

# Expression Analysis of Ribosome-inactivating Proteins From Cucumber\*

DANG Liu-Yi<sup>1)\*\*</sup>, Pierre Rougé<sup>2)</sup>, Els JM Van Damme<sup>1)\*\*</sup>

<sup>(1)</sup> Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology, Ghent University, Ghent, Belgium;

<sup>(2)</sup> UMR 152 PHARMA-DEV, Université de Toulouse, Toulouse, France)

## 通讯作者简介

Els J.M. Van Damme, 女, 博士, 比利时根特大学生物工程学院分子生物技术系全职教授。目前从事植物生物化学, 分子生物技术和糖生物学方面的研究, 主要涉及直接或间接与植物抗逆相关的多种蛋白质, 尤其是糖结合蛋白(凝集素)的生理功能的研究。目前已发表论文 300 余篇, 总被引用次数超过 10 000 次。

党刘毅, 男, 比利时根特大学分子生物技术系应用生物科学博士。主要从事黄瓜中两类凝集素家族的生物学功能研究, 即含有气单胞菌溶素结构域的 amaranthin 嵌合凝集素和核糖体失活蛋白。

**Abstract** Ribosome-inactivating proteins (RIPs) are a class of cytotoxic enzymes which possess highly specific rRNA N-glycosidase activity and are capable of catalytically inactivating prokaryotic or eukaryotic ribosomes. Due to their unique biological activities, RIPs have been considered to have great potential in medical and agricultural applications. The cucumber genome accommodates two genes encoding type 2 ribosome-inactivating proteins, further referred to as CumsaAB1 and CumsaAB2. Type 2 RIPs, represented by ricin, usually consist of two peptides linked by a disulfide bridge. A chain with N-glycosidase activity and B chain with carbohydrate-binding activity. In this study, the expression of the cucumber RIPs was analyzed. Sequence analysis showed that CumsaAB1 is synthesized with a signal peptide and subcellular localization studies further confirmed that the protein is expressed extracellularly, following the secretory pathway. Analyses of the transcript levels in various tissues during cucumber development showed that CumsaAB1 is present at extremely low levels in most tissues while the expression of CumsaAB2 is much higher, especially in leaves from plants at first-true-leaf stage and plants at the onset of flowering. Molecular modelling of the RIP sequences was performed to unravel the three-dimensional conformation of cucumber RIPs and their carbohydrate-binding sites. This study provided valuable information on the subcellular localization, the tissue-specific expression and the structure of RIPs from cucumber plants.

**Key words** lectin, ribosome-inactivating protein, subcellular localization, expression analysis

**DOI:** 10.16476/j.pibb.2017.0311

Through evolution plants have evolved to synthesize different kinds of toxic molecules to cope with all the challenges they face throughout their entire life cycle<sup>[1]</sup>. Among all toxic compounds, toxic proteins take up an important place and plants contain a wide range of proteins with different biological activities. Being ubiquitous proteins in the plant kingdom, lectins represent an important group of these toxic proteins, which are critical for the growth and development of plants<sup>[2]</sup>.

By definition, lectins are proteins with at least one non-catalytic domain that binds reversibly with

specific mono- or oligosaccharides<sup>[3]</sup>. Although the majority of lectins have been characterized from plants, these proteins have also been reported in

\*This work was supported by grants from Ghent University; Liuyi Dang is a recipient of a scholarship from the China Scholarship Council and received doctoral co-funding from the Special Research Council of Ghent University.

\*\*Corresponding author.

DANG Liu-Yi. Tel: 86-15529270964, E-mail: liuyi.dang@outlook.com

Els JM Van Damme. E-mail: Elsjm.VanDamme@UGent.be

Received: July 25, 2017 Accepted: September 30, 2017

animals, insects, viruses, fungi and bacteria<sup>[4]</sup>. They have drawn a lot of attention because of their possible agricultural and biomedical applications, *e.g.* their anti-tumor activities. The anti-tumor activities of different plant lectins has been shown for several cancer cell cultures, such as human hepatocarcinoma cells<sup>[5]</sup>, human bladder cancer cells<sup>[6]</sup>, human melanoma cells<sup>[7]</sup> and rat pancreatic cells<sup>[8]</sup>. Up to now, several hundreds of plant lectins have been identified, purified and at least partially characterized<sup>[9-10]</sup>.

Lectins are proteins endowed with one or more carbohydrate-binding sites, which enables them to specifically recognize and bind with certain glycan structures. Although some lectins recognize and interact with monosaccharides *e.g.* mannose, glucose, galactose, most plant lectins preferentially bind with more complex oligosaccharides like N- and O-linked glycans<sup>[11]</sup>. The carbohydrate-binding site of lectins typically consists of five to six amino acids, which bind the hydroxyls of the sugar residues mainly through hydrophobic interactions. The specific interaction between the lectin and the glycan involves the formation of a network of hydrogen bonds and can often be enhanced by a hydrophobic stacking of the pyranose ring of the sugar to the aromatic ring of aromatic residues (tyrosine, tryptophan or phenylalanine) located in the close vicinity of the carbohydrate-binding site<sup>[12]</sup>.

Since the family of lectins groups all proteins that specifically interact with glycan structures without altering the substrate, a large number of very diverse proteins complies with this definition. Currently, all plant lectins are classified into 12 families based on their carbohydrate-recognition domains. A detailed overview of these 12 plant lectin domains was described in several recent review papers<sup>[2, 13]</sup>. Here we focused on one of the most-well known lectin families, the ricin-B family in cucumber, an economically important and widely cultivated crop.

The ricin-B lectin family groups all proteins that show sequence homology to ricin, the most well-known type 2 RIP from *Ricinus communis*<sup>[14]</sup>. RIPs are a class of cytotoxic enzymes that possess highly specific rRNA N-glycosidase activity and are capable of catalytically inactivating prokaryotic or eukaryotic ribosomes<sup>[15]</sup>. Previous studies have categorized RIPs mainly in two types: type 1 RIPs with one single

N-glycosidase chain and type 2 RIPs consisting of one N-glycosidase chain and one ricin-B lectin chain. So far, the type 2 RIPs are widely distributed in plants, animals, bacteria and fungi<sup>[16]</sup>.

Our previous analysis of lectin domain sequences in the genome of cucumber (*Cucumis sativus* var. *sativus* L.) revealed the presence of 9 ricin B homologs with three different domain architectures<sup>[17]</sup>. Among them, 5 sequences contain a tandem array of ricin B domains, resembling the lectin chain of ricin. In 2 of the sequences this lectin domain is linked to a RIP domain, and thus these proteins can be classified as type 2 RIPs. The remaining 4 sequences contain a ricin B domain linked to a glycoside hydrolase family 5 domain (GH5).

In this study, the expression of the two type 2 RIPs from cucumber was analyzed. Their subcellular localization was investigated using stably expressed EGFP-fusion proteins in tobacco suspension cells. Furthermore, the transcript levels for type 2 RIP genes were quantified and compared in several tissues during development of cucumber plants. Finally, molecular modelling of the RIP sequences was performed to unravel the three-dimensional structure of cucumber RIPs and their carbohydrate-binding sites.

## 1 Materials and methods

### 1.1 Vectors for expression analysis

Expression vectors for cucumber RIP sequences C-terminally tagged with EGFP were constructed using the Gateway technology (Invitrogen). The coding sequences of the proteins under study were amplified as attB PCR products using cDNA obtained from leaves of 9 day-old plant leaves of *Cucumis sativus*. The primers used for the first PCR are shown in Table 1. The second PCR was performed using forward primer (5' GGGGACAAGTTTGTACAAAAAAGCAGGCT 3') and reverse primer (5' GGGGACCACTTTGTACAA GAAAGCTGGGT 3') and 1 : 10 diluted product from first PCR was used as template. The PCR program was as follows: 5 min 94 °C, 30 cycles (30 s 94 °C, 30 s 50 °C, 1.5 min 72 °C), 5 min 72 °C. Subsequently, the BP reaction was performed using the pDONR221 vector (Invitrogen). After sequencing of the entry clones, the LR reaction was done with the destination vector pK7WGF2 to fuse the EGFP sequence to the C-terminus of type 2 RIP sequence<sup>[18]</sup>.

Table 1 Primer list

	Gene	Forward (F) and Reverse (R) primers: 5'→3'
Subcellular localization	AB1	F: AAAAAAGCAGGCTTCACCATGACATCTTTAACGTTTAC R: AGAAAGCTGGGTCGTAGAATAATGTCCATTGTTG
	AB2	F: AAAAAAGCAGGCTTCACCATGAGAGTTTTATTAGCTTTC R: AGAAAGCTGGGTCGTACAAAGCAACCCAATGTT
qPCR analysis	CACS	F: TGGGAAGATTCTTATGAAGTGC R: CTCGTCAAATTTACACATTGGT
	PP2A	F: CAACAGGTGATATTGGATTATGATTATAC R: GCCAGCTCATCCTCATATAAG
	AB1	F: CCTCTTCAAACCGCTCAGG R: CATCCACGACCTTCTTCGTT
	AB2	F: ATCGAAGAGACCCGTTGTTG R: GAGCACTGGAATGCCATAGAG

## 1.2 Stable transformation of tobacco BY-2 cells

Stable transformation of BY-2 cells was achieved by co-cultivation with *Agrobacterium tumefaciens* containing the EGFP-fusion vector. In brief, the EGFP-fusion vector of CumsaAB1 and CumsaAB2 were transferred to *A. tumefaciens* strain LBA4404 by triparental mating. Four milliliter of a one-week old BY-2 cell culture was added to 40 ml of MS medium and grown for 4 days on a rotary shaker with speed of 150 r/min, at 25°C in the dark. For co-cultivation, 4 ml of the 4 day-old BY-2 cells were mixed with various concentrations of transformed LBA4404 cells and incubated for 2 days at 25°C without shaking. Then, the mixtures were transferred on MS agar plates containing kanamycin (100 mg/L), vancomycin (200 mg/L) and carbenicillin (500 mg/L), and kept at 25°C in the dark. After approximately 2 to 4 weeks, calli became visible and were transferred to fresh selection plates. The images of tobacco BY-2 cells expressing the fluorescent proteins were acquired using a Nikon A1R confocal system, mounted on a Nikon Ti microscope body. The acquired images were analyzed by Fiji software (<http://fiji.sc/>).

## 1.3 Preparation of protoplasts from stably transformed BY-2 cells

Stably transformed BY-2 cells from 4-day old cultures were harvested by vacuum filtration. Approximately 5 g of the collected cells was incubated for 3 h at 37°C with an enzyme solution (2% cellulase RS (Duchefa), 1% macerozyme (Yakult), 0.1% pectolyase (Duchefa), 0.4 mol/L mannitol, 40 mmol/L CaCl<sub>2</sub>, 10 mmol/L MES buffer, pH 5.5) to remove the cell wall and obtain protoplasts. Afterwards, protoplasts were filtered through a 100 micron cell strainer with washing buffer (10 mmol/L MES, 0.4 mol/L mannitol,

pH 6.7). The protoplasts were spun down for 5 min at 200 *g* and washed with washing buffer. The washed protoplasts are then suspended in 3 ml of BY-2 medium and quickly checked with the Nikon A1R confocal system.

## 1.4 Expression analysis of cucumber RIP genes in different tissues during plant development

Cucumber (*Cucumis sativus* L. cv. Vert Petit de Paris) seeds were germinated on moist filter paper in a petri dish for 2 days at 28°C in the dark. Germinated seedlings were transferred to pots containing commercial soil, and grown in a plant growth room at 28°C with a 16 h/8 h light/dark photoperiod. Samples of cotyledons, leaves, stems, roots, flower buds and fruits (collected at 0 days, 4 days, 8 days, 12 days, 16 days after pollination) were collected from plants at different stages of plant development. All experiments were performed between April 2013 and September 2014. The primers for target and reference genes for qPCR analysis are available in Table 1. The reference genes were chosen based on previous studies and their stabilities have been tested in cucumber<sup>[19]</sup>.

Total RNA was isolated from different samples using TRIzol reagent (Sigma-Aldrich) and treated with DNase I (Thermo Fisher Scientific) to remove any traces of genomic DNA according to the manufacturer's instructions. The first-strand cDNA synthesis was performed using M-MLV reverse transcriptase (Thermo Fisher Scientific). RNA concentrations were measured with Nanodrop. The quality of the cDNA was checked by standard RT-PCR using primers of reference genes.

Quantitative PCR analyses were performed using Rotor-Gene 3000 (Corbett Life Science, Qiagen) using Rotor Discs (Qiagen). The qPCR reactions were

carried out in a total volume of 20  $\mu$ l containing 10  $\mu$ l of SYBR Green PCR Sensi-Mix (Bioline), 1  $\mu$ l of each primer (10 mol/L), 1  $\mu$ l of cDNA template (20 mg/L) and 7  $\mu$ l of sterile distilled water. The conditions for qPCR reaction were 96 °C for 10 min followed by 45 cycles at 96 °C for 25 s, 58 °C for 25 s and 72 °C for 20 s. Gene-specific primers for qPCR were designed using Primer 3 (<http://primer3.ut.ee/>) for amplification of a 100 ~200 bp fragment. The results of qPCR were analyzed using Relative Expression Software Tool-384 version 2 (REST-384), which also determined the statistical significance of the results<sup>[20]</sup>. All experiments were performed with two independent biological replicates, each containing 3 technical replicates.

### 1.5 Molecular modelling of CumsaAB1 and CumsaAB2

Homology modeling of RIPs from cucumber was performed with the YASARA Structure program<sup>[21]</sup>, running on a 2.53 GHz Intel duo core Macintosh computer. Different models of CumsaAB1 (Csa5M428440)/CumsaAB2 (Csa6M040650) were built from the X-ray coordinates of bitter melon (*Momordica charantia*) RIP2 (PDB code 4ZGR)<sup>[22]</sup>, snake gourd (*Trichosanthes anguina*) RIP2 (PDB code 4HR6)<sup>[23]</sup>, ebulin I from dwarf elder (*Sambucus ebulus*) (PDB code 1HWM)<sup>[24]</sup>, agglutinin from jequirity (*Abrus precatorius*) (PDB code 2ZR1)<sup>[25]</sup>, and ricin from castor bean (*Ricinus communis*) (PDB code 3RTJ)<sup>[26]</sup>, used as templates. Finally, a hybrid model of CumsaAB1/CumsaAB2 was built from the different previous models. PROCHECK, ANOLEA<sup>[27]</sup>, and the calculated QMEAN scores<sup>[28-29]</sup>, were used to assess the geometric and thermodynamic qualities of the three-dimensional model. Only two residues (Asn79, Ala270) out of 518, occurred in the non-allowed regions in the Ramachandran plot. Both residues are mainly located in the loop regions connecting the  $\beta$ -sheets to the  $\alpha$ -helices in the model. Using ANOLEA to evaluate the model, only 13 residues (out of 518) of the CumsaAB1/CumsaAB2 model exhibited energy over the threshold value. The calculated QMEAN6 score of the model gave a value of 0.572.

### 1.6 Evolutionary relationships between different type 2 RIPs

A phylogenetic analysis was performed to check the evolutionary relationships of cucumber RIPs with other classical type 2 RIPs. Sequences of type 2 RIPs used in the analysis: CumsaAB1 and CumsaAB2 from

*C. sativus*, CitlaAB from *Citrullus lanatus* (Cucurbit Genomics Database: Cla007203), MaldoAB from *Malus domestica* (Genome database for Rosaceae: MDP0000711911) and sequences from NCBI database: ricin from *Ricinus communis* (XP\_002534649), Cinnamomin from *Cinnamomum camphora* (AAK82460), Ebulin from *Sambucus ebulus* (AJ400822), SNA-I from *Sambucus nigra* (P93543). Muscle (<http://www.ebi.ac.uk/Tools/msa/muscle/>) was used for the multiple sequence alignment and MEGA6 was used to generate the phylogenetic tree using bootstrap confidence values based on 500 iterations. Sequence alignments were represented by sequence logos as created by WebLogo 3<sup>[30]</sup>.

## 2 Results

### 2.1 Sequences encoding RIPs in cucumber genome

Two genes coding for type 2 RIPs, named CumsaAB1 and CumsaAB2, were identified in the cucumber genome. Sequence analysis with the Pfam database 30.0 showed that both RIPs have domain architectures similar to the classical type 2 RIPs, composed of a RIP domain (A chain) and lectin domain (B chain) (Figure 1). According to SignalP v4.1 CumsaAB1 and CumsaAB2 are synthesized with a signal peptide, suggesting that these proteins are synthesized following the secretory pathway. No transmembrane regions were predicted in the RIP sequences.

### 2.2 Subcellular localization of cucumber RIPs in BY-2 cells

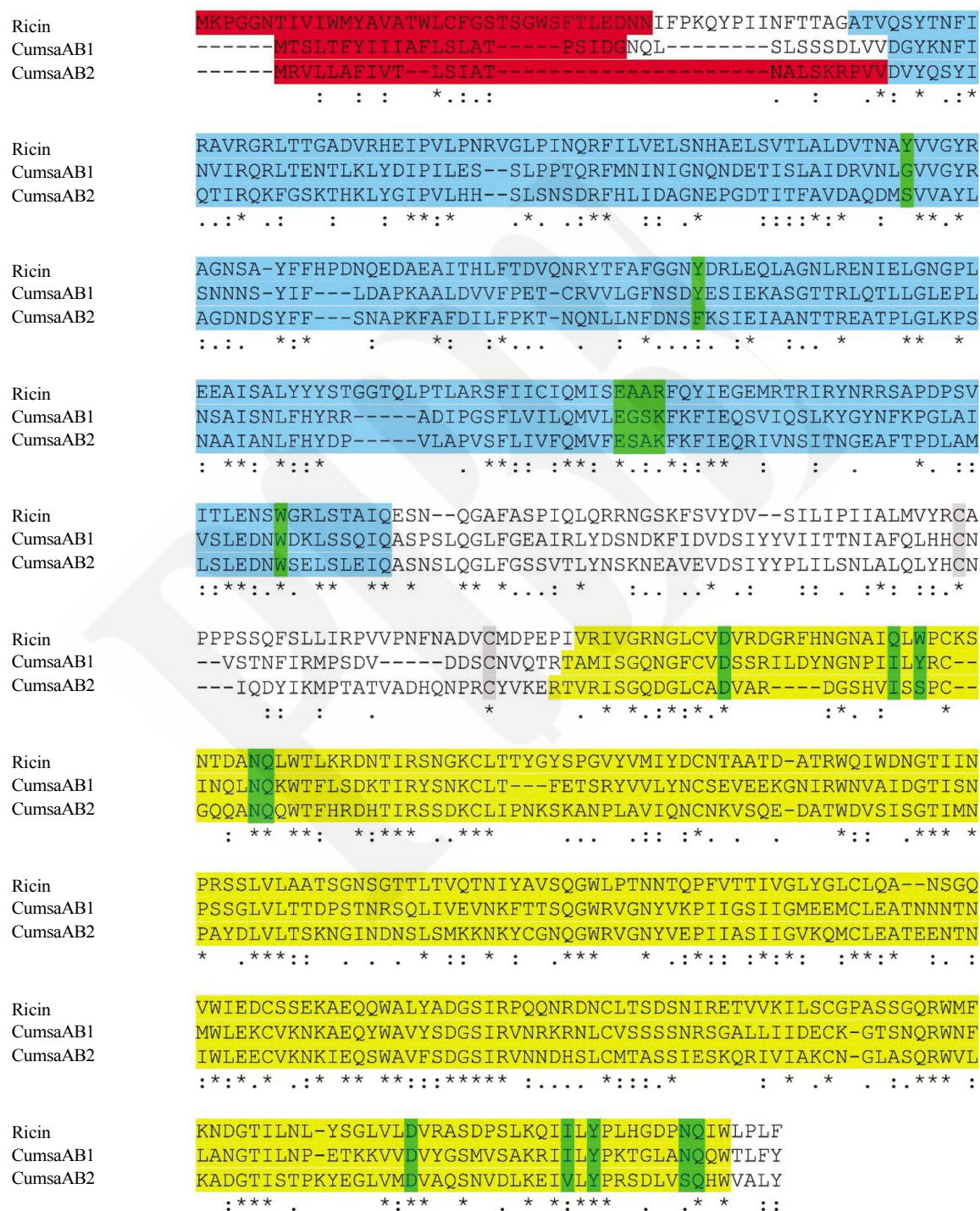
To validate their subcellular localization, the RIP sequences were C-terminally fused in-frame with EGFP and stably expressed in tobacco BY-2 cells. Images acquired by confocal microscopy revealed fluorescence of CumsaAB1-EGFP around the nucleus and at the cell periphery (Figures 2a & 2b). The plot profile showed the strongest fluorescence intensity at the edges of the cells (Figure 2c). More mature cells revealed a dotted fluorescence pattern at the edges of the cells (Figure 2d). No fluorescence was seen in the vacuole. Unfortunately, fluorescence was never observed for CumsaAB2-EGFP.

### 2.3 CumsaAB1 expression in protoplasts prepared from stably transformed BY-2 cells

To check in more detail the fluorescence at the edge of the cell, protoplasts were prepared by removing the cell wall from BY-2 cells stably

expressing EGFP-tagged CumsaAB1. Figure 3 shows different confocal images of a protoplast, from the middle of the protoplast to the plasma membrane. In middle of the protoplast (Figure 3a), the fluorescence

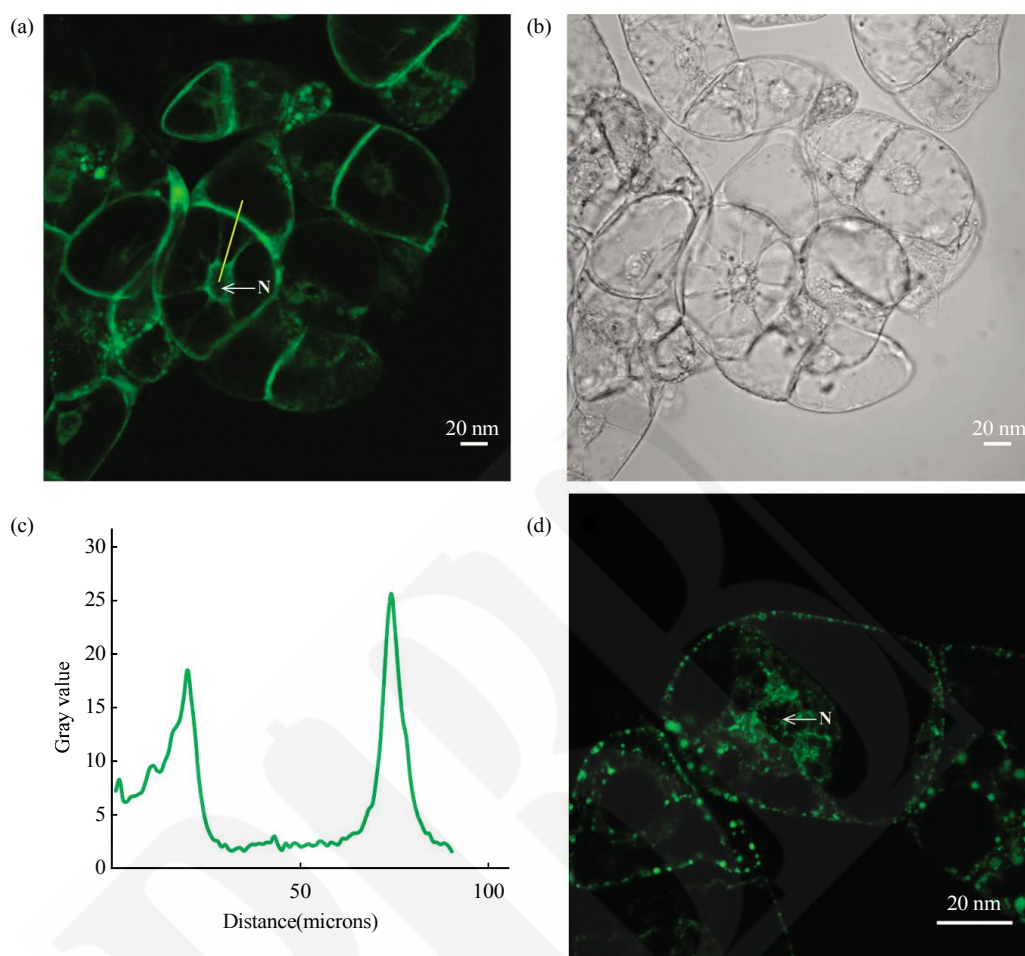
was mainly distributed around the nucleus and at the periphery of the cell. The fluorescence increased when the images were taken closer to the surface of the protoplast (Figures 3b & 3c). Figure 3D shows the



**Fig. 1 Multiple sequence alignment of CumsaAB1 and CumsaAB2 in comparison with ricin (UniProtKB/Swiss-Prot: P02879)**

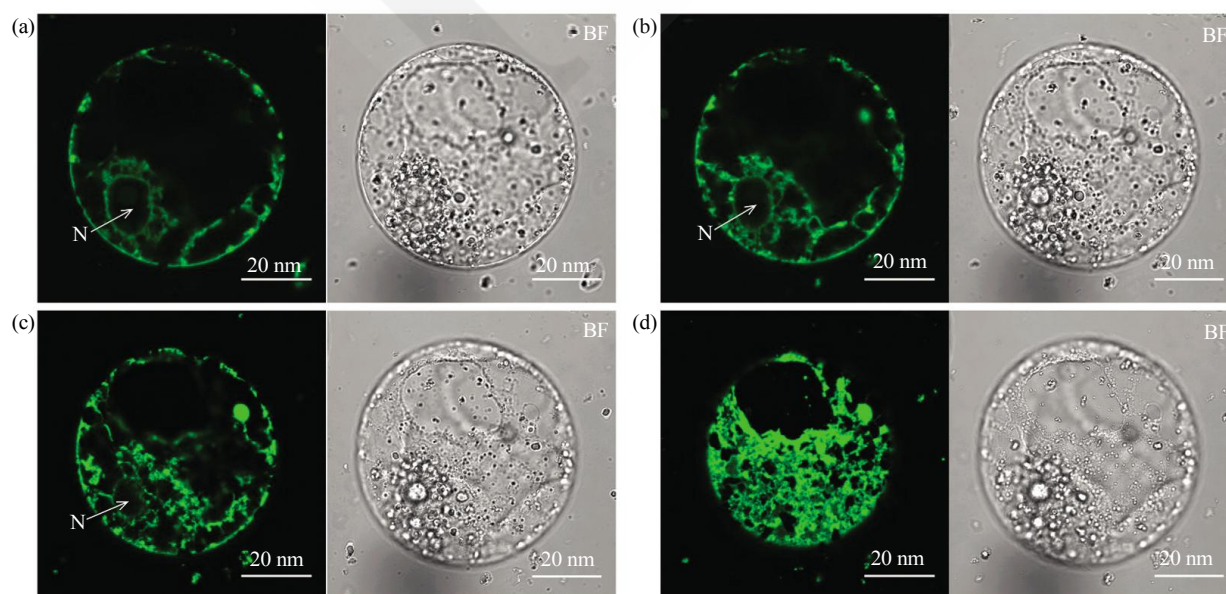
"\*" means identical amino acids in all sequences; "." means highly conserved amino acids and ":" refers to weakly conserved amino acids. Signal peptides are indicated in red; N-glycosidase domains (in blue) and ricin B domains (in yellow) are based on the Pfam database; cysteines for disulfide bridge formation between the A and B chains are marked in grey; residues important for N-glycosidase activity and carbohydrate-binding sites are marked in green.





**Fig. 2 CumsaAB1-EGFP expression in stably transformed tobacco BY-2 cells**

(a) Fluorescence of CumsaAB1-EGFP in BY-2 cells. (b) Bright field image of BY-2 cells. (c) Plot profile showing the fluorescence intensity across the cell (yellow line, starting from the nucleus to the edge of the cell). (d) CumsaAB1-EGFP expression in more mature BY-2 cells. N denotes the nucleus.



**Fig. 3 Confocal microscopy images of protoplasts from stably transformed BY-2 cells expressing CumsaAB1-EGFP**

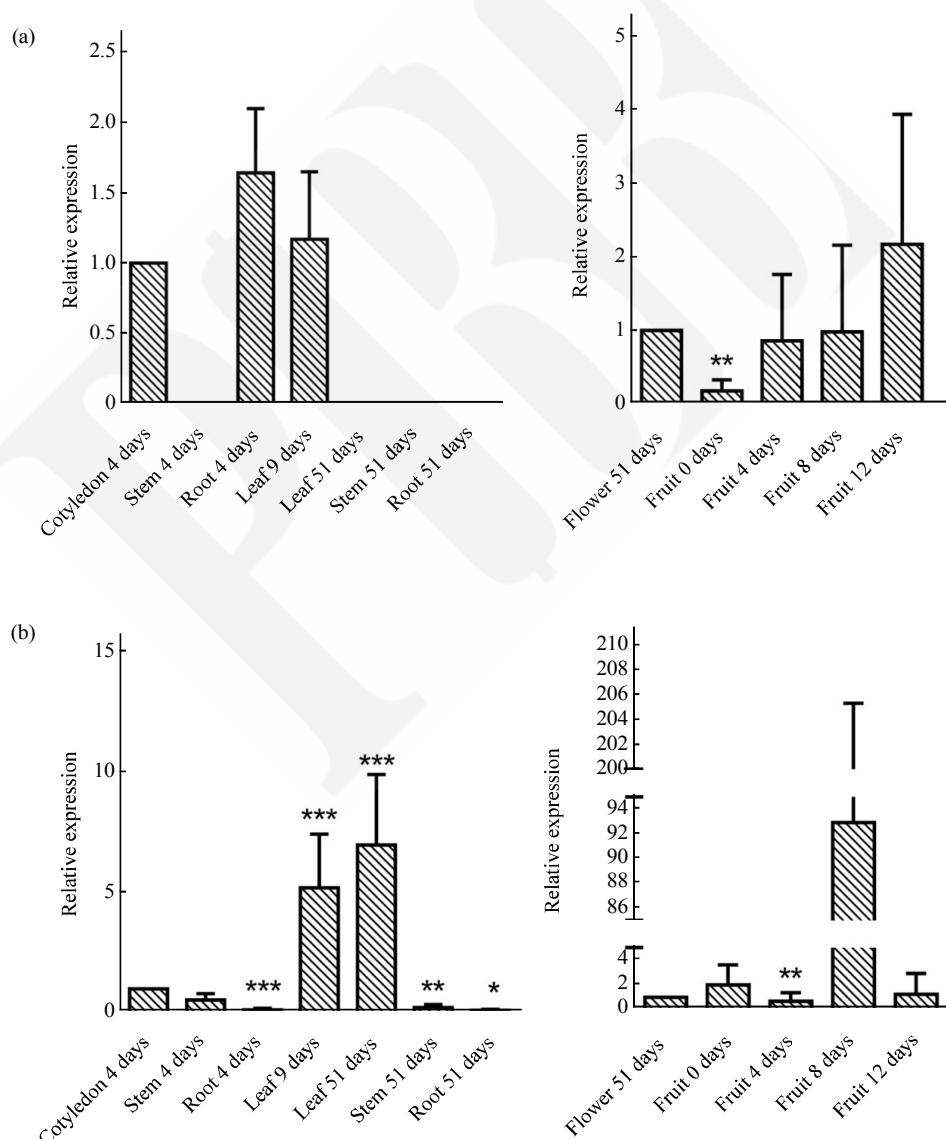
Panel (a)-(d) represent fluorescence and bright field images from different Z-stacks of one protoplast, starting from the middle of the cell (a) to the surface of the cell (d).

fluorescence at the surface of the protoplast. Considering the fluorescence distribution on the cell surface and the presence of vesicle-like dots, CumsaAB1 is probably secreted extracellularly.

## 2.4 Transcriptional analysis of RIP genes in different tissues throughout cucumber development

Transcript levels for CumsaAB1 and CumsaAB2 were analyzed in cucumber plants grown under normal conditions. Samples were collected at different developmental stages: seed germination (4 days), plants with first true leaf (9 days), plants that started flowering (51 days) and fruits at different stages of

maturation. For all samples, gene expression was normalized against reference genes (CACS and PP2A) and compared to 4-day cotyledons or flowers at day 51 of cucumber development. The expression profiles showed that CumsaAB2 is most abundant in 9-day and 51-day old leaves, with transcript levels being significantly higher than the transcript levels in the cotyledons (Figure 4b). The expression of CumsaAB1 was mainly detected in flowers and fruits but was too low to allow accurate quantification by qPCR in most tissues. It was estimated that the expression of CumsaAB1 in cotyledon is approximately 120 times lower than the expression of CusamAB2 (Figure 4a).



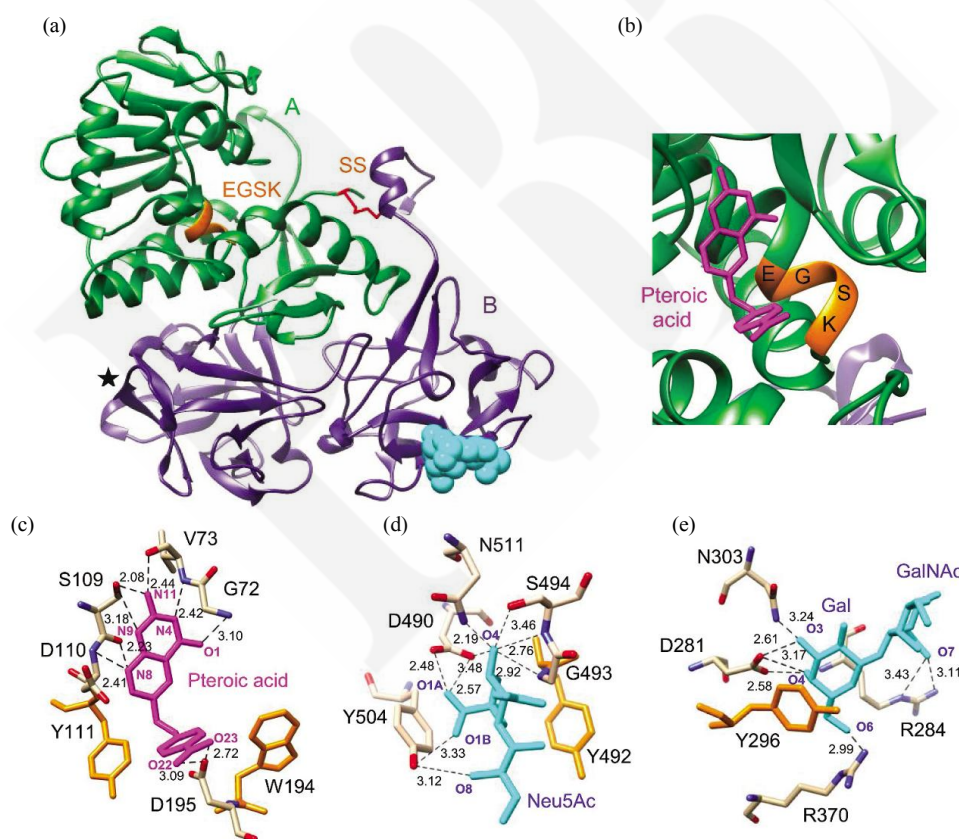
**Fig. 4 Transcriptional profiling of RIP genes in different tissues during cucumber plant development**

(a) AB1, (b) AB2. Relative expression levels of RIP genes from vegetative organs (leaves, stems, roots) are compared to the expression level in cotyledons while relative expression levels of RIP genes from reproductive organs (flowers and fruits) are compared to the expression level in flowers. Expression level comparison between cotyledon and flower: AB1 (cotyledon) : AB1 (flower) = 1 : 3.6; AB2 (cotyledon) : AB2 (flower) = 1 : 0.05. Bars represent means and standard errors from two biological replicates, each replicate containing a pool of 3 plants. Asterisks indicated statistically significant differences compared with the control tissue (\* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ ; REST analysis).

## 2.5 Molecular modelling of RIPs from cucumber

The cucumber CumsaAB1 protein consists of a polypeptide chain of 518 amino acids and is predicted to fold as a type 2 RIP, resulting from the covalent association of the A and B chains like in other canonical type 2 RIPs, *e.g.* ricin<sup>[31]</sup> or ebulin<sup>[24]</sup> (Figure 5a). The modeled A chain exhibits an overall three-dimensional fold similar to that found in other type 2 RIPs, built up of six strands of  $\beta$ -sheet clustered in a left-handed twisted bundle, associated with eight  $\alpha$ -helices. The B chain consists of two tandemly arrayed domains 1 and 2. A disulfide bridge occurs between two conserved Cys residues located at the C-terminal and N-terminal ends of the A (C246) and B chain (C263), respectively. Two asparagine residues, N247 and N251 in the sequence stretch connecting the

two cysteine residues are readily accessible to peptidases, suggesting a possible posttranslational cleavage by an asparagine endopeptidase, which would result in the separation of the A and B chains. However, the A and B chain would remain covalently bound by the C246-C263 disulfide bridge, after proteolytic cleavage. Two putative *N*-glycosylation sites 80NNS82 and 247NVS249, occur in the A chain, whereas 4 putative *N*-glycosylation sites are present in the B chain, 333NCS335 and 369NRS371 in domain 1, and 412NNT414 and 453NRS455 in domain 2. All of these *N*-glycosylation sites are readily exposed at the surface of CumsaAB1 and could actually be glycosylated according to the GlyProt server (<http://www.glycosciences.de>) prediction.



**Fig. 5 Molecular modelling of CumsaAB1**

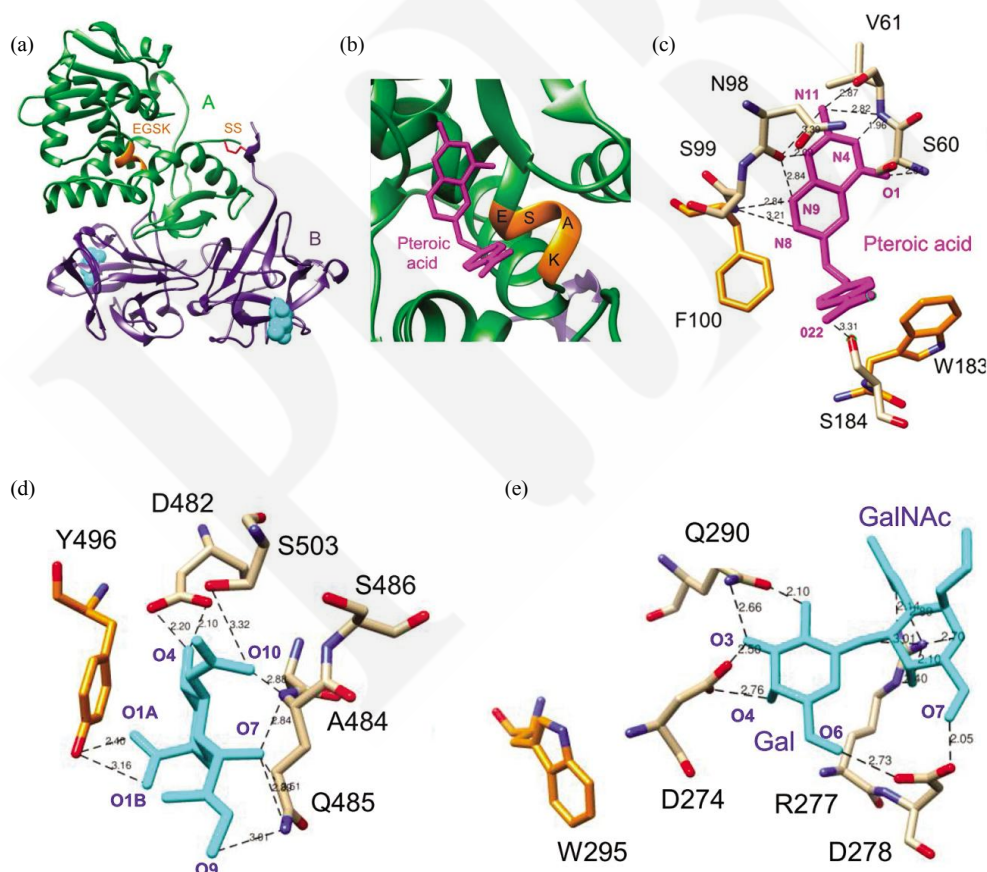
(a) Ribbon diagram showing the organization of modeled CumsaAB1 with two chains corresponding to the A (a) and the B chain (b), covalently linked by a disulfide bridge (SS) between the C-terminus of the A chain and the N-terminus of the B chain. The EGSK sequence, corresponding to the canonical EAAR sequence occurring in typical type 2 RIPs, is colored orange. A single carbohydrate-binding site of the B chain is occupied by the T-antigen disaccharide. (b) Docking of pterioic acid (magenta stick) into the active site cavity of the A chain, showing the EGSK sequence of the predicted active site. (c) Network of H-bonds (dashed black lines) anchoring pterioic acid to the active site of the A chain. Both the Y111 and W194 residues develop a stacking interaction with pterioic acid. (d) Network of hydrogen bonds (dashed black lines) potentially anchoring sialic acid (Neu5Ac) (cyan stick) to the amino acid residues forming the carbohydrate-binding site 2 of the CumsaAB1. The aromatic residues Y492 and Y504 interact through a stacking with the pyranose ring of Neu5Ac. (e) Network of hydrogen bonds (dashed black lines) anchoring the T-antigen disaccharide (Gal (β1-3) GalNAc) (cyan stick) to the amino acid residues forming the carbohydrate-binding site 1 of the cucumber type 2 RIP. The aromatic residue Y296 interacts through a stacking with the pyranose ring of Gal. The H-bond distances in panels (c-e) are indicated in Ångströms (Å).



The A chain exhibits a sequence stretch 160EGSK163, homologous to the sequence stretch 164EAAR167, which is known to play a key role in the N-glycosidase activity of the A chain in type 2 RIPs (Figure 5b). The key residues of the active site of the A chain (G72, V73, S109, D110, D195), readily accommodate pterioic acid (PTA), an A chain substrate analog, as shown from docking experiments (Figure 5c). Additional stacking interactions occur with the aromatic residues Y111 and W194. Two carbohydrate-binding sites (CBS) occur at both ends of subdomains 1 $\alpha$  (CBS1) and 2 $\gamma$  (CBS2) (Figure 5d, e). Docking experiments indicate that Neu5Ac readily anchors to CBS2 through a network of 9 hydrogen bonds with G493, S494, D490, Y504, and N511 residues

(Figure 5d). The two aromatic residues, Y492 and Y504, create a stacking interaction with the pyranose ring Neu5Ac, which reinforces the interaction with the sugar. Both Gal and T-antigen also dock properly to CBS2. CBS1 readily interacted with the T-antigen disaccharide (Gal $\beta$ 1,3GalNAc) in docking experiments (Figure 5e), through a network of 7 hydrogen bonds with D281, R284, N303, and R370. An additional stacking interaction with Y296 reinforces the binding of the sugar to the lectin moiety.

CumsaAB2 consists of a polypeptide chain of 510 amino acids and shares a very similar three dimensional structure with CumsaAB1 according to the modelling (Figure 6). In CumsaAB2, the A chain exhibits a sequence stretch 149ESAK152, homologous



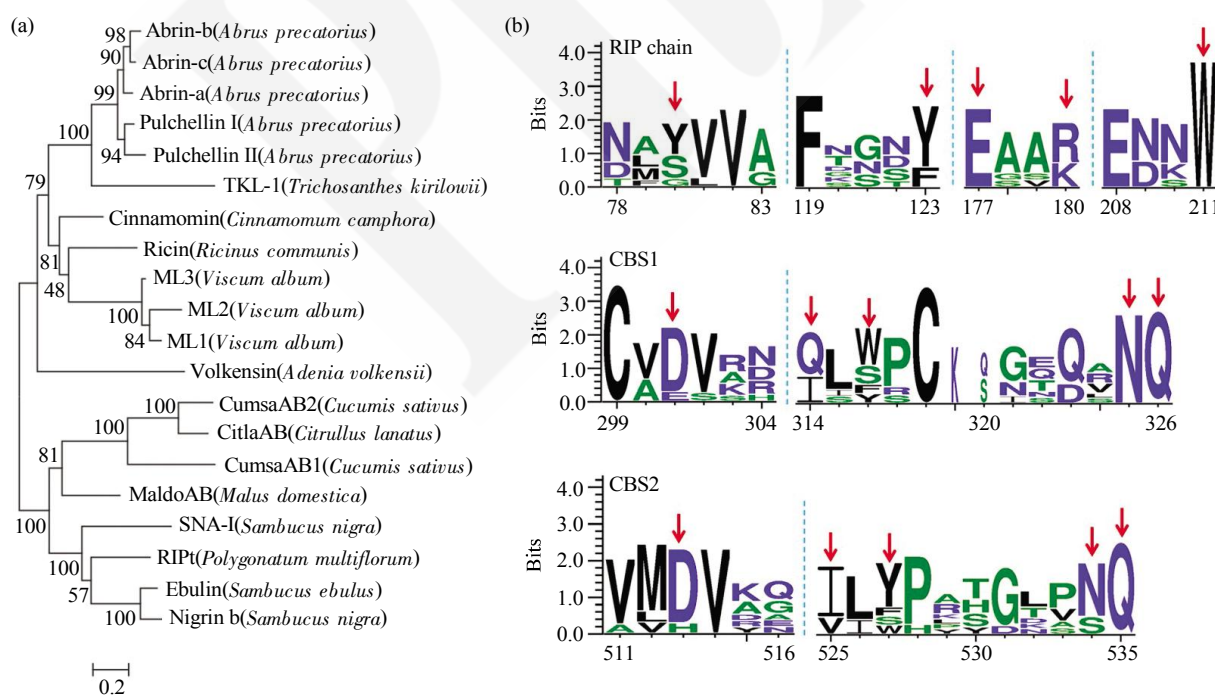
**Fig. 6 Molecular modelling of CumsaAB2**

(a) Ribbon diagram showing the organization of modelled CumsaAB2 with two chains corresponding to the A (a) and B chain (b), covalently linked by a disulfide bridge (SS) between the C-terminal end of A chain and the N-terminal end of B chain. The ESK sequence, corresponding to the canonical EAAR sequence occurring in typical RIP 2 RIPs, is colored orange. Both carbohydrate-binding sites of the B chain are occupied by galactose (balls colored cyan). (b) Docking of pterioic acid (magenta stick) into the active site cavity of the A chain, showing the ESK sequence of the active site. (c) Network of H-bonds (dashed black lines) anchoring pterioic acid to the active site of the A chain. Both the F100 and W183 residues develop a stacking interaction with pterioic acid. The H-bond distances are indicated in Ångströms (Å). (d) Network of hydrogen bonds (dashed black lines) anchoring sialic acid (Neu5Ac) (cyan stick) to the amino acid residues forming the carbohydrate-binding site 2 of CumsaAB2. The aromatic Y496 residue interacts through a stacking with the pyranose ring of Neu5Ac. The H-bond distances are indicated in Ångströms (Å). (e) Network of hydrogen bonds (dashed black lines) anchoring the T-antigen disaccharide (Gal (β1-3) GalNAc) (cyan stick) to the amino acid residues forming the carbohydrate-binding site 1 of the CumsaAB2. The aromatic W295 residue interacts through a stacking with the pyranose ring of Gal. The H-bond distances are indicated in Ångströms (Å).

to EGSK sequence of CumsaAB1 and the sequence stretch EAAR, which plays a key role in the N-glycosidase activity of the ricin A-chain. The key residues of the A chain active site (S60, V61, N98, S99, S184), readily accommodate pteric acid (PTA), a substrate analog as shown from docking experiments. Additional stacking interactions occur with the aromatic residues F100 and W183. Similar to CumsaAB1, two CBSs occur at both ends of subdomains of the lectin chain. Docking experiments indicate that Neu5Ac readily anchors to CBS2 through a network of hydrogen bonds with D490, Q493, S494, Y504, and S511 residues. The aromatic residue Y496 creates a stacking interaction with the pyranose ring Neu5Ac, which reinforces the interaction with the sugar. Both Gal and T-antigen also dock properly to CBS2. CBS1 readily interacted with the T-antigen disaccharide (Gal ( $\beta$ 1-3) GalNAc) in docking experiments, through a network of hydrogen bonds with D274, R277, D278, and Q290. An additional stacking interaction with W295 reinforces the binding of the sugar to the lectin moiety.

The molecular modelling study revealed a typical

type 2 RIP structure for the RIPs from cucumber. To demonstrate the evolutionary relationships between cucumber RIPs and other type 2 RIPs, a phylogenetic tree was built with type 2 RIP sequences from different plant species (Figure 7a). The dendrogram revealed two major clades. The two RIPs from cucumber are clustered together with the only RIP from watermelon, both species belong to the Cucurbitaceae family. Furthermore, the RIP from apple was also grouped into the same clade with the cucumber RIPs. The important residues for the N-glycosidase activity and CBSs are mostly conserved among all RIPs studied (Figure 7b). In the RIP domain, E241 and W248 are highly conserved in all sequences. Y117, Y160, R217 are also conserved in most sequences but some variations were observed (Y117S, Y160F, R217K), though amino acids are replaced by amino acids with similar properties in most cases. In both CBSs, 3 out of 5 amino acids (D357/D578, N381/N599, Q382/Q600) are highly conserved for each CBS while two amino acids at positions I319/I525 and I316/I527 are often replaced by other amino acids.



**Fig. 7** Sequence analyses of selected type 2 RIPs

(a) Phylogenetic tree of type 2 RIPs from different species. Sequences used in the analysis: CumsaAB1 (Csa5M428440.1); CumsaAB2 (Csa6M040650.1); CitlaAB (Cla007203); MaldoAB (MDP0000711911); Ricin (XP\_002534649.1); Cinnamomin (AAK82460.1); Ebulin (AJ400822); SNA- I (U27122); Abrin-a (AAA32624); Abrin-b (Q06077); Abrin-c (P28590); Pulchellin I (ABW23503); Pulchellin II (ABW23504); TKL-1 (1GGP); ML1 (P81446); ML2 (Q6H266); ML3 (P82683); RIPt (Q9M653); Nigrin b (P33183); Volkensin (CAD61022). (b) Analysis of conserved amino acids in the RIP domain and carbohydrate-binding domains of RIPs. Amino acids that are critical for the catalytic site (responsible for N-glycosidase activity) and carbohydrate-binding sites (CBS1, CSB2) are indicated by red arrows<sup>[32]</sup>. All position numbers are based on the ricin sequence.

### 3 Discussion

Currently, RIPs have been studied extensively, especially for their toxicity and their unique biological activities<sup>[33]</sup>. It was reported that some RIPs are toxic to pathogens and predators, such as fungi, bacteria, viruses, and insects<sup>[34]</sup>. Due to their unique biological activities, RIPs have been considered to have great potential in medical and agricultural applications<sup>[35]</sup>. In this study, the expression of the type 2 RIPs from cucumber was analyzed, with emphasis on their subcellular localization and tissue expression throughout development. Furthermore, molecular modelling was performed to get insight into the structure of the cucumber RIPs and the biological activities of the RIP and lectin domains.

Predicting the subcellular localization of proteins is an important aspect to elucidate the function(s) and interactions a protein can be involved in. Within a prokaryotic cell, basically only three locations are possible: inside or outside the plasma membrane, or inserted into the membrane<sup>[36]</sup>. In contrast, a eukaryotic cell contains various membrane-enveloped compartments and as a result the location of a protein can be more complex in eukaryotic cells compared to prokaryotic cells. The subcellular localization of the cucumber RIPs was studied with the use of EGFP-tagged proteins, which were stably expressed in BY-2 cells. In fact, different plant cells were tried as hosts for the expression of the EGFP-fused RIPs, including transient transformation of tobacco (*Nicotiana benthamiana*) leaves, stable transformation of *Arabidopsis* suspension PSB-D cells and tobacco BY-2 cells. However, fluorescence was observed only with CumsaAB1 expressed in stably transformed BY-2 cells.

Confocal microscopy analysis of CumsaAB1-EGFP expression in BY-2 cells showed the strongest fluorescence at the periphery of the cells and also surrounding the nucleus (most probably ER/Golgi apparatus) but no fluorescence was observed inside the nucleus or in the cytoplasm. Furthermore, vesicle-like fluorescent dots were observed at the surface of some cells, which can be validated by co-expressing RIPs with a vesicle marker protein. For a better illustration of the fluorescence on the cell surface, protoplasts were prepared from stably transformed BY-2 cells. Images taken at different levels in one cell showed the distribution of the fluorescence throughout the protoplasts. It is clear that the fluorescence was mainly

present unevenly at the surface of the protoplasts. Considering the absence of a transmembrane region and presence of the signal peptide in the sequence of CumsaAB1, the presence of fluorescence at the cell surface probably represents the final location of the protein before it is secreted extracellularly. It can be envisaged that CumsaAB1 synthesis follows the secretory pathway, RIPs are translocated from the ER to the cell surface and secreted, as reported before for some type 2 RIPs<sup>[37]</sup>.

Although RIPs have been identified in all kinds of tissues, the expression profile of RIPs can be vastly different between plant species. For example, ricin is expressed mainly in the seeds of *Ricinus* plants while pokeweed antiviral protein is expressed in different tissues, including leaves, seeds and roots<sup>[38]</sup>. Furthermore the level of expression can be highly variable. In cucumber, the expression levels for the two RIPs under study differ remarkably. In general, the transcript levels of CumsaAB1 were much lower than these of CumsaAB2 in almost all tissues so that the quantitative analysis of the qPCR data was not reliable for CumsaAB1. Our results indicate that CumsaAB2 is mainly expressed in vegetative tissues whereas CumsaAB1 is expressed in flowers and fruits, suggesting that both RIPs have complementary activities. Interestingly, the sequences of CumsaAB1 and CumsaAB2 are 47% identical (67% similarity) but their expressions levels are distinct, which might indicate different biological roles for these two type 2 RIPs.

Molecular modelling of cucumber RIPs predicted that the polypeptides can fold as reported for typical type 2 RIPs, consisting of an enzymatic A chain and a lectin ricin-B chain. In addition, docking experiments suggested that the active sites of the RIP domain and the lectin domain can accommodate the typical substrates known to interact with type 2 RIPs. It is generally accepted that the N-glycosidase activity of the RIP domain towards ribosomes from plants, bacteria, yeasts or animals on the one hand, and the carbohydrate-binding activity of the lectin domain on the other hand, both contribute to the biological activities of RIPs, such as their involvement in the plant defense against biotic stresses<sup>[39]</sup>. This study provided valuable information on the localization of cucumber RIPs at cell level as well as at tissue level, and yielded insight into the structural properties and biological activities of the proteins.

## References

- [1] Mithofer A, Boland W. Plant defense against herbivores: chemical aspects. *Annu Rev Plant Biol*, 2012, **63**: 431–450
- [2] Van Damme E J M, Lannoo N, Peumans W J. Plant lectins. *Adv Botanical Res*, 2008, **48**: 107–209
- [3] Peumans W J, Van Damme E J M. Lectins as plant defense proteins. *Plant Physiol*, 1995, **109**(2): 347–352
- [4] Van Damme E J M. History of plant lectin research. *Methods Mol Biol*, 2014, **1200**: 3–13
- [5] Lyu S Y, Choi S H, Park W B. Korean mistletoe lectin-induced apoptosis in hepatocarcinoma cells is associated with inhibition of telomerase *via* mitochondrial controlled pathway independent of p53. *Arch Pharm Res*, 2002, **25**(1): 93–101
- [6] Plattner V E, Wagner M, Ratzinger G, *et al.* Targeted drug delivery: binding and uptake of plant lectins using human 5637 bladder cancer cells. *Eur J Pharm Biopharm*, 2008, **70**(2): 572–576
- [7] Liu B, Cheng Y, Zhang B, *et al.* Polygonatum cyrtoneura lectin induces apoptosis and autophagy in human melanoma A375 cells through a mitochondria-mediated ROS-p38-p53 pathway. *Cancer Lett*, 2009, **275**(1): 54–60
- [8] Mikkat U, Damm I, Kirchhoff F, *et al.* Effects of lectins on CCK-8-stimulated enzyme secretion and differentiation of the rat pancreatic cell line AR42J. *Pancreas*, 2001, **23**(4): 368–374
- [9] Van Damme E J, Peumans W J, Pusztai A, *et al.* Handbook of plant lectins: properties and biomedical applications. Chichester, England: John Wiley & Sons, 1998
- [10] Van Damme E J M, Peumans W J, Barre A, *et al.* Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit Rev Plant Sci*, 1998, **17**(6): 575–692
- [11] Ghazarian H, Idoni B, Oppenheimer S B. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochemica*, 2011, **113**(3): 236–247
- [12] Fernandez-Alonso M D, Diaz D, Berbis M A, *et al.* Protein-carbohydrate interactions studied by NMR: from molecular recognition to drug design. *Curr Protein Pept Sci*, 2012, **13**(8): 816–830
- [13] Lannoo N, Van Damme E J M. Lectin domains at the frontiers of plant defense. *Front Plant Sci*, 2015, **5**: 397
- [14] Rutenber E, Ready M, Robertus J D. Structure and evolution of ricin B chain. *Nature*, 1987, **326**(6113): 624–626
- [15] Peumans W J, Hao Q, Van Damme E J M. Ribosome-inactivating proteins from plants: more than RNA N-glycosidases? *FASEB J*, 2001, **15**(9): 1493–1506
- [16] Shang C, Peumans W J, Van Damme E J M. Occurrence and taxonomical distribution of ribosome-inactivating proteins belonging to the ricin/shiga toxin superfamily. *Ribosome-inactivating proteins*. Oxford: John Wiley & Sons, Ltd. 2014: 11–27
- [17] Dang L, Van Damme E J M. Genome-wide identification and domain organization of lectin domains in cucumber. *Plant Physiol Biochem*, 2016, **108**: 165–176
- [18] Karimi M, Inze D, Depicker A. GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci*, 2002, **7**(5): 193–195
- [19] Migocka M, Papierniak A. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators. *Mol Breed*, 2011, **28**(3): 343–357
- [20] Pfaffl M W, Horgan G W, Dempfle L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res*, 2002, **30**(9): e36
- [21] Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA-a self-parameterizing force field. *Proteins*, 2002, **47**(3): 393–402
- [22] Chandran T, Sharma A, Vijayan M. Structural studies on a non-toxic homologue of type II RIPs from bitter melon: Molecular basis of non-toxicity, conformational selection and glycan structure. *J Biosci*, 2015, **40**(5): 929–941
- [23] Sharma A, Pohlentz G, Bobbili K B, *et al.* The sequence and structure of snake gourd (*Trichosanthes anguina*) seed lectin, a three-chain nontoxic homologue of type II RIPs. *Acta Crystallogr Sect D Biol Crystallogr*, 2013, **69**(8): 1493–1503
- [24] Pascal J M, Day P J, Monzingo A F, *et al.* 28-Å crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin I. *Proteins*, 2001, **43**(3): 319–326
- [25] Cheng J, Lu T H, Liu C L, *et al.* A biophysical elucidation for less toxicity of agglutinin than abrin-a from the seeds of *Abrus precatorius* in consequence of crystal structure. *J Biomed Sci*, 2010, **17**: 34
- [26] Monzingo A F, Robertus J D. X-ray analysis of substrate analogs in the ricin A-chain active site. *J Mol Biol*, 1992, **227**(4): 1136–1145
- [27] Laskowski R A, MacArthur M W, Moss D S, *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, 1993, **26**(2): 283–291
- [28] Melo F, Feytmans E. Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol*, 1998, **277**(5): 1141–1152
- [29] Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 2011, **27**(3): 343–350
- [30] Crooks G E, Hon G, Chandonia J M, *et al.* WebLogo: a sequence logo generator. *Genome Res*, 2004, **14**(6): 1188–1190
- [31] Rutenber E, Robertus J D. Structure of ricin B-chain at 25 Å resolution. *Proteins*, 1991, **10**(3): 260–269
- [32] Shang C, Rougé P, Van Damme E J M. Ribosome inactivating proteins from Rosaceae. *Molecules*, 2016, **21**(8): 1105
- [33] Stirpe F, Gilibert-Oriol R. Ribosome-inactivating proteins: an overview//Gopalakrishnakone P, Carlini C R, Ligabue-braun R. *Plant Toxins*. Dordrecht; Springer Netherlands, 2015: 1–29
- [34] Stirpe F, Battelli M G. Ribosome-inactivating proteins: progress and problems. *Cell Mol Life Sci*, 2006, **63**(16): 1850–1866
- [35] Stirpe F. Ribosome-inactivating Proteins: from toxins to useful



- proteins. *Toxicon*, 2013, **67**: 12–16
- [36] Emanuelsson O. Predicting protein subcellular localisation from amino acid sequence information. *Brief Bioinform*, 2002, **3** (4): 361–376
- [37] Frigerio L, Jolliffe N A, Di Cola A, *et al.* The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. *Plant Physiol*, 2001, **126**(1): 167–175
- [38] Stirpe F. *Ribosome-inactivating proteins*//Wiley R G, Lappi D A. *Molecular neurosurgery with targeted toxins*. Totowa, NJ; Humana Press, 2005: 9–29
- [39] De Virgilio M, Lombardi A, Caliandro R, *et al.* Ribosome-inactivating proteins: from plant defense to tumor attack. *Toxins* (Basel), 2010, **2**(11): 2699–2737

## 核糖体失活蛋白及其在黄瓜中的表达分析 \*

党刘毅<sup>1)\*\*</sup> Pierre Rougé<sup>2)</sup> Els JM Van Damme<sup>1)\*\*</sup>

<sup>(1)</sup> Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology, Ghent University, Ghent, Belgium;

<sup>(2)</sup> UMR 152 PHARMA-DEV, Université de Toulouse, Toulouse, France)

**摘要** 核糖体失活蛋白是一类具有高度特异性 rRNA N-糖苷酶活性的蛋白, 它们能够使原核或真核细胞的核糖体失活因而具有细胞毒性。由于其独特的生物学性质, 核糖体失活蛋白被认为在农业和医学中都有着巨大的应用潜力。我们之前的研究表明, 黄瓜的基因组中共包含 2 个 2 类核糖体失活蛋白基因, 分别命名为 CumsaAB1 和 CumsaAB2。以蓖麻毒蛋白 Ricin 为代表, 2 类核糖体失活蛋白通常由 2 条二硫键连接的肽链组成: 具有 N-糖苷酶活性的 A 链与具有凝集素活性的 B 链。本文研究了黄瓜中核糖体失活蛋白的表达情况。亚细胞定位研究表明 CumsaAB1 经过蛋白分泌通路表达于细胞外, 这与蛋白质序列分析显示的 CumsaAB1 包含一个信号肽而不含转膜区域相一致。对黄瓜的不同生长阶段的不同组织中的转录水平分析表明, CumsaAB1 在大部分组织中以极低的水平表达, 而 CumsaAB2 表达水平则明显更高, 尤其在第一片真叶阶段和刚开花的植物中。最后, 我们使用分子模拟对黄瓜中核糖体失活蛋白的结构及糖结合位点进行了分析。本研究对黄瓜中核糖体失活蛋白的亚细胞定位、表达水平和可能的蛋白质结构进行了研究, 为其进一步的生物学功能研究提供了重要信息。

**关键词** 凝集素, 核糖体失活蛋白, 亚细胞定位, 表达分析

**学科分类号** Q513

**DOI:** 10.16476/j.pibb.2017.0311

\* 比利时根特大学研究基金(FWO-Vlaanderen), 国家公派留学基金(201206970005)和根特大学 BOF-cofunding 特别研究基金资助项目。

\*\* 通讯联系人。

党刘毅. Tel: 86-15529270964, E-mail: liuyi.dang@outlook.com

Els JM Van Damme. E-mail: Elsjm.VanDamme@UGent.be

收稿日期: 2017-07-25, 接受日期: 2017-09-30