *Phytolacca americana*, etc. In *S. ebulus*, during evolution duplications and deletions have given rise to a variety of type 2 RIPs such as ebulin l, ebulin-RP or SEA and type 1 RIPs. In addition type 2 RIPs were converted in type B proteins by deletion of the RIP domain as is the case for the lectins SELlm, SElld, SELlf and SEAII. Some of these proteins are expressed at specific developmental states. The presence of all these proteins differing in their structure, enzymatic activity and sugar binding specificity may be related to different biological activities and could result in an advantage for the plant. The complexity and the large number of RIPs and lectins found in *Sambucus* makes this singular family a good model for studying the evolutionary process, the expression, distribution, and seasonal and developmental variations of these proteins with the goal of understanding their biological role.

Acknowledgements

This work was supported by grants BIO39/VA39/14 and BIO/VA17/15 to L.C. and funds from the University of Campania "Luigi Vanvitelli". We thank Chema Bassols for

helpful computer assistance and Judy Callaghan (Monash University, Melbourne, Australia) for correcting the manuscript.

LEGENDS OF FIGURES

Figure 1 Purification of ebulin-RP from *S. ebulus* leaves. Proteins isolated by ion-exchange chromatography onto CM-Sepharose and eluted with sodium chloride (panel a) were purified by Superdex 75 HiLoad chromatography (panel b). Ten micrograms of protein from the second peak eluted from Superdex 75 HiLoad chromatography (shaded area) were analysed by SDS-PAGE in the presence (+) and the absence (−) of 2-mercaptoethanol (2ME; inset) and further purified by affinity chromatography onto AT-Sepharose 6B (panel c) obtaining two fractions: an unbound fraction (ebulin-RP) and a bound fraction (ebulin l) which was eluted with 200 mM D-lactose. Both fractions were pooled, dialyzed and freeze-dried. The proteins were analyzed by SDS–PAGE in 12% gels either in the absence or the presence of 2-mercaptoethanol (2ME) and then stained with Coomassie brilliant blue (panel d) or blottet onto Immobilon membranes and then treated for glycan detection (panel e). The numbers indicate the corresponding size of the standards in kDa. eb: ebulin l. ebRP: ebulin-RP

Figure 2 Amino acid sequence alignment of ebulin-RP, SEA, ebulin l, SELlm and SELld preproteins. The leader peptides (blue, non-available for SEA), the A chains or A fragments (black), the connecting peptides (red) and the B chains (black) are indicated. Identical residues (*), conserved substitutions (:) and semiconserved substitutions (.) are reported. Key residues of the A-chain active site and the B-chain galactoside-binding clefts (represented in Figure 3e) are double underlined (green). The amino acid sequences obtained by Edman degradation (Table S1) as indicated in section

2.4 are underlined with a single line. Cysteines 44 (SEA) and 23 (SELld) involved in the linking between B-chains are boldfaced. Numbering refers to the position of the amino acids in the corresponding mature A and B chains. The cDNA sequence for ebulin RP was submitted to GenBank (accession number: MF170617).

Figure 3 Three-dimensional models of the sugar-binding sites from ebulin l, SEA and ebulin-RP. (a) The surfaces of the 1-alpha sugar-binding site from ebulin l (up left), SEA (down left) and ebulin-RP (down right) are represented. Electrostatic potential is indicated in red (negative charge), white (neutral) and blue (positive charge). For comparative purposes the position of D-galactose in the ebulin l 1-alpha site is indicated (sticks). (b) The 1-alpha site from ebulin l complexed with either D-galactose (sticks), sialic acid (black lines) or N-acetyl-D-glucosamine (red lines) is represented. The amino acids which bind the galactose molecule by either C-H-π interactions (W39) or both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks (see also Video S1). (c) The surfaces of the 2-gamma sugar-binding site from ebulin l (left), SEA (center) and ebulin-RP (right) are represented. Electrostatic potential is indicated in red (negative charge), white (neutral) and blue (positive charge). For comparative purposes the position of D-galactose in the ebulin l 2-gamma site is indicated (sticks). (d) The 2-gamma sugar-binding sites from ebulin l complexed with D-galactose (left, see also Video S2), SEA complexed with sialic acid (center, see also Video S3) and ebulin-RP complexed with N-acetyl-D-glucosamine (right, see also Video S4) are represented. The ligands and the amino acids which bind the sugar molecule by either C-H-π interactions (F249) or both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks. For comparative purposes the binding of the

other sugars is also shown: D-galactose (blue lines), sialic acid (black lines) and N-acetyl-D-glucosamine (red lines). (e) Comparison of sequence logo of sugar-binding sites 1-alpha (left) and 2-gamma (right) from type 2 RIP B-chains and lectins specific for D-galactose/N-acetyl-D-galactosamine, sialic acid (Neu5Ac) and ebulin-RP and SNLRPs. Letter height is proportional to the frequency of that amino acid at that position in the alignment respect to all the amino acids. Numbering refers to the position of the amino acids in ebulin l (Gal/GalNAc), SEA (Neu5Ac) and ebulin-RP (SNLRPs and ebulin-RP).

Figure 4 rRNA N-glycosidase activity of ebulin-RP on animal and yeast ribosomes. rRNA N-glycosidase activity was assayed as indicated in section 2.6. Each lane contained 3 μg of RNA isolated from either untreated (control) or RIP treated ribosomes from rabbit (a) or the yeast *Saccharomyces cerevisiae* (b). The arrows indicate the rabbit rRNAs and the fragments (Endo's fragment) released as a result of RIP action after acid aniline treatment (+). Numbers indicate the size of the standards in nucleotides.

Figure 5 Effects of ebulin-RP on nucleic acids. (a, b) Polynucleotide:adenine glycosidase activity of ebulin-RP, ebulin l, SEA and ricin, assayed on DNA and RNA. (a) Polynucleotide:adenine glycosidase activity of 3 μg of RIPs was assayed on salmon sperm DNA and Tobacco mosaic virus RNA as described in section 2.7 and the absorbance of the adenine released was measured at 260 nm. The data represents the mean of two duplicate experiments and the bars indicate the standard error of the mean. (b) Polynucleotide:adenine glycosidase activity of 3 μg of RIP was assayed on Tobacco mosaic virus RNA as indicated in section 2.7. Each lane contained 1 μg of RNA. Samples were treated (+) or not (−) with acid aniline. (c) Nicking activity of ebulin-RP,

ebulin l, SEA on pCR2.1 DNA. 200 ng/10 μL samples of plasmid DNA were incubated with 1 μg RIP in the absence or the presence of 5 mM Mg^{2+} as indicated in section 2.8. R, L, and S indicate relaxed, linear and supercoiled forms of pCR2.1, respectively. Linear: pCR2.1 DNA was linearized using EcoRI. The numbers indicate the size of the markers in nucleotides. C: control; eb: ebulin l; ebRP: ebulin-RP.

Figure 6 Binding and intracellular transport of ebulin-RP. (a) Effect of various sugars on viability of COLO 320 cells treated with ebulin-RP, ebulin l or SEA. Cells were preincubated with the different sugars for 1 h and then incubated with different concentrations of ebulin-RP, ebulin l and SEA for 24 h and cell viability was evaluated by a colorimetric assay as indicated in section 2.11. One representative experiment of three experiments performed in triplicate is shown. Symbols: (●), untreated; (■), D-glucose; (○), *N*-acetyl-D-glucosamine; (△), D-lactose; (▲), D-galactose; (□), *N*-acetyl-D-galactosamine. (b) Binding and uptake of fluorescent ebulin l and ebulin-RP in HeLa cells. HeLa cells were incubated at 4 ^{o}C for 1 h with Cy3-ebulin l or Cy3-ebulin RP to allow binding of the protein and then either fixed immediately (0 min) or incubated at 37 ^{o}C for 30 min before fixation. Bar, 50 μm. (c) Effect of Brefeldin A and ammonium chloride on viability of COLO320 cells treated with ricin, ebulin l or ebulin-RP. Cells were preincubated with Brefeldin A and ammonium chloride for 1 h and then incubated with different concentrations of ricin, ebulin l and ebulin-RP for 20 h and cell viability was evaluated by a colorimetric assay as indicated in section 2.11. One representative experiment of two experiments performed in triplicate is shown. Symbols: (●), untreated; (○), ammonium chloride; (△), Brefeldin A. Solid line, ricin; dotted line, ebulin l; dashed line, ebulin-RP.

Figure 7 Induction of cytotoxicity and apoptosis on COLO 320 and HeLa cells by ebulin-RP (a) rRNA N-glycosidase activity of ebulin-RP on RNA from COLO 320 cells. rRNA N-glycosidase activity was assayed as indicated in section 2.6. Each lane contained 1 μg of RNA isolated from either untreated cells or cells incubated with 1 μM ebulin-RP for 48 h. The arrow indicates the RNA fragment released as a result of RIP action upon acid aniline treatment. (b) Effect of ebulin-RP on internucleosomal DNA fragmentation. COLO 320 cells were incubated in the absence or presence of 1 μM of ebulin-RP for 48 and 72 h. The DNA was isolated and 4 μg was electrophoresed as indicated in section 2.9. The numbers indicate the corresponding size of the standards (λDNA HindIII/EcoRI) in Kb. (c) Caspase-3/7 activation in COLO 320 and HeLa cells treated with ebulin-RP for 48 h. Activity is expressed as the percentage of control values obtained from cultures grown in the absence of RIP. Data represent the mean ± SD of two experiments performed in duplicate. (d, e) Effect of Z-VAD and Nec-1 on cytotoxicity of ebulin-RP on HeLa cells. Cells were left untreated or preincubated with Z-VAD or Nec-1 for 3 h and then incubated with different concentrations of ebulin-RP for 48 h and cell viability was evaluated by a colorimetric assay as indicated in section 2.11 (d). Data represent the mean ± SD of two experiments performed in triplicate. Symbols: ●, untreated; ○, +Z-VAD and ■, +Nec-1. (e) Phase contrast microscopy and double staining with AnnexinV(green)/PI(red), followed by fluorescence microscopy at 48 h after treatment with 1 μM ebulin-RP. Bar, 100 μm.

Figura 8 Molecular phylogenetic analysis by the Maximum Likelihood method of the type 2 RIP B-chains and related lectins from the genus *Sambucus*. The evolutionary history was inferred as indicated in section 2.18. The sequences of the type 2 RIP B-chains from *Cinnamomum camphora* (accession code Q94BW5), *Abrus*

precatorius (AAA32624), *Ricinus communis* (XP002534649) and *Viscum album* (P81446) were used as outgroup. The name of the protein, the species and the accession number are indicated. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. All the sequences were retrieved and processed as indicated in section 2.14 and 2.15 Dimeric (A-B) and tetrameric (A-B-B-A) RIPs, and monomeric (B) and dimeric (B-B) lectins from *Sambucus*, and the specificity by sugars are also indicated.

Reference List

[1] J. Schrot, A. Weng, M.F. Melzig, Ribosome-inactivating and related proteins, Toxins 7 (2015) 1556-1615.

[2] A. Di Maro, L. Citores, R. Russo, R. Iglesias, J.M. Ferreras, Sequence comparison and phylogenetic analysis by the Maximum Likelihood method of ribosome-inactivating proteins from angiosperms, Plant Mol.Biol. 85 (2014) 575-588.

[3] Y. Endo, K. Tsurugi, The RNA N-glycosidase activity of ricin A-chain. The characteristics of the enzymatic activity of ricin A-chain with ribosomes and with rRNA, J.Biol.Chem. 263 (1988) 8735-8739.

[4] L. Barbieri, P. Valbonesi, E. Bonora, P. Gorini, A. Bolognesi, F. Stirpe, Polynucleotide:adenosine glycosidase activity of ribosome-inactivating proteins: effect on DNA, RNA and poly(A), Nucleic Acids Res. 25 (1997) 518-522.

[5] N.R. Shih, K.A. McDonald, A.P. Jackman, T. Girbes, R. Iglesias, Bifunctional plant defence enzymes with chitinase and ribosome inactivating activities from Trichosanthes kirilowii cell cultures, Plant Sci. 130 (1997) 145-150.

[6] A. Ruggiero, M.A. Di, V. Severino, A. Chambery, R. Berisio, Crystal structure of PD-L1, a ribosome inactivating protein from *Phytolacca dioica* L. leaves with the property to induce DNA cleavage, Biopolymers 91 (2009) 1135-1142.

[7] S. Lombard, M.E. Helmy, G. Pieroni, Lipolytic activity of ricin from *Ricinus sanguineus* and *Ricinus communis* on neutral lipids, Biochem J. 358 (2001) 773-781.

[8] A.N. Mak, Y.T. Wong, Y.J. An, S.S. Chan, K.H. Sze, S.W. Au, K.B. Wong, P.C. Shaw, Structure-function study of maize ribosome-inactivating protein: implications for the internal inactivation region and the sole glutamate in the active site, Nucleic Acids Research. 35 (2007) 6259-6267.

[9] J.D. Zaeytijd, E.J.M.Van Damme, Extensive Evolution of Cereal Ribosome-Inactivating Proteins Translates into Unique Structural Features, Activation Mechanisms, and Physiological Roles, Toxins 9 (2017) 123.

[10] M. Zeng, M. Zheng, D. Lu, J. Wang, W. Jiang, O. Sha, Anti-tumor activities and apoptotic mechanism of ribosome-inactivating proteins, Chin. J. Cancer 34 (2015) 30.

[11] S.D. Xiong, K. Yu, X.H. Liu, L.H. Yin, A. Kirschenbaum, S. Yao, G. Narla, A. DiFeo, J.B. Wu, Y. Yuan, S.M. Ho, Y.W. Lam, A.C. Levine, Ribosome-inactivating proteins isolated from dietary bitter melon induce apoptosis and inhibit histone deacetylase-1 selectively in premalignant and malignant prostate cancer cells, Int. J. Cancer 125 (2009) 774-782.
[12] E.F. Fang, T.B. Ng, P.C. Shaw, R.N. Wong, Recent progress in medicinal investigations on trichosanthin and other ribosome inactivating proteins from the plant genus *Trichosanthes*, Curr. Med. Chem. 18 (2011) 4410-4417.
[13] L. Polito, M. Bortolotti, M. Pedrazzi, A. Bolognesi, Immunotoxins and other conjugates containing saporin-s6 for cancer therapy, Toxins 3 (2011) 697-720.
[14] J.M. Ferreras, L. Citores, R. Iglesias, P. Jimenez, T. Girbes, Use of ribosome-inactivating proteins from *Sambucus* for the construction of immunotoxins and conjugates for cancer therapy, Toxins 3 (2011) 420-441.
[15] L. Polito, A. Djemil, M. Bortolotti, Plant toxin-based immunotoxins for cancer therapy: a short overview, Biomedicines 4 (2016) 12.
[16] R. Di, N.E. Tumer, Pokeweed antiviral protein: its cytotoxicity mechanism and applications in plant disease resistance, Toxins 7 (2015) 755-772.
[17] T. Girbés, L. Citores, R. Iglesias, J.M. Ferreras, R. Muñoz, M.A. Rojo, F.J. Arias, J.R. García, E. Méndez, M. Calonge, Ebulin 1, a nontoxic novel type 2 ribosome-inactivating protein from *Sambucus ebulus* L. leaves, J. Biol. Chem. 268 (1993) 18195-18199.
[18] T. Girbés, L. Citores, J. Miguel Ferreras, M. Angeles Rojo, R. Iglesias, R. Muñoz, F. Javier Arias, M. Calonge, J. Ramón García, E. Méndez, Isolation and partial characterization of nigrin b, a non-toxic novel type 2 ribosome-inactivating protein from the bark of.*Sambucus nigra* L, Plant Mol. Biol. 22 (1993) 1181-1186.
[19] E.J.M. Van Damme, S. Roy, A. Barre, P. Rougé, F. Van Leuven, W.J. Peumans, The major elderberry (*Sambucus nigra*) fruit protein is a lectin derived from a truncated type 2 ribosome-inactivating protein, Plant J. 12 (1997) 1251-1260.
[20] M.A. Rojo, L. Citores, F.J. Arias, J.M. Ferreras, P. Jimenez, T. Girbés, cDNA molecular cloning and seasonal acumulation of an ebulin l-related dimeric lectin of dwarf elder (*Sambucus ebulus* L.) leaves, Int. J. Biochem. Cell Biol. 35 (2003) 1061-1065.
[21] L. Citores, M.A. Rojo, P. Jiménez, J.M. Ferreras, R. Iglesias, I. Aranguez, T. Girbés, Transient occurrence of an ebulin-related d-galactose-lectin in shoots of *Sambucus ebulus* L, Phytochemistry 69 (2008) 857-864.
[22] E.J.M. Van Damme, S. Roy, A. Barre, L. Citores, K. Mostafapous, P. Rougé, F. Van Leuven, T. Girbés, I.J. Goldstein, W.J. Peumans, Elderberry (*Sambucus Nigra*) Bark Contains two Structurally Different Neusac(α2,6)Gal/Galnac-Binding Type 2 Ribosome-Inactivating Proteins, Eur. J. Biochem. 245 (1997) 648-655.
[23] E.J.M. Van Damme, A. Barre, P. Rougé, F. Van Leuven, W.J. Peumans, The NeuAc(α-2,6)-Gal/GalNAc-binding lectin from elderberry (*Sambucus Nigra*) bark, a type-2 Ribosome-Inactivating Protein with an unusual specificity and structure, Eur. J. Biochem. 235 (1996) 128-137.
[24] F.M. de Benito, L. Citores, R. Iglesias, J.M. Ferreras, E. Camafeita, E. Méndez, T. Girbés, Isolation and partial characterization of a novel and uncommon two-chain 64-kDa ribosome-inactivating protein from the bark of elder (*Sambucus nigra* L.), FEBS Lett. 413 (1997) 85-91.
[25] C. Shang, E.J.M. Van Damme, Comparative analysis of carbohydrate binding properties of *Sambucus nigra* lectins and ribosome-inactivating proteins, Glycoconj. J. 31 (2014) 345-354.

[26] J.M. Ferreras, L. Citores, R. Iglesias, P. Jiménez, T. Girbés, *Sambucus* ribosome-inactivating proteins and lectins, in: J.M. Lord, M.R. Hartley (Eds.) Toxic Plant Proteins-Series Plant Cell Monographs, Vol. 18, Springer, Heidelberg, 2010, pp. 107-131.

[27] J.M. Ferreras, L. Citores, R. Iglesias, A.M. Souza, P. Jimenez, M.J. Gayoso, T. Girbes, Occurrence of the type two ribosome-inactivating protein nigrin b in elderberry (*Sambucus nigra* L.) bark., Food Res. Int. 44 (2011) 2798-2805.

[28] M.G. Battelli, L. Citores, L. Buonamici, J.M. Ferreras, F.M. de Benito, F. Stirpe, T. Girbés, Toxicity and cytotoxicity of nigrin b, a two-chain ribosome-inactivating protein from Sambucus nigra: comparison with ricin, Arch. Toxicol. 71 (1997) 360-364.

[29] J.M. Pascal, P.J. Day, A.F. Monzingo, S.R. Ernst, J.D. Robertus, R. Iglesias, Y. Pérez, J.M. Férreras, L. Citores, T. Girbés, 2.8-Å crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin l, Proteins 43 (2001) 319-326.

[30] F.M. de Benito, L. Citores, R. Iglesias, J.M. Ferreras, F. Soriano, J. Arias, E. Mendez, T. Girbes, Ebulitins: A new family of type 1 ribosome-inactivating proteins (rRNA N-glycosidases) from leaves of *Sambucus ebulus* L. that coexist with the type 2 ribosome-inactivating protein ebulin 1, FEBS Lett. 360 (1995) 299-302.

[31] L. Citores, F.M. de Benito, R. Iglesias, J.M. Ferreras, P. Argüeso, P. Jiménez, E. Méndez, T. Girbés, Presence of polymerized and free forms of the non-toxic type 2 ribosome-inactivating protein ebulin and a structurally related new homodimeric lectin in fruits of *Sambucus ebulus* L, Planta, 204 (1998) 310-317.

[32] R. Iglesias, L. Citores, J.M. Ferreras, Y. Pérez, P. Jiménez, M.J. Gayoso, S. Olsnes, R. Tamburino, A. Di Maro, A. Parente, T. Girbés, Sialic acid-binding dwarf elder four-chain lectin displays nucleic acid N-glycosidase activity, Biochimie 92 (2010) 71-80.

[33] L. Citores, F.M. De Benito, R. Iglesias, J.M. Ferreras, P. Argüeso, P. Jiménez, A. Testera, E. Camafeita, E. Méndez, T. Girbés, Characterization of a new non-toxic two-chain ribosome-inactivating protein and a structurally-related lectin from rhizomes of dwarf elder (*Sambucus ebulus* L.), Cell. Mol. Biol. 43 (1997) 485-499.

[34] A. Parente, B. Conforto, M.A. Di, A. Chambery, L.P. De, A. Bolognesi, M. Iriti, F. Faoro, Type 1 ribosome-inactivating proteins from *Phytolacca dioica* L. leaves: differential seasonal and age expression, and cellular localization, Planta 228 (2008) 963-975.

[35] R. Iglesias, Citores, L. and Ferreras, J. M, Ribosomal RNA N-glycosylase activity assay of ribosome-inactivating proteins, Bio Protoc., 7 (2017) e2180.

[36] A. Di Maro, A. Chambery, A. Daniele, P. Casoria, A. Parente, Isolation and characterization of heterotepalins, type 1 ribosome-inactivating proteins from *Phytolacca heterotepala* leaves, Phytochemistry 68 (2007) 767-776.

[37] P.L. Huang, H.C. Chen, H.F. Kung, P.L. Huang, P. Huang, H.I. Huang, S. Lee-Huang, Anti-HIV plant proteins catalyze topological changes of DNA into inactive forms, Biofactors 4 (1992) 37-41.

[38] M.A. Rojo, M. Yato, N. Ishii-Minami, E. Minami, H. Kaku, L. Citores, T. Girbés, N. Shibuya, Isolation, cDNA cloning, biological properties, and carbohydrate binding specificity of sieboldin-b, a type II ribosome-inactivating protein from the bark of japanese elderberry (*Sambucus sieboldiana*), Arch. Biochem. Biophys 340 (1997) 185-194.

[39] H. Kaku, Y. Tanaka, K. Tazaki, E. Minami, H. Mizuno, N. Shibuya, Sialylated oligosaccharide-specific plant lectin from japanese elderberry (*Sambucus*

sieboldiana) bark tissue has a homologous structure to type II ribosome-inactivating proteins, ricin and abrin: cDNA cloning and molecular modeling study, J. Biol. Chem. 271 (1996) 1480-1485.
[40] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular evolutionary genetics analysis Version 6.0, Mol. Biol. Evol. 30 (2013) 2725-2729.
[41] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences, Nucleic Acids Res. 18 (1990) 6097-6100.
[42] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188-1190.
[43] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (2010) 725-738.
[44] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785-2791.
[45] S.M.D. Rizvi, S. Shakil, M. Haneef, A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians, Excli J. 12 (2013) 831-857.
[46] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Bioinformatics 8 (1992) 275-282.
[47] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, evolution 39 (1985) 783-791.
[48] E.J.M. Van Damme, A. Barre, P. Rougé, F. Van Leuven, W.J. Peumans, Isolation and molecular cloning of a novel type 2 ribosome-inactivating protein with an inactive B chain from elderberry (*Sambucus nigra*) bark, J. Biol. Chem. 272 (1997) 8353-8360.
[49] H. Kaku, H. Kaneko, N. Minamihara, K. Iwata, E.T. Jordan, M.A. Rojo, N. Minami-Ishii, E. Minami, S. Hisajima, N. Shibuya, Elderberry bark lectins evolved to recognize Neu5Acα2,6Gal/GalNAc sequence from a Gal/GalNAc binding lectin through the substitution of amino-acid residues critical for the binding to sialic acid, J. Biochem. 142 (2007) 393-401.
[50] M. de Sousa, L.M. Roberts, J.M. Lord, Restoration of lectin activity to an inactive abrin B chain by substitution and mutation of the 2γ subdomain, Eur. J. Biochem. 260 (1999) 355-361.
[51] X.P. Li, N. Tumer, Differences in ribosome binding and sarcin/ricin loop depurination by shiga and ricin holotoxins, Toxins 9 (2017) 133.
[52] R. Iglesias, L. Citores, A. Di Maro, J.M. Ferreras, Biological activities of the antiviral protein BE27 from sugar beet (*Beta vulgaris* L.), Planta 241 (2015) 421-433.
[53] S. Aceto, M.A. Di, B. Conforto, G.G. Siniscalco, A. Parente, B.P. Delli, L. Gaudio, Nicking activity on pBR322 DNA of ribosome inactivating proteins from *Phytolacca dioica* L. leaves, Biol.Chem. 386 (2005) 307-317.
[54] V.M. de, A. Lombardi, R. Caliandro, M.S. Fabbrini, Ribosome-inactivating proteins: from plant defense to tumor attack, Toxins 2 (2010) 2699-2737.
[55] M.K. Das, R.S. Sharma, V. Mishra, A cytotoxic type-2 ribosome inactivating protein (from leafless mistletoe) lacking sugar binding activity, Int. J. Biol. Macromol. 49 (2011) 1096-1103.
[56] R.A. Spooner, J.M. Lord, Ricin trafficking in cells, Toxins 7 (2015) 49-65.

Table 1 Effects of ebulin-RP, ebulin l, SEA and ricin on protein synthesis and cell viability. Translation assays were carried out using rabbit reticulocytes lysate as a cell-free system, as indicated in section 2.19. Cytotoxicity of ebulin-RP, ebulin l, SEA and ricin to cancer cells and hMSC cells was determined as loss of cell viability. $3 \cdot 10^3$ cells in 0.1 mL medium were seeded in 96-well plates and incubated for 48 h at 37°C under 5% CO_2 in the presence of varying concentrations of the toxins. Cell viability was measured in a colorimetric assay as indicated in section 2.11.

	IC_{50} (M)						
	Cell free	Cell cultures					
	Rabbit	HeLa	COLO320	HCT15	Raji	B16 mel4A5	hMSC
Ebulin-RP	$9.5 \cdot 10^{-11}$	$1 \cdot 10^{-6}$	$2.5 \cdot 10^{-6}$	$>5 \cdot 10^{-6}$	$2.2 \cdot 10^{-6}$	$>5 \cdot 10^{-6}$	$8 \cdot 10^{-6}$
Ebulin l	$8 \cdot 10^{-11}$	$1.1 \cdot 10^{-7}$	$2 \cdot 10^{-8}$	$9 \cdot 10^{-8}$	$1.5 \cdot 10^{-7}$	$7.5 \cdot 10^{-7}$	$8 \cdot 10^{-6}$
SEA	$1 \cdot 10^{-9}$	$6 \cdot 10^{-7}$	$2 \cdot 10^{-7}$	nd	$7 \cdot 10^{-7}$	$1.3 \cdot 10^{-6}$	$>5.10^{-6}$
Ricin	$2 \cdot 10^{-11}$	$6 \cdot 10^{-13}$	$1.4 \cdot 10^{-13}$	$2 \cdot 10^{-12}$	$6 \cdot 10^{-12}$	$8 \cdot 10^{-10}$	$1 \cdot 10^{-13}$

Figure 1

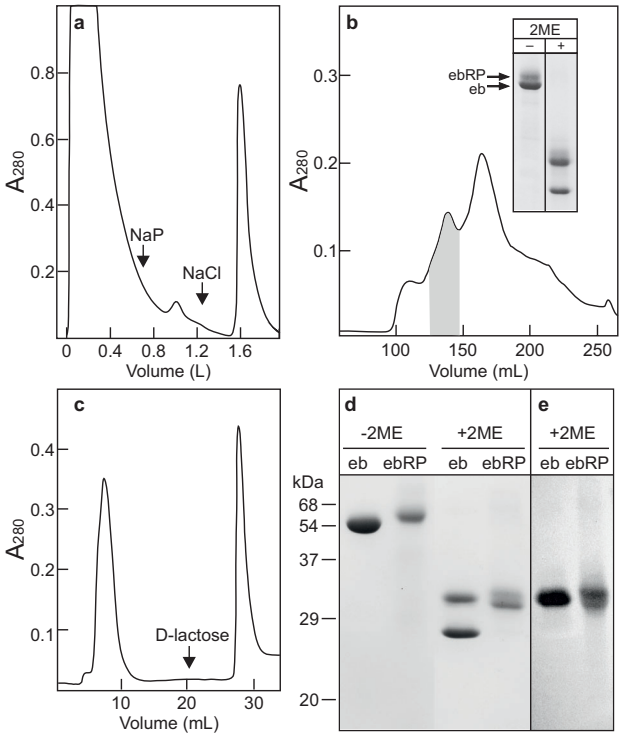

a
0.8
0.6
0.4
0.2
A280
NaP
NaCl
0
0.4
0.8
1.2
1.6
Volume (L)
b
0.3
0.2
0.1
A280
2ME
–
+
ebRP
eb
100
150
200
250
Volume (mL)
c
0.4
0.3
0.2
0.1
A280
D-lactose
10
20
30
Volume (mL)
d
-2ME
+2ME
eb
ebRP
eb
ebRP
kDa
68
54
37
29
20
e
+2ME
eb
ebRP

Figure 2


Ebulin-RP MRVVA---TILYLVVLSICGLGTHGSRVTAPAYPSISLNLAGAQWISYRNFLGALQDLVTRRSDTALDLPVLKPERQVSVENRFVLTRLTN---PSGDTV 68
SEA XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTPPVYPSVSFNLAGAD--TYGPFLRELREKVILGNHTAFDLPVLNPESQILESDRFVLVPLTNSTNSSGDTV 70
Ebulin l MRVVKAAMLYLHIVVLAIYSVGIQG-----IDYPSVSFNLAGAKSTTYRDFLKNLRDRVATGTYEVNGLPVLRRESEVQVKNRFVLVRLTN---YNGDTV 67
SELlm MRVVSGAMLHLYIVVFAICSVGIQG-----IDYPSVSFNWAGALSATYRDFLR----------
SELld MRVVAAAMLGLYIVVVAICSVGIQG-----RDYPSVSFKF----------
TLAIDVVNLYVVAFRANGTSYFFKDSTKIENDNLFQDTTRKNLTFTGNYISLESQAGTHRESISLGPYPLAQAILSLSRYK-SGGDTKSLAKALLVVIQM 167
SEA TLAIDVVNLNVVAFSSNGRSYFFSGSSAVQRDNLFVDTTQEDLNFTGNYTSLELHVGVGRVDIPLGLNSLAQAISSLWTYTLSAGDTKPLARGLLVVIQM 170
Ebulin l TSAVDVTNLYLVAFSANGNSYFFKDATELQKSNLFLGTTQHTLSFTGNYDNLETAAGTRRESIELGPNPLDGAITSLWYDG-------GVARSLLVLIQM 160
SELlm
SELld
VSEAARFRYIELRIWTSITDANEFTPDPLMLSMENNWSSKSKEIQGATPG-GTFAQALQLKDQGNNPINVTNFKRLFQLTYIAVLLYGCRPTTS---SSY 261
SEA VSEAARFRYIELRILTSITDASEFTPDPLMLSMENNWSSMSSEIQQTQPG-GIFGGVVQLRDQRNNSIEVTNSRRLFQLTYVAVLLHGCLAVTT---SSY 264
Ebulin l VPEAARFRYIEQEVRRSLQQLTSFTPNALMLSMENNWSSMSLEVQLSGDNVSPFSGTVQLQNYDHTPRLVDNFEELYKITGIAILLFRCVATKTTHNAIR 254
SELlm
SELld NAIR
SNNAIAAQIIKMPVFRVGEYDEVCTV-VDVTRRISGRDGLCVGVRSGQYNDGTPVQLWSCGQQSNQQWTFRTDRTIRSLGKCLTNSGGSYGNSAVIYNCD 80
SEA SNNAIAGQKIKMSVFRGVEYEKVCSV-VEVTRRISGWDGLCVDVSDDHYIDGNPVQLSPCGNQCTQQWTFRTDGTIRWLGKCLTAS-----SSVMINDCN 75
Ebulin l MPHVLVGEDNKFN-----DGET-CAIPAPFTRRIVGRDGLCVDVRNGYDTDGTPIQLWPCGTQRNQQWTFYNDKTIRSMGKCMTANGLNSGSYIMITDCS 81
SELlm MPNVFVGEDIKYN-----DGVTY-TVSASFTGNIIGRDGLCVDVRNGYDTDGTPIQLWPCGSQRNQRWTFYEDGTIQSMRKCMTANGLNSGSYIMIFNCS 81
SELld MPHDFVGEDNKFN-----DGE-------RFTRQIIGRDGLCVDARTGCD-----IQLRPCGSQTSQQWTFYEDGTIRSMGKCMTANGFNSGSYIMIFDCS 70
TAIPGATKWVLSIDGTITNPASGLVLTAPQAAQGTTLLLQNNVHAASQSWSVG-NVKPLVTFIVGYNQMCSQGNTENNPVRLEDCVLNRSEQKWALYGDG 179
SEA TVPPEDTKWVVSIDGTITNPSSGLVLTAPQTLEGTALSPENNIHAAIQGWTVG-DVEPLVTSIVGYKQMCLREDGENNFVWMEHCVLNRTEQEWALYGDG 174
Ebulin l TAAEDATKWEVLIDGSIINPSSGLVMTAPSGASRTTLLLENNIHAASQGWTVSNDVQPIATLIVGYNEMCLQANGENNNVWMEDCDVTSVQQQWALFDDR 181
SELlm TAAEDAIKWEVTIDGSIINPSSGLVMTAPRAASRTILLLENNILAASQGWTVSNDIQPNVTSIVGYKEMCLQANGENNNVWMENCDG-SVQQQWALFGDR 180
SELld SATENATKWEVTIDGSIINPSSGLVMTAPSGASGTTLVLENNILAASQGWTVSNDVQPNVTLMVGYNNMCLKANGENNKVWMENCVSTSVQQQWALFGDR 170
TIRVNSNRSLCVTTEGHSTSDLIIILKCQGLSNQRWVFNTNGTISNPNAKLVMEVRQSNVSLRQVIIYHPTGNANQQWITSTHQP 264
SEA TIRVNSNRSLCVTSQYHEPSDLIIILKCQGSSNQRWVFNTNGTISNPNTTKVMDVAQANVSLRKIILYRPNGESNHQWITTTHPA 259
Ebulin l TIRVNNSRGLCVTSNGYVSKDLIVIRKCQGLATQRWFFNSDGSVVNLKSTRVMDVKESDVSLQEVIIFPATGNPNQQWRTQVPQI 266
SELlm TIRVNSSRGLCVTSNGYVSKDLIIILKCQGLASQRWLFNSDGSVVNPNTTLVMDVRRSNVSLREIILYPSTGNPNQKWRTEVLPS 265
SELld TIRVNSSRDLCLTSRGYVSKDIIISFTCQGLPTQRWFFKSDGTIVNPNTTLVMDVKGSDVSLREIIIYPSVGSSNQQWKTEVLPS 255

Figure3

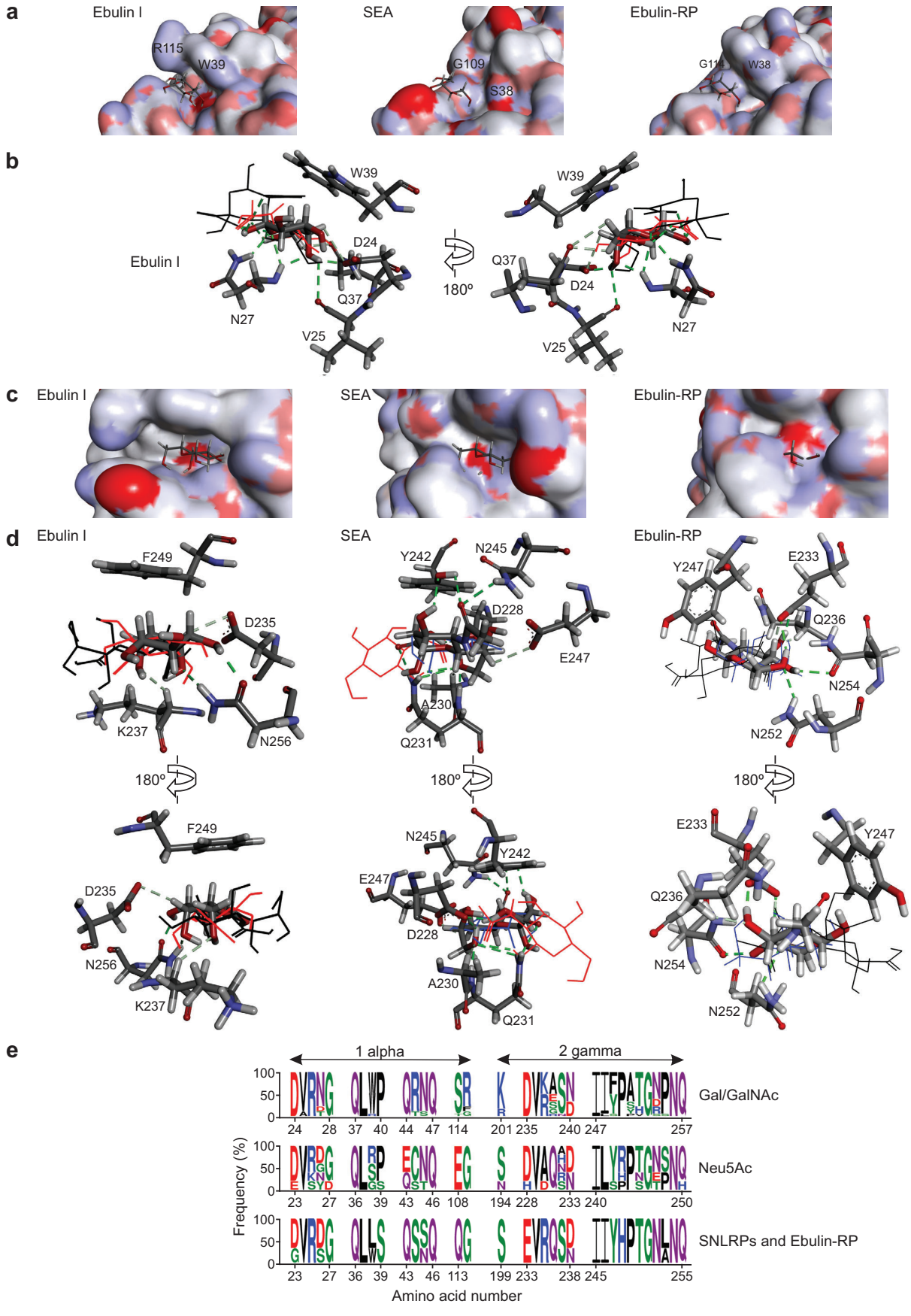

a
Ebulin I
R115
W39
SEA
G109
S38
Ebulin-RP
G110
W38
b
Ebulin I
W39
D24
Q37
N27
V25
180°
W39
Q37
D24
N27
V25
c
Ebulin I
SEA
Ebulin-RP
d
Ebulin I
F249
D235
K237
N256
180°
F249
D235
N256
K237
SEA
Y242
N245
D228
E247
A230
Q231
180°
N245
Y242
E247
D228
A230
Q231
Ebulin-RP
Y247
E233
Q236
N254
N252
180°
E233
Y247
Q236
N254
N252
e
1 alpha
2 gamma
Gal/GalNAc
Neu5Ac
SNLRPs and Ebulin-RP
Frequency (%)
Amino acid number

Figure 4

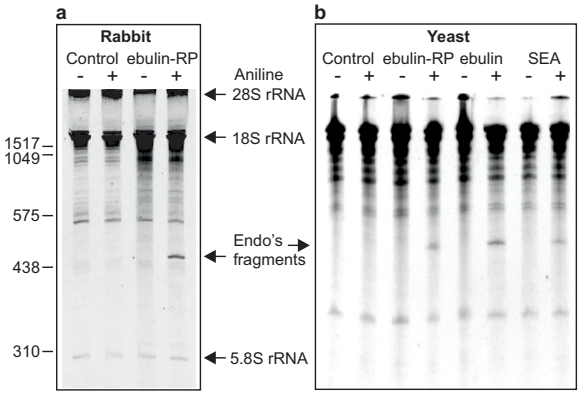

a
Rabbit
Control ebulin-RP
- + - +
1517
1049
575
438
310
Aniline
28S rRNA
18S rRNA
Endo's
fragments
5.8S rRNA
b
Yeast
Control ebulin-RP ebulin SEA
- + - + - + - +

Figure 5

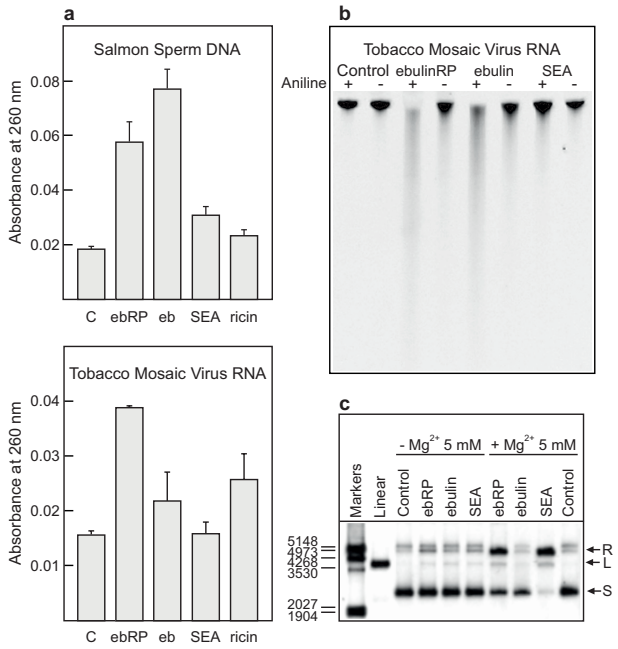

a
Salmon Sperm DNA
Absorbance at 260 nm
0.08
0.06
0.04
0.02
C ebRP eb SEA ricin
Tobacco Mosaic Virus RNA
Absorbance at 260 nm
0.04
0.03
0.02
0.01
C ebRP eb SEA ricin
b
Tobacco Mosaic Virus RNA
Aniline
Control ebulinRP ebulin SEA
+ - + - + - + -
c
- Mg2+ 5 mM
+ Mg2+ 5 mM
Markers
Linear
Control
ebRP
ebulin
SEA
ebRP
ebulin
SEA
Control
5148
4973
4268
3530
2027
1904
R
L
S

Figure6

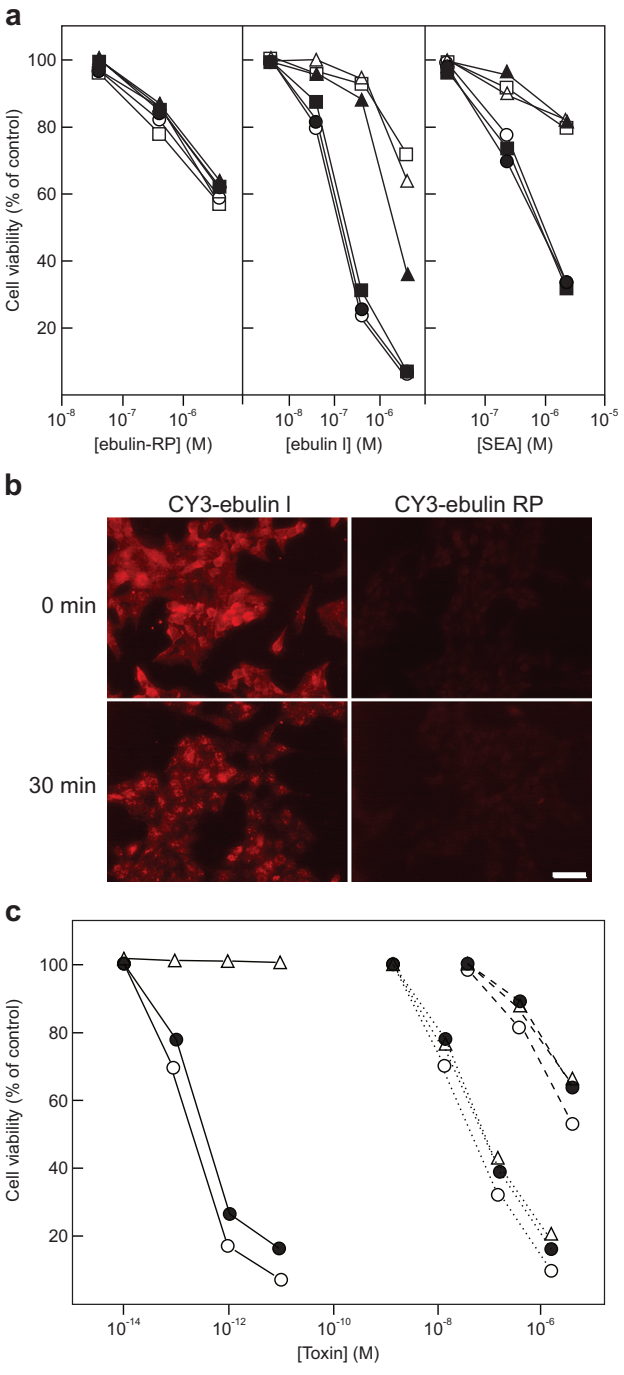

a
Cell viability (% of control)
100
80
60
40
20
10^{-8} 10^{-7} 10^{-6}
[ebulin-RP] (M)
10^{-8} 10^{-7} 10^{-6}
[ebulin l] (M)
10^{-7} 10^{-6} 10^{-5}
[SEA] (M)
b
CY3-ebulin l
CY3-ebulin RP
0 min
30 min
c
Cell viability (% of control)
100
80
60
40
20
10^{-14} 10^{-12} 10^{-10} 10^{-8} 10^{-6}
[Toxin] (M)

Figure 7

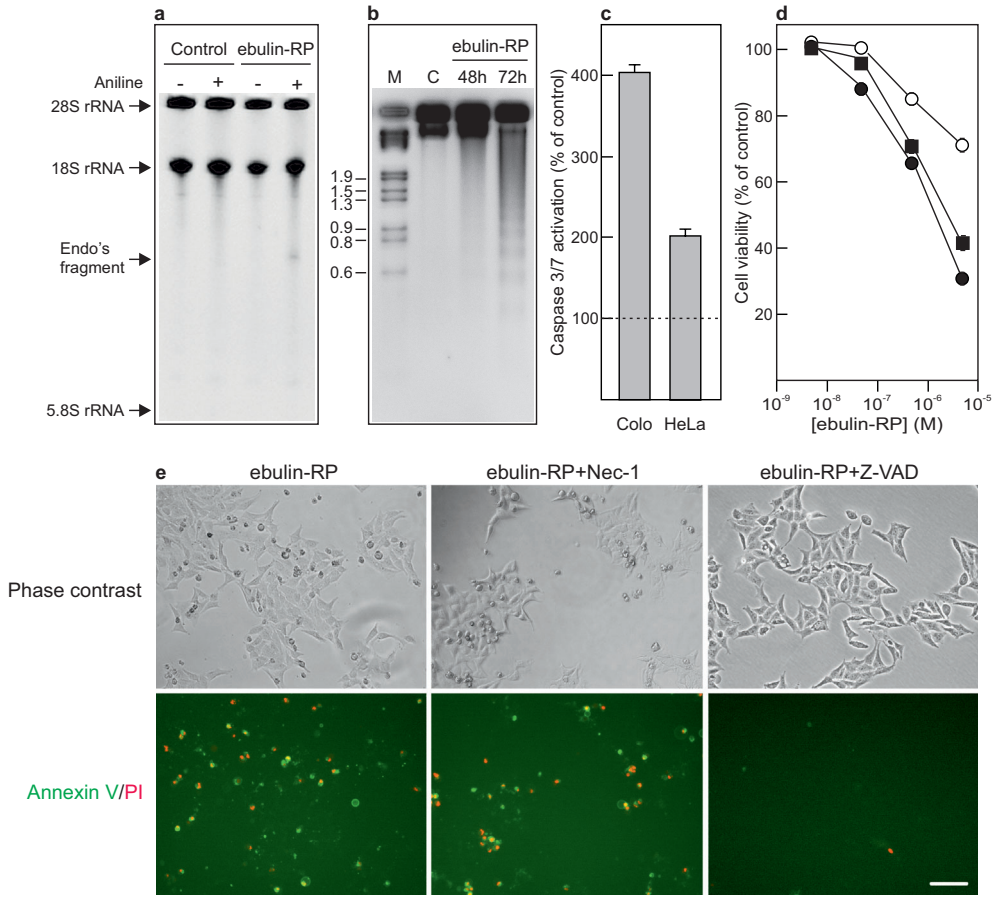
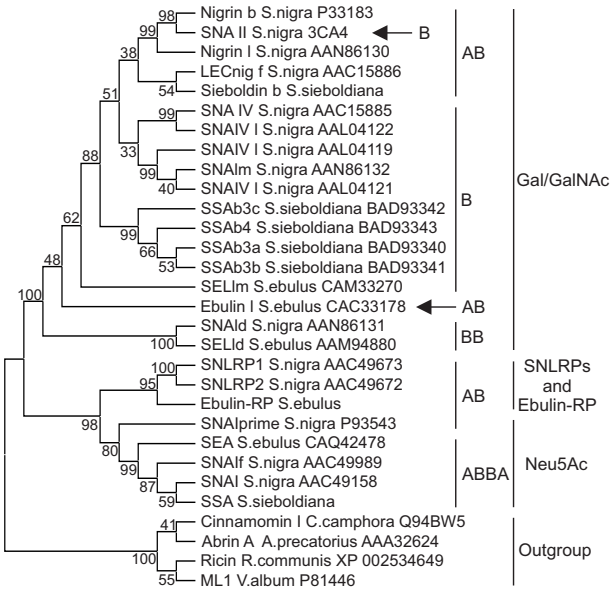

a
Control
ebulin-RP
Aniline
- + - +
28S rRNA
18S rRNA
Endo's fragment
5.8S rRNA
b
ebulin-RP
M C 48h 72h
1.9
1.5
1.3
0.9
0.8
0.6
c
Caspase 3/7 activation (% of control)
400
300
200
100
Colo HeLa
d
Cell viability (% of control)
100
80
60
40
20
10^{-9} 10^{-8} 10^{-7} 10^{-6} 10^{-5}
[ebulin-RP] (M)
e
ebulin-RP
ebulin-RP+Nec-1
ebulin-RP+Z-VAD
Phase contrast
Annexin V/PI

Figure 8


Nigrin b S.nigra P33183
SNA II S.nigra 3CA4
B
Nigrin I S.nigra AAN86130
LECnig f S.nigra AAC15886
Sieboldin b S.sieboldiana
SNA IV S.nigra AAC15885
SNAIV I S.nigra AAL04122
SNAIV I S.nigra AAL04119
SNAIm S.nigra AAN86132
SNAIV I S.nigra AAL04121
SSAb3c S.sieboldiana BAD93342
SSAb4 S.sieboldiana BAD93343
SSAb3a S.sieboldiana BAD93340
SSAb3b S.sieboldiana BAD93341
SELIm S.ebulus CAM33270
Ebulin I S.ebulus CAC33178
SNAId S.nigra AAN86131
SELId S.ebulus AAM94880
SNLRP1 S.nigra AAC49673
SNLRP2 S.nigra AAC49672
Ebulin-RP S.ebulus
SNAIprime S.nigra P93543
SEA S.ebulus CAQ42478
SNAIf S.nigra AAC49989
SNAI S.nigra AAC49158
SSA S.sieboldiana
Cinnamomin I C.camphora Q94BW5
Abrin A A.precatorius AAA32624
Ricin R.communis XP 002534649
ML1 V.album P81446
AB
B
AB
BB
AB
ABBA
Gal/GalNAc
SNLRPs and Ebulin-RP
Neu5Ac
Outgroup

SUPPLEMENTARY MATERIAL

EBULIN-RP, A NOVEL MEMBER OF THE EBULIN GENE FAMILY WITH LOW CYTOTOXICITY AS A RESULT OF DEFICIENT SUGAR BINDING DOMAINS

Rosario Iglesias[a], J. Miguel Ferreras[a], Antimo Di Maro[b], Lucía Citores [a]

[a] Department of Biochemistry and Molecular Biology and Physiology, Faculty of Sciences, University of Valladolid, E-47011 Valladolid, Spain

[b] Department of Environmental, Biological and Pharmaceutical Sciences and Technologies, University of Campania "Luigi Vanvitelli", I-81100 Caserta, Italy

List of contents:

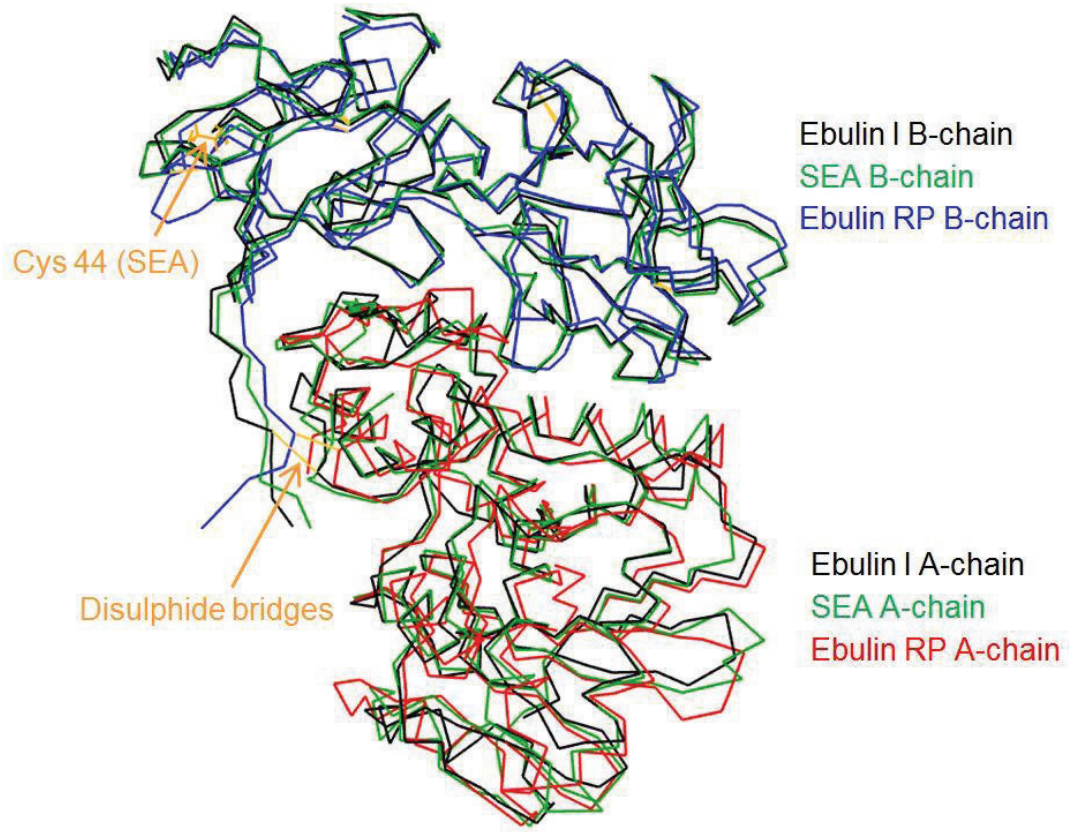


Figure S1 Structure of ebulin-RP compared with ebulin l and SEA. The three-dimensional structural modelling was carried out on the I-TASSER server and the figure was generated with DS Visualizer 3.5. The superposition of α-carbon traces of ebulin l (Accession number 1HWN) (black), SEA (green), ebulin-RP A-chain (red) and ebulin-RP B-chain (blue) is represented. The disulphide bridges which link the A and B chains and the cysteine 44 that links two B chains from SEA are also represented (orange).

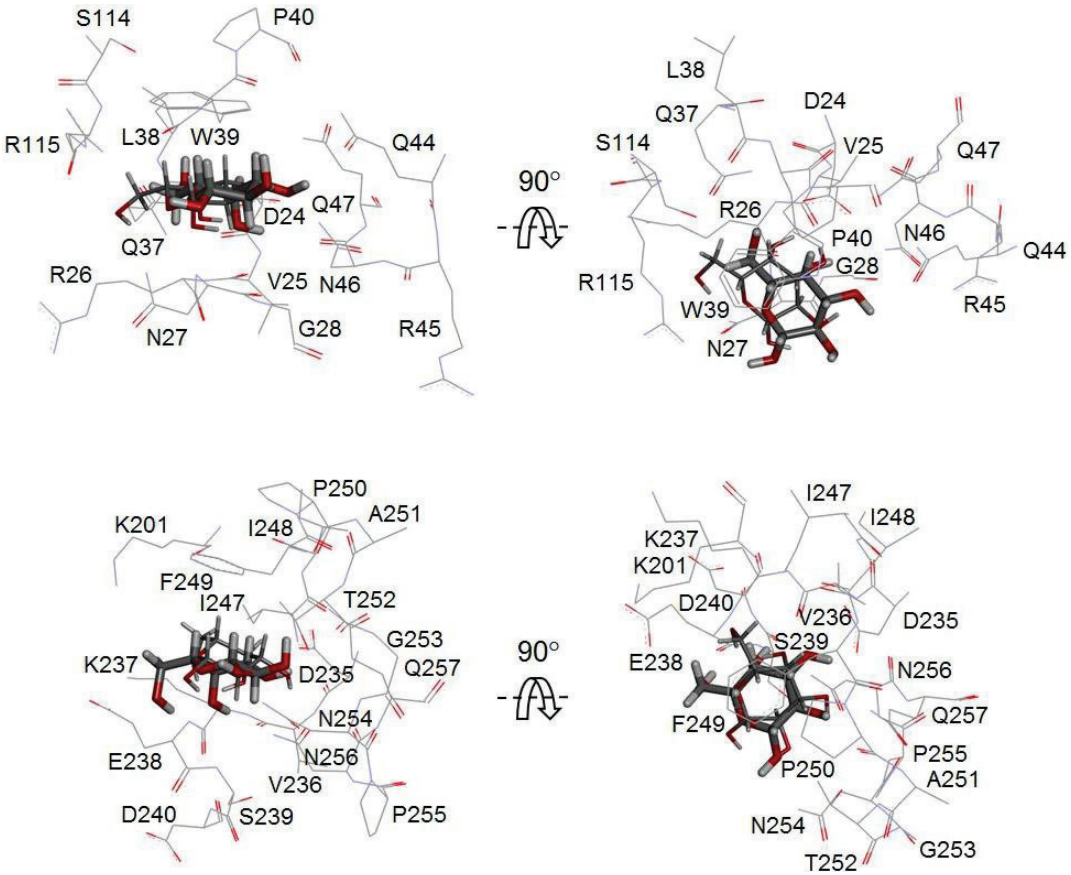


Figure S2 Comparison of the three-dimensional models of ebulin l (accession code 1HWN) sugar-binding sites 1 alpha (up) and 2 gamma (down) bound to D-galactose. The results obtained using AutoDock 4.2 (thin sticks) are compared with those obtained by X-ray diffraction (thick sticks) as reported previously (Pascal et al. 2001). The amino acids of the sugar-binding sites are represented by lines.

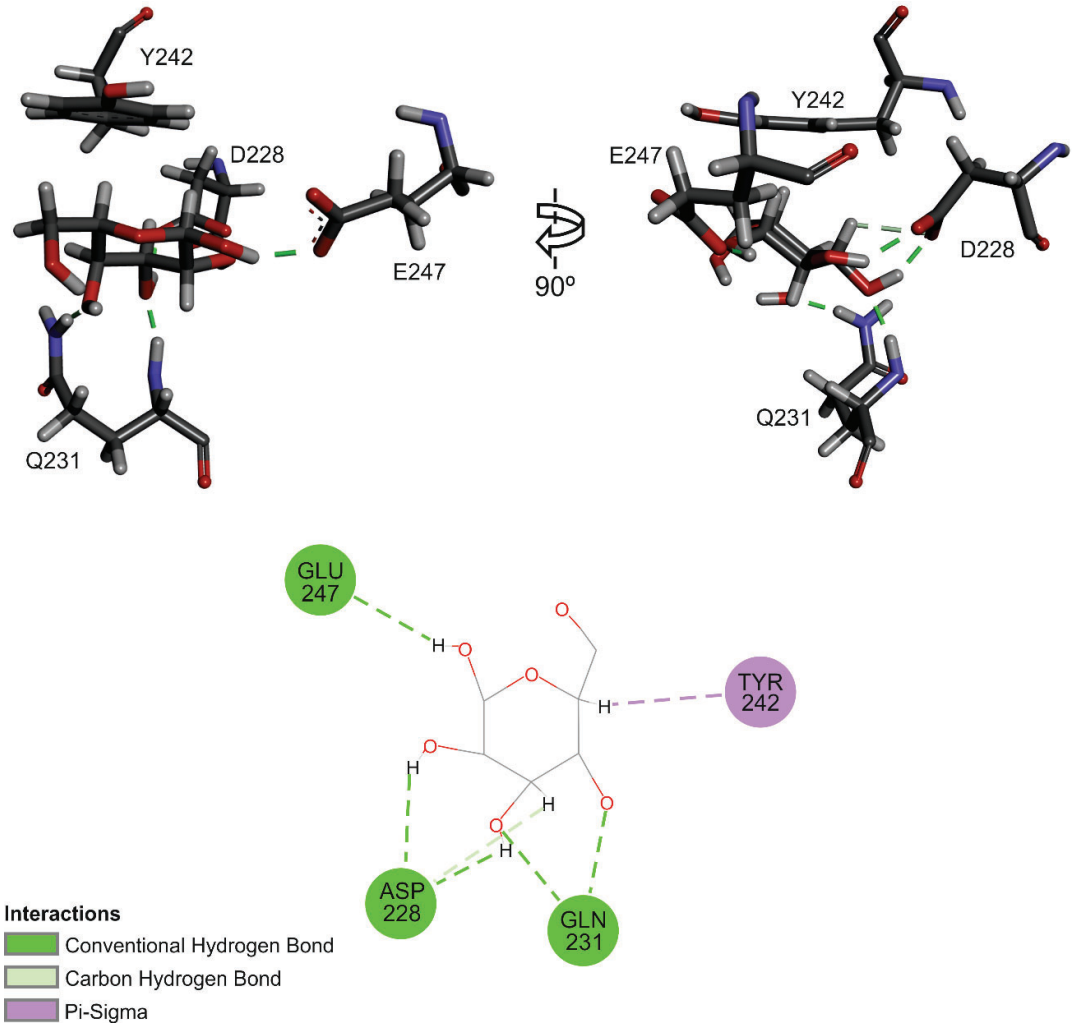


Figure S3 Three-dimensional model of SEA 2-gamma site (up) bound to D-galactose. The ligands and the amino acids which bind the sugar molecule by either C-H-π interactions (Y242) or both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks. Receptor-ligand interactions are also shown in a 2D diagram (down).

B chain

A chain

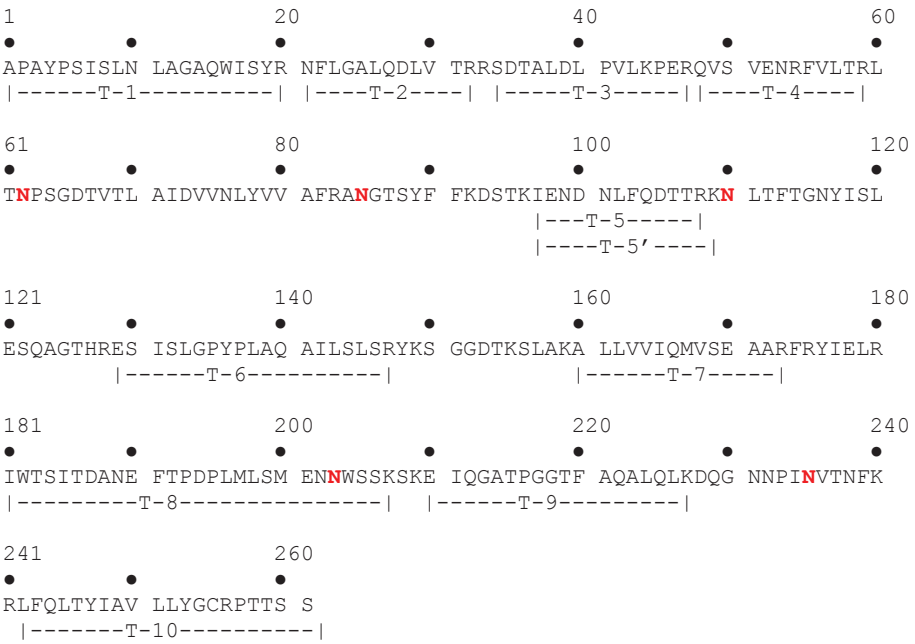


Figure S4 Amino acid sequence mapping of B and A chains of ebulin-RP mapped on deduced amino acid sequence from ebulin-RP gene. The M*r* of tryptic peptides are reported in **Tables S3 and S4**. Asparaginyl residues present in N-linked glycosylation consensus sequence are highlighted in red. Residues involved in alpha and gamma sites are reported in blue and green, respectively.

Table S1. **N-terminal amino acid sequences of native ebulin-RP, its A or B chains, and its CNBr peptides.** N-terminal amino acid sequencing was carried out as indicated in section 2.4.

Protein/peptide	Sequence	Sequence position
Native		
Ebulin-RP	APAYPSISLNLAGAQ	1-15 (A-chain)
SDS-PAGE		
A-chain	APAYPSISLNLAGAQWISYR	1-20 (A-chain)
B-chain	(Y)DEVCTVVDVTRRISG*	1-16 (B-chain)
RP-HPLC CNBr fragmentation		
Peak 1	SKEIQGATPGGTFAQALQLK	208-227 (A-chain)
Peak 2	EVRQSNVSLRQVIIYHPTGNANQQWITSTHQP	233-264 (B-chain)
Peak 3	VSEAARFRYIELRIWTSITD	168-187 (A-chain)

*B-chain showed to different sequences: the first sequence, accounting for approximately 70% of the total protein sample, was YDEVCTVVDVTRRISG and the second one was DEVCTVVDVTRRISG.

Table S2 Protein structures chosen by I-TASSER as the templates in the modelling of ebulin-RP and SEA.

Protein	**Templates used by I-TASSER**	**Figure**
Pro ebulin RP	2VLC (cinnamomin III)	S1
Ebulin RP A-chain	1HWN (ebulin l) 1HWM (ebulin l) 2PJO (recombinant ricin A-chain) 5GU4 (recombinant ricin A-chain)	Cited in section 3.3
Ebulin RP B-chain	1HWP (ebulin l) 1ONK (mistletoe lectin I) 3C9Z (SNAII) 4ZA3 (*Momordica charantia* type 2 RIP) 2VLC (cinnamomin III) 2ZR1 (*Abrus precatorius* agglutinin)	3
SEA A-chain	1BR6 (recombinant ricin A-chain) 1HWM (ebulin l) 3KTZ (GAP31)	S1
SEA B-chain	1HWP (ebulin l) 1GGP (*Trichosanthes kirilowii* lectin-1) 3CA6 (SNAII) 2VLC (cinnamomin III) 3C9Z (SNAII) 2ZR1 (*Abrus precatorius* agglutinin)	S1, 3, S3

Table S3 Molecular mass values of tryptic (T) peptides from ebulin-RP B-chain determined by MALDI-TOF mass spectrometry. The theoretical molecular masses were obtained from the deduced B-chain amino acid sequence from ebulin-RP gene.

Peptide [a]	Sequence position	Experimental molecular mass	Theoretical molecular mass	Δ (Da)	Missed cleavage at	Notes
T-1	1-12	1455.65	1455.68	0.03		Cys-CAM [b]
T-1'	1-13	1611.75	1611.78	0.03	R12	Cys-CAM [b]
T-2	14-17	432.83	432.26	0.57		-
T-3	18-25	875.40	875.44	0.04		Cys-CAM [b]
T-4	26-51	3072.89	3072.36	0.53		Cys-CAM [b]
T-5	62-88	2791.35	2791.27	0.08		Cys-CAM [b]
T-6	89-160	7653.79	7652.62	1.17		Cys-CAM [b]
T-7	173-182	1551.67	1151.58	0.09		-
T-8	188-206	2087.04	2087.10	0.06		Cys-CAM [b]
T-9	230-235	746.49	746.42	0.07		-
T-10	236-242	803.59	803.44	0.15		-
T-11	243-264	2533.25	2533.26	0.01		*C-terminal*

[a] [M+H]+ experimental molecular mass values obtained by MALDI-TOF MS. The monoisotopic molecular masses have been considered, except for the T-6 peptide for which the average molecular mass is reported. [b] cysteinyl residues have been treated with iodoacetamide to form carbamidomethyl-cysteine (Cys-CAM).

Table S4 Molecular mass values of tryptic (T) peptides from ebulin-RP A-chain determined by MALDI-TOF mass spectrometry. The theoretical molecular masses were obtained from the deduced A-chain amino acid sequence from ebulin-RP gene.

Peptide[a]	Sequence position	Experimental molecular mass	Theoretical molecular mass	Δ (Da)	Missed cleavage at	Notes
T-1	1-20	2178.16	2178.13	0.03		-
T-2	21-32	1346.78	1346.74	0.04		-
T-3	34-47	1553.87	1553.85	0.02	K44	-
T-4	48-59	1447.84	1447.80	0.04	R54	-
T-5	97-108	1465.74	1465.69	0.05		-
T-5'	97-109	1593.80	1593.79	0.01		-
T-6	129-147	2015.12	2015.12	0.00		-
T-7	160-173	1499.89	1499.86	0.03		-
T-8	181-207	3130.11	3129.48	0.63		-
T-9	210-227	1830.00	1829.97	0.03		-
T-10	242-261	2246.19	2246.19	0.00		Cys-CAM [b] *C-terminal*

[a] [M+H]+ experimental molecular mass values obtained by MALDI-TOF MS. The monoisotopic molecular masses have been considered, except for the T-8 peptide for which the average molecular mass is reported. [b] cysteinyl residues have been treated with iodoacetamide to form carbamidomethyl-cysteine (Cys-CAM).

Experimental procedures for tables S3 and S4

In-gel tryptic digestion

Due to insolubility of the A- and B-chains of ebulin-RP after reduction, tryptic peptides were obtained after A- and B-chains separation by SDS-PAGE and in-gel tryptic digestion (Ref. 1). Briefly, 10 μg of ebulin-RP was subjected to SDS-PAGE with β-mercaptoethanol. Furthermore, A- and B-chain bands were excised from gels and destained by washing twice with 200 μL of water, followed by a further washing step with 50% acetonitrile. The gel pieces were then dried in a SpeedVac Vacuum (Savant Instruments, Holbrook, NY) and rehydrated with 200 μL of 50 mM ammonium bicarbonate, pH 8.0. Samples were reduced at 55 °C for 15 min by adding 10 mM dithiothreitol (final concentration) and alkylated in the dark at room temperature with 40 mM iodoacetamide (final concentration), for 15 min. Following two washes with 200 μL of water and then with 50% acetonitrile, gel pieces were dried and rehydrated with 200 μL of 50 mM ammonium bicarbonate pH 8.0. Enzymatic digestions were performed by incubation at 37 °C for 3 h following the addition of 2.0 μL of 70 ng/μL TPCK-treated bovine trypsin (Sigma Aldrich, Milan, Italy). Peptides were extracted in two steps by sequential addition of 200 μL of 1% TFA and then 200 μL of 2% TFA/ 50% acetonitrile for 10 min in a sonication bath. The combined supernatants dried in the SpeedVac Vacuum, were resuspended in 5.0 μL of 0.1% TFA/ 50% acetonitrile for MALDI-TOF MS analyses.

Peptide mass fingerprinting by MALDI-TOF mass spectrometry

For MALDI-TOF analysis, 1.0 μL of digestion mixtures or peptide solution was mixed with 1.0 μL of saturated α-cyano-4- hydroxycinnamic acid matrix solution [10 mg/mL in acetonitrile: 0.1% TFA (1 : 1; v/v)] or sinapinic acid [10 mg/mL in acetonitrile/0.1% TFA (2 : 3; v/v)]. Thus, a droplet of the resulting mixture (1 μL) was placed on the mass spectrometer's sample target and dried at room temperature. Once the liquid was completely evaporated, samples were loaded into the mass spectrometer [MALDI-TOF micro MX spectrometer (Waters, Manchester, UK)] and analysed. The instrument was externally calibrated using a tryptic alcohol dehydrogenase digest (Waters) in reflectron mode. For linear mode, a four-point external calibration was applied using an appropriate mixture (10 pmol/mL) of insulin, cytochrome C, horse Mb and trypsinogen as standard proteins (Sigma). A mass accuracy near to the nominal (50 and 300 ppm in reflectron and linear modes, respectively), was achieved for each standard. All spectra were processed and analysed using MassLynx4.0 software (Ref. 1).

Reference 1

Severino, V., Chambery, A., Vitiello, M., Cantisani, M., Galdiero, S., Galdiero, M., Malorni, L., Di Maro, A., Parente, A. Proteomic analysis of human U937 cell line activation mediated by Haemophilus influenzae type b P2 porin and its surface-exposed loop 7 (2010) Journal of Proteome Research, 9 (2), pp. 1050-1062.

Video S1 The 1-alpha site from ebulin l complexed with either D-galactose (sticks), sialic acid (black lines) or N-acetyl-D-glucosamine (red lines) is represented. The amino acids which bind the galactose molecule by either C-H-π interactions (W39) or both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks (see also Figure 3b).

Video S2 The 2-gamma sugar-binding sites from ebulin l complexed with D-galactose is represented (see also Figure 3d). The ligand and the amino acids which bind the sugar molecule by either C-H-π interactions (F249) or both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks. For comparative purposes the binding of the other sugars is also shown: sialic acid (black lines) and N-acetyl-D-glucosamine (red lines).

Video S3 The 2-gamma sugar-binding sites from SEA complexed with sialic acid is represented (see also Figure 3d). The ligand and the amino acids which bind the sugar molecule by both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks. For comparative purposes the binding of the other sugars is also shown: D-galactose (blue lines) and N-acetyl-D-glucosamine (red lines).

Video S4 The 2-gamma sugar-binding sites from ebulin-RP complexed with N-acetyl-D-glucosamine is represented (see also Figure 3d). The ligand and the amino acids which bind the sugar molecule by both conventional (dashed green lines) and non-conventional (dashed light green lines) hydrogen bonds are represented by sticks. For comparative purposes the binding of the other sugars is also shown: D-galactose (blue lines) and sialic acid (black lines).

PDB file S1 (Ebulin 1alpha GAL.pdb). Ebulin l (accession code 1HWN) sugar-binding site 1-alpha bound to D-galactose. Docking was carried out using Autodock 4.2 as indicate in section 2.17 of Materials and methods. The file only includes the galactose and the amino acids represented in figure 3b.

PDB file S2 (Ebulin 2gamma GAL.pdb). Ebulin l (accession code 1HWN) sugar-binding site 2-gamma bound to D-galactose. Docking was carried out using Autodock 4.2 as indicate in section 2.17 of Materials and methods. The file only includes the galactose and the amino acids represented in figure 3d.

PDB file S3 (SEA 2gamma SIA.pdb). SEA sugar-binding site 2-gamma bound to sialic acid. The three-dimensional structural modelling was carried out on the I-TASSER server as indicate in section 2.16 of Materials and methods. Docking was carried out using Autodock 4.2 as indicate in section 2.17 of Materials and methods. The file only includes the sialic acid and the amino acids represented in figure 3d.

PDB file S4 (EbulinRP 2gamma NAG.pdb). Ebulin-RP sugar-binding site 2-gamma bound to N-acetyl-D-glucosamine. The three-dimensional structural modelling was

carried out on the I-TASSER server as indicate in section 2.16 of Materials and methods. Docking was carried out using Autodock 4.2 as indicate in section 2.17 of Materials and methods. The file only includes the N-acetyl-D-glucosamine and the amino acids represented in figure 3d.